Statistical Physics for Communication and Computer Science 15/5/2011

Lecture Notes 10: Bethe free energy and its relation to BP

Lecturer: Nicolas Macris *Scribe: Marc Vuffray*

1 Introduction

We want to compute the free energy and phase transition thresholds of graphical models such as those occurring 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 locally tree-like, a form of this theory developed by Bethe & Peierls is very well adapted, and will be developed 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 occurring 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.

Their 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 development 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 $\varepsilon_{BP} < \varepsilon < \varepsilon_{MAP}$.

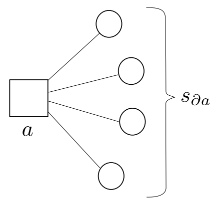
Concerning K-SAT we will see that pure Bethe-Peierls method is not exact. To improve upon it one has to go one-step further and develop an even more sophisticated mean field theory called "cavity method" or "replica method". This will be undertaken last two lectures of this course.

2 Free energy of graphical models on trees

Take

$$\mu\left(s_1,...,s_n\right) = \frac{1}{Z}\prod_a f_a\left(s_{\partial a}\right),$$

on a tree.



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

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

Claim 1 *On a tree we have*

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

where d_i is the degree of node i.

Proof. By induction over number M of clauses a.

- $M = 1$ clause

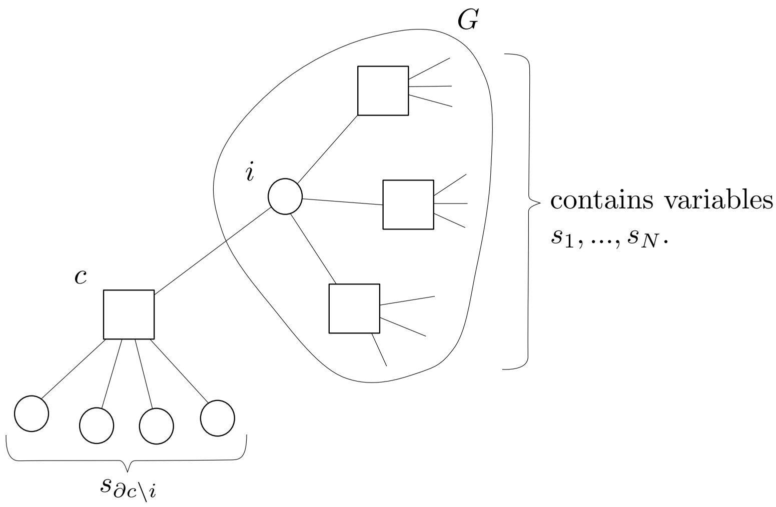
  $\mu(s_{\partial a}) = \mu_a(s_{\partial a}) \, \mu_i(s_i)^{1-1}$ trivially.

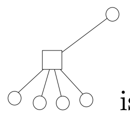
- Induction hypothesis: for any tree like graphical model with M clause we have the factorization

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

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

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



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

$$\Pr\left(s_{\partial c\setminus i}, s_1, ..., s_N\right) = \Pr\left(s_{\partial c\setminus i} \mid s_1, ..., s_N\right) \Pr\left(s_1, ..., s_N\right).$$

Here Pr is computed with the new graphical model measure

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

Now we have

$$\begin{aligned}
\Pr\left(s_{\partial c\setminus i} \mid s_1, ..., s_N\right) &= \Pr\left(s_{\partial c\setminus i} \mid s_i\right) \\
&= \frac{\Pr\left(s_{\partial c}\right)}{\Pr\left(s_i\right)} = \frac{\mu_c^{new}\left(s_{\partial c}\right)}{\mu_i^{new}\left(s_i\right)}.
\end{aligned}$$

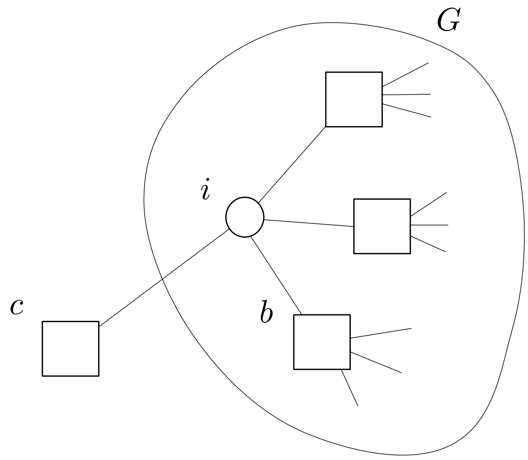
So:

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

$\Pr\left(s_1, ..., s_N\right)$ is the marginalization of new model with respect to variables $s_{\partial c\setminus i}$. Thus

$$\begin{aligned}
\Pr\left(s_1, ..., s_N\right) &= \frac{1}{Z_{new}} \underbrace{\sum_{s_{\partial c\setminus i}} f_c\left(s_{\partial c}\right)}_{\widetilde{f}_c(s_i)} \prod_a f_a\left(s_{\partial a}\right) \\
&\equiv \frac{1}{Z_{new}} \widetilde{f}_c\left(s_i\right) \prod_a f_a\left(s_{\partial a}\right).
\end{aligned}$$

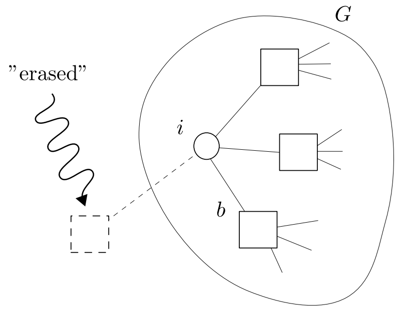
This distribution corresponds to a graphical model of the type:



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

$$\begin{aligned}
\Pr\left(s_1,...,s_N\right) &= \frac{1}{Z_{new}}\widetilde{f}_c\left(s_i\right)\prod_a f_a\left(s_{\partial a}\right) \\
&= \frac{1}{Z_{new}}\widetilde{f}_c\left(s_i\right) f_b\left(s_{\partial b}\right)\prod_{a\neq b} f_a\left(s_{\partial a}\right) \\
&= \frac{1}{Z_{new}}\widetilde{\widetilde{f}}_b\left(s_{\partial b}\right)\prod_{a\neq b} f_a\left(s_{\partial a}\right).
\end{aligned}$$

This graphical model now has M clauses:



So we can apply to it the induction hypothesis

$$\Longrightarrow \Pr\left(s_1,...,s_N\right) = \prod_a \mu_a\left(s_{\partial a}\right)\prod_i \left(\mu_i\left(s_i\right)\right)^{1-d_i}.$$

Here μ_a & μ_i are the marginals of $\Pr\left(s_1,...,s_N\right)$. But those marginals are also those of $\mu_{new}\left(s_{\partial c\backslash i}, s_1,...,s_N\right)$. So in the above formula we have $\mu_c = \mu_a^{new}$ & $\mu_i = \mu_i^{new}$. Finally we get:

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

■

Claim 2 *Take any Gibbs distribution in the form:*

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

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

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

Precisely,

$$\begin{aligned}
\left\langle\mathcal{H}\right\rangle_\mu &= \sum_{s_1,...,s_N}\mathcal{H}\left(s_1,...,s_N\right)\mu\left(s_1,...,s_N\right) \\
S\left[\mu\right] &= -\sum_{s_1,...,s_N}\mu\left(s_1,...,s_N\right)\ln\mu\left(s_1,...,s_N\right).
\end{aligned}$$

Proof.

$$\begin{aligned} S[\mu] &= -\frac{1}{Z}\sum_{s_1,\ldots,s_N} e^{-\mathcal{H}(s_1,\ldots,s_N)} \ln \frac{\exp(-\mathcal{H}(s_1,\ldots,s_N))}{Z} \\ &= \frac{1}{Z}\sum_{s_1,\ldots,s_N} \mathcal{H}(s_1,\ldots,s_N)\exp(-\mathcal{H}(s_1,\ldots,s_N)) + \ln Z \\ &\Longrightarrow \ln Z = S[\mu] - \langle \mathcal{H}\rangle_\mu . \end{aligned}$$

■

Remark 3 *at this point it is worth mentioning the "variational principle" that states that the Gibbs distribution minimizes the Gibbs functional*

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

If μ is the Gibbs distribution associated to the Hamiltonian $\mathcal{H}$, it satisfies

$$F[\mu] \le F[\nu] \text{ for all } \nu.$$

For our graphical model the Hamiltonian is

$$\prod_a f_a(s_{\partial a}) = \exp(-\mathcal{H}(s_1,\ldots,s_N)).$$

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

$$\begin{aligned} F &= \langle \mathcal{H}\rangle_\mu - S[\mu] \\ &= -\sum_a \sum_{s_1,\ldots,s_N} (\ln f_a(s_{\partial a}))\,\mu(s_1,\ldots,s_N) \\ &+ \sum_{s_1,\ldots,s_N} \mu(s_1,\ldots,s_N) \ln\left(\prod_a \mu_a(s_{\partial a}) \prod_i \mu_i(s_i)^{1-d_i}\right) \\ &= -\sum_a \sum_{s_1,\ldots,s_N} (\ln f_a(s_{\partial a}))\,\mu(s_{\partial a}) \\ &+ \sum_a \sum_{s_1,\ldots,s_N} \mu(s_1,\ldots,s_N)(\ln(\mu_a(s_{\partial a}))) \\ &+ \sum_i (1-d_i) \sum_{s_1,\ldots,s_N} \mu(s_1,\ldots,s_N)(\ln(\mu_i(s_i))) \\ &= \sum_a \sum_{s_{\partial a}} \mu_a(s_{\partial a}) \ln \mu_a(s_{\partial a}) + \sum_i (1-d_i) \sum_{s_i} \mu_i(s_i) \ln \mu_i(s_i) \\ &- \sum_a \sum_{s_{\partial a}} \mu(s_{\partial a}) \ln f_a(s_{\partial a}). \end{aligned}$$

Corollary 4 *on a tree graphical model the free energy can be expressed as*

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

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

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

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 equation

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

Some algebra leads to the expression in terms of messages:

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

We get three contributions to the total free energy:

$$F_a = \ln\left(\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{i \in \partial a} \nu_{i\to a}(s_i)\right)$$
$$F_i = \ln\left(\sum_{s_i} \prod_{b \in \partial i} \nu_{b\to i}(s_i)\right)$$
$$F_{ia} = \ln\left(\sum_{s_i} \nu_{i\to a}(s_i)\,\nu_{a\to i}(s_i)\right)$$

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 "messages") called $\nu_{j\to a}(x_j)$ & $\nu_{a\to j}(x_j)$. For the moment these are not necessarily the BP messages. The only constraint on these distributions are that they are normalized to 1.

Definition 5 *The Bethe free energy functional is by definition:*

$$F_{Bethe}[\{\nu_{j\to a}, \nu_{a\to j}\}] \equiv F_a[\{\nu_{j\to b}\}] + F_j[\{\nu_{j\to b}\}] - F_{aj}[\{\nu_{j\to a,} \nu_{a\to j}\}] .$$

with

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

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

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

Proposition 6 *The stationary points of the Bethe free energy satisfy the BP fixed point equation and conversely fixed points of BP are stationary points of the Bethe free energy.*

Remark 7 *In this proposition by "stationary points" we mean "interior stationary points".*

Proof. Introduce the Lagrangian (we consider only interior stationary points).

$$L\left(\nu,\widehat{\nu},\lambda,\widehat{\lambda}\right) = F(\nu,\widehat{\nu}) - \sum_{ai} \lambda_{a\to i}\left(\sum_{s_i} \widehat{\nu}_{a\to i}(s_i) - 1\right) - \sum_{ai} \lambda_{i\to a}\left(\sum_{s_i} \nu_{i\to a}(s_i) - 1\right).$$

Look at stationary points of L

$$\begin{aligned}
\frac{\delta L}{\delta \widehat{\nu}_{a\to i}(s_i)} &= 0 \Longrightarrow \widehat{\lambda}_{a\to i} = \frac{\widehat{\nu}_{a\to i}(s_i)}{\sum_{s_i} \nu_{i\to a}(s_i)\,\widehat{\nu}_{a\to i}(s_i)} - \frac{\sum_{\sim s_i} f_a(s_{\partial a}) \prod_{j\in\partial a\setminus i} \nu_{j\to a}(s_j)}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{j\in\partial a} \nu_{j\to a}(s_j)}.\\
\frac{\delta L}{\delta \nu_{i\to a}(s_i)} &= 0 \Longrightarrow \lambda_{i\to a} = \frac{\nu_{i\to a}(s_i)}{\sum_{s_i} \nu_{i\to a}(s_i)\,\widehat{\nu}_{a\to i}(s_i)} - \frac{\prod_{b\in\partial i\setminus a} \widehat{\nu}_{b\to i}(s_i)}{\sum_{s_i} \prod_{b\in\partial i} \widehat{\nu}_{b\to i}(s_i)}.\\
\frac{\delta L}{\delta \widehat{\lambda}_{a\to i}} &= 0 \Longrightarrow \sum_{s_i} \widehat{\nu}_{a\to i}(s_i) = 1.\\
\frac{\delta L}{\delta \lambda_{i\to a}} &= 0 \Longrightarrow \sum_{s_i} \nu_{i\to a}(s_i) = 1.
\end{aligned}$$

Let us show that $\widehat{\lambda}_{a\to i} = \lambda_{i\to a} = 0$ (so that the normalization constraint is trivially enforced). Multiply the first two equations by $\nu_{i\to a}(s_i)$ & $\widehat{\nu}_{a\to i}(s_i)$. Then sum over s_i. This implies:

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

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

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

$$\widehat{\lambda}_{a\to i} = 0; \lambda_{i\to a} = 0 \text{ \& } \nu,\widehat{\nu} \text{ satisfy BP equations.}$$

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

Since λ & $\widehat{\lambda} = 0$ at a stationary point we have

$$0 = \frac{\delta L}{\delta \nu_{BP}} = \frac{\delta F}{\delta \nu_{BP}} - \underbrace{\sum_{ai} \widehat{\lambda}_{ai}^{sta} - \sum_{i} \lambda_i^{sta}}_{=0}$$

$\Longrightarrow$ idem for $\frac{\delta F}{\delta \widehat{\nu}_{BP}}$. Thus

$$\left.\frac{\delta F}{\delta \nu}\right|_{\nu=\nu_{BP}} = 0 \ \& \ \left.\frac{\delta F}{\delta \widehat{\nu}}\right|_{\widehat{\nu}=\widehat{\nu}_{BP}} = 0$$

■

It is interesting to consider the special case of binary variables. We work directly in the space of binary distribution by parametrizing:

$$\nu_{i\to a}\left(s_i\right) = \frac{e^{h_{i\to a} s_i}}{2\cosh h_{i\to a}} = \frac{1}{2}\left(1 + s_i \tanh h_{i\to a}\right)$$
$$\widehat{\nu}_{a\to i}\left(s_i\right) = \frac{e^{\widehat{h}_{a\to i} s_i}}{2\cosh \widehat{h}_{a\to i}} = \frac{1}{2}\left(1 + s_i \tanh \widehat{h}_{a\to i}\right)$$

The Bethe functional now becomes:

$$F_{Bethe}\left[h, \widehat{h}\right] \equiv F_a + F_i - F_{ai}.$$
$$F_a = \ln\left(\sum_{s_{\partial a}} f_a\left(s_{\partial a}\right) \prod_{j\in\partial a} \frac{1}{2}\left(1 + s_j \tanh h_{j\to a}\right)\right)$$
$$F_i = \ln\left(\sum_{s_j} \prod_{b\in\partial j} \frac{1}{2}\left(1 + s_j \tanh \widehat{h}_{b\to j}\right)\right)$$
$$F_{ai} = \ln\left(\sum_{s_j} \frac{1}{2}\left(1 + s_j \tanh \widehat{h}_{a\to j}\right) \frac{1}{2}\left(1 + s_j \tanh h_{j\to a}\right)\right)$$
$$= \ln\left(\frac{1}{2}\left(1 + \tanh \widehat{h}_{a\to j} \tanh h_{j\to a}\right)\right).$$

Stationary conditions.

$$\frac{\delta F}{\delta \tanh h_{j\to a}} = 0 \ \& \ \frac{\delta F}{\delta \tanh \widehat{h}_{a\to j}} = 0$$

Proof.

A)

$$\frac{\delta F}{\delta \tanh h_{j\to a}} = \frac{\sum_{s_{\partial a}} s_j f_a(s_{\partial a}) \prod_{i\in\partial a\setminus j} \frac{1}{2}(1+s_i \tanh h_{i\to a})}{\sum_{s_{\partial a}} f_a(s_{\partial a}) \prod_{i\in\partial a} \frac{1}{2}(1+s_i \tanh h_{i\to a})} - \frac{\frac{1}{2}\tanh \widehat{h}_{a\to j}}{\frac{1}{2}\left(1+\tanh \widehat{h}_{a\to j}\tanh h_{j\to a}\right)}$$

Thus

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

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

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

Since the left hand side is equal to

$$\sum_j s_j \widehat{\nu}_{a\to j}(s_j).$$

and the right hand side 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:

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

one of the two BP equations.

B) $\frac{\delta F}{\delta \tanh \widehat{h}_{a\to j}} = 0$ the calculation is similar.

$$\frac{\delta F}{\delta \tanh \widehat{h}_{a\to j}} = \frac{\sum_{s_j} s_j \prod_{b\in\partial j\setminus a} \frac{1}{2}\left(1+s_j \tanh \widehat{h}_{b\to j}\right)}{\sum_{s_j} \prod_{b\in\partial j} \frac{1}{2}\left(1+s_j \tanh \widehat{h}_{b\to j}\right)} - \frac{\tanh h_{j\to a}}{1+\tanh \widehat{h}_{a\to j}\tanh h_{j\to a}}.$$

Thus

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

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

$$\tanh h_{j\to a} = \frac{\sum_{s_j} s_j \prod_{b\in\partial j\smallsetminus a} \frac{1}{2}\left(1 + s_j \tanh \widehat{h}_{b\to j}\right)}{\sum_{s_j} \prod_{b\in\partial j\smallsetminus a} \frac{1}{2}\left(1 + s_j \tanh \widehat{h}_{b\to j}\right)}.$$

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:

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

■