New Generalizations of the Bethe Approximation via Asymptotic Expansion

Ryuhei Mori   Toshiyuki Tanaka

Kyoto University

35th Symposium of Information Theory and Its Application
Beppu, Oita, Japan

13 December 2012

The Bethe approximation

- Successful approximation for low-density parity-check codes, compressed sensing, etc.
- Efficient message passing algorithm belief propagation (BP).
- A fixed point of BP is a stationary point of the Bethe free energy [Yedidia et al. 2005].

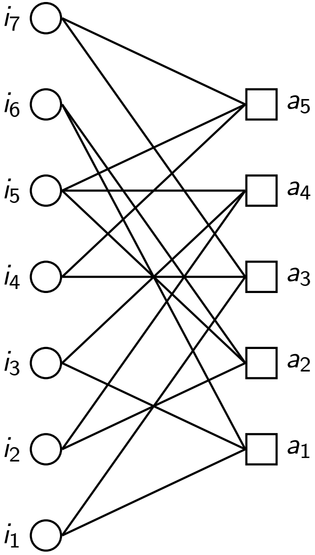
Factor graph and partition function

For a factor graph G.

- V: the set of variable nodes
- F: the set of factor nodes
- $\mathcal{X}$: the alphabet set
- N: the number of variables
- d_o: the degree of a node for $o \in V \cup F$
- f_a: a non-negative function in $\mathcal{X}^{d_a} \to \mathbb{R}_{\geq 0}$.

$$p(\mathbf{x}; G) := \frac{1}{Z(G)} \prod_{a \in F} f_a(\mathbf{x}_{\partial a})$$

$$Z(G) := \sum_{\mathbf{x} \in \mathcal{X}^N} \prod_{a \in F} f_a(\mathbf{x}_{\partial a})$$



The Legendre transformation

$$-\log Z(G) = \inf_{q \in \mathcal{P}(\mathcal{X}^N)} \left\{ -\sum_{\mathbf{x} \in \mathcal{X}^N} q(\mathbf{x}) \log \prod_{a \in F} f_a(\mathbf{x}_{\partial a}) - H(q) \right\}$$

where $H(q)$ is the Shannon entropy.

$\log Z(G)$ and $-H(q)$ are dual in the sense of Legendre transformation.

$$\log Z(G) \longleftrightarrow -H(q)$$

The Bethe free energy

$$-\log Z(G) = \inf_{q \in \mathcal{P}(\mathcal{X}^N)} \left\{ -\sum_{\mathbf{x} \in \mathcal{X}^N} q(\mathbf{x}) \log \prod_{a \in F} f_a(\mathbf{x}_{\partial a}) - H(q) \right\}$$

$$\begin{aligned} -\log Z_{\mathrm{Bethe}}(G) = \inf_{(b_i \in \mathcal{P}(\mathcal{X}))_{i \in V}, (b_a \in \mathcal{P}(\mathcal{X}^{d_a}))_{a \in F}} \Bigg\{ \\ -\sum_{a \in F} \sum_{\mathbf{x} \in \mathcal{X}^{d_a}} b_a(\mathbf{x}_{\partial a}) \log f_a(\mathbf{x}_{\partial a}) - H_{\mathrm{Bethe}}((b_i)_{i \in V}, (b_a)_{a \in F}) \Bigg\} \end{aligned}$$

where

$$H_{\mathrm{Bethe}}((b_i)_{i \in V}, (b_a)_{a \in F}) := \sum_{a \in F} H(b_a) - \sum_{i \in V} (d_i - 1) H(b_i).$$

Charactrizations of the Bethe free energy

- Loop calculus [Chertkov and Chernyak 2006, 2007]

$$Z(G) = Z_{\mathrm{Bethe}} \left(1 + \sum_{C:\,\text{generalized loop}} r(C) \right).$$

$\longrightarrow$ generalized to non-binary alphabet [This work]

Charactrizations of the Bethe free energy

- Loop calculus [Chertkov and Chernyak 2006, 2007]

$$Z(G) = Z_{\mathrm{Bethe}} \left(1 + \sum_{C:\,\text{generalized loop}} r(C) \right).$$

$\longrightarrow$ generalized to non-binary alphabet [This work]

- Method of graph cover [Vontobel 2010]

$$\frac{1}{M} \log \langle Z_{\Sigma_M} \rangle \to \log Z_{\mathrm{Bethe}}$$

$\longrightarrow$ generalized to the second-order analysis [This work]

Loop calculus for the binary alphabet

Lemma (Chertkov and Chernyak 2006, Sudderth et al., 2008)

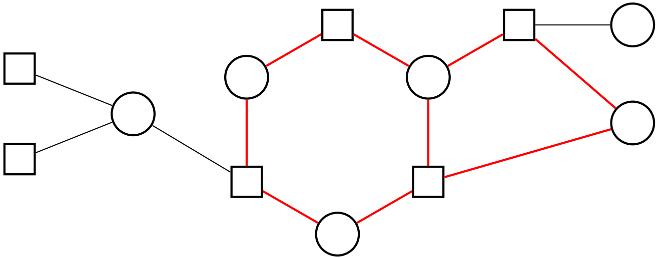
Assume that the alphabet is binary, i.e., $\mathcal{X} = \{0, 1\}$*. Let* $\eta_i := \langle X_i \rangle_{b_i} = b_i(1)$*. For any stationary point* $((b_i), (b_a))$ *of the Bethe free energy,*

$$Z(G) = Z_{\text{Bethe}}((b_i)_{i \in V}, (b_a)_{a \in F}) \sum_{E' \subseteq E} \mathcal{Z}(E')$$

where

$$\mathcal{Z}(E') := \prod_{i \in V} \left\langle \left(\frac{X_i - \eta_i}{\sqrt{\langle (X_i - \eta_i)^2 \rangle_{b_i}}} \right)^{d_i(E')} \right\rangle_{b_i} \cdot \prod_{a \in F} \left\langle \prod_{i \in \partial a,\, (i,a) \in E'} \frac{X_i - \eta_i}{\sqrt{\langle (X_i - \eta_i)^2 \rangle_{b_i}}} \right\rangle_{b_a} .$$

Generalized loop



$$\mathcal{G} := \{E' \subseteq E \mid d_o(E') \neq 1 \text{ for } o \in V \cup F\}$$

$$Z(G) = Z_{\mathrm{Bethe}}((b_i)_{i\in V}, (b_a)_{a\in F}) \left(1 + \sum_{E' \in \mathcal{G}\setminus\{\varnothing\}} \mathscr{Z}(E')\right).$$

Loop calculus for a non-binary alphabet 1/2

Theorem (This work)

For any stationary point $((b_i),(b_a))$ *of the Bethe free energy,*

$$Z(G) = Z_{\mathrm{Bethe}}((b_i)_{i \in V}, (b_a)_{a \in F}) \sum_{E' \subseteq E} \mathcal{Z}(E')$$

where

$$\mathcal{Z}(E') := \sum_{\mathbf{y} \in (\mathcal{X} \setminus \{0\})^{|E'|}} \prod_{i \in V} \left\langle \prod_{a \in \partial i, (i,a) \in E'} \frac{\partial \log b_i(X_i)}{\partial \eta_{i, y_{i,a}}} \right\rangle_{b_i} \cdot \prod_{a \in F} \left\langle \prod_{i \in \partial a, (i,a) \in E'} \frac{\partial \log b_i(X_i)}{\partial \theta_{i, y_{i,a}}} \right\rangle_{b_a}.$$

Coordinate systems the natural parameters $(\theta_{i,y})_{y \in \mathcal{X} \setminus \{0\}}$ and the expectation parameters $(\eta_{i,y})_{y \in \mathcal{X} \setminus \{0\}}$.

Loop calculus for a non-binary alphabet 2/2

The Jacobian matrix $\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}}$ is the Fisher information matrix.

Theorem (This work)

If one chooses a sufficient statistic $\mathbf{t}_i(x_i)$ for $i \in V$ such that the Fisher information matrix is diagonal at b_i, it holds

$$\mathcal{Z}(E') = \sum_{\mathbf{y} \in (\mathcal{X} \setminus \{0\})^{|E'|}} \prod_{i \in V} \left\langle \prod_{a \in \partial i, (i,a) \in E'} \frac{t_{i,y_{i,a}}(X_i) - \eta_{i,y_{i,a}}}{\sqrt{\left\langle \left(t_{i,y_{i,a}}(X_i) - \eta_{i,y_{i,a}} \right)^2 \right\rangle_{b_i}}} \right\rangle_{b_i}$$

$$\cdot \prod_{a \in F} \left\langle \prod_{i \in \partial a, (i,a) \in E'} \frac{t_{i,y_{i,a}}(X_i) - \eta_{i,y_{i,a}}}{\sqrt{\left\langle \left(t_{i,y_{i,a}}(X_i) - \eta_{i,y_{i,a}} \right)^2 \right\rangle_{b_i}}} \right\rangle_{b_a}.$$

Acknowledgment: P. Vontobel for insightful discussion about normal factor graph.

Loop calculus for expectations

Theorem (This work; it can be simplified like the previous theorem)

Let $C \subseteq V$, $F_C := \{a \in F \mid \partial a \subseteq C\}$ and $g : \mathcal{X}^{|C|} \to \mathbb{R}$. For any $((b_i),(b_a)) \in \mathcal{A}$, it holds

$$Z\langle g(\mathbf{X}_C)\rangle_p = Z_{\text{Bethe}}((b_i)_{i\in V},(b_a)_{a\in F}) \sum_{E' \subseteq E \setminus E(F_C)} \mathcal{Z}(E')$$

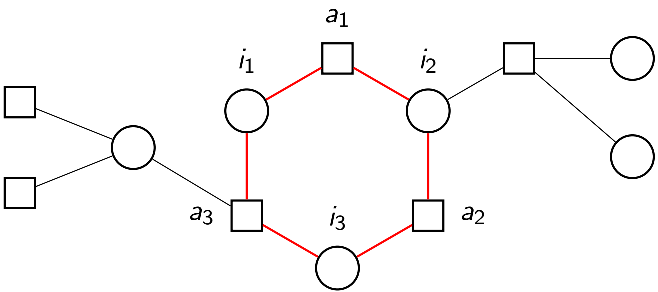
where

$$\mathcal{Z}(E') := \sum_{\mathbf{y}\in(\mathcal{X}\setminus\{0\})^{|E'|}} \prod_{i\in V\setminus C} \left\langle \prod_{a\in\partial i,(i,a)\in E'} \frac{\partial \log b_i(X_i)}{\partial \eta_{i,y_{i,a}}} \right\rangle_{b_i} \prod_{a\in F\setminus F_C} \left\langle \prod_{i\in\partial a,(i,a)\in E'} \frac{\partial \log b_i(X_i)}{\partial \theta_{i,y_{i,a}}} \right\rangle_{b_a} \cdot \left\langle g(\mathbf{X}_C) \prod_{i\in C,(i,a)\in E'} \frac{\partial \log b_i(X_i)}{\partial \eta_{i,y_{i,a}}} \right\rangle_{b_C}.$$

Here, $\langle\cdot\rangle_{b_C}$ is a pseudo expectation with respect to

$$b_C(\mathbf{x}_C) = \prod_{i\in C} b_i(x_i) \prod_{a\in F_C} \frac{b_a(\mathbf{x}_{\partial a})}{\prod_{i\in\partial a} b_i(x_i)}.$$

Loop calculus for single-cycle graph

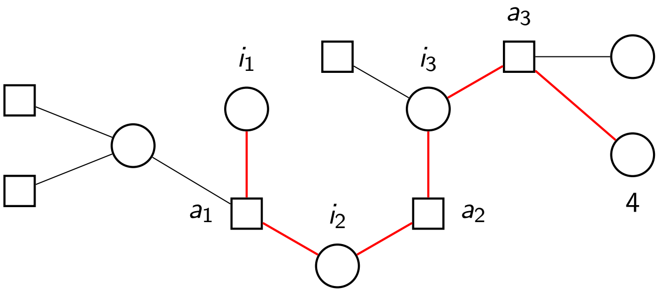


$$\mathrm{Cor}_{b_{a_k}}[\mathbf{t}_{i_k}(X_i), \mathbf{t}_{i_{k+1}}(X_{i_{k+1}})]$$
$$:= \mathrm{Var}_{b_k}[\mathbf{t}_{i_k}(X_{i_k})]^{-\frac{1}{2}} \mathrm{Cov}_{b_{a_k}}[\mathbf{t}_{i_k}(X_{i_k}), \mathbf{t}_{i_{k+1}}(X_{i_{k+1}})] \mathrm{Var}_{b_{k+1}}[\mathbf{t}_{i_{k+1}}(X_{i_{k+1}})]^{-\frac{1}{2}}.$$

Corollary (Partition function of single-cycle factor graph)

$$Z(G) = Z_{\mathrm{Bethe}}((b_i)_{i \in V}, (b_a)_{a \in F})$$
$$\cdot \left(1 + \mathrm{tr}\left(\mathrm{Cor}_{b_{a_1}}[\mathbf{t}_{i_1}(X_{i_1}), \mathbf{t}_{i_2}(X_{i_2})] \mathrm{Cor}_{b_{a_2}}[\mathbf{t}_{i_2}(X_{i_2}), \mathbf{t}_{i_3}(X_{i_3})] \cdots \mathrm{Cor}_{b_{a_n}}[\mathbf{t}_{i_n}(X_{i_n}), \mathbf{t}_{i_1}(X_{i_1})]\right)\right).$$

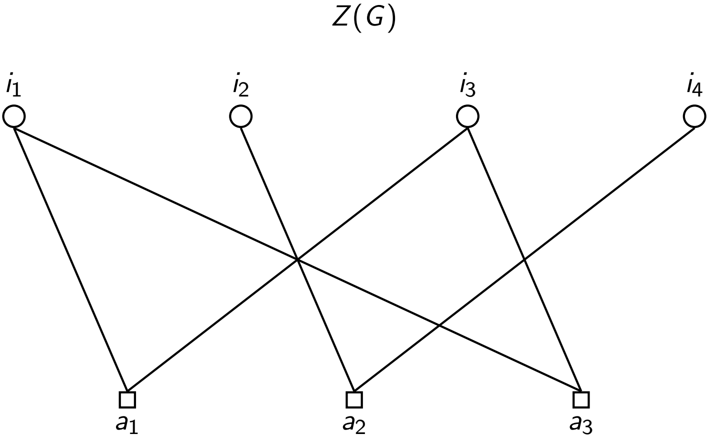
Correlation matrix on a tree factor graph



Corollary (Correlation matrix on a tree factor graph; Watanabe 2010)

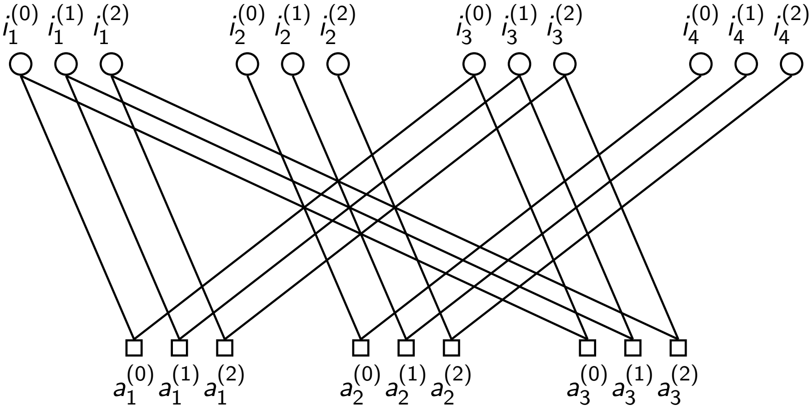
$$\mathrm{Cor}_p[X_1, X_n] \\ = \mathrm{Cor}_p[\mathbf{t}_1(X_1), \mathbf{t}_2(X_2)]\mathrm{Cor}_p[\mathbf{t}_2(X_2), \mathbf{t}_3(X_3)] \cdots \mathrm{Cor}_p[\mathbf{t}_{n-1}(X_{n-1}), \mathbf{t}_n(X_n)]$$

Graph cover



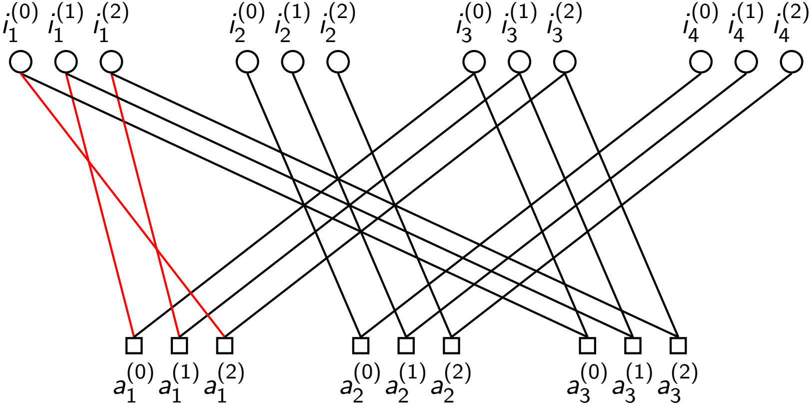
Graph cover

$Z(G)^M$



Graph cover

$$Z(G_{\boldsymbol{\sigma}}) \overset{?}{\approx} Z(G)^M$$



The method of graph cover

Lemma (Vontobel 2010)

$$\log\langle Z_{\Sigma_M}\rangle = M \log Z_{\text{Bethe}} + o(M)$$

Sketch of the proof.
The method of types and Laplace method. □

The second-order analysis for graph cover

Lemma (This work)

$$\log\langle Z_{\Sigma_M}\rangle = M \log Z_{\text{Bethe}} + \log \sqrt{\zeta(\mathbf{u})} + o(1)$$

where $\zeta(\mathbf{u})$ *is the edge zeta function and*
$u^a_{i\to j} = \mathrm{Cor}_{b_a}[\mathbf{t}_i(X_i), \mathbf{t}_j(X_j)]$.

Sketch of the proof.

Laplace method with the central approximation. □

Interpretation of Legendre transformation by large deviation

$$\log Z(G) = \frac{1}{M} \log Z(G)^M = \lim_{M\to\infty} \frac{1}{M} \log Z(G)^M$$
$$= -\inf_{p \in \mathcal{P}(\mathcal{X}^N)} \left\{ -\sum_{\mathbf{x}\in\mathcal{X}^N} p(\mathbf{x}) \log \prod_{a\in F} f_a(\mathbf{x}_{\partial a}) - H(p) \right\}$$

From more detailed analysis (asymptotic expansion)

$$\log Z(G)^M = M \log Z(G) + \underbrace{\log \sqrt{\frac{\det\left(\mathcal{J}(\boldsymbol{\theta})\right)}{\prod_{\mathbf{x}} p(\mathbf{x})}}}_{=0} + \frac{1}{M}0 + \frac{1}{M^2}0 + \cdots$$

Asymptotic expansion and asymptotic Bethe approximation

$$\log Z(G)^{M} = M \log Z(G) + \underbrace{\log \sqrt{\frac{\det(\mathcal{J}(\boldsymbol{\theta}))}{\prod_{\mathbf{x}} p(\mathbf{x})}}}_{=0} + \frac{1}{M} 0 + \frac{1}{M^2} 0 + \cdots$$

$$\log\langle Z_{\Sigma_M}\rangle = M \log Z_{\text{Bethe}} + \underbrace{\log \sqrt{\frac{\det(\nabla F_{\text{Bethe}})^{-1}}{\prod_i \prod_{x_i} b_i(x_i)^{1-d_i} \prod_{a\in F} \prod_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a})}}}_{=\log\sqrt{\zeta(\mathbf{u})}}$$

[Watanabe and Fukumizu 2010]

$$+ \frac{1}{M} g_1 + \frac{1}{M^2} g_2 + \cdots .$$

By letting $M = 1$,

Definition (Asymptotic Bethe approximation)

For $m = 1, 2, \ldots,$

$$\log Z_{\text{AB}}^{(m)} := \log Z_{\text{Bethe}} + \log\sqrt{\zeta(\mathbf{u})} + g_1 + g_2 + \cdots + g_{m-1}.$$

Edge zeta function

Definition (Prime cycle)

A closed walk $e_1 \rightharpoonup e_2 \cdots \rightharpoonup e_n \rightharpoonup e_1$ is a prime cycle $\iff$ it is backtrackless and cannot be expressed as power of another walk.

Definition (Edge zeta function)

$$\zeta(\mathbf{u}) = \prod_{\substack{(e_1 \rightharpoonup e_2 \cdots \rightharpoonup e_n \rightharpoonup e_1) \\ \text{is a prime cycle}}} \frac{1}{\det\left(I - u_{e_1,e_2} u_{e_2,e_3} \cdots u_{e_n,e_1}\right)}.$$

Lemma (Watanabe-Fukumizu formula; 2010)

$$\zeta(\mathbf{u})^{-1} = \det(\nabla^2 F_{\text{Bethe}}((\eta_i), (\eta_{\langle a \rangle}))) \cdot \prod_{i \in V} \det(\mathrm{Var}_{b_i}[\mathbf{t}_i(X_i)])^{1-d_i} \prod_{a \in F} \det(\mathrm{Var}_{b_a}[\mathbf{t}_a(X_{\partial a})])$$

where $u^a_{i \to j} = \mathrm{Cor}_{b_a}[\mathbf{t}_i(X_i), \mathbf{t}_j(X_j)]$.

Single-cycle graph

Let

$$A := \mathrm{Cor}_{b_{a_1}}[\mathbf{t}_{i_1}(X_{i_1}), \mathbf{t}_{i_2}(X_{i_2})]\mathrm{Cor}_{b_{a_2}}[\mathbf{t}_{i_2}(X_{i_2}), \mathbf{t}_{i_3}(X_{i_3})] \cdots \mathrm{Cor}_{b_{a_n}}[\mathbf{t}_{i_n}(X_{i_n}), \mathbf{t}_{i_1}(X_{i_1})]$$

Then, the true partition function Z and the asymptotic Bethe approximation $Z^{(1)}_{\mathrm{AB}}$ are

$$Z = Z_{\mathrm{Bethe}}((b_i)_{i\in V}, (b_a)_{a\in F})\left(1 + \mathrm{tr}(A)\right).$$

$$Z^{(1)}_{\mathrm{AB}} = Z_{\mathrm{Bethe}}((b_i)_{i\in V}, (b_a)_{a\in F})\frac{1}{\det(I - A)}.$$

$$= Z_{\mathrm{Bethe}}((b_i)_{i\in V}, (b_a)_{a\in F})\left(1 + \mathrm{tr}(A) + O(\rho(A)^2)\right)$$

where $\rho(A)$ is the spectrum radius of A.

The asymptotic Bethe approximation is accurate when $A \approx 0$.

General factor graph

$$Z(G) = Z_{\mathrm{Bethe}}((b_i)_{i \in V}, (b_a)_{a \in F}) \sum_{E' \in \mathcal{G}} \mathcal{Z}(E')$$

Generalized loop

$$\mathcal{G} := \{E' \subseteq E \mid d_o(E') \neq 1 \text{ for } o \in V \cup F\}$$

(Simple) loop [Gomez et al. 2006], [Chertkov and Chernyak 2007]

$$\mathcal{L} := \{E' \subseteq E \mid d_o(E') = 0, 2 \text{ for } o \in V \cup F, \textit{connected}\}$$

For $E' \in \mathcal{L}$

$$\mathcal{Z}(E') = \mathrm{tr}(A).$$

Roughly speaking, $Z_{\mathrm{AB}}^{(m)}$ enumerates the weights of $Z(E')$ for $E' \in \mathcal{L}$.

Numerical calculation: Ising model

$$Z = \sum_{\mathbf{x} \in \{+1,-1\}^N} \exp\left\{ \beta \left(\sum_{(i,j) \in E} x_i x_j + h \sum_{i=1}^{N} x_i \right) \right\}$$

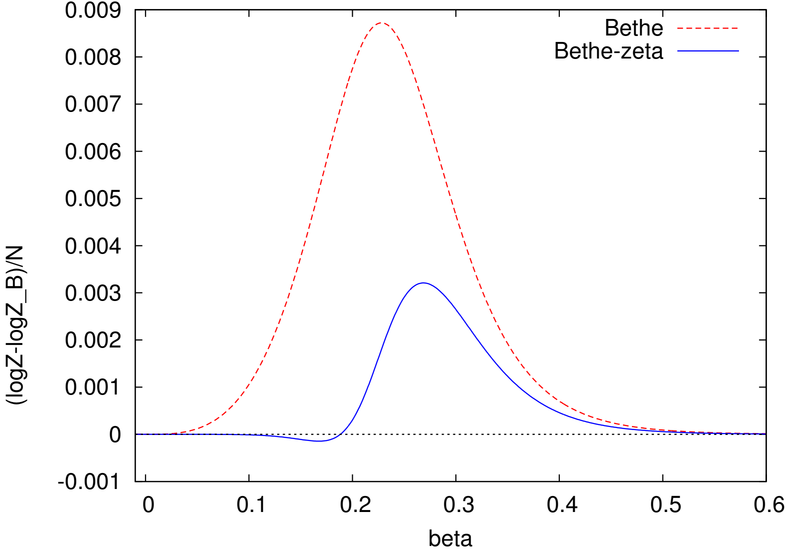
For a locally tree-like graph, if $\beta \geq 0$,
the Bethe approximation is asymptotically exact, i.e.,

$$\lim_{N \to \infty} \frac{1}{N} \log Z = \lim_{N \to \infty} \frac{1}{N} \log Z_{\text{Bethe}}$$

[Dembo and Montanari 2010].

$$|\mathrm{Cor}_{b_a}(X_i, X_j)| \leq \tanh(|\beta|) \ .$$

Results of numerical calculation: Ising model



$N = 16,\ d_{avg} = 4.375,\ h = 0.5.$

Summary and future works

Summary:

- Chertkov and Chernyak's loop calculus is generalized to non-binary alphabets by using tangent vectors for information manifold of exponential family.
- New generalization of the Bethe free energy is obtained by Vontobel's method of graph cover and Watanabe-Fukumizu formula.

Future works about asymptotic Bethe approximation:

- Rigorous proof of the accuracy for sparse factor graphs.
- Higher order approximations.
- Relationship with the Plefka expansion.