Belief Propagation, Cavity Method and Pure Gibbs States in Combinatorial Problems: a (Personal) Survey

Andrea Montanari

Stanford University

March 21, 2007

Outline

Outline

Outline

Outline

Outline

1. Introduction
2. Bethe-Peierls approximation
3. Generic scenarios
   - Relation with correlation decay
   - Relation with pure state/cluster decomposition
4. The cavity method with many pure states
5. Conclusion

Outline

Outline

Introduction

Sources

On the cavity method:
→ M. Mézard and G. Parisi, 'The Bethe lattice spin glass revisited'

On random k-SAT:
→ M. Mézard, G. Parisi, and R. Zecchina, 'Analytic and Algorithmic Solution of Random Satisfiability Problems'
→ F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova 'Gibbs States and the Set of Solutions of Random Constraint Satisfaction Problems'

Formalization:
→ A. Dembo and A.Montanari, *In preparation* [DM07]

General:
→ M. Mézard and A. Montanari, *Upcoming book (check online)*
→ google ee374

Structure of the presentation

Discuss general ideas on a standard model

Check relevance/meaning on random k-SAT

Ask whatever you want

Forgive me if I'll not explain everything is interesting

Structure of the presentation

Discuss general ideas on a standard model

Check relevance/meaning on random k-SAT

Ask whatever you want

Forgive me if I'll not explain everything is interesting

Structure of the presentation

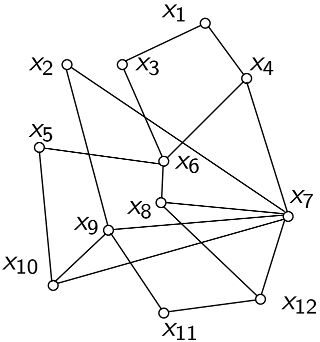
Discuss general ideas on a standard model

Check relevance/meaning on random k-SAT

Ask whatever you want

Forgive me if I'll not explain everything is interesting

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

'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.

‘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.

k-satisfiability

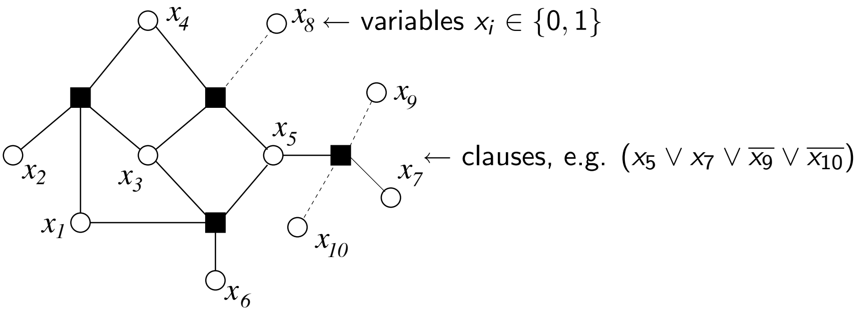
n variables: $\underline{x} = (x_1, x_2, \dots, x_n)$, $x_i \in \{0, 1\}$

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



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

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

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.

Does not *really* satisfy assumptions 1-3 above but . . .

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.

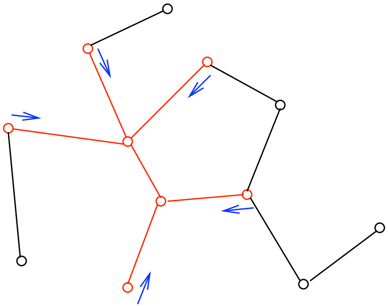
Does not *really* satisfy assumptions 1-3 above but . . .

Bethe-Peierls approximation

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

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

Bethe states

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* $\mathrm{diam}(F) \le 2r$

$$||\rho_U - \nu_U||_{TV} \le \varepsilon\,.$$

Consistency Condition $\rightarrow$ Bethe Equations

Proposition (DM07)

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

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

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

Belief Propagation

For $t = 0, 1, \ldots$

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

Consistency Condition $\rightarrow$ 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 $i \to j$

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

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

Belief Propagation

For $t = 0, 1, \ldots$

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

Generic scenarios

Generic Scenarios

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

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

1. $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.
['replica symmetric - RS']

2. $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states.
['one-step replica symmetry breaking - 1RSB']

3. $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states.
['dynamical' 1RSB]

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

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

1. $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.
['replica symmetric - RS']

2. $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states.
['one-step replica symmetry breaking - 1RSB']

3. $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states.
['dynamical' 1RSB]

Generic Scenarios

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

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

1. $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.
['replica symmetric - RS']

2. $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states.
['one-step replica symmetry breaking - 1RSB']

3. $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states.
['dynamical' 1RSB]

Generic Scenarios

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

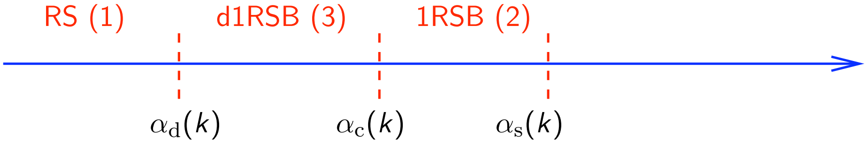
[consider a sequence of models with $n \to \infty$]

1. $\mu(\,\cdot\,)$ is a Bethe state and cannot be further decomposed.
['replica symmetric - RS']

2. $\mu(\,\cdot\,)$ is not a Bethe state but is a convex combination of Bethe states.
['one-step replica symmetry breaking - 1RSB']

3. $\mu(\,\cdot\,)$ is a Bethe state but can also be decomposed as a convex combination of Bethe states.
['dynamical' 1RSB]

What happens in k-SAT?



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

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

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

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. Gamarnik, Nair, Tatikonda. . .]

Extremality:

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

[cf. Roch, Vera. . .]

Concentration:

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

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 (DM07)

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

Theorem (Tatikonda-Jordan 02)

If μ is unique 'with rate $\delta(\,\cdot\,)$' then it 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.

First steps

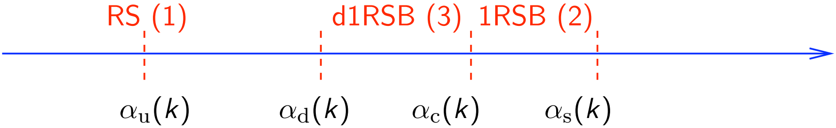
Theorem (DM07)

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

Theorem (Tatikonda-Jordan 02)

If μ is unique 'with rate $\delta(\,\cdot\,)$' then it 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.

What happens in k-SAT?



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

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

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

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

Relation with pure state/cluster decomposition

Pure states/cluster decomposition

Definition

It is a 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.

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

The $\mu_q(\,\cdot\,)$ are Bethe states.

Pure states: Generic scenarios

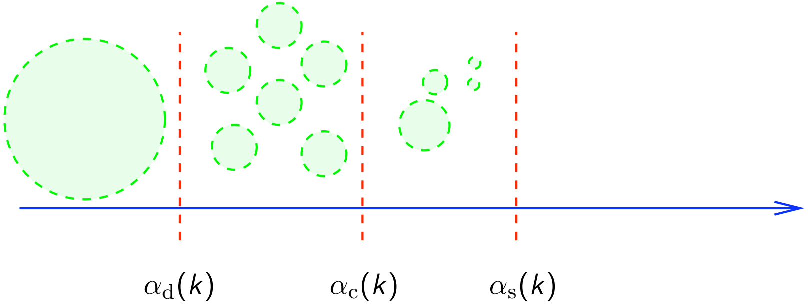
Let $N(\delta)$ the minimal number of states with measure $\geq 1-\delta$

RS$\Rightarrow$ $\quad N(\delta) = 1$

d1RSB$\Rightarrow$ $\quad N = e^{n(\Sigma \pm \varepsilon)}$

1RSB$\Rightarrow$ $\quad N(\delta) = \Theta(1) \quad [\to$ unbounded random variable]

Pure states decomposition in k-SAT



[cf. Mora, Achlioptas]

The cavity method with many pure states

Many Bethe states (d1RSB)

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

with $\mu_q(\,\cdot\,)$ Bethe wrt message set $\{\nu^{(q)}_{i\to j}\}$

Let $\{\nu_{i\to j}\}$ be the random message set defined by

$$\{\nu_{i\to j}\} = \{\nu^{(q)}_{i\to j}\} \quad \text{with probability } w_q.$$

for $q = 1, \dots, N$, and $\mathrm{M}(\nu)$ denote its distribution.

Many Bethe states (d1RSB)

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

with $\mu_q(\,\cdot\,)$ Bethe wrt message set $\{\nu^{(q)}_{i\to j}\}$

Let $\{\nu_{i\to j}\}$ be the random message set defined by

$$\{\nu_{i\to j}\} = \{\nu^{(q)}_{i\to j}\} \quad \text{with probability } w_q.$$

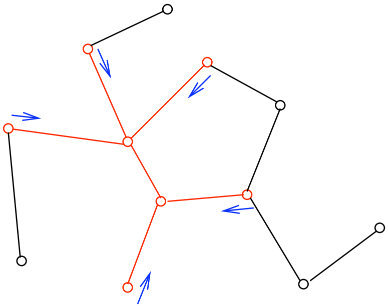
for $q = 1, \dots, N$, and $\mathrm{M}(\nu)$ denote its distribution.

The structure of M(ν): 1RSB messages

Definition

A 'set of 1RSB messages' is a collection $\{Q_{i\to j}(\,\cdot\,)\}$ *indexed by directed edges in* G*, where* $Q_{i\to j}(\,\cdot\,)$ *is a distribution over the set of probability measures over* $\mathcal{X}$*.*

The structure of M(ν): induced distribution



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

$$Q_U(\nu_U) \equiv \frac{1}{Z_U} \prod_{i \to j} \mathbb{I}(\nu_{i\to j} = \mathrm{T}\nu_{i\to j})\, W(\nu_U) \prod_{i \in \partial F} Q_{i\to j(i)}(\nu_{i\to j(i)})\,.$$

. . . where

$$W(\nu_F) = \prod_{(ij)\in F} W_{ij}(\nu_{i\to j}, \nu_{j\to i}) \prod_{i\in F} W_i(\{\nu_{l\to i}, l\in\partial i\})$$

is the partition function on F with b.c. $\{\nu_{i\to j(i)}\}$
[log $W(\nu_F)$ is the Bethe free energy]

1RSB Bethe States

Almost a Definition

A probability distribution $\mathrm{M}(\nu)$ is an (ϵ, r) 1RSB Bethe state, if there exists a set of 1RSB messages $\{Q_{i\to j}(\,\cdot\,)\}$ such that, for any $F \subseteq G$ with $\mathrm{diam}(F) \le 2r$

$$||\mathrm{M}_U - Q_U||_{\mathrm{TV}} \le \varepsilon\,.$$

1RSB consistency equations

$$Q_{i\to j}(\,\cdot\,) \propto \int z\{\nu_{l\to i}\}\, \mathbb{I}(f(\nu_{l\to i}) \in \,\cdot\,) \prod_{l\in\partial i\setminus j} \mathrm{d}Q_{l\to i}(\nu_{l\to i})$$

Formally $Q_{i\to j} = \mathrm{T}^* Q_{i\to j}$

General Survey Propagation/1RSB Message Passing

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

[1RSB bounds, cf. Franz]

1RSB consistency equations

$$Q_{i\to j}(\,\cdot\,) \propto \int z\{\nu_{l\to i}\}\,\mathbb{I}(f(\nu_{l\to i}) \in \,\cdot\,) \prod_{l\in\partial i\setminus j} \mathrm{d}Q_{l\to i}(\nu_{l\to i})$$

Formally $Q_{i\to j} = \mathrm{T}^* Q_{i\to j}$

General Survey Propagation/1RSB Message Passing

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

[1RSB bounds, cf. Franz]

Conclusion

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