Lecture 10 : Bethe free energy and its relation to Belief Propagation.

1. Introduction.

We want to compute the free energy and phase transition thresholds of graphical models such as those occuring in Coding & K-SAT problem. Note that for Coding this corresponds to analyze MAP decoding. For K-SAT this corresponds to analyzing the SAT-UNSAT threshold behavior.

For general graphical models (or statistical mechanics models) this is an impossible task. An important approximation philosophy is the so-called "mean-field theory". For models defined on sparse graphs, that are localy tree-like, a form of this theory developped by Bethe & Peierls is very well adapted, and will be developped in this lecture. Note that this is already a "sophisticated" version of the most basic mean field theory

. As we will see the Bethe-Peierls theory involves fixed point equations that are the same as those occuring in Belief-Propagation. Their use and to some extent interpretation are however different. Note that there is clash of initials (BP) that is solely due to an historical accident and may create a lot of confusion.

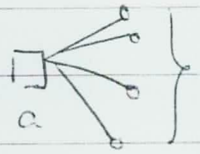
There exist class of models for which the "mean field theory" (a priori an approximate method) gives the exact solution. We have already seen such a model namely the CW model. Such models are commonly called "mean-field models". Such models are important because they offer a useful guide for the developpement of more ambitious mean field approximations for more complicated models.

It is conjectured, and partly proven, that the Bethe-Peierls mean field theory, is exact for LDPC codes on BMS channels. In other words it allows to correctly calculate the MAP threshold and performance curves. Since the Bethe-Peierls theory involves the same fixed point equations than Belief-Propagation a consequence is that there is a very intimate connection between MAP and BP decoding. This will be explained further in this lecture. We immediately stress that this connection does not make the decoding task any easier in the noise region $\epsilon_{BP} < \epsilon < \epsilon_{MAP}$.

<u>Concerning K-SAT</u> we will see that the pure Bethe-Peierls <u>method is not exact</u>. To improve upon it one has to go one-step further and develop an even more sophisticated mean field theory called "cavity method" or "replica method". This will be undertaken in the last two lectures of this course.

2. Free energy of graphical models on trees.

Take $\mu(s_1 \ldots s_N) = \frac{1}{Z} \prod_a f_a(s_{\partial a})$ over a tree



$s_{\partial a}$.

We want to express $f = -\frac{1}{N} \ln Z$ as a functional of the marginals as for the CW model.

As we will see the marginals that are involved are

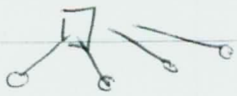
$$\mu_i(s_i) = \sum_{\sim s_i} \mu(s_1 \ldots s_N) \quad \& \quad \mu_a(s_{\partial a}) = \sum_{\sim s_{\partial a}} \mu(s_1 \ldots s_N).$$

Claim 1: on a tree we have

$$\mu(s_1 \ldots s_N) = \prod_a \mu_a(s_{\partial a}) \prod_i (\mu_i(s_i))^{1-di}.$$

where di = degree of node i.

Proof: By induction over number M of clauses a.

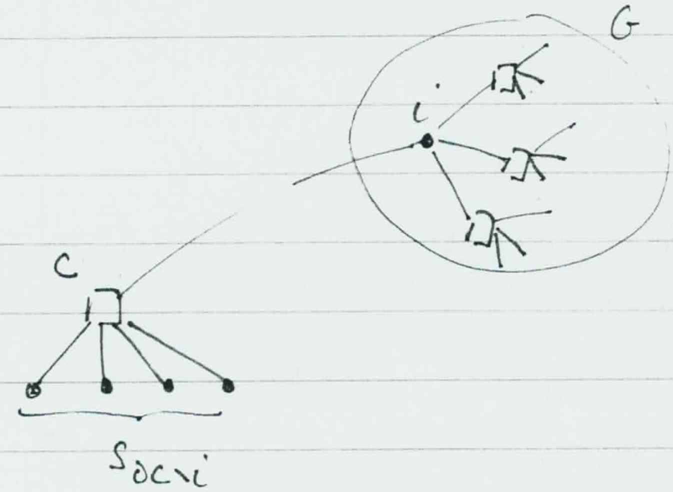
- $M = 1$ clause:  $\mu(s_{\partial a}) = \mu_a(s_{\partial a}) \mu_i(s_i)^{1-1}$ trivially.

- Induction hypothesis: for any tree like graphical model with M clauses we have the factorization

$$\mu(s_1 \dots s_N) = \prod_a \mu_a(s_{\partial a}) \prod_i \mu_i(s_i)^{1-d_i}$$

where $\mu_a(s_{\partial a})$ & $\mu_i(s_i)$ are the marginal of $\mu(s_1 \dots s_N)$.

- Add one clause in such a way that new model with $M+1$ clauses is a tree.



G contains variables $s_1 \dots s_N$.

(cases where $s_{\partial c \setminus i}$ are absent or where [clause] is not connected to old graph are treated in the same way than what follows).

$$\text{Prob}(s_{\partial c \setminus i}, s_1 \dots s_N) = \text{Prob}(s_{\partial c \setminus i} | s_1 \dots s_N)\, \text{Prob}(s_1 \dots s_N)$$

Here Prob is computed with new graphical model measure that is

$$\mu_{new}(s_{\partial c \setminus i}, s_1 \dots s_N) = \frac{1}{Z_{new}} f_c(s_{\partial c}) \prod_a f_a(s_{\partial a})$$

Now we have $\text{Prob}(s_{\partial c \setminus i} \mid s_1 \ldots s_N) = \text{Prob}(s_{\partial c \setminus i} \mid s_i)$.

$$= \frac{\text{Prob}(s_{\partial c})}{\text{Prob}(s_i)} = \frac{\mu_c^{new}(s_{\partial c})}{\mu_i^{new}(s_i)}.$$

So :

$$\mu_{new} = \mu_c^{new}(s_{\partial c}) \left(\mu_i^{new}(s_i)\right)^{-1} \text{Prob}(s_1 \ldots s_N).$$

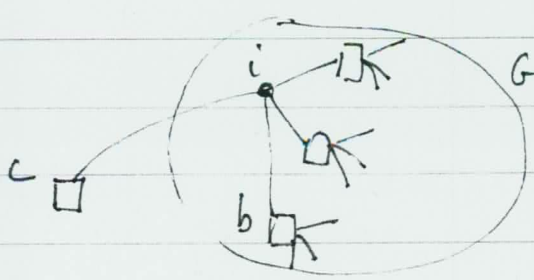
$\text{Prob}(s_1 \ldots s_N)$ is the marginalization of new model with respect to variables $s_{\partial c \setminus i}$. Indeed $\sum_{s_{\partial c \setminus i}} \text{Prob}(s_{\partial c \setminus i} \mid s_i) = 1$.

Thus

$$\text{Prob}(s_1 \ldots s_N) = \frac{1}{Z_{new}} \underbrace{\sum_{s_{\partial c \setminus i}} f_c(s_{\partial c})}\ \prod_a f_a(s_{\partial a}).$$

$$\equiv \frac{1}{Z_{new}} \tilde{f}_c(s_i) \prod_a f_a(s_{\partial a}).$$

This distribution corresponds to a graphical model of the type



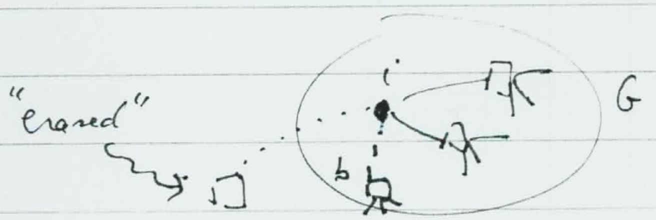
This model still has $M+1$ clauses. However clause c can be absorbed in clause b:

$$\mathrm{Prob}(s_1 \dots s_N) = \frac{1}{Z_{new}} \tilde{f}_c(s_i) \prod_a f_a(s_{\partial a})$$

$$= \frac{1}{Z_{new}} \tilde{f}_c(s_i)\, f_b(s_{\partial b}) \prod_{a \neq b} f_a(s_{\partial a})$$

$$= \frac{1}{Z_{new}} \tilde{\tilde{f}}_b(s_{\partial b}) \prod_{a \neq b} f_a(s_{\partial a})$$

This graphical model has M clauses



So we can apply to it the induction hypothesis

$$\Rightarrow \quad \mathrm{Prob}(s_1 \dots s_N) = \prod_a \mu_a(s_{\partial a}) \prod_i \left(\mu_i(s_i)\right)^{1-d_i}$$

Here μ_a & μ_i are the marginals of $\mathrm{Prob}(s_1 \dots s_N)$. But these marginal are also those of $\mu_{new}(s_{\partial c \setminus i}, s_1 \dots s_N)$. So in the above formula we have $\mu_a = \mu_a^{new}$ & $\mu_i = \mu_i^{new}$.

<u>Finaly we get:</u>

$$\mu_{new} = \mu_c^{new}(s_{\partial c})\, \mu_i^{new}(s_i)^{-1} \prod_{a \in G} \mu_a^{new}(s_{\partial a}) \prod_{j \in G} \mu_j^{new}(s_j)^{1-d_j}$$

$$= \mu_c^{new}(s_{\partial c}) \left(\mu_i^{new}(s_i)\right)^{1-(d_i+1)} \prod_{a \in G} \mu_a^{new}(s_{\partial a}) \prod_{j \neq c} \left(\mu_j^{new}(s_j)\right)^{1-d_j}$$

Claim 2: Take any Gibbs distr in the form:

$$\mu(s_1 \dots s_N) = \frac{1}{Z} \exp\left(-\mathcal{H}(s_1 \dots s_N)\right).$$

The free energy is equal to the difference of the average energy and the average entropy:

$$F = -\log Z = \langle \mathcal{H} \rangle_\mu - S[\mu] .$$

Precisely,

$$\langle \mathcal{H} \rangle_\mu = \sum_{s_1 \dots s_N} \mathcal{H}(s_1 \dots s_N)\, \mu(s_1 \dots s_N)$$

$$S[\mu] = -\sum_{s_1 \dots s_N} \mu(s_1 \dots s_N) \ln \mu(s_1, s_2 \dots s_N).$$

Proof:

$$S[\mu] = -\frac{1}{Z} \sum_{s_1 \dots s_N} e^{-\mathcal{H}(s_1 \dots s_N)} \ln \frac{\exp(-\mathcal{H}(s_1 \dots s_N))}{Z}.$$

$$= \frac{1}{Z} \sum_{s_1 \dots s_N} \mathcal{H}(s_1 \dots s_N) \exp - \mathcal{H}(s_1 \dots s_N) + \ln Z$$

$$\Rightarrow \quad \ln Z = S[\mu] - \langle \mathcal{H} \rangle_\mu .$$

Remark: at this point it is worth mentioning the "variational principle" that states that the Gibbs distr minimizes the Gibbs functional

$$F[\nu] = \langle \mathcal{H} \rangle_\nu - S[\nu] .$$

$F[\mu] \leq F[\nu]$ for all ν.

For our graphical model the Hamiltonian is s.t.

$$\prod_a f_a(s_{\partial a}) = \exp\left(-\mathcal{H}(s_1 \cdots s_N)\right).$$

Thus $$\mathcal{H}(s_1 \ldots s_N) = -\sum_a \ln\left(f_a(s_{\partial a})\right).$$

$$F = \langle \mathcal{H} \rangle_\mu - S[\mu]$$

$$= -\sum_a \sum_{s_1 \cdots s_N} \left(\ln f_a(s_{\partial a})\right) \mu(s_1 \cdots s_N)$$

$$+ \sum_{s_1 \cdots s_N} \mu(s_1 \cdots s_N) \ln\left(\prod_a \mu_a(s_{\partial a}) \prod_i \mu_i(s_i)^{1-d_i}\right)$$

$$= -\sum_a \sum_{s_{\partial a}} \left(\ln f_a(s_{\partial a})\right) \mu(s_{\partial a}) \qquad \leftarrow \text{energy term}$$

$$+ \sum_a \sum_{s_1 \cdots s_N} \mu(s_1 \cdots s_N) \left(\ln \mu_a(s_{\partial a})\right)$$

$$+ \sum_i (1-d_i) \sum_{s_1 \cdots s_N} \mu(s_1 \cdots s_N) \left(\ln \mu_i(s_i)\right)$$

$$= \sum_a \sum_{s_{\partial a}} \mu_a(s_{\partial a}) \ln \mu_a(s_{\partial a}) + \sum_i (1-d_i) \sum_{s_i} \mu_i(s_i) \ln \mu_i(s_i)$$

$$- \sum_a \sum_{s_{\partial a}} \mu(s_{\partial a}) \ln f_a(s_{\partial a}).$$

Corollary : on a tree graphical model the free energy can be expressed as

$$\boxed{F = \sum_a \sum_{s_{\partial a}} \mu_a(s_{\partial a}) \ln \frac{\mu_a(s_{\partial a})}{f_a(s_{\partial a})} + \sum_i (1-d_i) \sum_i \mu_i(s_i) \ln \mu_i(s_i)}$$

This can also be expressed in terms of edge messages. Indeed we have seen that on a tree the marginals are exactly given

$$\begin{cases} \mu_i(s_i) \propto \prod_{a \in \partial i} \nu_{a \to i}(s_i) \\ \mu_a(s_{\partial a}) \propto f_a(s_{\partial a}) \prod_{i \in \partial a} \nu_{i \to a}(s_i) \end{cases}$$

where $\nu_{a \to i}$, $\nu_{i \to a}$ are a set of messages associated to edges of the graph. These messages are the (unique on a tree) solution of BP equations:

$$\begin{cases} \nu_{i \to a}(s_i) = \prod_{b \in \partial i \setminus a} \nu_{b \to i}(s_i) \\ \nu_{a \to i}(s_i) = \sum_{\sim s_i} f_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{j \to a}(s_j) . \end{cases}$$

Some algebra leads to the expression in terms of messages:

$$F = \sum_a F_a + \sum_i F_i - \sum_{(ia)} F_{ia}.$$

we get three contributions to the total free energy.

Check node contribution:

$$F_a = \ln \left\{ \sum_{S_{\partial a}} f_a(S_{\partial a}) \prod_{i \in \partial a} \nu_{i\to a}(S_i) \right\}$$

$$F_i = \ln \left\{ \sum_{S_i} \prod_{b \in \partial i} \nu_{b\to i}(S_i) \right.$$

$$F_{ai} = \ln \left\{ \sum_{S_i} \nu_{i\to a}(S_i)\, \nu_{a\to i}(S_i). \right\}.$$

3. Notion of Bethe free energy for general graphical models.

Consider a general graphical model, not necessarily tree like. Consider the set of all edges E and associate to each edge $(j\,a)$ distributions (or "message") called $\nu_{j\to a}(x_j)$ & $\hat{\nu}_{a\to j}(x_j)$.

...

For the moment these are not necessarily the BP messages. The only constraint on these distr are that they are normalized to 1.

Definition. The Bethe free energy functional is by definition:

$$F_{\text{Bethe}}[\{\nu_{i\to a}, \hat{\nu}_{a\to i}\}] \equiv + F_a[\{\nu_{j\to s}\}] + F_j[\{\hat{\nu}_{j\to s}\}]$$

$$- F_{aj}[\nu_{j\to a}, \hat{\nu}_{a\to j}].$$

with

$$F_a[\nu] = \ln\left\{\sum_{S_{\partial a}} f_a(S_{\partial a}) \prod_{j\in\partial a} \nu_{i\to a}(s_i)\right\}$$

$$F_j[\hat{\nu}] = \ln\left\{\sum_{s_j} \prod_{b\in\partial j} \hat{\nu}_{b\to j}(s_j)\right\}$$

$$F_{aj}[\nu, \hat{\nu}] = \ln\left\{\sum_{s_j} \nu_{j\to a}(s_j)\, \hat{\nu}_{a\to j}(s_j)\right\}.$$

<u>Proposition</u>. The stationary points of the Bethe free energy satisfy the BP fixed point equations and conversely fixed pts of BP are stationary points of Bethe free energy.

<u>Remark</u>: in this proposition by "stat pts" we mean "interior stationary pt"

<u>Proof.</u>

Introduce the Lagrangian (we consider only int stat pts).

$$L(\nu, \hat{\nu}, \lambda, \hat{\lambda}) = F(\nu, \hat{\nu}) - \sum_{ai} \hat{\lambda}_{a\to i}\Big(\sum_{s_i} \hat{\nu}_{a\to i}(s_i) - 1\Big)$$

$$- \sum_{ai} \lambda_{a\to i}\Big(\sum_{s_i} \nu_{i\to a}(s_i) - 1\Big).$$

Look at stationary points of L.

$$\frac{\delta L}{\delta \hat{\nu}_{a\to i}(s_i)} = 0 \Rightarrow \hat{\lambda}_{a\to i} = \frac{\hat{\nu}_{a\to i}(s_i)}{\sum_{s_i} \nu_{i\to a}(s_i)\hat{\nu}_{a\to i}(s_i)} - \frac{\sum_{\sim s_i} f_a(s_{\partial a}) \prod_{j\in\partial a\setminus i} \nu_{j\to a}(s_j)}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{j\in\partial a} \nu_{j\to a}(s_j)}$$

$$\frac{\delta L}{\delta \nu_{i\to a}(s_i)} = 0 \Rightarrow \lambda_{i\to a} = \frac{\nu_{i\to a}(s_i)}{\sum_{s_i} \nu_{i\to a}(s_i)\hat{\nu}_{a\to i}(s_i)} - \frac{\prod_{b\in\partial i\setminus a} \hat{\nu}_{b\to i}(s_i)}{\sum_{s_i} \prod_{b\in\partial i} \hat{\nu}_{b\to i}(s_i)}$$

$$\frac{\delta L}{\delta \hat{\lambda}_{a\to i}} = 0 \Rightarrow \sum_{s_i} \hat{\nu}_{a\to i}(s_i) = 1.$$

$$\frac{\delta L}{\delta \lambda_{i\to a}} = 0 \Rightarrow \sum_{s_i} \nu_{i\to a}(s_i) = 1$$

Let us show that $\hat{\lambda}_{a\to i} = \lambda_{i\to a} = 0$ (so that the $\underbrace{\text{constraint}}_{\text{normalization}}$ is trivially enforced)

Multiply first two equalities by $\nu_{i\to a}(s_i)$ & $\hat{\nu}_{a\to i}(s_i)$.
Then sum over s_i. This implies:

$$\hat{\lambda}_{a\to i} \cdot \sum_{s_i} \nu_{i\to a}(s_i) = 0$$

$$\lambda_{i\to a} \cdot \sum_{s_i} \hat{\nu}_{a\to i}(s_i) = 0.$$

Because of the last two equations we get $\hat{\lambda}_{a\to i} = \lambda_{i\to a} = 0$.
Thus stationary points of $L(\nu, \hat{\nu}, \lambda, \hat{\lambda})$ are:

$\hat{\lambda}_{a\to i} = 0$; $\lambda_{i\to a} = 0$ & $\nu, \hat{\nu}$ satisfy BP equs.

$$\begin{cases} \hat{\nu}_{a\to i}(s_i) \propto \displaystyle\sum_{\sim s_i} f_a(s_{\partial a}) \prod_{j\in\partial a\setminus i} \nu_{j\to a}(s_j) \\ \quad\text{Normalized to } 1. \\ \nu_{i\to a}(s_i) = \displaystyle\prod_{b\in\partial i\setminus a} \hat{\nu}_{b\to i}(s_i) \end{cases}$$

Since λ & $\hat{\lambda} = 0$ at stat point we have

$$0 = \frac{\delta L}{\delta \nu_{BP}} = \frac{\delta F}{\delta \nu_{BP}} - \underbrace{\sum_{ai} \hat{\lambda}_{ai}^{\,s.t} - \sum_{i} \lambda_{ai}^{\,s.t}}_{0}$$

$\Rightarrow$ idem for $\dfrac{\delta F}{\delta \hat{\nu}_{BP}}$. Thus $\left.\dfrac{\delta F}{\delta \nu}\right|_{\nu=\nu_{BP}} = 0$ & $\left.\dfrac{\delta F}{\delta \hat{\nu}}\right|_{\hat{\nu}=\hat{\nu}_{BP}} = 0$

Proof of Proposition page 13 by binary variables.

We work directly in the space of binary distributions by parametrizing:

$$\nu_{i\to a}(s_i) = \frac{e^{h_{i\to a}(s_i)}}{2\cosh h_{i\to a}} = \frac{1}{2}(1 + s_i \tan h_{i\to a})$$

$$\hat{\nu}_{a\to i}(s_i) = \frac{e^{\hat{h}_{a\to i} s_i}}{2\cosh \hat{h}_{a\to i}} = \frac{1}{2}(1 + s_i \tan \hat{h}_{i\to a}).$$

The Bethe functional now becomes:

$$F_{Bethe}[h, \hat{h}] = F_a + F_i - F_{ai},$$

$$F_a = \ln\left\{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{j\in\partial a} \frac{1}{2}(1 + s_j \tan h_{j\to a})\right\}.$$

$$F_j = \ln\left\{\sum_{s_j} \prod_{b\in\partial j} \frac{1}{2}(1 + s_j \tan \hat{h}_{b\to j})\right\}$$

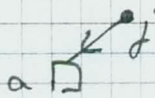
$$F_{aj} = \ln\left\{\sum_{s_j} \frac{1}{2}(1 + s_j \tan \hat{h}_{a\to j}) \cdot \frac{1}{2}(1 + s_j \tan h_{j\to a})\right\}.$$

$$= \ln\left\{\frac{1}{2}(1 + \tan \hat{h}_{a\to j} \cdot \tan h_{j\to a})\right\}.$$

Stationarity conditions.

A). $\dfrac{\delta F}{\delta \tanh_{j\to a}} = 0$ & $\dfrac{\delta F}{\delta \tan \hat{h}_{a\to j}} = 0$.

$$\frac{\delta F}{\delta \tanh_{j\to a}} = \frac{\sum_{s_{\partial a}} s_j\, f_a(s_{\partial a}) \prod_{i\in\partial a\setminus j} \frac{1}{2}(1+s_i \tanh h_{i\to a})}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{i\in\partial a} \frac{1}{2}(1+s_i \tanh h_{i\to a})}$$



$$- \frac{\frac{1}{2}\tan \hat{h}_{a\to j}}{\frac{1}{2}(1+\tan \hat{h}_{a\to j} \tanh h_{j\to a})}$$

Thus

$$\tan \hat{h}_{a\to j} = \left\{ \frac{1+\tan \hat{h}_{a\to j} \tanh h_{j\to a}}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{i\in\partial a} \frac{1}{2}(1+s_i \tanh h_{i\to a})} \right\}$$

$$\cdot \left\{ \sum_{s_{\partial a}} s_j\, f_a(s_{\partial a}) \prod_{i\in\partial a\setminus j} \frac{1}{2}(1+s_i \tanh h_{i\to a}) \right.$$

From this equation, by bringing $(\tan \hat{h}_{a\to j})$ on the same side we obtain:

$$\tan \hat{h}_{a\to j} = \frac{\sum_{s_{\partial a}} s_j\, f_a(s_{\partial a}) \prod_{i\in\partial a\setminus j} \frac{1}{2}(1+s_i \tanh h_{i\to a})}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{i\in\partial a\setminus j} \frac{1}{2}(1+s_i \tanh h_{i\to a})}.$$

Since the left hand side is equal to

$$\sum_j s_j \, \hat{\nu}_{a\to j}(s_j)$$

and the r.h.s is equal to

$$\sum_{s_j} s_j \left\{ \frac{\sum_{\sim s_j} f_a(s_{\partial a}) \prod_{i \in \partial a \setminus j} \nu_{i\to a}(s_i)}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{i\in \partial a \setminus j} \nu_{i\to a}(s_i)} \right\}.$$

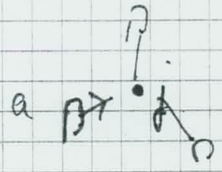
We see that the stationarity condition is equivalent to:

$$\hat{\nu}_{a\to j}(s_j) \propto \sum_{\sim s_j} f_a(s_{\partial a}) \prod_{i\in\partial a\setminus j} \nu_{i\to a}(s_i).$$

one of the two BP equations.

B) $\dfrac{\delta F}{\delta \tan \hat{h}_{a\to j}} = 0$ the calculation is similar.

$$\frac{\delta F}{\delta \tan \hat{h}_{a\to j}} = \frac{\sum_{s_j} s_j \prod_{b\in \partial j\setminus a} \frac{1}{2}(1+s_j \tan \hat{h}_{b\to j})}{\sum_{s_j} \prod_{b\in\partial j} \frac{1}{2}(1+s_j \tan \hat{h}_{b\to j})}.$$



$$- \frac{\tan h_{j\to a}}{1+\tan \hat{h}_{a\to j} \tanh_{j\to a}}$$

Thus

$$\tan h_{j\to a} = \left\{ \frac{1 + \tan \hat{h}_{a\to j} \tan h_{j\to a}}{\sum_{s_j} \prod_{b\in\partial j} (1 + s_j \tan \hat{h}_{b\to j})} \right\} \cdot \sum_{s_j} s_j \prod_{b\in\partial j\setminus a} (1 + s_j \tan \hat{h}_{b\to j})$$

Now isolate $(\tan h_{j\to a})$ on one side and deduce :

$$\tan h_{j\to a} = \frac{\sum_{s_j} s_j \prod_{b\in\partial j\setminus a} \frac{1}{2}(1 + s_j \tan \hat{h}_{b\to j})}{\sum_{s_j} \prod_{b\in\partial j\setminus a} \frac{1}{2}(1 + s_j \tan \hat{h}_{b\to j})}$$

The left hand side is $\sum_{s_j} s_j \nu_{j\to a}(s_j)$. Thus this equation is equivalent to the second BP equation :

$$\nu_{j\to a}(s_j) = \frac{\prod_{b\in\partial j\setminus a} \hat{\nu}_{b\to j}(s_j)}{\sum_{s_j} \prod_{b\in\partial j\setminus a} \hat{\nu}_{b\to j}(s_j)}.$$