Theoretical Analysis of Accuracy of Gaussian Belief Propagation

Yu Nishiyama[1] and Sumio Watanabe[2]

[1] Department of Computational Intelligence and Systems Science,
Tokyo Institute of Technology,
4259, Nagatuta, Midori-ku, Yokohama, 226-8503 Japan
nishiyudesu@cs.pi.titech.ac.jp
[2] Precision and Intelligence Laboratory, Tokyo Institute of Technology,
4259, Nagatsuta, Midori-ku, Yokohama, 226-8503 Japan
swatanab@pi.titech.ac.jp

Abstract. Belief propagation (BP) is the calculation method which enables us to obtain the marginal probabilities with a tractable computational cost. BP is known to provide true marginal probabilities when the graph describing the target distribution has a tree structure, while do approximate marginal probabilities when the graph has loops. The accuracy of loopy belief propagation (LBP) has been studied. In this paper, we focus on applying LBP to a multi-dimensional Gaussian distribution and analytically show how accurate LBP is for some cases.

1 Introduction

Belief propagation (BP) is the calculation method which enables us to obtain the marginal probabilities with a tractable computational cost and has been widely studied in the areas such as probabilistic inference for artificial intelligence, turbo codes, code division multiple access (CDMA) systems, low-density parity check (LDPC) codes, and probabilistic image processing. BP is known to provide true marginal probabilities when the graph describing the target distribution has a tree (singly connected) structure, while do approximate marginal probabilities when the graph has loops (cycles).

The accuracy of loopy belief propagation (LBP) has been theoretically studied when the target distributions are discrete distributions [1]-[3] and Gaussian Markov random fields mainly about the means [4]. The accuracy has also been studied from the viewpoint of information geometry [5][6]. In probabilistic image processing, the accuracy has been numerically studied as the accuracy of estimated hyperparameters based on Gaussian graphical models [7][8].

In this paper, we focus on applying loopy belief propagation (LBP) to a multi-dimensional Gaussian distribution whose inverse covariance matrix corresponds to the graph with loops. Then, we mathematically show the differences between true marginal densities and the approximate marginal densities calculated by LBP. To be more specific, we give the exact solutions of messages and approximate marginal densities calculated by LBP and give the Kullback-Leibler (KL) distances when the inverse covariance matrix corresponds to the graph with a single loop. Next, we develop the results to the more general case that the graph has a single loop and an arbitrary tree structure. Furthermore, we give the expansions of inverse variances of approximate marginal densities up to third-order when the graph is allowed to have an arbitrary structure (i.e., multiloops).

The situation in which LBP is applied to a multi-dimensional Gaussian distribution can be seen for example in probabilistic image processing based on Gaussian graphical models [7][8] and Markov random fields [4]. Besides, clarifying the theoretical properties of LBP in the standard models such as a Gaussian distribution helps to know the properties of more complex LBP and to design LBP algorithms efficiently.

This paper is organized as follows. In section **2**, we review the scheme for belief propagation. In section **3**, we give the main results of this paper. In section **4**, we give the conclusion and future works.

2 Belief Propagation

Here, we review the scheme for Belief propagation. We review the belief propagation for tree structures in section **2.1**, loopy belief propagation in section **2.2**, and the LBP when we apply it to a multi-dimensional Gaussian distribution in section **2.3**.

2.1 Belief Propagation for tree structures

Let the target distribution to which we apply LBP be the that of pairwise form

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{\{ij\} \in B} W_{ij}(x_i, x_j), \tag{1}$$

where Z is the normalization constant and B is the set which shows the existences of correlations between x_i and x_j. For example, when $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ and correlations exist between $\{x_1, x_2\}$, $\{x_2, x_3\}$, and $\{x_1, x_4\}$ respectively, the set B is expressed as $B = \{\{12\}, \{23\}, \{14\}\}$. When the graph given by B is a tree graph, the marginal distributions $\{p_i(x_i)\}$, $\{p_{ij}(x_i, x_j)\}$ are exactly calculated as follows by using normalized messages $\{\mathcal{M}_{i \to j}\}$, $\{\mathcal{M}_{j \to i}\}$, $(\{ij\} \in B)$.

$$\begin{aligned} p_i(x_i) &= \frac{1}{Z_i} \prod_{k \in \mathcal{N}_i} \mathcal{M}_{k \to i}(x_i), \\ p_{ij}(x_i, x_j) &= \frac{1}{Z_{ij}} \Bigg(\prod_{k \in \mathcal{N}_i \setminus \{j\}} \mathcal{M}_{k \to i}(x_i) \Bigg) W_{ij}(x_i, x_j) \Bigg(\prod_{k \in \mathcal{N}_j \setminus \{i\}} \mathcal{M}_{k \to j}(x_j) \Bigg). \end{aligned} \tag{2}$$

Here, both Z_i and Z_{ij} are the normalization constants and $\mathcal{N}_i$ is the subset of random variables which directly correlate with random variables x_i. $\mathcal{N}_i$ is so-called the set of nearest neighbor variables of x_i.

2.2 Loopy Belief Propagation

Eqs.(2) are exactly correct when the graph given by B is a tree graph. In loopy belief propagation, we also compose marginal distributions by eqs.(2) when the graph given by B has loops. This leap of logic yields the problem of convergence of LBP and gives marginal distributions approximately. Since the marginal distributions $\{p_i(x_i)\}$ and $\{p_{ij}(x_i, x_j)\}$ should satisfy the constraints $\int p_{ij}(x_i, x_j)dx_j = p_i(x_i)$ for consistency, messages $\{\mathcal{M}_{i\to j}\}$ satisfy the equations given by

$$\mathcal{M}_{i\to j}(x_j) = \frac{1}{\tilde{Z}_{ij}} \int_{-\infty}^{\infty} W_{ij}(x_i, x_j) \prod_{k\in\mathcal{N}_i\setminus\{j\}} \mathcal{M}_{k\to i}(x_i)dx_i,$$
$$\mathcal{M}_{j\to i}(x_i) = \frac{1}{\tilde{Z}_{ji}} \int_{-\infty}^{\infty} W_{ij}(x_i, x_j) \prod_{k\in\mathcal{N}_j\setminus\{i\}} \mathcal{M}_{k\to j}(x_j)dx_j. \quad (3)$$

Note that there are $2|B|$ equations in total so that eqs.(3) are the decision equations for $2|B|$ messages $\{\mathcal{M}_{i\to j}\}$.

2.3 Gaussian Belief Propagation

We assume that the target probability density in eq.(1) is a multi-dimensional Gaussian probability density whose mean vector is $\mathbf{0}$ and $\mathbf{x} \in R^d$:

$$p(\mathbf{x}) = \sqrt{\frac{\det S}{(2\pi)^d}} \exp\Big\{-\frac{1}{2}\mathbf{x}^T S\mathbf{x}\Big\}. \quad (4)$$

Here, S is an inverse covariance matrix and we denote the components of the matrix S by $(S)_{ij} = s_{i,j}$. Then, $W_{ij}(x_i, x_j)$ in eq.(1) can be expressed as

$$W_{ij}(x_i, x_j) = \exp\Big\{-\frac{1}{2}(\frac{s_{i,i}}{|\mathcal{N}_i|}x_i^2 + s_{i,j}x_ix_j + \frac{s_{j,j}}{|\mathcal{N}_j|}x_j^2)\Big\}, \quad (5)$$

where $|\mathcal{N}_i|$ is the number of elements of subset $\mathcal{N}_i$. For simplicity, we put $\frac{s_{i,i}}{|\mathcal{N}_i|} = \tilde{s}_{i,i}$. We set the condition $|\mathcal{N}_i| > 0$ for $^\forall i \in \{1, \cdots, d\}$.

We assume that the probability densities of messages $\{\mathcal{M}_{i\to j}\}$ are Gaussian densities:

$$\mathcal{M}_{i\to j}(x_j) = \sqrt{\frac{\lambda_{i\to j}}{2\pi}} \exp\Big\{-\frac{\lambda_{i\to j}}{2}x_j^2\Big\}. \quad (6)$$

Then, by substituting eqs.(6) into eqs.(3), we reduce eqs.(3) to the decision equations for the parameters of inverse variances $\{\lambda_{i\to j}\}$ as follows.

$$\tilde{\lambda}_{i\to j} = -\frac{s_{i,j}^2}{s_{i,i} + \sum_{k\in\mathcal{N}_i\setminus\{j\}} \tilde{\lambda}_{k\to i}}. \quad (7)$$

Here, $\{\tilde{\lambda}_{i\to j}\}$ are $\tilde{\lambda}_{i\to j} \equiv \lambda_{i\to j} - \tilde{s}_{j,j}$. After obtaining the values of $\{\tilde{\lambda}_{i\to j}\}$ which satisfy eqs.(7), we compose messages $\{\mathcal{M}_{i\to j}\}$ by eqs.(6). Next, we compose marginal densities $\{\tilde{p}_i(x_i)\}$ and $\{\tilde{p}_{ij}(x_i, x_j)\}$ approximately by substituting eqs.(6) into eqs.(2). That is, we obtain marginal densities by substituting $\{\lambda_{i\to j}\}$ which satisfy eqs.(7) into

$$\tilde{p}_i(x_i) \propto \exp\Big\{-\frac{\sum_{k\in\mathcal{N}_i}\lambda_{k\to i}}{2}x_i^2\Big\}, \quad \tilde{p}_{ij}(x_i,x_j) \propto \exp\Big\{-(x_i,x_j)\frac{\tilde{S}_{i,j}}{2}\begin{pmatrix} x_i \\ x_j \end{pmatrix}\Big\},$$

$$\tilde{S}_{i,j} \equiv \begin{pmatrix} \tilde{s}_{i,i} + \sum_{k\in\mathcal{N}_i\backslash\{j\}}\lambda_{k\to i} & s_{i,j} \\ s_{i,j} & \tilde{s}_{j,j} + \sum_{k\in\mathcal{N}_j\backslash\{i\}}\lambda_{k\to j} \end{pmatrix}. \tag{8}$$

Throughout this paper, we put the inverse variances of approximate marginal densities calculated by LBP as $\Lambda_i(\equiv \sum_{k\in\mathcal{N}_i}\lambda_{k\to i})$.

3 Main results of this paper

We consider the differences between true and approximate marginal densities when we apply LBP to a multi-dimensional Gaussian density. We consider the differences in each case (**3.1**, **3.2**, **3.3**). In section **3.1**, we give the exact solutions of messages $\{\mathcal{M}_{i\to j}\}$ (equally the inverse variances $\{\lambda_{i\to j}\}$), approximate marginal densities $\tilde{p}_i(x_i)$ and $\tilde{p}_{ij}(x_i,x_j)$ (equally $\{\Lambda_i\}$ and $\{\tilde{S}_{i,i+1}\}$), and the KL distances when inverse covariance matrix S corresponds to the graph with a single loop. In section **3.2**, we develop the results of **3.1** to the graph with a single loop and tree structures. In section **3.3**, we give the expansions of inverse covariances of marginal densities $\{\Lambda_i\}$ when the covariances are very small. All the proofs in section **3** are shown in [9].

3.1 On the Graphs with a Single Loop

We consider the case in which the inverse covariance matrix S can be described as the graph which have only a single loop as illustrated in Fig.1. Then, without loss of generality, we can consider the case that the set B is $B = \{\{12\},\{23\},\cdots,\{d-1d\},\{d1\}\}$. For the graph with a single loop, we obtain the following theorem.

Theorem 1. *When the inverse covariance matrix S corresponds to the graph given by $B = \{\{12\},\{23\},\cdots,\{d-1d\},\{d1\}\}$, the inverse variances of messages $\{\lambda_{i\to j}\}$ are given by*

$$\lambda_{i\to i+1} = \frac{s_{i,i+1}\Delta_{i,i+1} - s_{i+1,i+2}\Delta_{i+1,i+2} \pm \sqrt{\mathcal{D}}}{2\Delta_{i+1,i+1}},$$

$$\lambda_{i\to i-1} = \frac{s_{i,i-1}\Delta_{i,i-1} - s_{i-1,i-2}\Delta_{i-1,i-2} \pm \sqrt{\mathcal{D}}}{2\Delta_{i-1,i-1}},$$

$$\mathcal{D} \equiv (\det S)^2 + (-1)^d 4 s_{1,2}s_{2,3}\cdots s_{d-1,d}s_{d,1}\det S, \tag{9}$$

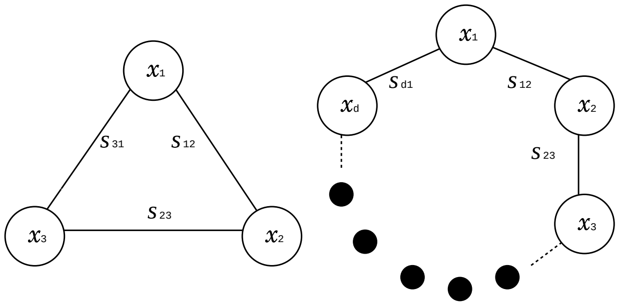

Fig. 1. examples of a single loop

where $i \in \{1, 2, \cdots, d\}$ *and periodic boundary conditions* $s_{d,d+1} \equiv s_{d,1}$, $s_{1,0} \equiv s_{1,d}$ *hold.* $\Delta_{i,j}$ *are the cofactors of the matrix* S.

Theorem 1 gives immediately the following corollary.

Corollary 1. *When the inverse covariance matrix* S *corresponds to the graph given by* $B = \{\{12\}, \{23\}, \cdots, \{d-1d\}, \{d1\}\}$*, the inverse variances* $\{\Lambda_i\}$ *and the inverse covariance matrices* $\{\tilde{S}_{i,i+1}\}$ *of approximate marginal densities are given by*

$$\Lambda_i = \frac{\det S}{\Delta_{i,i}}\left(1 + \frac{(-1)^d 4 s_{1,2} \cdots s_{d,1}}{\det S}\right)^{\frac{1}{2}}, \quad \tilde{S}_{i,i+1} = \begin{pmatrix} \frac{E_{i,i+1}}{\Delta_{i,i}} & s_{i,i+1} \\ s_{i,i+1} & \frac{E_{i,i+1}}{\Delta_{i+1,i+1}} \end{pmatrix}, \quad (10)$$

where

$$E_{i,i+1} \equiv \frac{\det S + \sqrt{\mathcal{D}}}{2} - s_{i,i+1}\Delta_{i,i+1}. \quad (11)$$

In corollary 1, if $s_{1,2}, s_{2,3}, \cdots, s_{d,1}$ approach 0, we know that Λ_i and $\tilde{S}_{i,i+1}$ approach

$$\Lambda_i \to \frac{\det S}{\Delta_{i,i}}, \quad \tilde{S}_{i,i+1} \to \begin{pmatrix} \frac{\det S}{\Delta_{i,i}} & 0 \\ 0 & \frac{\det S}{\Delta_{i+1,i+1}} \end{pmatrix} \quad (12)$$

respectively since $E_{i,i+1} \to \det S$. These results are equivalent to the inverse covariance matrices of the true marginal densities. Taking the limit as $s_{1,2}, s_{2,3}, \cdots$, $s_{d,1} \to 0$ means that the graph with a single loop (e.g., Fig.1) turns to a tree graph and LBP can calculate marginal densities exactly. Hence, corollary 1 agrees with and explains the well-known fact that belief propagation is exactly correct when the graph provided by a target distribution is a tree graph but incorrect when the graph has loops.

When we calculate the KL distances between true and approximate marginal densities in order to know the differences between both densities, we obtain the following theorem. Here, KL and symmetrized KL (SKL) distances are defined as follows.

$$\mathrm{KL}(q(\mathbf{x})||p(\mathbf{x})) \equiv \int q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}, \tag{13}$$

$$\mathrm{SKL}(q(\mathbf{x}), p(\mathbf{x})) \equiv \mathrm{KL}(q(\mathbf{x})||p(\mathbf{x})) + \mathrm{KL}(p(\mathbf{x})||q(\mathbf{x})). \tag{14}$$

We note that a SKL distance satisfies the definition of distance although a KL distance does not (since $\mathrm{KL}(q||p) \neq \mathrm{KL}(p||q)$).

Theorem 2. *The KL and SKL distances between true marginal densities $p_i(x_i)$ and approximate marginal densities $\tilde{p}_i(x_i)$ calculated by LBP are given by*

$$\begin{aligned}\mathrm{KL}(p_i||\tilde{p}_i) &= -\frac{1}{2} + \frac{1}{2}(1+\epsilon)^{\frac{1}{2}} - \frac{1}{4}\log\Big(1+\epsilon\Big), \\ \mathrm{SKL}(p_i, \tilde{p}_i) &= -1 + \frac{1}{2}(1+\epsilon)^{-\frac{1}{2}} + \frac{1}{2}(1+\epsilon)^{\frac{1}{2}}.\end{aligned} \tag{15}$$

where ϵ is given by

$$\epsilon \equiv \frac{(-1)^d 4 s_{1,2} \cdots s_{d,1}}{\det S}. \tag{16}$$

Similarly, the KL and SKL distances between $p_{ij}(x_i, x_j)$ and $\tilde{p}_{ij}(x_i, x_j)$ are given by

$$\begin{aligned}\mathrm{KL}(p_{i,i+1}||\tilde{p}_{i,i+1}) &= -1 - \frac{1}{2}\log(1-\bar{\Delta}_{i,i+1}) + \frac{1+\sqrt{1+\epsilon}}{2} + \frac{s_{i,i+1}\Delta_{i+1,i}}{\det S} \\ &\quad -\frac{1}{2}\log\left\{ \left(\frac{1+\sqrt{1+\epsilon}}{2}\right)^2 - \frac{s_{i,i+1}^2 \Delta_{i,i}\Delta_{i+1,i+1}}{(\det S)^2} \right\}, \\ \mathrm{SKL}(p_{i,i+1}, \tilde{p}_{i,i+1}) &= -2 + \Big[1 + \frac{1}{1-\bar{\Delta}_{i,i+1}} \Big\{ \Big(\frac{1+\sqrt{1+\epsilon}}{2} - \frac{s_{i,i+1}\Delta_{i,i+1}}{\det S}\Big)^2 \\ &\quad - \frac{s_{i,i+1}^2 \Delta_{i,i}\Delta_{i+1,i+1}}{(\det S)^2} \Big\}^{-1} \Big] \frac{1+\sqrt{1+\epsilon}}{2},\end{aligned} \tag{17}$$

where we put $\bar{\Delta}_{i,i+1} \equiv \frac{\Delta_{i,i+1}^2}{\Delta_{i,i}\Delta_{i+1,i+1}}$.

From the above theorem 2, we can know that both KL and SKL distances go to 0 as ϵ tends to 0 since $\bar{\Delta}_{i,i+1}$ tends to 0. We can say that ϵ is the control parameter which decides the distances between both marginal densities. The parameter ϵ decides the accuracy of LBP in the case that the target distribution is a multi-dimensional Gaussian density whose inverse covariance matrix S forms the graph with a single loop.

3.2 On the Graphs with a Single Loop and Tree Structures

From the results for a single loop in section **3.1**, we can immediately develop the results to the more general case that the graph given by B has a single

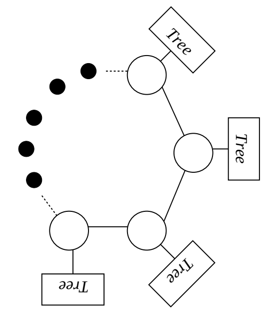

Fig. 2. graphs with a single loop and tree structures

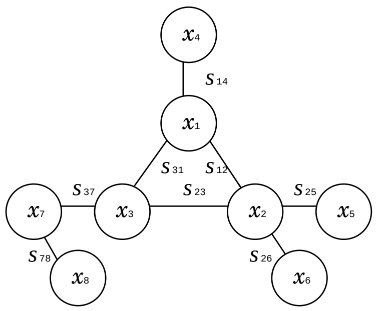

Fig. 3. an example

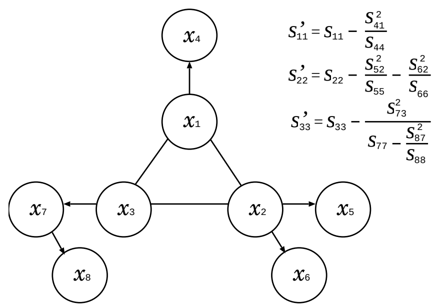

Fig. 4. the same graph as Fig.3

loop and an arbitrary tree structure as illustrated in Fig. 2. For example, we consider the graph shown by Fig.3. Then, in practice, we calculate the inverse variances $\{\Lambda_i\}$. The variances $\{\Lambda_i\}$ are obtained by solving eqs.(7) when $d = 8$ and $B = \{\{12\}, \{23\}, \{31\}, \{14\}, \{25\}, \{26\}, \{37\}, \{78\}\}$. The graph in Fig.3 can be equivalent to the graph in Fig.4 after replacing inverse variances $\{s_{i,i}\}$ as

$$s_{1,1} \to s'_{1,1} \left(\equiv s_{1,1} - \frac{s_{4,1}^2}{s_{4,4}} \right), \quad s_{2,2} \to s'_{2,2} \left(\equiv s_{2,2} - \frac{s_{5,2}^2}{s_{5,5}} - \frac{s_{6,2}^2}{s_{6,6}} \right),$$
$$s_{3,3} \to s'_{3,3} \left(\equiv s_{3,3} - \frac{s_{7,3}^2}{s_{7,7} - \frac{s_{8,7}^2}{s_{8,8}}} \right) \tag{18}$$

(These replacements are understandable by solving eqs.(7) directly). In Fig.4, the directed edges mean that there exist messages in those directions but not in reverse directions (e.g., message $\mathcal{M}_{1\to 4}(x_4)$ exists but message $\mathcal{M}_{4\to 1}(x_1)$ does not exist between x_1 and x_4). If we focus on variable nodes x_1, x_2, and x_3, these variable nodes form a closed single loop. Then, referring to the result of corollary 1, we obtain

$$\Lambda'_i = \frac{\det S'}{\Delta'_{i,i}} \left(1 - \frac{4 s_{1,2} s_{2,3} s_{3,1}}{\det S'}\right)^{\frac{1}{2}} \tag{19}$$

for $i \in \{1, 2, 3\}$. Here, S' and $\Delta'_{i,i}$ are the inverse covariance matrix and the cofactor respectively composed by $\{s'_{i,i}\}$ instead of $\{s_{i,i}\}$. The others $\{\Lambda_i\}$, $i \in \{4, \cdots, 8\}$ are calculated as follows by using the notation of continued fractions.

$$\Lambda'_4 = s_{4,4} - \frac{s_{1,4}^2}{\Lambda'_1 +}\frac{s_{4,1}^2}{s_{4,4}}, \quad \Lambda'_5 = s_{5,5} - \frac{s_{2,5}^2}{\Lambda'_2 +}\frac{s_{5,2}^2}{s_{5,5}}, \quad \Lambda'_6 = s_{6,6} - \frac{s_{2,6}^2}{\Lambda'_2 +}\frac{s_{6,2}^2}{s_{6,6}},$$
$$\Lambda'_8 = s_{8,8} - \frac{s_{7,8}^2}{\Lambda'_7 +}\frac{s_{8,7}^2}{s_{8,8}}, \quad \Lambda'_7 = s_{7,7} - \frac{s_{8,7}^2}{s_{8,8}} - \frac{s_{3,7}^2}{\Lambda'_3 +}\frac{s_{7,3}^2}{s_{7,7} -}\frac{s_{8,7}^2}{s_{8,8}} \tag{20}$$

(These equations are also understandable by solving eqs.(7)).

In a similar way, we obtain $\{\lambda_{ij}\}$, $\{\Lambda_i\}$, $\{\tilde{S}_{i,j}\}$, KL, and SKL distances in various connected graphs which have at most a single loops.

3.3 Expansions of $\{\Lambda_i\}$ at Small Covariances

In sections **3.1** and **3.2**, we address the inverse covariance matrix S whose graph structure is restricted to the graph with a single loop. In this section, we address the inverse covariance matrix S whose graph is allowed to have an arbitrary structure (i.e. the graph has multiloops) but the covariance is very small. To achieve that, we introduce a new parameter s and change the off-diagonal components of the matrix S to $(S)_{ij} = ss_{i,j}$, where s satisfy $0 \leq s \leq 1$. If $s = 0$, the matrix S turns to a diagonal matrix and if $s = 1$, the matrix S turns to the original matrix S in eq.(4). Then, our aim is to obtain the expansions of the inverse variances $\{\Lambda_i\}$ with respect to s:

$$\Lambda_i(s) = a_{i0} - a_{i1}s - a_{i2}s^2 - a_{i3}s^3 + \cdots \tag{21}$$

when s satisfies $s \ll 1$.

In this paper, for simplicity, we derive the terms up to third-order. Before that, we prepare the following theorem for simultaneous equations which $\{\Lambda_i\}$ satisfy.

Theorem 3. *The approximate inverse variances $\{\Lambda_i\}$ calculated by LBP satisfy either of $\sum_{i=1}^{d} 2^{|\mathcal{N}_i|}$ simultaneous equations*

$$|\mathcal{N}_i| - 2 + \frac{2s_{i,i}}{\Lambda_i} = \sum_{j \in \mathcal{N}_i} \pm\sqrt{1 + \frac{4s^2 s_{j,i}^2}{\Lambda_j \Lambda_i}}, \quad i \in \{1, \cdots, d\}, \tag{22}$$

where $\pm$ takes an arbitrary sign for each term in the summation.

There exist $\sum_{i=1}^{d} 2^{|\mathcal{N}_i|}$ simultaneous equations since the sign $\pm$ assigns either of positive or negative for each term. In addition, we note that there exist extra simultaneous equations which have no solution. For example, there is no solution when $|\mathcal{N}_i| \geq 2$ for $^{\exists}i$ and all the signs in the summation are negative.

When $s = 0$ in the matrix S, approximate inverse variance Λ_i should be equal to the true inverse variance $s_{i,i}$ (i.e., $\Lambda_i(0) = s_{i,i}$ for $i \in \{1, \cdots, d\}$) since the matrix S becomes a diagonal matrix and LBP algorithm turns to BP algorithm for tree structures. By imposing these conditions, we obtain the following theorem.

Theorem 4. *For $\sum_{i=1}^{d} 2^{|\mathcal{N}_i|}$ simultaneous equations in Theorem 3, conditions $\Lambda_i(0) = s_{i,i}$, $i \in \{1, \cdots, d\}$ holds if and only if all the signs in the summation are positive. That is, the simultaneous equations are given by*

$$|\mathcal{N}_i| - 2 + \frac{2s_{i,i}}{\Lambda_i} = \sum_{j \in \mathcal{N}_i} \sqrt{1 + \frac{4s^2 s_{j,i}^2}{\Lambda_j \Lambda_i}}, \quad i \in \{1, \cdots, d\}. \tag{23}$$

Then, the solutions of the inverse variances $\{\Lambda_i\}$ which satisfy eqs.(23) are expanded with respect to s as follows.

$$\Lambda_i(s) = s_{i,i} - \sum_{j=1(\neq i)}^{d} \frac{s_{j,i}^2}{s_{j,j}} s^2 + O(s^4). \tag{24}$$

Compared with the result of theorem 4, the inverse variances of true marginal densities are expanded as follows.

$$\frac{\det S}{\Delta_{i,i}} = s_{i,i} - \sum_{j=1(\neq i)}^{d} \frac{s_{j,i}^2}{s_{j,j}} s^2 + \frac{s_{i,i}}{3} [\mathrm{tr}(S_d^{-1} S_o)^3 - \mathrm{tr}\{(S_d)_{i,i}^{-1} (S_o)_{i,i}\}^3] s^3$$
$$+O(s^4). \tag{25}$$

Here, matrices S_d and S_o are the diagonal and off-diagonal matrices which satisfy $S = S_d + sS_o$ respectively. Matrices $(S_d)_{i,i}$ and $(S_o)_{i,i}$ are the minor matrices of S_d and S_o with the i-th row and column omitted. From the eqs.(24)(25), we can know that the approximate inverse variances $\{\Lambda_i\}$ calculated by LBP give exact coefficients up to second-order term but yield the differences from third-order term. When we calculate the KL distances, we obtain the following theorem.

Theorem 5. *KL distances from true marginal densities $p_i(x_i)$ to approximate marginal densities $\tilde{p}_i(x_i)$ calculated by LBP are expanded as follows with respect to s for $i \in \{1, 2, \cdots, d\}$.*

$$KL(p_i || \tilde{p}_i) = \Big(tr \frac{(S_d^{-1} S_o)^3 - \{(S_d)_{i,i}^{-1} (S_o)_{i,i}\}^3}{6} \Big)^2 s^6 + O(s^7). \tag{26}$$

Theorem 5 tells us that approximate marginal densities calculated by LBP are how close to the true marginal densities when covariances in S are very small.

4 Conclusion and Future Works

In this paper, we analytically show how accurate LBP is when the target distribution is a multi-dimensional Gaussian distribution. In particular, we calculate the fixed point, approximate marginal densities, and KL distances exactly when matrix S forms the graph with a single loop. Then, we know that a control parameter ϵ decides the accuracy. Next, we develop the result to the more general case. Furthermore, we show the series expansions of inverse variances $\{\Lambda_i(s)\}$ and KL distances up to third-order with respect to s when the matrix S forms an arbitrary graph. These results contribute to the foundation for understanding theoretical properties underlying more complex LBP algorithms and designing LBP algorithms efficiently. We have several future works, some of which are to derive the convergence rate to a fixed point, to obtain the higher order terms in eqs.(24), and to compare the theoretical results with numerical experiments.

Acknowledgement

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for JSPS Fellows and for Scientific Research 18079007, 2007.

References

1. Y. Weiss, "Belief propagation and revision in networks with loops", Technical Report 1616, MIT AI lab, 1997.
2. S.M. Aji, G.B. Horn, and R.J. McEliece, "On the convergence of iterative decoding on graphs with a single cycle", In *proc. 1998 ISIT*, 1998.
3. Y. Weiss, "Correctness of local probability propagation in graphical models with loops", *Neural Computation* **12**(1), 1-41, 2000.
4. Y. Weiss, W. Freeman, "Correctness of belief propagation in graphical models with arbitrary topology", *Neural Computation* **13**(10), 2173-2200, 2001.
5. S. Ikeda, T. Tanaka, and S. Amari, "information geometry of turbo and low-density parity-check codes", IEEE Trans. Inf. Theory, vol.50, no.6, pp.1097-1114.
6. S. Ikeda, T. Tanaka, and S. Amari, "Stochastic reasoning, free energy, and information geometry", *Neural Computation* **16**(9), 1779-1810, 2004.
7. K. Tanaka, H Shouno, M. Okada, "Accuracy of the Bethe approximation for hyperparameter estimation in probabilistic image processing", J. Phys. A, Math. Gen., vol.37, no.36, pp.8675-8696, sept, 2004.
8. K. Tanaka, "Generalized Belief Propagation Formula in Probabilistic Information Processing Based on Gaussian Graphical Model", IEICE D-II, vol.J88-D-II, no.12, pp.2368-2379, 2005.
9. Y. Nishiyama, S. Watanabe, "Theoretical Analysis of Accuracy of Belief Propagation in Gaussian Models", IEICE Technical Report, vol.107, no.50, pp.23-28, 2007.