

Loop series expansion with propagation diagrams

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2009 J. Phys. A: Math. Theor. 42 045001

(<http://iopscience.iop.org/1751-8121/42/4/045001>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.155

The article was downloaded on 03/06/2010 at 08:11

Please note that [terms and conditions apply](#).

Loop series expansion with propagation diagrams

Yusuke Watanabe and Kenji Fukumizu

Institute of Statistical Mathematics, 4-6-7, Minami-Azabu, Minato-Ku, Tokyo 106-8569, Japan

E-mail: watay@ism.ac.jp

Received 8 August 2008, in final form 7 November 2008

Published 17 December 2008

Online at stacks.iop.org/JPhysA/42/045001

Abstract

The Bethe approximation is a successful method for approximating partition functions of probabilistic models associated with a graph. Recently, Chertkov and Chernyak derived an interesting formula called ‘loop series expansion’, which is an expansion of the partition function. The main term of the series is the Bethe approximation while other terms are labelled by subgraphs called generalized loops. In this paper, we derive a loop series expansion of binary pairwise Markov random fields with ‘propagation diagrams’, which describes rules as to how ‘first messages’ and ‘secondary messages’ propagate. Our approach allows us to express the loop series in the form of a polynomial with coefficients positive integers. Using the propagation diagrams, we establish a new formula that shows a relation between the exact marginal probabilities and their Bethe approximations.

PACS numbers: 05.50.+q, 02.10.Ox

(Some figures in this article are in colour only in the electronic version)

1. Introduction

A Markov random field (MRF) associated with a graph is given by a joint probability distribution over a set of variables. In the associated graph, the nodes represent variables and the edges represent probabilistic dependence between variables. A typical example of a MRF is a Gibbs distribution of the Ising model on a finite lattice. The joint distribution is often given in an unnormalized form, and the normalization factor of a MRF is called a partition function.

The main topic of this paper is computation of the partition function and the marginal distributions of a MRF with discrete variables. This problem is in general computationally intractable for a large number of variables, and some approximation method is required. Among many approximation methods, the Bethe approximation [1] has attracted renewed interest of computer scientists; it is equivalent to the loopy belief propagation (LBP) algorithm

[2, 3], which has been successfully used for many applications such as error correcting codes, inference on graphs, image processing and so on [4–6].

Chertkov and Chernyak [7, 8] give a new and interesting formula called loop series expansion, which expresses the partition function in terms of a finite series. The first term is the Bethe approximation, and the others are labelled by so-called generalized loops. The Bethe approximation can be corrected with this formula, though summing up all the terms requires computational efforts exponential to the number of linearly independent cycles.

In this paper we propose an alternative diagram-based method for deriving the loop series expansion formula. In our approach, we define secondary messages, which are orthogonal to the messages used in the LBP algorithm, and show that they satisfy a set of rules as they propagate. For each node and edge, we associate parameters $\{\gamma_i\}$ and $\{\beta_{ij}\}$, respectively; γ_i is related to the approximated marginal of a node i , and β_{ij} to the approximated correlation of adjacent nodes i and j . The loop series is represented by a polynomial of these variables with coefficients positive integers. This positivity is useful for deriving a bound on the number of generalized loops.

The main result of this paper is theorem 4, with which we can calculate the true marginal probabilities in terms of the beliefs at the convergence of the LBP, $\{\gamma_i\}$, and $\{\beta_{ij}\}$. The terms in the formula of the marginals depend on the topological structure of the graph.

This paper is organized as follows. In section 2, we briefly review the definition of a pairwise MRF, the Bethe approximation and the LBP algorithm. In section 3, we characterize the Bethe approximation as the fixed points of the LBP and deduce the fixed point equation in theorem 1. In section 4, we define first and secondary messages, and study their propagation rules. These rules are fundamental tools for our analysis. In section 5, we derive the loop series formula, and compare it with the results of Chertkov and Chernyak. We deduce consequences of our representation of the expansion: the connection to the partition function of the Ising model (with uniform coupling constant and external field) and the upper bound on the number of generalized loops. In section 6, we prove an expansion formula for the true marginal probability, and provide some examples.

2. Bethe approximation and loopy belief propagation algorithm

2.1. Pairwise Markov random field

We introduce a probabilistic model considered in this paper, a MRF of binary states with pairwise interactions. Let $G := (V, E)$ be a connected undirected graph, where $V = \{1, \dots, N\}$ is a set of nodes and $E \subset \{(i, j); 1 \leq i < j \leq N\}$ is a set of undirected edges. Each node $i \in V$ is associated with a binary space $\chi_i = \{\pm 1\}$. We make a set of directed edges from E by $\vec{E} = \{(i, j), (j, i); (i, j) \in E\}$. The neighbour of i is denoted by $N(i) \subset V$, and $d_i = |N(i)|$ is called the degree of i . A joint probability distribution on the graph G is given by the form

$$p(x) = \frac{1}{Z} \prod_{ij \in E} \psi_{ij}(x_i, x_j) \prod_{i \in V} \phi_i(x_i), \quad (1)$$

where $\psi_{ij}(x_i, x_j) : \chi_i \times \chi_j \rightarrow \mathbb{R}_{>0}$ and $\phi_i : \chi_i \rightarrow \mathbb{R}_{>0}$ are positive functions called compatibility functions. The normalization factor Z is called the partition function. A set of random variables which has a probability distribution in the form of (1) is called a Markov random field (MRF) or an undirected graphical model on the graph G . This class of probability distributions is equivalent to the Ising model with arbitrary coupling constants and local magnetic fields. In traditional literatures of statistical physics, a graph G is often

given by an infinite lattice, but as per recent interest, especially in computer science, G has an arbitrary topology with finite nodes.

Without loss of generality, univariate compatibility functions ϕ_i can be neglected because they can be included in bivariate compatibility functions ψ_{ij} . This operation does not affect the Bethe approximation and the LBP algorithm given below; we assume it as per the following.

2.2. Loopy belief propagation algorithm

The LBP algorithm computes the Bethe approximation of the partition function and the marginal distribution of each node with the message passing method [2, 3, 9]. This algorithm is summarized as follows.

- (i) Initialization: For all $(j, i) \in \vec{E}$, the message from i to j is a vector $m_{(j,i)}^0 \in \mathbb{R}^2$. Initialize as

$$m_{(j,i)}^0(x_j) = 1 \quad \forall x_j \in \chi_j. \quad (2)$$

- (ii) Message Passing: For each $t = 0, 1, \dots$, update the messages by

$$m_{(j,i)}^{t+1}(x_j) = \omega \sum_{x_i \in \chi_i} \psi_{ji}(x_j, x_i) \prod_{k \in N(i) \setminus \{j\}} m_{(i,k)}^t(x_i), \quad (3)$$

until it converges. Finally we obtain $\{m_{(j,i)}^*\}_{(j,i) \in \vec{E}}$.

- (iii) Approximated marginals and the partition function are computed by the following formulae:

$$b_i(x_i) := \omega \prod_{j \in N(i)} m_{(i,j)}^*(x_i), \quad (4)$$

$$b_{ji}(x_j, x_i) := \omega \psi_{ji}(x_j, x_i) \prod_{k \in N(j) \setminus \{i\}} m_{(j,k)}^*(x_j) \prod_{k' \in N(i) \setminus \{j\}} m_{(i,k')}^*(x_i), \quad (5)$$

$$\begin{aligned} \log Z_B := & \sum_{(j,i) \in E} \sum_{x_j, x_i} b_{ji}(x_j, x_i) \log \psi_{ji}(x_j, x_i) - \sum_{(j,i) \in E} \sum_{x_j, x_i} b_{ji}(x_j, x_i) \log b_{ji}(x_j, x_i) \\ & + \sum_{i \in V} (d_i - 1) \sum_{x_i} b_i(x_i) \log b_i(x_i), \end{aligned} \quad (6)$$

where ω are appropriate normalization constants, b_i are called beliefs and Z_B is called the Bethe approximation of the partition function.

In step (ii), there is an ambiguity as to the order of updating the messages. We do not specify the order, because the fixed points of the LBP algorithm do not depend on its choice. Note that this LBP algorithm does not necessarily converge, and there may be more than one fixed points unless the interactions are sufficiently weak [10].

3. The fixed-point equation of LBP

When the LBP converges, any converged messages $\{m_{(j,i)}^*\}_{(j,i) \in \vec{E}}$ satisfy a certain equation shown in theorem 1. By this theorem we show that the converged messages can be normalized simultaneously and we define $\{\mu_{(j,i)}\}_{(j,i) \in \vec{E}}$ called first messages.

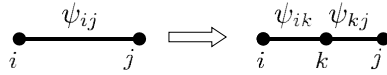


Figure 1. A new node can be added on an edge.

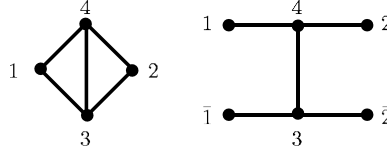


Figure 2. An example of G and \hat{G} .

3.1. Graph operations

First, we remark that we can always add a new node without changing the marginals and beliefs of the others. For an edge (i, j) , we can add a node k between i and j as in figure 1 with new compatibility functions ψ_{ik}, ψ_{kj} satisfying

$$\psi_{ij}(x_i, x_j) = \sum_{x_k} \psi_{ik}(x_i, x_k) \psi_{kj}(x_k, x_j). \tag{7}$$

This operation will be used implicitly many times in this paper. Adding new nodes, if necessary, we can always assume that ‘there are sufficiently many nodes of degree two’.

Next we define a graph \hat{G} by G . Let $L := |E| - |V| + 1$, the number of linearly independent cycles. Cutting and duplicating L nodes of degree two appropriately, we obtain a connected tree \hat{G} , since we assume that there are sufficiently many nodes of degree two [11]. See figure 2. Renumbering the nodes of V , we assume that the cut nodes are numbered by $\{1, \dots, L\}$. We define $\hat{G} = (\hat{V}, \hat{E})$ by $\hat{V}_1 = \{1, \dots, L\} \cup \{\bar{1}, \dots, \bar{L}\}$, $\hat{V}_2 = \{L + 1, \dots, N\}$ and $\hat{V} = \hat{V}_2 \cup \hat{V}_1$. \hat{E} is also naturally defined. We call \hat{V}_1 leaf nodes.

3.2. Belief propagation equations

Using the converged messages $\{m_{(j,i)}^*\}$, we define messages coming into the leaf nodes of the graph \hat{G} as follows. Let s be a cut node with the neighbour $N(s) = \{u, v\}$ in G , and let $(s, v), (\bar{s}, u) \in \hat{E}$ be the edges at the duplicated nodes. We define $\mu_s \propto m_{(s,u)}^*, \mu_{\bar{s}} \propto m_{(s,v)}^*$ and normalize them by $\sum_{x_s} \mu_s(x_s) \mu_{\bar{s}}(x_s) = 1$. Generalizations of the transfer matrices are defined by

$$T_{x_1, \dots, x_L}^{x_{\bar{1}}, \dots, x_{\bar{L}}} = \sum_{x_{L+1}, \dots, x_N} \prod_{(i,j) \in \hat{E}} \psi_{ij}(x_i, x_j), \tag{8}$$

$$T_s^{x_s} = \sum_{x_{s+1}, \dots, x_L} \sum_{x_{\bar{s}+1}, \dots, x_{\bar{L}}} T_{x_1, \dots, x_L}^{x_{\bar{1}}, \dots, x_{\bar{L}}} \prod_{\substack{m=1 \\ m \neq s}}^L \mu_m(x_m) \mu_{\bar{m}}(x_{\bar{m}}). \tag{9}$$

Since $\{m_{(j,i)}^*\}$ is the convergence point of the LBP algorithm, it is easy to see that there is $\alpha_s > 0$ that satisfies $\alpha_s \mu_s(x_s) = \sum_{x_{s'}} T_s^{x_s'} \mu_s(x_{s'})$. The following theorem states that all α_s are equal to Z_B .

Theorem 1. For $s = 1, \dots, L$

$$Z_B \mu_s(x_{\bar{s}}) = \sum_{x_s} T_s^{x_s}_{x_{\bar{s}}} \mu_s(x_s) \quad x_{\bar{s}} \in \chi_{\bar{s}}, \tag{10}$$

$$Z_B \mu_{\bar{s}}(x_s) = \sum_{x_{\bar{s}}} T_s^{x_s}_{x_{\bar{s}}} \mu_{\bar{s}}(x_{\bar{s}}) \quad x_s \in \chi_s. \tag{11}$$

Proof. From $\alpha_s \mu_s(x_s) = \sum_{x_s'} T_s^{x_s'} \mu_s(x_s')$, α_s and $\alpha_{\bar{s}}$ are the Perron–Frobenius eigenvalues of the matrix $T_s^{x_s'}_{x_s}$ and its transpose, respectively. Therefore $\alpha_s = \alpha_{\bar{s}}$ for $1 \leq s \leq L$. Next, we prove $\alpha_s = \alpha_l$ for $1 \leq s, l \leq L$. As we normalize $\sum_{x_s} \mu_s(x_s) \mu_{\bar{s}}(x_s) = 1$,

$$\begin{aligned} \alpha_s &= \sum_{x_s} \{\alpha_s \mu_s(x_s)\} \mu_{\bar{s}}(x_s) \\ &= \sum_{x_1, \dots, x_L} \sum_{x_1, \dots, x_L} T^{x_1, \dots, x_L}_{x_1, \dots, x_L} \prod_{m=1}^L \mu_m(x_m) \mu_{\bar{m}}(x_{\bar{m}}) \\ &= \alpha_l, \end{aligned}$$

which shows $\alpha_l = \alpha_s = \alpha$. Finally we prove $\alpha = Z_B$. By dividing some ψ function by α from the first, we can assume that $\alpha = 1$. Then it is sufficient to prove $\log Z_B = 0$. By distributing the messages $\{\mu_s\}_{s \in \hat{V}_1}$ on the tree \hat{G} from its leaf nodes \hat{V}_1 without normalization, we define a message $\mu_{(j,i)}$ at each edge $(j,i) \in \vec{E}$. Because \hat{G} is a tree, the messages $\{\mu_{(j,i)}\}_{(j,i) \in \vec{E}}$ are uniquely defined step by step from the leaf nodes. By the assumption $\alpha = 1$, we have $\mu_s = \mu_{(s,u)}$. Therefore,

$$\mu_{(j,i)}(x_j) = \sum_{x_i \in \chi_i} \psi_{ji}(x_j, x_i) \prod_{s \in N(i) \setminus \{j\}} \mu_{(i,s)}(x_i), \quad \forall (j,i) \in \vec{E}. \tag{12}$$

By the relation

$$\sum_{x_i} \prod_{j \in N(i)} \mu_{(i,j)}(x_i) = \sum_{x_1, \dots, x_L} \sum_{x_1, \dots, x_L} T^{x_1, \dots, x_L}_{x_1, \dots, x_L} \prod_{m=1}^L \mu_m(x_m) \mu_{\bar{m}}(x_{\bar{m}}) = 1, \tag{13}$$

we obtain

$$b_i(x_i) = \prod_{j \in N(i)} \mu_{(i,j)}(x_i), \tag{14}$$

$$b_{ji}(x_j, x_i) = \psi_{ji}(x_j, x_i) \prod_{s \in N(j) \setminus \{i\}} \mu_{(j,s)}(x_j) \prod_{s' \in N(i) \setminus \{j\}} \mu_{(i,s')}(x_i). \tag{15}$$

The assertion follows from putting (14) and (15) into the definition of Z_B (6). □

As used in the proof of the above theorem, the normalization $Z_B = 1$ is convenient; in the rest of this paper we assume the following.

Assumption. By normalizing one of $\{\psi_{ij}\}$, we assume $Z_B = 1$.

In the proof of the above theorem, we defined the messages $\mu_{(j,i)} \propto m_{(j,i)}^*$ satisfying (12), (14) and (15). We call $\mu_{(j,i)}$ (normalized) first messages. While these conditions are similar to (3), (4) and (5), an important difference is disappearance of the normalization constants ω .

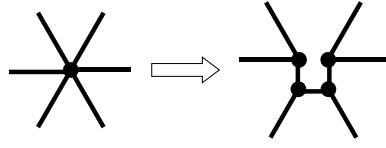


Figure 3. A node of degree n can be split into $n - 2$ nodes of degree 3.

By the above assumption, the first messages satisfy (12). Note that we define $\{\mu_{(j,i)}\}_{(j,i) \in \bar{E}}$ on the graph G , not only on the graph \hat{G} .

This equation is a generalization of the recursive expression for calculating free energy of the Bethe lattice [12]. For each fixed point of the LBP algorithm $\{m_{(j,i)}^*\}$, there is a solution of equation (10), (11). On the other hand, for each solution of (10) and (11), there is a LBP fixed point.

4. Propagation diagrams

We proved in the previous section, if we normalize $\sum_{x_s} \mu_s(x_s) \mu_{\bar{s}}(x_s) = 1$, the Bethe-approximated partition function is given by

$$Z_B = \sum_{x_1, \dots, x_L} \sum_{x_{\bar{1}}, \dots, x_{\bar{L}}} T^{x_1, \dots, x_L}_{x_{\bar{1}}, \dots, x_{\bar{L}}} \prod_{m=1}^L \mu_m(x_m) \mu_{\bar{m}}(x_{\bar{m}}), \tag{16}$$

while the true partition function is

$$Z = \sum_{x_1, \dots, x_L} T^{x_1, \dots, x_L}_{x_1, \dots, x_L} = \sum_{x_1, \dots, x_L} \sum_{x_{\bar{1}}, \dots, x_{\bar{L}}} T^{x_1, \dots, x_L}_{x_{\bar{1}}, \dots, x_{\bar{L}}} \prod_{m=1}^L \delta_{x_m, x_{\bar{m}}}. \tag{17}$$

Let us define vectors v_s and $v_{\bar{s}}$ for $s = 1, \dots, L$ so as to satisfy

$$\sum_{x_s} \mu_s(x_s) v_{\bar{s}}(x_s) = 0, \quad \sum_{x_s} \mu_{\bar{s}}(x_s) v_s(x_s) = 0, \quad \sum_{x_s} v_s(x_s) v_{\bar{s}}(x_s) = 1. \tag{18}$$

Then, we have a decomposition of the unit matrix

$$\delta_{x_s, x_{\bar{s}}} = \mu_s(x_s) \mu_{\bar{s}}(x_{\bar{s}}) + v_s(x_s) v_{\bar{s}}(x_{\bar{s}}). \tag{19}$$

We can expand (17) using (19) in a sum of 2^L terms. The first term is obviously the Bethe approximated partition function (16). But the explicit form of the remaining $2^L - 1$ terms is not obvious. In this section we define secondary messages $\{v(i, j)\}_{(i,j) \in \bar{E}}$ and derive rules, which describe how these messages propagate, for deriving the remaining terms.

4.1. Definition of secondary messages

By splitting nodes as in figure 3, any graphs can be transformed so that every node is of degree, at most, three. We make the compatibility functions of new edges infinitely strong: $\psi_{ij}(x_i, x_j) = \delta_{x_i, x_j}$. This change does not affect the fixed points of the LBP algorithm, hence the Bethe approximation. In subsections 4.1 and 4.2 we assume that the graph G has undergone this transformation. The same transformation also appears in [13, 14]. The first messages μ are defined on this transformed graph.

We define ‘secondary messages’ $\{v_{(i,j)}\}$ on the transformed graph using $\{\mu_{(i,j)}\}$. First, let j be a node of degree two such that at least one adjacent node, denoted by k , has degree two (figure 4). We define v at the node j by the following conditions:

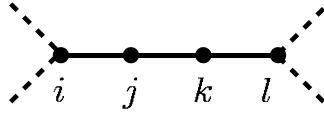


Figure 4. Node j is degree 2.

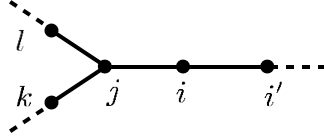


Figure 5. Node j is degree 3, node i is degree 2.



Figure 6. Rules at degree 2 and 3 nodes.

$$\sum_{x_j} v_{(j,i)}(x_j) \mu_{(j,k)}(x_j) = 0, \quad \sum_{x_j} \mu_{(j,i)}(x_j) v_{(j,k)}(x_j) = 0, \quad (20)$$

$$\sum_{x_j} v_{(j,i)}(x_j) v_{(j,k)}(x_j) = 1. \quad (21)$$

Similarly, let j be a node of degree two such that at least one adjacent node, denoted by i , has degree two (figure 5). We define v at node j by the following conditions:

$$\begin{aligned} \sum_{x_j} v_{(j,i)}(x_j) \mu_{(j,k)}(x_j) \mu_{(j,l)}(x_j) &= 0, & \sum_{x_j} \mu_{(j,i)}(x_j) v_{(j,k)}(x_j) \mu_{(j,l)}(x_j) &= 0, \\ \sum_{x_j} \mu_{(j,i)}(x_j) \mu_{(j,k)}(x_j) v_{(j,l)}(x_j) &= 0, \end{aligned} \quad (22)$$

$$\begin{aligned} \sum_{x_j} v_{(j,i)}(x_j) v_{(j,k)}(x_j) \mu_{(j,l)}(x_j) &= 1, & \sum_{x_j} \mu_{(j,i)}(x_j) v_{(j,k)}(x_j) v_{(j,l)}(x_j) &= 1, \\ \sum_{x_j} v_{(j,i)}(x_j) \mu_{(j,k)}(x_j) v_{(j,l)}(x_j) &= 1. \end{aligned} \quad (23)$$

These conditions determine $v_{(j,i)}$ uniquely up to a scalar factor at nodes of degree two, and up to a sign at nodes of degree three. We assume that the first component $v_{(i,j)}(+)$ is negative without loss of generality. The above relations and $\sum_{x_j} \prod_{i \in N(j)} \mu_{(j,i)}(x_j) = 1$ (13) are pictorially summarized in figure 6. These cases are sufficient because we can add a node of degree two at any edge if necessary. We call such diagrams *propagation diagrams*. The dashed and the light arrows express μ and v , respectively.

A condition which is similar to (20) and (22) is imposed in [15] to deduce the update rule of the LBP algorithm, though they consider a graphical model with variables on the edges.

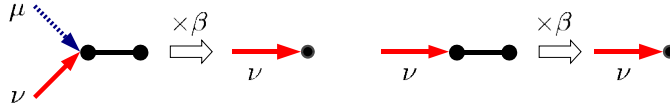


Figure 7. Propagation rules.

4.2. Propagation rules

The rules depicted in figure 6 describe what happens when messages collide at a node. The next two lemmas show how the first and secondary messages propagate.

Lemma 1. *See figure 4. Suppose nodes j and k are of degree 2. Then, there is $\beta_{jk} \in \mathbb{R}$ such that*

$$\beta_{jk} v_{(j,k)}(x_j) = \sum_{x_k} \psi_{jk}(x_j, x_k) v_{(k,l)}(x_k), \quad \beta_{jk} v_{(k,j)}(x_k) = \sum_{x_j} \psi_{kj}(x_k, x_j) v_{(j,i)}(x_j). \quad (24)$$

Proof. Using (12) and (20), $\sum_{x_j, x_k} \mu_{(j,i)}(x_j) \psi_{(jk)}(x_j, x_k) v_{(k,l)}(x_k) = \sum_{x_k} \mu_{(k,j)}(x_k) v_{(k,l)}(x_k) = 0$. Hence $\sum_{x_k} \psi_{jk}(x_j, x_k) v_{(k,l)}(x_k) \propto v_{(j,k)}(x_j)$. The proportion is equal in both directions (j, k) and (k, j) because of the condition (21). \square

Next we proceed to a degree three node.

Lemma 2. *See figure 5. Suppose node j is of degree three and i is of degree two. Then, there is $\beta_{ji} \in \mathbb{R}$ such that*

$$\beta_{ji} v_{(j,i)}(x_j) = \sum_{x_i} \psi_{ji}(x_j, x_i) v_{(i,i')} (x_i), \quad (25)$$

$$\beta_{ji} v_{(i,j)}(x_i) = \sum_{x_j} \psi_{ji}(x_j, x_i) v_{(j,l)}(x_j) \mu_{(j,k)}(x_j), \quad (26)$$

$$\beta_{ji} v_{(i,j)}(x_i) = \sum_{x_j} \psi_{ji}(x_j, x_i) \mu_{(j,l)}(x_j) v_{(j,k)}(x_j). \quad (27)$$

Proof. We can show in the same way as the previous lemma. \square

These lemmas say that the secondary messages ν propagate with rate β for both directions, though the first messages μ propagate without variation in scales. Equations (26) and (27) hold when the adjacent nodes i and j have degree three as in figure 8. We associate numbers β_{ij} for all undirected edges in E . Propagation diagrams of these results are summarized in figure 7.

4.2.1. Two more rules. In addition to the rules in figure 6, we have to give a rule for the case in which three secondary messages ν collide at a node.

Lemma 3. *Let $N(j) = \{i, k, l\}$, then*

$$\sum_{x_i} v_{(j,i)}(x_j) v_{(j,k)}(x_j) v_{(j,l)}(x_j) = \frac{1}{\sqrt{b_j(+)} b_j(-)} (b_j(+)-b_j(-)) =: \gamma_j. \quad (28)$$

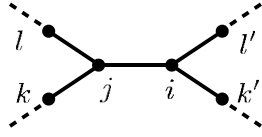


Figure 8. Nodes j and i have degree three.

Proof. We need a direct computation of the message vectors. By the orthogonal condition (22), $v_{(j,i)}$, $v_{(j,k)}$ and $v_{(j,l)}$ are determined up to scalar factors. The scales are determined by (23). Since we assume the first components of v are negative, v must have the following form:

$$v_{(j,i)}(x_j) = -x_j \sqrt{\frac{\mu_{(j,i)}(1)\mu_{(j,i)}(-1)}{\mu_{(j,k)}(1)\mu_{(j,k)}(-1)\mu_{(j,l)}(1)\mu_{(j,l)}(-1)}} \mu_{(j,k)}(-x_j)\mu_{(j,l)}(-x_j). \quad (29)$$

From (14) the result follows. \square

We use the next lemma when we split a node as in figure 3.

Lemma 4. See figure 8. If the nodes j and i are of degree three and $\psi_{ij}(x_i, x_j) = \delta_{x_i, x_j}$, then $\beta_{ij} = 1$.

Proof. By (27), (29) and the remark after lemma 2,

$$\begin{aligned} \beta_{ij} v_{(i,j)}(x_i) &= \sum_{x_j} \psi_{ij}(x_i, x_j) \mu_{(j,l)}(x_j) v_{(j,k)}(x_j) \\ &= \mu_{(j,l)}(x_i) v_{(j,k)}(x_i) \\ &= -x_i \sqrt{\frac{\mu_{(j,k)}(1)\mu_{(j,k)}(-1)\mu_{(j,l)}(1)\mu_{(j,l)}(-1)}{\mu_{(j,i)}(1)\mu_{(j,i)}(-1)}} \mu_{(j,i)}(-x_i). \end{aligned}$$

Since $\mu_{(j,i)}(x_i) = \mu_{(i,l')}(x_i)\mu_{(i,k')}(x_i)$ and $\mu_{(i,j)}(x_i) = \mu_{(j,l)}(x_i)\mu_{(j,k)}(x_i)$, we can show $\beta_{ij} = 1$ by using (29) for (i, j) . \square

4.3. At general nodes

We have defined v on the transformed graph $G_T = (V_T, E_T)$ in which the degree of every node is at most three. By shrinking added edges, we define v on the original graph $G = (V, E)$. In other words the injection $E \rightarrow E_T$ induces the messages on the original graph from the transformed graph. At each node of the original graph, we show that the following theorem holds. This theorem generalizes the rules in figure 6 and lemma 3.

Theorem 2. Let $j \in V$ and $N(j) = \{i_1, \dots, i_{d_j}\}$. Then, we have

$$f_n(\gamma_j) = \sum_{x_j} \prod_{s=1}^n v_{(j,i_s)}(x_j) \prod_{t=1}^{d_j-n} \mu_{(j,i_{n+t})}(x_j), \quad (30)$$

where $\{f_n(x)\}_{n=0}^\infty$ is a set of polynomials defined inductively by the relations $f_0(x) = 1$, $f_1(x) = 0$ and $f_{n+1}(x) = x f_n(x) + f_{n-1}(x)$.

Proof. We can reduce to the case of $n = d_j$ by splitting node j and using the propagation rules in figure 7. See figure 9 for example. The proof is done by induction; the cases of $n = 1, 2, 3$ is obtained by definition. In the case of n we split the node j into j' and j'' where 2 and $n - 2$ secondary messages join at j' and j'' , respectively. Adding a node k between j' and j'' , we

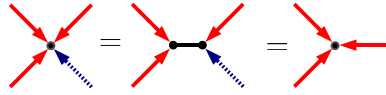


Figure 9. Reduction to the case of $n = d_j$.

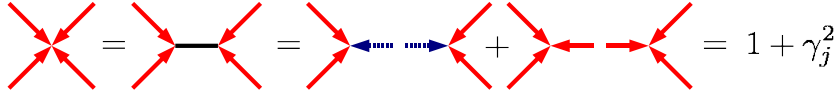


Figure 10. Derivation of the inductive relation in the case of $n = 4$.

use the relation $\mu_{(k,j)}(x_k)\mu_{(k,j'')}(x_{k'}) + v_{(k,j')}(x_k)v_{(k,j'')}(x_{k'}) = \delta_{x_k,x_{k'}}$. By lemma 4 and the induction hypothesis, we obtain $f_n(\gamma_j) = f_2(\gamma_j)f_{n-2}(\gamma_j) + f_3(\gamma_j)f_{n-1}(\gamma_j)$. The assertion follows from the definition of the polynomials f_n . Figure 10 illustrates this procedure. \square

It is surprising that the right-hand side is determined by the value γ_j which depends only on the belief b_j , while the messages $\mu_{(j,i_s)}$ and $v_{(j,i_s)}$ are not determined by b_j .

In this section we have derived a set of rules. In addition to the diagrams to show the rules, diagrams in which these rules are successively applied are also called propagation diagrams.

4.4. In the case of one-dimensional systems

In the case of one-dimensional spin systems, results of section 3 and 4 are reduced to easy observations. Let the graph G be a cycle of length N . We cut the node 1 as discussed in section 3.1 to obtain a string. The partition function is

$$Z = \sum_{x_1 = \pm 1} T^{x_1} \tag{31}$$

where the transfer matrix $T = T_1$ is defined by (8) and (9). Therefore Z is equal to the sum of the first and second eigenvalues of the matrix T . By theorem 1, the first messages μ_1 and $\mu_{\bar{1}}$ are the first right and left eigenvectors of the transfer matrix, where the first eigenvalue is Z_B . By lemma 1, we see that v_1 and $v_{\bar{1}}$ are the second right and left eigenvectors, and the second eigenvalue is $Z_B \prod \beta_{ji}$. The conditions of (20) are regarded as orthogonality of eigenvectors. The results of these sections are a generalization to the transfer tensor associated with more complicated graphs with nodes of degree three.

5. Loop series expansion formula

5.1. Derivation of loop series expansion formula

Let $G = (V, E)$ be a graph, and D be a subset of the edge-set E . An edge-induced subgraph of D is a subgraph of the graph G whose edge-set is D and whose node-set consists of all nodes that are incident with at least one edge in D .

We are now ready to prove the expansion formula of the partition function.

Theorem 3. For each solution of (10), (11), define $\{\beta_{ij}\}, \{\gamma_i\}$ by the above lemmas. Then the following expansion formula holds:

$$Z = Z_B \left(1 + \sum_{\phi \neq C \in \mathcal{G}} r(C) \right), \tag{32}$$

$$r(C) := \prod_{ij \in E_C} \beta_{ij} \prod_{i \in V_C} f_{d_i(C)}(\gamma_i),$$

where \mathcal{G} is the set of all edge-induced subgraphs of G and $C = (V_C, E_C)$.

Proof. Adding a node k on each edge (i, j) of G , we expand the partition function by the relation $\mu_{(k,i)}(x_k)\mu_{(k,j)}(x_k') + \nu_{(k,i)}(x_k)\nu_{(k,j)}(x_k') = \delta_{x_k, x_k'}$. We have $2^{|E|}$ terms in this expansion, because for each edge (i, j) there is a choice: $\mu_{(k,i)}(x_k)\mu_{(k,j)}(x_k')$ or $\nu_{(k,i)}(x_k)\nu_{(k,j)}(x_k')$. If we regard the ν -edges as a edge-set, each term can be identified with an edge-induced graph. By theorem 2 and lemmas 1 and 2, the term corresponding to C is equal to $r(C)$. \square

Since $f_1(x) = 0$, C makes a contribution to the sum only if C does not have a node of degree one in C . Such C is called a generalized loop. In the case that G is a tree, there is no generalized loop, therefore $Z = Z_B$. This is well known [9].

Note that $f_n(0) = 1$ if n is even, and $f_n(0) = 0$ if n is odd. If $\gamma_i = 0$ for all $i \in V$, only generalized loops with even degrees contribute to the sum. This is the reminiscent of the high temperature expansion of the Ising model without the magnetic field. We further discuss this point in section 5.3. In addition, if the graph is planar, these terms are summed by a single Pfaffian [13, 14].

We can choose and expand some of the edges step by step with propagation diagrams as in appendix A.1, while we expand at all edges in the above proof.

Let $\Theta(\{\beta_{ij}\}, \{\gamma_i\}) := 1 + \sum_{\phi \neq C \in \mathcal{G}} r(C)$. This is a polynomial of indeterminates $\{\beta_{ij}\}$ and $\{\gamma_i\}$, and its coefficients are positive integers, because $f_n(x)$ are polynomials with coefficients positive integers. We can assign other quantities to nodes and edges such as $\{m_i\}$ and $\{\tau_{ij}\}$ discussed later in section 5.2, and represent the formula in different manner with these quantities, but such assignment may cause large positive and negative coefficients. Advantage of using γ and β is that the coefficients are not huge because they are all positive and the total sum is determined by L as discussed later in 5.4.2. Note that the method of propagation diagrams gives an algorithm for computing Θ .

The messages μ and ν are explicitly given if we change the compatibility functions. By the definitions of the beliefs (4) and (5), we see that

$$\prod_{(i,j) \in E} \psi_{ij}(x_i, x_j) \propto \prod_{(i,j) \in E} b_{ij}(x_i, x_j) \prod_{i \in V} \frac{1}{b_i(x_i)^{d_i-1}}. \tag{33}$$

Therefore, we can retake compatibility functions as

$$\psi_{ij}(x_i, x_j) = \frac{b_{ij}(x_i, x_j)}{b_i(x_i)^{(d_i-1)/d_i} b_j(x_j)^{(d_j-1)/d_j}} \tag{34}$$

without changing the joint probability distribution. Moreover, this does not cause any change to the result of the LBP algorithm, namely beliefs. In the rest of section 5.1 and section 5.2, we assume this representation of compatibility functions. By (12), for all $(j, i) \in \vec{E}$ the first messages have the following forms: $\mu_{(j,i)}(x_j) = b_j(x_j)^{1/d_j}$. The secondary messages are determined by theorem 2 as

$$\nu_{(j,i)}(x_j) = \frac{-x_j b_j(-x_j)^{(d_j-1)/d_j}}{b_j(+)^{(d_j-2)/2d_j} b_j(-)^{(d_j-2)/2d_j}} \quad \forall (j, i) \in \vec{E}. \tag{35}$$

Using lemmas 1 and 2 on the transformed graph, we see that

$$\beta_{ij} = \sum_{x_i, x_j} \frac{b_{ij}(x_i, x_j)}{b_i(x_i)^{(d_i-1)/d_i} b_j(x_j)^{(d_j-1)/d_j}} v_{(i,i_1)}(x_i) v_{(j,j_1)}(x_j) \prod_{s=2}^{d_i-1} \mu_{(i,i_s)}(x_i) \prod_{t=2}^{d_j-1} \mu_{(j,j_t)}(x_j), \quad (36)$$

where $N(i) = \{j, i_1, \dots, i_{d_i-1}\}$ and $N(j) = \{i, j_1, \dots, j_{d_j-1}\}$. A direct computation shows that

$$\beta_{ij} = \frac{b_{ij}(+, +) b_{ij}(-, -) - b_{ij}(+, -) b_{ij}(-, +)}{\sqrt{b_i(+)} b_i(-) \sqrt{b_j(+)} b_j(-)} \quad \forall (i, j) \in E. \quad (37)$$

Equation (37) implies that $|\beta_{ij}| \leq 1$. By (5), $\beta_{ij} = 0$ if and only if ψ_{ij} can be factorized by some functions as $\psi_{ij}(x_i, x_j) = \psi_i(x_i) \psi_j(x_j)$; no interaction between nodes i and j .

5.2. Relation to the result of Chertkov and Chernyak

Chertkov and Chernyak [8] show the loop series expansion formula for general vertex models and factor graph models, which are more general than pairwise interaction models considered in this paper. Focusing on pairwise interaction models, however, we found the further relations of the first and secondary messages in theorem 2, which derives our representation of $r(C)$. In this section we show that the expansion formula given in theorem 3 is equivalent to the result of Chertkov and Chernyak in [8]. Let us briefly review their result in the case of a pairwise MRF:

$$\begin{aligned} Z &= Z_B \left(1 + \sum_{C \in \mathcal{G}} \check{r}(C) \right), & \check{r}(C) &= \prod_{ij \in E_C} \tau_{ij} \prod_{i \in V_C} \rho_i(C), \\ \rho_i(C) &= \frac{(1 - m_i)^{d_i(C)-1} + (-1)^{d_i(C)} (1 + m_i)^{d_i(C)-1}}{2(1 - m_i^2)^{d_i(C)-1}}, & (38) \\ \tau_{ij} &= \sum_{x_i, x_j} b_{ij}(x_i, x_j) (x_i - m_i) (x_j - m_j), & m_i &= b_i(1) - b_i(-1). \end{aligned}$$

It suffices to prove $r(C) = \check{r}(C)$ for all generalized loops C .

By the inductive definition of the polynomials f_n , we see that

$$f_n(x) = \frac{\lambda_1^{n-1} - \lambda_2^{n-1}}{\lambda_1 - \lambda_2}, \quad (39)$$

where λ_1, λ_2 are the roots of the quadratic equation $\lambda^2 - x\lambda - 1 = 0$. Using the definition of γ_i , direct calculation derives

$$\rho_i(C) (2\sqrt{b_i(1)b_i(-1)})^{d_i(C)} = f_{d_i(C)}(\gamma_i). \quad (40)$$

Using (37), we see that

$$\tau_{ij} = \beta_{ij} (2\sqrt{b_i(1)b_i(-1)}) (2\sqrt{b_j(1)b_j(-1)}). \quad (41)$$

Combining (40) and (41) gives the claim.

We append two comments on the difference between our approach and theirs. First, we derived the expansion from the viewpoint of message passing operation. The quantities β_{ij} and γ_i are characterized by propagation of messages. On the other hand, Chertkov and Chernyak used covariances and means of the beliefs for the expansion. Secondly, we interpreted the recursion relation of f_n by transformations of the graphs in the proof of theorem 2, though the corresponding relation is not clear in their choice of variables τ_{ij} and ρ_i . The recursion is effectively used for upper bounding the number of generalized loops in section 5.4.2.

5.3. Ising partition function on a regular graph

In this section we briefly discuss the connection between the polynomial $\theta(\beta, \gamma) := \Theta(\{\beta_{ij} = \beta\}, \{\gamma_i = \gamma\})$ and the partition function of the Ising model on a regular graph G . A graph G is called regular if all of the degrees of nodes are the same. We see in corollary 1 that θ can be regarded as a transform of the partition function on the basis of the Bethe approximation, and apply it to the derivation of the susceptibility formula. In this subsection, we assume $\gamma_i = \gamma, \beta_{ij} = \beta$ and G is regular graph of degree d .

Since we have relations (28) and (37), we solve b_{ij} by β and γ as

$$b_{ij}(x_i, x_j) = \frac{1}{4} \left(1 + \frac{x_i \gamma}{\sqrt{4 + \gamma^2}} + \frac{x_j \gamma}{\sqrt{4 + \gamma^2}} + \frac{x_i x_j \gamma^2}{4 + \gamma^2} \right) + \frac{x_i x_j \beta}{4 + \gamma^2}. \tag{42}$$

By (32) and (34), the polynomial θ admits the following identity:

$$\theta(\beta, \gamma) = \sum_{x_i} \prod_{(i,j)} b_{ij}(x_i, x_j) \prod_i b_i(x_i)^{1-d} \tag{43}$$

where b_{ij} is defined by (42) and b_i is defined by $\sum_{x_j=\pm 1} b_{ij}$. Let $Z(K, h) = \sum_{x_i} \exp(K \sum_{ij} x_i x_j + h \sum_i x_i)$ be a partition function. As a function of y and z , we obtain the following identity.

Corollary 1. For a regular graph G of degree d ,

$$\theta(\beta, \gamma) = Z(K, h) \left(\frac{\sqrt{1 - z^2(1 + y^2 z^2)}}{1 - y^2 z^2} \right)^{|E|} \left(\frac{\sqrt{(1 - y^2)(1 - y^2 z^2)}}{2(1 + y^2 z)} \right)^{|V|} \tag{44}$$

where $\beta = (1 - y^2)z/(1 - y^2 z^2)$ and $\gamma = 2y(1 + z)/\sqrt{(1 - y^2)(1 - y^2 z^2)}$. Furthermore, $K = \tanh^{-1} z$ and

$$\exp(2h) = \left(\frac{1 + y}{1 - y} \right) \left(\frac{1 + yz}{1 - yz} \right)^{1-d}. \tag{45}$$

Proof. The proof is accomplished by calculations based on (43). Let $b_{ij}(x_i, x_j) = \exp(K x_i x_j + h' x_i + h' x_j + C)$ and $b_i(x_i) = \exp(h'' x_i + D)$. We define $z = \tanh K$ and $y = \tanh h'$. By (42), β, γ and C are solved by y and z . The condition $b_i = \sum_{x_j=\pm 1} b_{ij}$ determines h'' and D . Let $h = dh' + (1 - d)h''$, then the product of b_{ij} and b_i is proportional to $\exp(K \sum_{ij} x_i x_j + h \sum_i x_i)$. \square

If $y = 0$, then $h = 0, \gamma = 0$ and $\beta = z$. This theorem is reduced to the well-known high temperature expansion. Therefore this formula is an extension of the high temperature expansion of the Ising model with an external field.

We proceed to obtain a formula of zero field susceptibility which is defined by

$$\chi(K) := \frac{d}{dh} \log Z(K, h) \Big|_{h=0}. \tag{46}$$

By the differentiation of (43), we have

$$(1 + z - dz)^2 \chi(K) = (1 + z)(1 + z - dz) + 2z(z^2 - 1) \frac{\partial}{\partial z} \log \theta(z, 0) + 8(1 + z)^2 \frac{\partial}{\partial \gamma^2} \log \theta(z, \gamma) \Big|_{\gamma=0}. \tag{47}$$

If we substitute 1 for θ , which corresponds to the Bethe approximation, this formula reduces to the well-known formula of the Bethe approximation of susceptibility [16]. A higher order approximation can be obtained by enumerating generalized loops that appear in θ . Comparison of traditional ways of enumeration of subgraphs [17, 18] and our expansion is an interesting future research topic.

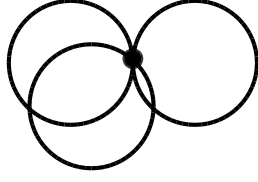


Figure 11. $L = 3$ case.

5.4. Miscellaneous topics

5.4.1. Representation of T_k . Using the first and secondary messages and $\{\beta_{ij}\}$, $T_k^{x_k, x_{\bar{k}}}$ defined in (9) admits a simple representation. Let k be a leaf node of the tree \hat{G} and $N(k) = \{i, j\}$ on the original graph G , and let i_0, i_1, \dots, i_l be the unique path from $i_0 = k$ to $i_l = \bar{k}$ on \hat{G} , where $i_1 = j, i_{l-1} = i$. It is easy to see that $T_k \mu_{(k,i)} = \mu_{(k,i)}$ and $T_k \nu_{(k,i)} = \prod_{t=1}^l \beta_{i_{t-1}i_t} \nu_{(k,i)}$ with propagation diagrams on \hat{G} . Therefore

$$T_k^{x_k, x_{\bar{k}}} = \mu_{(k,i)}(x_{\bar{k}}) \mu_{(k,j)}(x_k) + \prod_{t=1}^l \beta_{i_{t-1}i_t} \nu_{(k,i)}(x_{\bar{k}}) \nu_{(k,j)}(x_k). \tag{48}$$

This shows that $\nu_{(k,j)}$ and $\nu_{(k,i)}$ are the left and right eigenvectors of the matrix T_k and their eigenvalue is $\prod_{t=1}^l \beta_{i_{t-1}i_t}$.

5.4.2. The number of generalized loops. We first show that the polynomial $\theta(1, \gamma)$ depends on the graph G only through L , the number of linearly independent cycles. Since $\beta = 1$, we can shrink any edges without changing corresponding polynomial θ . Any graph with L independent cycles can be reduced to a graph in which only one node has degree more than two. See figure 11: L rings are joined at one point. Therefore, cutting L loops, we obtain

$$\begin{aligned} \theta(1, \gamma) &= \sum_{k=0}^L \binom{L}{k} f_{2k}(\gamma) \\ &= \left(\frac{2\sqrt{4+\gamma^2}}{\gamma + \sqrt{4+\gamma^2}} \right)^{L-1} + \left(\frac{2\sqrt{4+\gamma^2}}{-\gamma + \sqrt{4+\gamma^2}} \right)^{L-1}. \end{aligned} \tag{49}$$

This equality shows that the sum of coefficients of Θ is equal to

$$\theta(1, 1) = \left(\frac{5 - \sqrt{5}}{2} \right)^{L-1} + \left(\frac{5 + \sqrt{5}}{2} \right)^{L-1}. \tag{50}$$

Moreover, we obtain a bound for the number of generalized loops.

Corollary 2. Let \mathcal{G}_0 be the set of all generalized loops of G including an empty set. Then,

$$|\mathcal{G}_0| \leq \left(\frac{5 - \sqrt{5}}{2} \right)^{L-1} + \left(\frac{5 + \sqrt{5}}{2} \right)^{L-1}. \tag{51}$$

This bound is attained if and only if every node of a generalized loop has the degree at most three.

Proof. Since $f_n(1) > 1$ for all $n > 4$ and $f_2(1) = f_3(1) = 1$, we have $r(C)|_{\beta=\gamma=1} \geq 1$ for all $C \in \mathcal{G}_0$, and the equality holds if and only if $d_i(C) \leq 3$ for all $i \in V_C$. This shows $|\mathcal{G}_0| \leq \theta(1, 1)$ and the equality condition. \square

6. Marginal expansion formula

6.1. Derivation of the marginal expansion formula

We show relations between the approximated marginal b_k and the true marginal p_k . In this section, we take compatibility functions as (34) once again. Without loss of generality, we assume $k = 1$ and $N(1) = \{i, j\}$, i.e. node 1 has degree two. Indeed, split of nodes as in figure 3 gives the same marginal probability to the added nodes as the original one. Because

$$\mu_{(1,i)} = \mu_{(1,j)} = \begin{pmatrix} \sqrt{b_1(+)} \\ \sqrt{b_1(-)} \end{pmatrix}, \quad \nu_{(1,i)} = \nu_{(1,j)} = \begin{pmatrix} -\sqrt{b_1(-)} \\ \sqrt{b_1(+)} \end{pmatrix},$$

we see the following equations by a direct computation:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = b_1(+)\mu_{(1,i)}\mu_{(1,j)}^T + b_1(-)\nu_{(1,i)}\nu_{(1,j)}^T - \sqrt{b_1(+)}\sqrt{b_1(-)}\mu_{(1,i)}\nu_{(1,j)}^T - \sqrt{b_1(+)}\sqrt{b_1(-)}\nu_{(1,i)}\mu_{(1,j)}^T, \quad (52)$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = b_1(-)\mu_{(1,i)}\mu_{(1,j)}^T + b_1(+)\nu_{(1,i)}\nu_{(1,j)}^T + \sqrt{b_1(+)}\sqrt{b_1(-)}\mu_{(1,i)}\nu_{(1,j)}^T + \sqrt{b_1(+)}\sqrt{b_1(-)}\nu_{(1,i)}\mu_{(1,j)}^T. \quad (53)$$

By the definition of marginal probability, we see that

$$\frac{Z}{Z_B} p_1(\pm) = \sum_{x_1, x_{\bar{1}}} T_{x_1, x_{\bar{1}}}^{x_1} I_{\{x_1 = x_{\bar{1}} = \pm\}}, \quad (54)$$

where $T_{x_1, x_{\bar{1}}}^{x_1} := \sum_{x_2, \dots, x_L} T_{x_1, \dots, x_L}^{x_1, \dots, x_L}$ and I is the indicator function. Using these equations, we can show the following theorem.

Theorem 4.

$$\begin{aligned} \frac{Z}{Z_B} p_1(\pm) &= b_1(\pm) \left(\sum_{x_1} T_{x_1, x_{\bar{1}}}^{x_1} \mu_{(1,i)}(x_1) \mu_{(1,j)}(x_1) \right) \\ &\quad + b_1(\mp) \left(\sum_{x_1} T_{x_1, x_{\bar{1}}}^{x_1} \nu_{(1,i)}(x_1) \nu_{(1,j)}(x_1) \right) \mp \sqrt{b_1(+)}\sqrt{b_1(-)} \\ &\quad \times \left(\sum_{x_1} T_{x_1, x_{\bar{1}}}^{x_1} \mu_{(1,i)}(x_1) \nu_{(1,j)}(x_1) + \sum_{x_1} T_{x_1, x_{\bar{1}}}^{x_1} \nu_{(1,i)}(x_1) \mu_{(1,j)}(x_1) \right). \end{aligned} \quad (55)$$

The four summation terms appeared in (55) can be expanded with propagation diagrams.

This expansion is computationally intractable if L is large, in a similar way to the partition function Z . For relatively small graphs, however, we may be able to expand these terms. The terms in the expansion of the first two summations are labelled by the generalized loops, while the other terms in expansion of the last two summations are labelled by other subgraphs: each subgraph does not have nodes of degree one except the node 1. The expansion may be heuristically used for an approximate computation of marginal probability distributions, namely we can correct beliefs using terms corresponding to major subgraphs in expanded representation of (55).

With this theorem, an already known fact is easily deduced [19].

$$\begin{aligned}
 p_1(+)-p_1(-) &\propto \gamma_1 \left(\text{Diagram 1} - \text{Diagram 2} \right) \\
 &\quad - 2 \left(\text{Diagram 3} + \text{Diagram 4} \right) \\
 &= \gamma_1(1 - \beta_{12}\beta_{23}\beta_{13}) - 2(0 + 0)
 \end{aligned}$$

Figure 12. An example of expansion with propagation diagrams in one loop cases.

Corollary 3. Letting $L = 1$ and node 1 is on the unique cycle in G , $p_1(+)-p_1(-)$ and $b_1(+)-b_1(-)$ have the same sign.

Proof. By theorem 4,

$$\begin{aligned}
 \frac{Z}{Z_B} \frac{p_1(+)-p_1(-)}{\sqrt{b_1(+)}b_1(-)} &= \gamma_1 \left(\sum_{x_1} T_1^{x_1} \mu_{(1,i)}(x_1)\mu_{(1,j)}(x_1) - \sum_{x_1} T_1^{x_1} \nu_{(1,i)}(x_1)\nu_{(1,j)}(x_1) \right) \\
 &\quad - 2 \left(\sum_{x_1} T_1^{x_1} \mu_{(1,i)}(x_1)\nu_{(1,j)}(x_1) + \sum_{x_1} T_1^{x_1} \nu_{(1,i)}(x_1)\mu_{(1,j)}(x_1) \right). \tag{56}
 \end{aligned}$$

The right-hand side can be expanded with propagation diagrams as in figure 12. The first summation is 1, the second is a product of β , and the third and fourth is equal to 0. Since $|\beta_{ij}| \leq 1$ the result follows. \square

A problem of finding an assignment that maximizes the marginal probability p_1 is called the maximum marginal assignment problem [19]. This corollary asserts that the assignment that maximizes the belief b_1 is the solution of this problem.

6.2. Example

Consider a graph in figure 2. In this case theorem 4 turns out to be

$$\begin{aligned}
 \frac{Z}{Z_B} p_1(\pm) &= b_1(\pm)(1 + \beta_{23}\beta_{34}\beta_{24}) + b_1(\mp)\beta_{14}\beta_{13}(\beta_{34} + \beta_{34}\beta_{24}\beta_{23}\gamma_3\gamma_4) \\
 &\quad \mp \sqrt{b_1(+)}b_1(-)(\beta_{13}\beta_{34}\beta_{23}\beta_{24}\gamma_3 + \beta_{14}\beta_{34}\beta_{23}\beta_{24}\gamma_4). \tag{57}
 \end{aligned}$$

By (57), we have

$$\begin{aligned}
 \frac{Z}{Z_B} \frac{p_1(+)-p_1(-)}{\sqrt{b_1(+)}b_1(-)} &= \gamma_1(1 + \beta_{23}\beta_{34}\beta_{24} - \beta_{14}\beta_{13}\beta_{34} - \beta_{14}\beta_{13}\beta_{34}\beta_{24}\beta_{23}\gamma_3\gamma_4) \\
 &\quad - 2(\beta_{13}\beta_{34}\beta_{23}\beta_{24}\gamma_3 + \beta_{14}\beta_{34}\beta_{23}\beta_{24}\gamma_4). \tag{58}
 \end{aligned}$$

The right-hand side discriminates which state is more plausible. Let $\beta_{34} = 0$ in (58), this expression is reduced to the case of $L = 1$ and consistent with the result of corollary 3. Let $\beta_{13} = 0$. In this case we see that $p_1(+)-p_1(-)$ does not necessarily have the same sign as γ_1 [19].

If, for example, $|\beta_{ij}| \leq 1/2$, $|\gamma_3|, |\gamma_4| \leq 1$ and $|\gamma_3|/2, |\gamma_2|/2 \leq |\gamma_1|$, then we see from (58) that $p_1(+)-p_1(-)$ and γ_1 have the same sign. The first condition requires weakness of the interactions, and the second condition requires that the beliefs at the nodes of degree three

$$\begin{aligned}
 &= \text{Diagram 1} + \text{Diagram 2} \\
 &+ \text{Diagram 3} + \text{Diagram 4} \\
 &= 1 + \beta_{14}\beta_{34}\beta_{13} + \beta_{23}\beta_{34}\beta_{24} \\
 &+ \text{Diagram 5} + \text{Diagram 6} \\
 &= 1 + \beta_{14}\beta_{34}\beta_{13} + \beta_{23}\beta_{34}\beta_{24} \\
 &\quad + \beta_{13}\beta_{14}\beta_{23}\beta_{24} + \beta_{13}\beta_{14}\beta_{23}\beta_{24}\beta_{34}\gamma_3\gamma_4
 \end{aligned}$$

Figure A1. Expansion of the partition function with the method of propagation diagrams. It is not necessary to expand all edges as in the proof of theorem 3.

are not too much biased. The last condition is satisfied if $p_1(+)$ and $p_1(-)$ are not too close to each other.

If we take variables m_i and τ_{ij} in section 5.2 instead of γ_i and β_{ij} , the expressions (57) and (58) become more complicated in general, and it is hard to find simple conditions.

7. Concluding remarks

We introduced propagation diagrams that enable us to compute loop series expansion of a partition function and marginal distributions with a set of simple rules. In this method, parameters β_{ij} and γ_i are naturally assigned to each edge and node.

Accuracy of the Bethe approximation depends both on the strength of interactions and the topology of the underlying graph. The effect of the interactions is captured by the values of β and γ . The topological aspect of the graph, in the sense of the Bethe approximation, is extracted in the polynomial Θ .

We suggest future research topics. First, understanding of the structure of the polynomial Θ is important to construct efficient approximation algorithms exploiting graph topology. The properties of Θ should be investigated further. Secondly, on the basis of the results of this paper, it is interesting to understand the empirically known fact: if the LBP does not converge, the quality of the Bethe approximation is low [20]. Since we show a direct relation between the message passing operation and the expansion variables β and γ , convergence of the LBP algorithm can be analysed using them.

Acknowledgments

This work was supported in part by Grant-in-Aid for JSPS Fellows 20-993 and Grant-in-Aid for Scientific Research (C) 19500249.

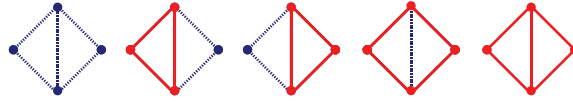


Figure A2. Light parts are generalized loops.

Appendix

A.1. Example of expansion

We consider the graph in figure 2. Normalizing $Z_B = 1$, we can calculate the loop expansion of the partition function as in the following figure, using propagation diagrams.

Five terms in the final expression of figure A1 correspond to the subgraphs in figure A2 respectively.

References

- [1] Bethe H 1935 *Proc. R. Soc. A* **150** 552–75
- [2] Murphy K, Weiss Y and Jordan M 1999 *Proc. of Uncertainty in AI* 467–75
- [3] Yedidia J, Freeman W and Weiss Y 2001 *Adv. Neural Inf. Process. Syst.* **13** 689–95
- [4] Gallager R 1962 *IEEE Trans. Inf. Theory* **8** 21–8
- [5] McEliece R and Cheng D 1998 *IEEE J. Sel. Areas Commun.* **16** 140–52
- [6] Freeman W, Pasztor E and Carmichael O 2000 *Int. J. Comput. Vis.* **40** 25–47
- [7] Chertkov M and Chernyak V 2006 *Phys. Rev. E* **73** 65102
- [8] Chertkov M and Chernyak V 2006 *J. Stat. Mech: Theory Exp.* **6** P06009
- [9] Pearl J 1988 *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Mateo, CA: Morgan Kaufmann Publishers)
- [10] Tatikonda S and Jordan M 2002 *Proc. Uncertainty in AI* vol 18 pp 493–500
- [11] Yellen J and Gross J 1998 *Graph Theory and Its Applications* (Boca Raton, FL: CRC Press) p 152
- [12] Baxter R 1982 *Exactly Solved Models in Statistical Mechanics* (London: Academic) p 47
- [13] Fisher M 1966 *J. Math. Phys.* **7** 1776–81
- [14] Chertkov M, Chernyak V and Teodorescu R 2008 *J. Stat. Mech: Theory Exp.* **5** P05003
- [15] Chernyak V and Chertkov M 2007 *Proc. of ISIT* arXiv:cs.IT/0701086
- [16] Domb C and Sykes M 1957 *Proc. R. Soc. A* **240** 214–28
- [17] Domb C and Sykes M 1957 *Phil. Mag.* **2** 733–49
- [18] Sykes M 1961 *J. Math. Phys.* **2** 52–62
- [19] Weiss Y 2000 *Neural Comput.* **12** 1–41
- [20] Mooij J and Kappen H 2005 *J. Stat. Mech: Theory Exp.* **11** P11012