Gibbs States and Message Passing Algorithms in Random k-SAT and Graphical Models

Andrea Montanari

Stanford University

May 3, 2007

Outline

Outline

Outline

Outline

1. Introduction
2. Pure state/cluster decomposition
3. Relation with Bethe-Peierls approximation
4. Relation with correlation decay
5. Message passing algorithms
6. Conclusion

Outline

1. Introduction
2. Pure state/cluster decomposition
3. Relation with Bethe-Peierls approximation
4. Relation with correlation decay
5. Message passing algorithms
6. Conclusion

Outline

Introduction

Structure of the presentation

Explore (some) interesting phenomena in random k-SAT

Infer general ideas (and some theorem) for a standard model

Ask whatever you want

Structure of the presentation

Explore (some) interesting phenomena in random k-SAT

Infer general ideas (and some theorem) for a standard model

Ask whatever you want

Structure of the presentation

Explore (some) interesting phenomena in random k-SAT

Infer general ideas (and some theorem) for a standard model

Ask whatever you want

Sources

On random k-SAT:

$\longrightarrow$ M. Mézard, G. Parisi, and R. Zecchina, 'Analytic and Algorithmic Solution of Random Satisfiability Problems', Science 2002

$\longrightarrow$ A. Montanari, D. Shah, 'Counting good truth assignments of random k-SAT formulae', SODA 2007

$\longrightarrow$ F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova 'Gibbs States and the Set of Solutions of Random Constraint Satisfaction Problems', PNAS 2007

Formalization:

$\longrightarrow$ A. Dembo and A.Montanari, *In preparation* [DM07]

k-satisfiability

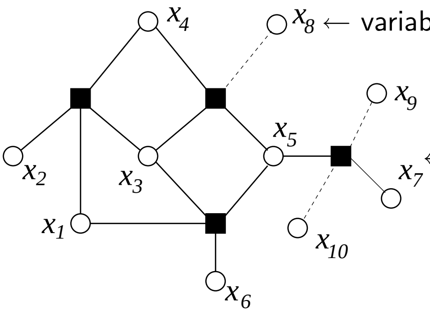
n variables: $\underline{x} = (x_1, x_2, \dots, x_n)$, $x_i \in \{0, 1\}$

m k-clauses

$$(x_1 \vee \overline{x_5} \vee x_7) \wedge (x_5 \vee x_8 \vee \overline{x_9}) \wedge \cdots \wedge (\overline{x_{66}} \vee \overline{x_{21}} \vee \overline{x_{32}})$$

Hereafter $k \geq 4$ (ask me why at the end)

Uniform measure over solutions



$\leftarrow$ variables $x_i \in \{0,1\}$

$\leftarrow$ clauses, e.g. $(x_5 \vee x_7 \vee \overline{x_9} \vee \overline{x_{10}})$

$$F = \cdots \wedge \underbrace{(x_{i_1(a)} \vee \overline{x}_{i_2(a)} \vee \cdots \vee x_{i_k(a)})}_{a\text{-th clause}} \wedge \cdots$$

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{a=1}^{M} \psi_a(x_{i_1(a)}, \ldots, x_{i_k(a)})$$

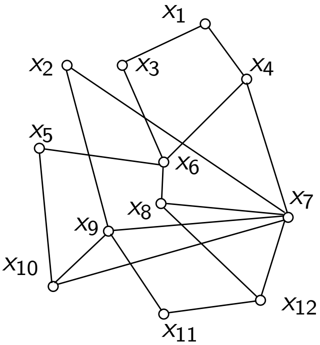
$\mu(\underline{x}) \Leftrightarrow$ Set of solutions $\mathcal{S}$

Random k-satisfiability

Each clause is uniformly random among the $2^k \binom{n}{k}$ possible ones.

$n, m \to \infty$ with $\alpha = m/n$ fixed.

'Standard model'



$$G = (V, E),\ V = [n],\ \underline{x} = (x_1, \dots, x_n),\ x_i \in \mathcal{X}$$

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{(ij) \in G} \psi_{ij}(x_i, x_j)\,.$$

'Standard model' (assumptions)

1. G has bounded degree.

2. G has girth larger than 2ℓ (with $\ell = \ell(n) \to \infty$).

3. $\psi_{\min} \le \psi_{ij}(x_i, x_j) \le \psi_{\max}$ uniformly.

Not *really* fulfilled by random k-SAT but . . .

'Standard model' (assumptions)

1. G has bounded degree.

2. G has girth larger than 2ℓ (with $\ell = \ell(n) \to \infty$).

3. $\psi_{\min} \le \psi_{ij}(x_i, x_j) \le \psi_{\max}$ uniformly.

Not *really* fulfilled by random k-SAT but ...

'Standard model' (assumptions)

1. G has bounded degree.

2. G has girth larger than 2ℓ
(with $\ell = \ell(n) \to \infty$).

3. $\psi_{\min} \leq \psi_{ij}(x_i, x_j) \leq \psi_{\max}$ uniformly.

Not *really* fulfilled by random k-SAT but ...

'Standard model' (assumptions)

1. G has bounded degree.

2. G has girth larger than 2ℓ
(with $\ell = \ell(n) \to \infty$).

3. $\psi_{\min} \le \psi_{ij}(x_i, x_j) \le \psi_{\max}$ uniformly.

Not *really* fulfilled by random k-SAT but . . .

Pure state/cluster decomposition

'Exponentially many clusters'

What does this mean?
[Mossel, Mézard/Palassini/Rivoire (2005), ...]

‘Exponentially many clusters’

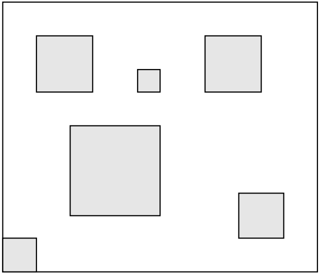
What does this mean?
[Mossel, Mézard/Palassini/Rivoire (2005), ...]

A toy model: Random sub-cubes in $\{0,1\}^n$

[from an idea by Dimitris Achlioptas]

$N = 2^{n\Sigma_0}$ clusters: $\quad \mathcal{S} = \cup_{a=1}^{N} \mathcal{S}_a$

$\{\mathcal{S}_a\}$ iid cubes with 'centers' $\underline{x}^{(a)} \in \{0,1,*\}^n$:

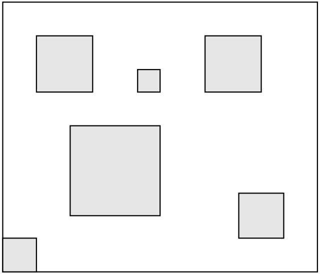


A toy model: Random sub-cubes in $\{0,1\}^n$

[from an idea by Dimitris Achlioptas]

$N = 2^{n\Sigma_0}$ clusters: $\quad \mathcal{S} = \cup_{a=1}^{N} \mathcal{S}_a$

$\{\mathcal{S}_a\}$ iid cubes with 'centers' $\underline{x}^{(a)} \in \{0,1,*\}^n$:

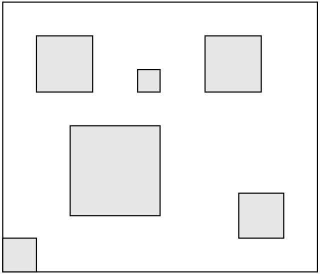


A toy model: Random sub-cubes in $\{0,1\}^n$

[from an idea by Dimitris Achlioptas]

$N = 2^{n\Sigma_0}$ clusters: $\quad \mathcal{S} = \cup_{a=1}^{N} \mathcal{S}_a$

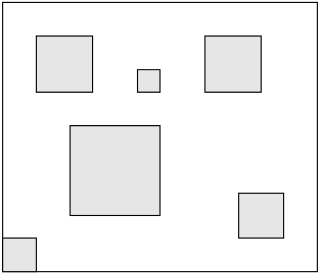
$\{\mathcal{S}_a\}$ iid cubes with 'centers' $\underline{x}^{(a)} \in \{0, 1, *\}^n$:



How shall I construct one cluster?

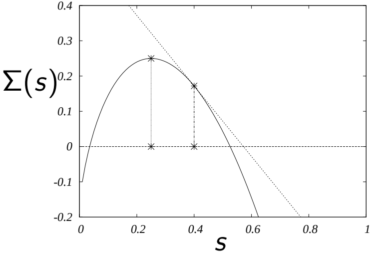
$$\mathcal{S}_a = \left\{ \underline{x} \in \{0,1\}^n : \ x_i \ = \ x_i^{(a)} \right\}$$

$$x_i^{(a)} = \begin{cases} * & \text{prob } p \ , \\ 1 & \text{prob } (1-p)/2 \ , \\ 0 & \text{prob } (1-p)/2 \ , \end{cases}$$



Most of clusters have size 2^{np}, but...

$$\#\{\text{clusters of size } 2^{ns}\} \doteq 2^{n\Sigma(s)}$$
$$\Sigma(s) = \Sigma_0 - D(s||p) \qquad \text{if} \geq 0 \text{ and...}$$

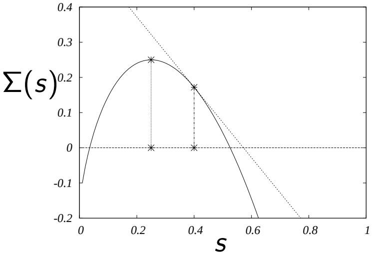


(d1RSB)

Most of solutions are in $2^{n\Sigma(s_*)}$ clusters of size 2^{ns_*}, $s_* > p$.

Most of clusters have size 2^{np}, but...

$$\#\{\text{clusters of size } 2^{ns}\} \doteq 2^{n\Sigma(s)}$$
$$\Sigma(s) = \Sigma_0 - D(s||p) \qquad \text{if} \geq 0 \text{ and}\ldots$$



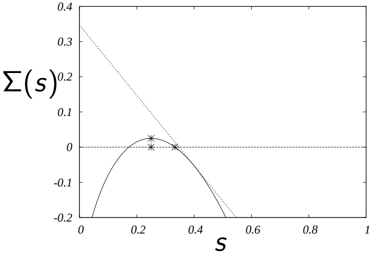
(d1RSB)

Most of solutions are in $2^{n\Sigma(s_*)}$ clusters of size 2^{ns_*}, $s_* > p$.

Most of clusters have size 2^{np}, but...

$$\#\{\text{clusters of size } 2^{ns}\} \doteq 2^{n\Sigma(s)}$$

$$\Sigma(s) = \Sigma_0 - D(s||p) \quad \text{ if } \geq 0 \text{ and}\ldots$$



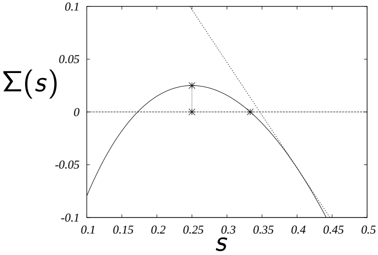
(1RSB)

Most of solutions are in $2^{o(n)}$ clusters of size $2^{ns_{\max}}$, $s_{\max} \in (p, s_*)$.

Most of clusters have size 2^{np}, but...

$$\#\{\text{clusters of size } 2^{ns}\} \doteq 2^{n\Sigma(s)}$$

$$\Sigma(s) = \Sigma_0 - D(s||p) \qquad \text{if} \geq 0 \text{ and} \ldots$$



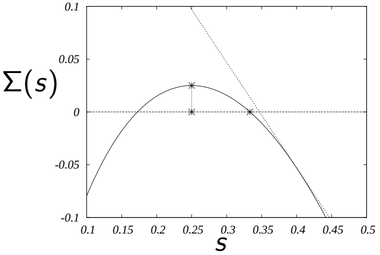
(1RSB)

Most of solutions are in $2^{o(n)}$ clusters of size $2^{ns_{\max}}$, $s_{\max} \in (p, s_*)$.

Most of clusters have size 2^{np}, but...

$$\#\{\text{clusters of size } 2^{ns}\} \doteq 2^{n\Sigma(s)}$$

$$\Sigma(s) = \Sigma_0 - D(s||p) \quad \text{if} \geq 0 \text{ and}\ldots$$

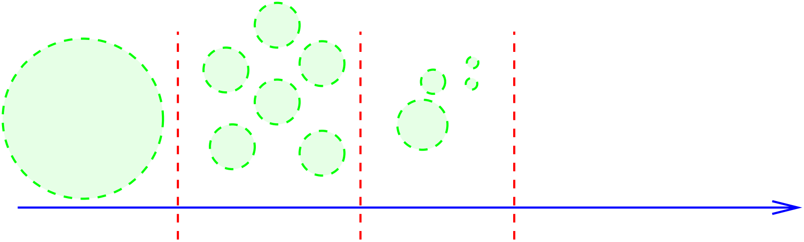


(1RSB)

Most of solutions are in $2^{o(n)}$ clusters of size $2^{ns_{\max}}$, $s_{\max} \in (p, s_*)$.

Enough with toys. . .

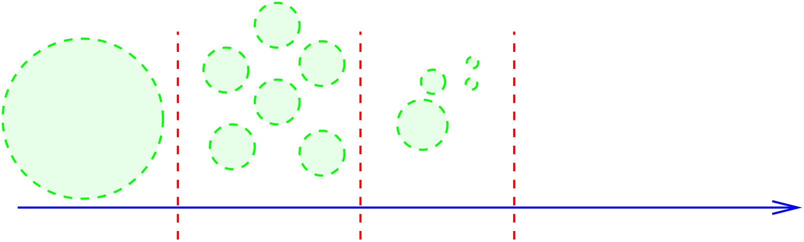
Pure states decomposition in k-SAT



$\alpha_{\rm d}(k)$ $\alpha_{\rm c}(k)$ $\alpha_{\rm s}(k)$

[Biroli et al. 01, Mézard et al. 02, Mézard et al. 05, Achlioptas et al. 06, KMRSZ (us) 06]

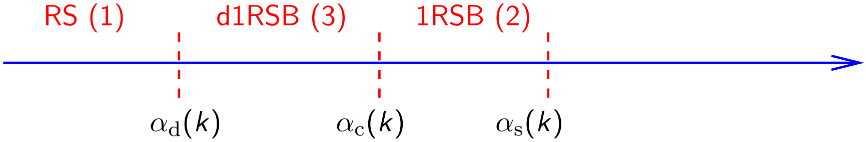
The 3 scenarios seem universal (coloring, codes, ...)



$\alpha_{\rm d}(k)$ $\alpha_{\rm c}(k)$ $\alpha_{\rm s}(k)$

[Biroli et al. 01, Mézard et al. 02, Mézard et al. 05, Achlioptas et al. 06, KMRSZ (us) 06]

The 3 scenarios seem universal (coloring, codes, . . .)



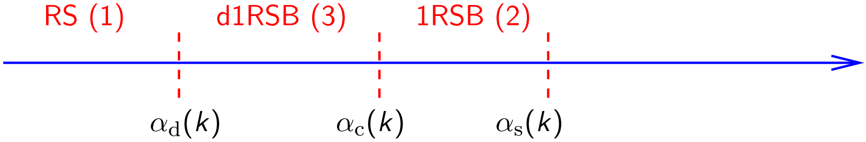
$$\alpha_{\rm d}(k) = (2^k \log k)/k + \dots \qquad (\alpha_{\rm d}(4) \approx 9.38)$$

$$\alpha_{\rm c}(k) = 2^k \log 2 - \tfrac{3}{2} \log 2 + \dots \qquad (\alpha_{\rm c}(4) \approx 9.547)$$

$$\alpha_{\rm s}(k) = 2^k \log 2 - \tfrac{1}{2}(1 + \log 2) + \dots \qquad (\alpha_{\rm s}(4) \approx 9.93)$$

[Achlioptas, Naor, Peres, 2005, $\alpha_{\rm s}(k) = 2^k \log 2 + O(k)$]

Pure states decomposition in k-SAT



$$\alpha_{\rm d}(k) = (2^k \log k)/k + \dots \qquad (\alpha_{\rm d}(4) \approx 9.38)$$

$$\alpha_{\rm c}(k) = 2^k \log 2 - \frac{3}{2}\log 2 + \dots \qquad (\alpha_{\rm c}(4) \approx 9.547)$$

$$\alpha_{\rm s}(k) = 2^k \log 2 - \frac{1}{2}(1+\log 2) + \dots \qquad (\alpha_{\rm s}(4) \approx 9.93)$$

[Achlioptas, Naor, Peres, 2005, $\alpha_{\rm s}(k) = 2^k \log 2 + O(k)$]

Howx to formalize this in general?

Definition

It is the 'finer' partition $\Omega_1 \cup \cdots \cup \Omega_N = \mathcal{X}^n$, such that

$$\frac{\mu(\partial_\epsilon \Omega_q)}{(1-\mu(\Omega_q))\mu(\Omega_q)} \leq \exp\{-C(\epsilon)n\}\,.$$

where $C(\epsilon) > 0$ for ϵ small enough.

[the conductance of μ is exponentially small]

$$\mu(\,\cdot\,) = \sum_{q=1}^{N} w_q \mu_q(\,\cdot\,)\,.$$

Howx to formalize this in general?

Definition

It is the 'finer' partition $\Omega_1 \cup \cdots \cup \Omega_N = \mathcal{X}^n$, such that

$$\frac{\mu(\partial_\epsilon \Omega_q)}{(1-\mu(\Omega_q))\mu(\Omega_q)} \le \exp\{-C(\epsilon)n\}\,.$$

where $C(\epsilon) > 0$ for ϵ small enough.

[the conductance of μ is exponentially small]

$$\mu(\,\cdot\,) = \sum_{q=1}^{N} w_q \mu_q(\,\cdot\,)\,.$$

Pure states: 3 scenarios

Let $N(\delta)$ the minimal number of states with measure $\geq 1 - \delta$

[RS] $N(\delta) = 1$

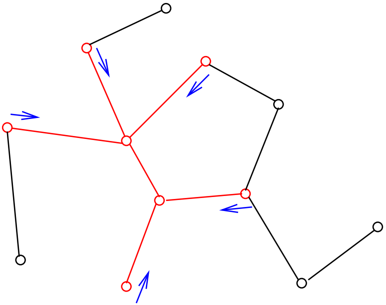
[d1RSB] $N(\delta) = e^{n(\Sigma \pm \varepsilon)}$

[1RSB] $N(\delta) = \Theta(1)$ [$\rightarrow$ unbounded random variable]

Relation with Bethe-Peierls approximation

Definition

A 'set of messages' (aka cavity fields) is a collection $\{\nu_{i\to j}(\,\cdot\,)\}$ *indexed by directed edges in* G*, where* $\nu_{i\to j}(\,\cdot\,)$ *is a distribution over* $\mathcal{X}$*.*



Given $F \subseteq G$, $\mathrm{diam}(F) \le 2\ell$, such that $\deg_F(i) = \deg_G(i)$ or ≤ 1

$$\nu_F(\underline{x}_F) \equiv \frac{1}{W(\nu_F)} \prod_{(ij)\in F} \psi_{ij}(x_i, x_j) \prod_{i\in\partial F} \nu_{i\to j(i)}(x_i)\,.$$

Definition

A probability distribution ρ *on* $\mathcal{X}^V$ *is an* (ε, r) *Bethe state, if there exists a set of messages* $\{\nu_{i\to j}(\,\cdot\,)\}$ *such that, for any* $F \subseteq G$ *with* $\mathrm{diam}(F) \le 2r$

$$||\rho_F - \nu_F||_{TV} \le \varepsilon\,.$$

Proposition (DM07)

If ρ is a $(\varepsilon, 2)$-Bethe state with respect to the message set $\{\nu_{i\to j}(\,\cdot\,)\}$, then, for any $i \to j$

$$||\nu_{i\to j} - \mathrm{T}\nu_{i\to j}||_{TV} \le C\varepsilon\,,$$

$$\mathrm{T}\nu_{i\to j}(x_i) = \frac{1}{z_{i\to j}} \prod_{l\in\partial i\setminus j} \sum_{x_l} \psi_{il}(x_i, x_l)\nu_{l\to i}(x_l)\,.$$

Belief Propagation

For $t = 0, 1, \ldots$

$$\nu^{(t+1)}_{i\to j} = \mathrm{T}\nu^{(t)}_{i\to j}$$

Consistency Condition $\longrightarrow$ Bethe Equations

Proposition (DM07)

If ρ *is a* $(\varepsilon, 2)$*-Bethe state with respect to the message set* $\{\nu_{i\to j}(\,\cdot\,)\}$*, then, for any* $i \to j$

$$||\nu_{i\to j} - \mathrm{T}\nu_{i\to j}||_{TV} \le C\varepsilon\,,$$

$$\mathrm{T}\nu_{i\to j}(x_i) = \frac{1}{z_{i\to j}} \prod_{l\in\partial i\setminus j} \sum_{x_l} \psi_{il}(x_i, x_l)\nu_{l\to i}(x_l)\,.$$

Belief Propagation

For $t = 0, 1, \ldots$

$$\nu^{(t+1)}_{i\to j} = \mathrm{T}\nu^{(t)}_{i\to j}$$

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{(ij)\in G} \psi_{ij}(x_i, x_j)\,.$$

[consider a sequence of models with $n \to \infty$]

(RS) $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.

(1RSB) $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states ($\leftrightarrow$ clusters).

(d1RSB) $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states($\leftrightarrow$ clusters).

3 Scenarios

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{(ij)\in G} \psi_{ij}(x_i, x_j)\,.$$

[consider a sequence of models with $n \to \infty$]

(RS) $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.

(1RSB) $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states ($\leftrightarrow$ clusters).

(d1RSB) $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states($\leftrightarrow$ clusters).

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{(ij)\in G} \psi_{ij}(x_i, x_j)\,.$$

[consider a sequence of models with $n \to \infty$]

(RS) $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.

(1RSB) $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states ($\leftrightarrow$ clusters).

(d1RSB) $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states($\leftrightarrow$ clusters).

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{(ij)\in G} \psi_{ij}(x_i, x_j)\,.$$

[consider a sequence of models with $n \to \infty$]

(RS) $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.

(1RSB) $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states ($\leftrightarrow$ clusters).

(d1RSB) $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states($\leftrightarrow$ clusters).

Relation with correlation decay

Relation with correlation decay: Notation

- $i \in \{1, \dots, N\}$ uniformly at random.
- $\mathrm{B}(i, r)$ ball of radius r and center i.
- $x_{\sim i,r} = \{\, x_j \ : \ j \notin \mathrm{B}(i, r) \,\}$.

Relation with correlation decay: Definitions

Uniqueness:

$$\sup_{x,x'} \sum_{x_i} \left|\mu(x_i|x_{\sim i,r}) - \mu(x_i|x'_{\sim i,r})\right| \to 0$$

[cf. Tatikonda, Gamarnik, Bayati,. . .]

Extremality:

$$\sum_{x_i, x_{\sim i,\ell}} |\mu(x_i, x_{\sim i,r}) - \mu(x_i)\mu(x_{\sim i,r})| \to 0$$

[cf. Peres, Mossel]

Concentration:

$$\sum_{x_{i(1)}\cdots x_{i(k)}} \left|\mu(x_{i(1)}, \dots, x_{i(k)}) - \mu(x_{i(1)}) \cdots \mu(x_{i(k)})\right| \to 0$$

Relation with correlation decay

RS $\Leftrightarrow$ Extremality

d1RSB $\Leftrightarrow$ No extremality; Concentration

1RSB $\Leftrightarrow$ No extremality; No concentration

[First rigorous under a suitable (WEAK) interpretation of two sides]

Relation with correlation decay

RS $\Leftrightarrow$ Extremality

d1RSB $\Leftrightarrow$ No extremality; Concentration

1RSB $\Leftrightarrow$ No extremality; No concentration

[First rigorous under a suitable (WEAK) interpretation of two sides]

Relation with correlation decay

RS $\Leftrightarrow$ Extremality

d1RSB $\Leftrightarrow$ No extremality; Concentration

1RSB $\Leftrightarrow$ No extremality; No concentration

[First rigorous under a suitable (WEAK) interpretation of two sides]

Relation with correlation decay

RS $\Leftrightarrow$ Extremality

d1RSB $\Leftrightarrow$ No extremality; Concentration

1RSB $\Leftrightarrow$ No extremality; No concentration

[First rigorous under a suitable (WEAK) interpretation of two sides]

First steps

Theorem (Tatikonda-Jordan 02)

If μ is unique 'with rate $\delta(\,\cdot\,)$' then it is an (ε, r) Bethe state for any $r < \ell$ and $\varepsilon \geq C\delta(\ell - r)$, with respect to the message set output by belief propagation.

Theorem (DM07)

If μ is extremal 'with rate $\delta(\,\cdot\,)$' then it is an (ε, r) Bethe state for any $r < \ell$ and $\varepsilon \geq C\delta(\ell - r)$.

First steps

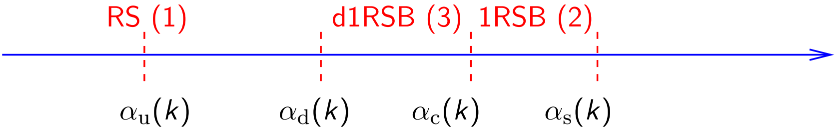
Theorem (Tatikonda-Jordan 02)

If μ is unique 'with rate $\delta(\,\cdot\,)$' then it is an (ε, r) Bethe state for any $r < \ell$ and $\varepsilon \geq C\delta(\ell - r)$, with respect to the message set output by belief propagation.

Theorem (DM07)

If μ is extremal 'with rate $\delta(\,\cdot\,)$' then it is an (ε, r) Bethe state for any $r < \ell$ and $\varepsilon \geq C\delta(\ell - r)$.

What happens in k-SAT?



$\alpha_{\rm u}(k) = (2\log k)/k + \dots$ [rigorous!, MS07]

$\alpha_{\rm d}(k) = (2^k \log k)/k + \dots$ $(\alpha_{\rm d}(4) \approx 9.38)$

$\alpha_{\rm c}(k) = 2^k \log 2 - \frac{3}{2}\log 2 + \dots$ $(\alpha_{\rm c}(4) \approx 9.547)$

$\alpha_{\rm s}(k) = 2^k \log 2 - \frac{1}{2}(1 + \log 2) + \dots$ $(\alpha_{\rm s}(4) \approx 9.93)$

Message passing algorithms

BP can work in the RS and d1RSB regimes.

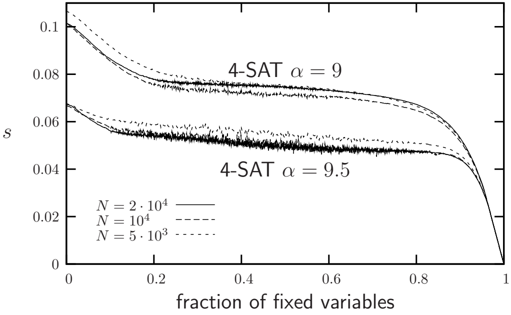
BP cannot work in the 1RSB regime.

Implications

BP can work in the RS and d1RSB regimes.

BP cannot work in the 1RSB regime.

Sequential BP search



Finds a solution with positive probability for $\alpha < \alpha_c(k)$.

Conclusion

- Many (difficult!) open problems.
- Theory of Gibbs measures (locally Markov processes) on (a class of) *finite* graphs.