Random structures, spin glasses and efficient algorithms

Amin Coja-Oghlan

University of Warwick

Overview

The general theme

Probability/Statistical Mechanics $\leftrightarrow$ Algorithms and Complexity.

Overview

The general theme

Probability/Statistical Mechanics $\leftrightarrow$ Algorithms and Complexity.

Outline

- Sherrington-Kirkpatrick
- Random k-SAT
- Why do the people 'on the other side of the bridge' care?
- Replica symmetry breaking in random k-SAT
- Dynamics and efficient algorithms

Spin glasses

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.

Spin glasses

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.
- A *configuartion* σ assigns spins to sites.

Spin glasses

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.
- A *configuartion* σ assigns spins to sites.
- The *Hamiltonian* H maps σ to *energy levels*.

Spin glasses

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.
- A *configuartion* σ assigns spins to sites.
- The *Hamiltonian* H maps σ to *energy levels*.
- This gives rise to the *Gibbs measure*

$$\begin{aligned} \mu(\{\sigma\}) &= Z^{-1}\exp(-\beta \cdot H(\sigma)), && \text{where} \\ \beta &\in (0,\infty) && \text{“inverse temperature”}, \\ Z &= \sum_{\sigma} \exp(-\beta H(\sigma)) && \text{“partition function”}. \end{aligned}$$

Spin glasses

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.
- A *configuartion* σ assigns spins to sites.
- The *Hamiltonian* H maps σ to *energy levels*.
- This gives rise to the *Gibbs measure*

$$\begin{aligned} \mu(\{\sigma\}) &= Z^{-1}\exp(-\beta \cdot H(\sigma)), && \text{where} \\ \beta &\in (0,\infty) && \text{“inverse temperature”}, \\ Z &= \sum_{\sigma} \exp(-\beta H(\sigma)) && \text{“partition function”}. \end{aligned}$$

- The twist is that H itself is *random*.
- Thus, there are *two levels* of randomness.

Spin glasses (ctd.)

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.
- A *configuartion* σ assigns spins to sites.
- The *Hamiltonian* H maps configurations to *energy levels*.
- $\rightsquigarrow$ *Gibbs measure* μ, *partition function* Z.
- H is random.

Spin glasses (ctd.)

A model of disordered systems

- There are n *sites*, . . .
- . . . each of which can be assigned one of q *spins*.
- A *configuartion* σ assigns spins to sites.
- The *Hamiltonian* H maps configurations to *energy levels*.
- $\rightsquigarrow$ *Gibbs measure* μ, *partition function* Z.
- H is random.

General question

- How do μ and Z "look like" with probability $1 - o(1)$ as $n \to \infty$?
- *Thermodynamic limit:* think of $n \to \infty$ (while q is fixed).
- We are interested in *low temperature*, i.e., $\beta \to \infty$.

The SK model

The Sherrington-Kirkpatrick model

- There are n *sites* $v_1, \ldots, v_n$.
- The *spins* are $\{-1, 1\}$.
- We have $e_{ij} = \Phi(0, 1)$ *mutually independent* Gaussians.
- The *Hamiltonian* is

$$H(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} \frac{1 - \sigma(v_i)\sigma(v_j)}{2} \cdot e_{ij}.$$

The SK model

The Sherrington-Kirkpatrick model

- There are n *sites* $v_1, \ldots, v_n$.
- The *spins* are $\{-1, 1\}$.
- We have $e_{ij} = \Phi(0, 1)$ *mutually independent* Gaussians.
- The *Hamiltonian* is

$$H(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} \frac{1 - \sigma(v_i)\sigma(v_j)}{2} \cdot e_{ij}.$$

Max Cut

- Let $G = (V, E)$ be the *complete graph* with edge weights e_{ij}.
- The *Max Cut* of G is $-\sqrt{n} \min_{\sigma} H(\sigma)$.
- As $\beta \to \infty$, we have $\min_{\sigma} H(\sigma) \sim \mathrm{E}_{\mu}[H(\sigma)]$.

The SK model

The Sherrington-Kirkpatrick model

$$H(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \le i < j \le n} \frac{1 - \sigma(v_i)\sigma(v_j)}{2} \cdot e_{ij}.$$

The SK model

The Sherrington-Kirkpatrick model

$$H(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \le i < j \le n} \frac{1 - \sigma(v_i)\sigma(v_j)}{2} \cdot e_{ij}.$$

The Parisi formula

- *Key quantity*: $\lim_{\beta \to \infty} \lim_{n \to \infty} \frac{1}{n} \cdot \mathrm{E}\left[\log Z\right].$

The SK model

The Sherrington-Kirkpatrick model

$$H(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \le i < j \le n} \frac{1 - \sigma(v_i)\sigma(v_j)}{2} \cdot e_{ij}.$$

The Parisi formula

- *Key quantity*: $\lim_{\beta \to \infty} \lim_{n \to \infty} \frac{1}{n} \cdot \mathrm{E}\left[\log Z\right]$.
- The answer is the Parisi formula (1980),...
- ...as shown by Talagrand (*Annals of Math 2006*).

The SK model

The Sherrington-Kirkpatrick model

$$H(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \le i < j \le n} \frac{1 - \sigma(v_i)\sigma(v_j)}{2} \cdot e_{ij}.$$

The Parisi formula

- *Key quantity*: $\lim_{\beta \to \infty} \lim_{n \to \infty} \frac{1}{n} \cdot \mathrm{E}\left[\log Z\right]$.
- The answer is the Parisi formula (1980),...
- ...as shown by Talagrand (*Annals of Math 2006*).
- It is *easy* to compute $\frac{1}{n} \cdot \log \mathrm{E}\left[Z\right]$...
- ...but the result is different due to *large deviations*.

Rigorous vs. non-rigorous

- The physics methods are *sophisticated* and *insightful*...
- ...but mathematically highly non-rigorous.

The k-SAT problem

k-CNF formulas

- Let $k \geq 3$.
- Let $V = \{x_1, \ldots, x_n\}$: *propositional variables*.
- A *k-clause* is an expression

$$C = l_1 \vee l_2 \vee \cdots \vee l_k, \quad \text{where } l_i \in \{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}.$$

- An expression $F = C_1 \wedge \cdots \wedge C_m$ is called a *k-CNF*.
- An *assignment* σ maps V to $\{\text{true,false}\}$.
- σ is *satisfying* if all clauses C_i evaluate to true.

The k-SAT problem

k-CNF formulas

- Let $k \geq 3$.
- Let $V = \{x_1, \ldots, x_n\}$: *propositional variables*.
- A *k-clause* is an expression

$$C = l_1 \vee l_2 \vee \cdots \vee l_k, \quad \text{where } l_i \in \{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}.$$

- An expression $F = C_1 \wedge \cdots \wedge C_m$ is called a *k-CNF*.
- An *assignment* σ maps V to $\{\text{true,false}\}$.
- σ is *satisfying* if all clauses C_i evaluate to true.

The (computational) k-SAT problem

Given F, *find* a satisfying assignment σ (if there is one).

The k-SAT problem (ctd.)

The k-SAT problem

Given F, *find* a satisfying assignment σ (if there is one).

k-SAT is "hard"

- There are 2^n assignments *in total*.
- To find a satisfying one, we could *enumerate* them all.
- However, even for $n = 100$ this is prohibitive (besides, in practice $n = 100,000$).
- Yet, no better algorithm is known.

The k-SAT problem (ctd.)

The k-SAT problem

Given F, *find* a satisfying assignment σ (if there is one).

NP-completeness

- k-SAT is *computationally equivalent* to a great variety of problems:
  - Graph coloring,
  - Travelling salesman,
  - Max Cut,
  - ...
- *NP-complete* problems.
- Conjecture: these problems do *not* have efficieny algorithms.
- *(One of the harder ways to make $1m.)*

The k-SAT problem (ctd.)

The k-SAT problem

Given F, *find* a satisfying assignment σ (if there is one).

NP-completeness

- k-SAT is *NP-complete*...
- ...hence, there should (?) exist "hard" instances.
- However, complexity theory *fails* to produce any.
- It also *fails* to identify easy cases.
- Can ideas from *statistical mechanics* help?

Random k-SAT

The random k-SAT model

- $k \geq 3$ is *fixed*.
- $V = \{x_1, \ldots, x_n\}$: *propositional variables* (sites).
- Choose k-clauses $C_1, \ldots, C_m$ *uniformly and independently*.
- Let $\mathbf{\Phi} = \mathbf{\Phi}_k(n, m) = C_1 \wedge \cdots \wedge C_m$.
- The *Hamiltonian* is $H(\sigma) = \#\text{clauses } C_i \text{ that are false under } \sigma$.

Random k-SAT

The random k-SAT model

- $k \geq 3$ is *fixed*.
- $V = \{x_1, \ldots, x_n\}$: *propositional variables* (sites).
- Choose k-clauses $C_1, \ldots, C_m$ *uniformly and independently*.
- Let $\mathbf{\Phi} = \mathbf{\Phi}_k(n, m) = C_1 \wedge \cdots \wedge C_m$.
- The *Hamiltonian* is $H(\sigma) = \#\text{clauses } C_i \text{ that are false under } \sigma$.

Zero temperature

- We are interested in *satisfying assignments*, i.e., $H(\sigma) = 0$.
- This corresponds to exactly *zero temperature*, i.e., $\beta = \infty$.
- Thus, $Z = \#\text{satisfying assignments}$.

Random k-SAT

The random k-SAT model

- $k \geq 3$ is *fixed*.
- $V = \{x_1, \ldots, x_n\}$: *propositional variables* (sites).
- Choose k-clauses $C_1, \ldots, C_m$ *uniformly and independently*.
- Let $\mathbf{\Phi} = \mathbf{\Phi}_k(n, m) = C_1 \wedge \cdots \wedge C_m$.
- The *Hamiltonian* is $H(\sigma) = \#\text{clauses } C_i \text{ that are false under } \sigma$.

"With high probability" / thermodynamic limit

Goal: prove statements that hold with probability $1 - o(1)$ as $n \to \infty$.

Random k-SAT

The random k-SAT model

- $k \geq 3$ is *fixed*.
- $V = \{x_1, \ldots, x_n\}$: *propositional variables* (sites).
- Choose k-clauses $C_1, \ldots, C_m$ *uniformly and independently*.
- Let $\mathbf{\Phi} = \mathbf{\Phi}_k(n, m) = C_1 \wedge \cdots \wedge C_m$.
- The *Hamiltonian* is $H(\sigma) = \#$clauses C_i that are false under σ.

Objectives

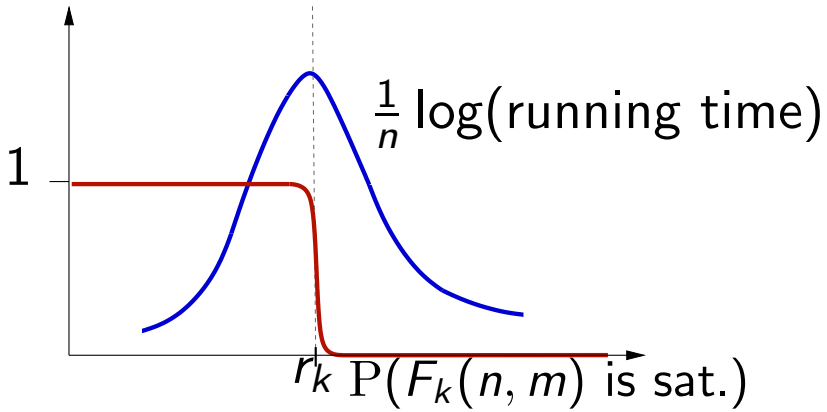
- For which m/n does $\mathbf{\Phi}_k(n, m)$ have a satisfying assignment w.h.p.?
- For what ranges of parameters is finding one *hard/easy*? Why?

The k-SAT threshold

Theorem (Friedgut *JAMS 1999*)

For each k there is a *threshold* r_k so that w.h.p.

- $\mathbf{\Phi}_k(n, m)$ is *satisfiable* if $m/n < r_k - \epsilon$,
- $\mathbf{\Phi}_k(n, m)$ is *unsatisfiable* if $m/n > r_k + \epsilon$.



Running time of "practical" algorithms is *exponential* and *peaks at* r_k.

The k-SAT threshold

Theorem (Friedgut *JAMS 1999*)

For each k there is a *threshold* r_k so that w.h.p.

- $\mathbf{\Phi}_k(n, m)$ is *satisfiable* if $m/n < r_k - \epsilon$,
- $\mathbf{\Phi}_k(n, m)$ is *unsatisfiable* if $m/n > r_k + \epsilon$.

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2$.

Proof

Non-constructive, via the 2nd moment method.

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (upper bound)

- Let $Z = \#\text{satisfying assignments}$.
- Then
$$\begin{aligned} \mathrm{E}Z &= \sum_{\sigma \in \{0,1\}^n} \mathrm{P}\left[\sigma \text{ satisfying}\right] = 2^n \cdot \mathrm{P}\left[\sigma_0 \text{ satisfying}\right] \\ &= 2^n \cdot \mathrm{P}\left[\sigma_0 \text{ satisfies one random clause}\right]^m \\ &= 2^n \cdot (1 - 2^{-k})^m. \end{aligned}$$

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (upper bound)

- Let $Z = \#$satisfying assignments.
- Then
$$\begin{aligned} \mathrm{E}Z &= \sum_{\sigma \in \{0,1\}^n} \mathrm{P}\left[\sigma \text{ satisfying}\right] = 2^n \cdot \mathrm{P}\left[\sigma_0 \text{ satisfying}\right] \\ &= 2^n \cdot \mathrm{P}\left[\sigma_0 \text{ satisfies one random clause}\right]^m \\ &= 2^n \cdot (1 - 2^{-k})^m. \end{aligned}$$
- Hence, if $r = m/n > 2^k \ln 2$, then
$$\frac{\ln \mathrm{E}Z}{n} \leq \ln 2 + \frac{m}{n} \ln(1 - 2^{-k}) \leq \ln 2 - r/2^k < 0.$$
- Consequently, $\mathrm{P}\left[Z > 0\right] = o(1)$.

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.
- $$\begin{aligned} \mathrm{E}(Z^2) &= \sum_{\sigma,\tau} \mathrm{P}\left[\text{both } \sigma, \tau \text{ satisfying}\right] \\ &= 2^n \mathrm{P}\left[\sigma_0 \text{ satisfying}\right] \sum_{\tau} \mathrm{P}\left[\tau \text{ satisfying} | \sigma_0 \text{ satisfying}\right] \\ &\sim 2^n \mathrm{P}\left[\sigma_0 \text{ satisfying}\right] \max_{\alpha \in [0,1]} \sum_{\tau : \langle \sigma_0, \tau \rangle = (1-\alpha)n} \mathrm{P}\left[\tau \text{ sat} | \sigma_0 \text{ sat}\right]. \end{aligned}$$

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.
- $$\begin{aligned} \mathrm{E}(Z^2) &= \sum_{\sigma,\tau} \mathrm{P}\left[\text{both } \sigma, \tau \text{ satisfying}\right] \\ &= 2^n \mathrm{P}\left[\sigma_0 \text{ satisfying}\right] \sum_{\tau} \mathrm{P}\left[\tau \text{ satisfying} | \sigma_0 \text{ satisfying}\right] \\ &\sim 2^n \mathrm{P}\left[\sigma_0 \text{ satisfying}\right] \max_{\alpha \in [0,1]} \sum_{\tau : \langle \sigma_0, \tau \rangle = (1-\alpha)n} \mathrm{P}\left[\tau \text{ sat} | \sigma_0 \text{ sat}\right]. \end{aligned}$$
- $\rightsquigarrow \frac{1}{n} \ln \mathrm{E}(Z^2) \sim \ln 2 + \max_\alpha h(\alpha) + r \ln \left[1 - 2^{1-k} + 2^{-k}(1-\alpha)^k\right]$.

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.
- $\frac{1}{n} \ln \mathrm{E}(Z^2) \sim \ln 2 + \max_\alpha h(\alpha) + r \ln \left[1 - 2^{1-k} + 2^{-k}(1-\alpha)^k\right]$.

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2$.

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.
- $\frac{1}{n} \ln \mathrm{E}(Z^2) \sim \ln 2 + \max_\alpha h(\alpha) + r \ln \left[1 - 2^{1-k} + 2^{-k}(1-\alpha)^k\right]$.
- We have $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$ iff the max is *attained at* $\alpha = 1/2$.

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2.$

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.
- $\frac{1}{n} \ln \mathrm{E}(Z^2) \sim \ln 2 + \max_\alpha h(\alpha) + r \ln \left[1 - 2^{1-k} + 2^{-k}(1-\alpha)^k\right]$.
- We have $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$ iff the max is *attained at* $\alpha = 1/2$.
- However, the max occurs at $\alpha < 1/2$:(

The second moment argument

Theorem (Achlioptas, Peres *JAMS 2004*)

$2^k \ln 2 - k \leq r_k \leq 2^k \ln 2$.

Proof (lower bound)

- Let $Z = \#$satisfying assignments.
- Goal: derive that $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$.
- $\frac{1}{n} \ln \mathrm{E}(Z^2) \sim \ln 2 + \max_\alpha h(\alpha) + r \ln \left[1 - 2^{1-k} + 2^{-k}(1-\alpha)^k\right]$.
- We have $\mathrm{E}(Z^2) \leq C \cdot \mathrm{E}(Z)^2$ iff the max is *attained at* $\alpha = 1/2$.
- However, the max occurs at $\alpha < 1/2$:(
- Fix: design a (slightly) *modified* variable $Z_b < Z \ldots$
- $\ldots$ and apply 2nd moment to Z_b.

The cavity method

Mertens, Mezard, Zecchina *RSA 2006*

- *Non-rigorous* computation of r_k for any $k \geq 3 \ldots$
- $\ldots$ via the *cavity method*.

The cavity method

Mertens, Mezard, Zecchina *RSA 2006*

- *Non-rigorous* computation of r_k for any $k \geq 3$...
- ...via the *cavity method*.

Franz, Leone *J Stat Phys 2003*

- These computations yield a *rigorous upper bound*...
- ...modulo 'population dynamics'.
- Pachenko, Talagrand *PTRF 2004*: generalized version.

The cavity method

Mertens, Mezard, Zecchina *RSA 2006*

- *Non-rigorous* computation of r_k for any $k \geq 3$...
- ...via the *cavity method*.

Franz, Leone *J Stat Phys 2003*

- These computations yield a *rigorous upper bound*...
- ...modulo 'population dynamics'.
- Pachenko, Talagrand *PTRF 2004*: generalized version.

Research problem

Determine r_k *rigorously*.

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

The factor graph

- $\mathbf{\Phi}$ =random k-CNF with n var and m clauses.
- Set up an *auxiliary graph*:
  - Vertices: variables and clauses.
  - Edges: connect each clause with the variables it contains.
- This is a *sparse* graph w/out *short cycles*.

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

Krzakala, Montanari, Ricci, Semerijan, Zdeborova *PNAS 2007*

Non-rigorous work, based on cavity method.

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

Krzakala, Montanari, Ricci, Semerijan, Zdeborova *PNAS 2007*

Non-rigorous work, based on cavity method.

Unique phase. $\mathrm{E} \sup_{x_\omega, y_\omega} \|\mu(\cdot|x_\omega) - \mu(\cdot|y_\omega)\| \to 0$.

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

Krzakala, Montanari, Ricci, Semerijan, Zdeborova *PNAS 2007*

Non-rigorous work, based on cavity method.

RS phase. $\mathrm{E} \sum_{x_\omega} \mu(x_\omega) \, \|\mu(\cdot|x_\omega) - \mu(\cdot)\| \to 0.$

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

Krzakala, Montanari, Ricci, Semerijan, Zdeborova *PNAS 2007*

Non-rigorous work, based on cavity method.

dynamic RSB. Gibbs measure shatters into *exponentially small* components.

Replica symmetry breaking

Question

- How does the Gibbs measure *evolve* as we increase m/n?
- Impact on *dynamics* (or *algorithms*)?

Krzakala, Montanari, Ricci, Semerijan, Zdeborova *PNAS 2007*

Non-rigorous work, based on cavity method.

Condensation. Gibbs measure concentrated on a *sub-exponential* number of components.

Replica symmetry breaking (rigorous)

Theorem (Achlioptas, ACO 2008)

For $r = m/n > 2^k \ln(k)/k$ the Gibbs measure

- shatters into exponentially small components,
- which are far apart in the Hamming cube,
- and separted by high-energy barriers.

Replica symmetry breaking (rigorous)

Theorem (Achlioptas, ACO 2008)

For $r = m/n > 2^k \ln(k)/k$ the Gibbs measure

- shatters into exponentially small components,
- which are far apart in the Hamming cube,
- and separated by high-energy barriers.

- Rigorous version of dRSB.
- Barrier for, e.g., Glauber dynamics.

Replica symmetry breaking (rigorous)

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

x is frozen if for any satisfying assignment τ

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

Replica symmetry breaking (rigorous)

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

x is liquid if there is a satisfying assignment τ s.t.

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

Replica symmetry breaking (rigorous)

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

x is liquid if there is a satisfying assignment τ s.t.

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

Theorem (Achlioptas, ACO 2008)

- If $m/n < 2^k \ln k/k$, then almost all variables are *liquid* w.h.p.
- If $m/n > 2^k \ln k/k$, then almost all variables are *frozen* w.h.p.

Replica symmetry breaking (rigorous)

Theorem (ACO, Gerke 2010+)

- If $m/n < 2^k \ln k/k$, then

$$\lim_{k,\omega,n\to\infty} \mathrm{E} \sum_{x_\omega} \mu(x_\omega) \, \|\mu(\cdot|x_\omega) - \mu(\cdot)\| = 0.$$

- If $m/n > 2^k \ln k/k$, then

$$\lim_{k,\omega,n\to\infty} \mathrm{E} \sum_{x_\omega} \mu(x_\omega) \, \|\mu(\cdot|x_\omega) - \mu(\cdot)\| = \frac{1}{2}.$$

Replica symmetry breaking (rigorous)

Theorem (ACO, Gerke 2010+)

- If $m/n < 2^k \ln k/k$, then

$$\lim_{k,\omega,n\to\infty} \mathrm{E} \sum_{x_\omega} \mu(x_\omega) \, \|\mu(\cdot|x_\omega) - \mu(\cdot)\| = 0.$$

- If $m/n > 2^k \ln k/k$, then

$$\lim_{k,\omega,n\to\infty} \mathrm{E} \sum_{x_\omega} \mu(x_\omega) \, \|\mu(\cdot|x_\omega) - \mu(\cdot)\| = \frac{1}{2}.$$

Aka reconstruction threshold.

Proof techniques

The Gibbs distribution

1. Generate a random formula $\mathbf{\Phi}$.
2. Choose a satisfying assignment σ uniformly.
3. *Result:* the pair $(\mathbf{\Phi}, \sigma)$.

Proof techniques

The Gibbs distribution

1. Generate a random formula $\mathbf{\Phi}$.
2. Choose a satisfying assignment σ uniformly.
3. *Result:* the pair $(\mathbf{\Phi}, \sigma)$.

The planted model

1. Generate a random assignment $\sigma : V = \{x_1, \ldots, x_n\} \to \{\text{true}, \text{false}\}$.
2. Choose m random clauses that are *satisfied by* $\sigma \rightsquigarrow \mathbf{\Phi}$.
3. *Result:* the pair $(\mathbf{\Phi}, \sigma)$.

Planted vs. uniform

Key lemma

There is $\xi_k \to 0$ such that *any* property that holds with probability $1 - \exp(-\xi_k n)$ in the *planted model* holds in the *Gibbs distribution* w.h.p.

Planted vs. uniform

Key lemma

There is $\xi_k \to 0$ such that *any* property that holds with probability $1 - \exp(-\xi_k n)$ in the *planted model* holds in the *Gibbs distribution* w.h.p.

Proof

- Study the number Z of satisfying assignments of $\mathbf{\Phi}$.
- *2nd moment method* + *sharp threshold result* yield

$$0 < \ln \mathrm{E} Z - \mathrm{E} \ln Z < \xi_k n.$$

- Hence, the planted model favors solution-rich formulas – but only "slightly".

Planted vs. uniform

Key lemma

There is $\xi_k \to 0$ such that *any* property that holds with probability $1 - \exp(-\xi_k n)$ in the *planted model* holds in the *Gibbs distribution* w.h.p.

Proof of the main result

- *Large deviations analysis* of a branching process $\Rightarrow$ *fluid* vars.
- *Discrepancy properties* of $\mathbf{\Phi}$ (+large deviations) $\Rightarrow$ *frozen* vars.

Planted vs. uniform (ctd.)

Corollary (Achlioptas, ACO 2008)

There is $\zeta_k > 0$ such that w.h.p.

$$Z < \exp(-\zeta_k n) \cdot \mathrm{E}Z.$$

Research Problem

Compute the *median* of Z.

Algorithms

Question

- The *satisfiability threshold* is $r_k \sim 2^k \ln 2$.
- *Dynamic RSB* occurs at $r = m/n \sim 2^k \ln(k)/k$.
- For what r can we *find* a satisfying assignment *efficiently*?

Algorithms

Question

- The *satisfiability threshold* is $r_k \sim 2^k \ln 2$.
- *Dynamic RSB* occurs at $r = m/n \sim 2^k \ln(k)/k$.
- For what r can we *find* a satisfying assignment *efficiently*?

Intuition (?)

- For $r = m/n < 2^k \ln(k)/k$ the problem should be *easy*:
  - correlations are purely local,
  - solutions are abundant.

Algorithms

Question

- The *satisfiability threshold* is $r_k \sim 2^k \ln 2$.
- *Dynamic RSB* occurs at $r = m/n \sim 2^k \ln(k)/k$.
- For what r can we *find* a satisfying assignment *efficiently*?

Intuition (?)

- For $r = m/n < 2^k \ln(k)/k$ the problem should be *easy*:
  - correlations are purely local,
  - solutions are abundant.
- For $r = m/n > 2^k \ln(k)/k$ things may become *hard*:
  - there are long-range correlations due to *frozen variables*,
  - the solution space resembles an error-correcting code.

Algorithms

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

Algorithms

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

Algorithm	*Density* $m/n < \cdots$	
Pure Literal	$o(1)$ as $k \to \infty$	Kim 2006
Walksat, rigorous	$\frac{1}{6} \cdot 2^k/k^2$	CFFKV 2009
Walksat, non-rigorous	$2^k/k$	Monasson 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
BP+decimation (non-rigorous)	$e \cdot 2^k/k$	Montanari 2007

Algorithms

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

In summary,

. . . *efficient algorithms* are known to succeed up to $m/n = c \cdot 2^k/k$.

Algorithms

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

In summary,

. . . *efficient algorithms* are known to succeed up to $m/n = c \cdot 2^k/k$,

Problem (Chvatal, Reed 1992; Achlioptas, Peres *JAMS 2004*)

Devise an algorithm that succeeds up to $m/n = 2^k \omega(k)/k$, $\omega(k) \to \infty$.

Algorithms for random k-SAT (ctd.)

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

Algorithms for random k-SAT (ctd.)

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

Theorem (ACO 2010^+)

An efficient algorithm `Fix` succeeds up to $m/n \sim 2^k \cdot \ln k/k$ w.h.p.

Algorithms for random k-SAT (ctd.)

Question

- *Dynamic RSB* threshold: $r = m/n \sim 2^k \ln(k)/k$.
- Can we *find* satisfying assignments up to *dRSB*?

Theorem (ACO 2010^+)

An efficient algorithm `Fix` succeeds up to $m/n \sim 2^k \cdot \ln k/k$ w.h.p.

- `Fix` is a deterministic *local search* algorithm.
- The *analysis* is via a blend of probabilistic and combinatorial methods.

Conclusion

- Random k-SAT is a *spin glass* problem (with a combinatorial flavor).
- There are many others such as *random graph coloring*, ...
- A plethora of open problems:
  - *precise* thresholds,
  - existence of the *condensation phase*,
  - *volume* of the set of satisfying assignments,
  - study of *dynamics*,
  - and the physicist's *algorithms* (Mezard, Parisi, Zecchina *Science 2002*),
  - *positive temperature* regime,
  - *general* methods.