社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

ガウシアン確率伝搬法の近似精度に対する理論解析

西山　　悠[†]　　渡辺　澄夫[††]

† 東京工業大学総合理工学研究科知能システム科学専攻 〒 226-8503 横浜市緑区長津田 4259
†† 東京工業大学精密工学研究所 〒 226-8503 横浜市緑区長津田 4259
E-mail: †nishiyudesu@cs.pi.titech.ac.jp, ††swatanab@pi.titech.ac.jp

あらまし　確率伝搬法は与えられた確率分布の周辺分布を効率的な計算量で求めることのできる計算手法であり，ループの入った確率伝搬法 (LBP) の場合には，(収束した場合に) 近似的な周辺分布を与えることが知られている．本稿では，平均 0 の多次元ガウス分布に対して，LBP によって近似的に周辺分布を求めた場合に，真の周辺分布に対する近似周辺分布のずれを理論的に考察する．具体的には，ガウス分布の分散共分散行列が 1 重ループのグラフ構造に対応する場合に，厳密にメッセージの精度，近似周辺分布の精度，真の分布と近似周辺分布との間の KL 距離を理論的に導出する．さらに，分散分散共分散行列が任意のグラフ構造をとり共分散が小さいときの場合に，近似周辺分布の精度，KL 距離について，共分散の低次からの展開式を理論的に導出する．

キーワード　確率伝搬法，カルバック距離，1 重ループ

Theoretical Analysis of Accuracy of Belief Propagation in Gaussian Models

Yu NISHIYAMA[†] and Sumio WATANABE[††]

† Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama, 226-8503 Japan
†† Precision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta Midori-ku, Yokohama, 226-8503 Japan
E-mail: †nishiyudesu@cs.pi.titech.ac.jp, ††swatanab@pi.titech.ac.jp

Abstract　Belief propagation (BP) is an algorithm which can compute marginal probability distributions with a tractable computational cost. Loopy belief propagation (LBP) applied to the graphs containing loops is known to provide marginal distributions approximately if LBP converges. In this paper, we apply LBP to a multi-dimensional Gaussian distribution that has loops and analytically show how accurate LBP is in some cases. Specifically, we analytically show messages, approximate marginal densities, and the KL distances at fixed points of LBP when the graph corresponding to a Gaussian distribution has at most a single loop. Besides, for the graphs which have arbitrary structures, we derive the expansions of approximate marginal densities when covariances are small.

Key words　Loopy Belief Propagation, Kullback Leibler Distance, Single Loop

1. introduction

Belief propagation (BP) is an algorithm which can compute marginal probability distributions of a target distribution with a tractable computational cost and has widely been studied in the areas such as probabilistic inference for artificial intelligence, turbo codes, low-density parity check (LDPC) codes, code division multiple access (CDMA), and probabilistic image processing. It is known that belief propagation provides exact marginal probability distribution of a tree graph but not of a graph with loops. Loopy belief propagation (LBP), which is the BP applied to the graphs containing loops, has a problem whether it converges or not and if it does, LBP provides approximate marginal probability distributions. The conditions for the uniqueness of LBP fixed points have been studied [1], which is considered to have a close connection to convergence of LBP. The approximate accuracy of LBP has been theoretically studied when the target distribution is defined on the discrete random variables [2] [3] and studied in graphical models such as Markov random fields [4]. From the viewpoint of information geometry, the approximate accuracy has also been studied [5] [6]. In probabilistic image processing, the accuracy has numerically been assessed through comparing the value of a estimated hyperparameter given by LBP with the exact value based on Gaussian graphical models [7] [8].

In this paper, we show the approximate accuracy of LBP analytically when we apply it to a multi-dimensional Gaussian distribution that has loops and clarify the quantities that decide degrees of the approximation and know the difference between approximate and true distributions. The situation in which we apply LBP to a multi-dimensional Gaussian distribution can be seen in Gaussian Markov random fields [4] and probabilistic image processing based on Gaussian graphical models [7] [9]. In addition, clarifying the theoretical properties of LBP in the simple and standard models such as a Gaussian distribution may help to know the properties of more complex LBP and to design LBP algorithm efficiently. In this paper we analytically clarify the messages which appear in LBP, approximate marginal probability densities, and the Kullback-Leibler (KL) divergence between approximate and true marginal densities at the fixed points of LBP. We exactly calculate these quantities when the inverse covariance matrix of a Gaussian density corresponds to a graph with a single loop. After that, we generalize the results to the graph with a single loop and arbitrary tree structures. Moreover, when the graph has arbitrary structures (i.e., the graph has multiloops), we derive the expansions of inverse variances of approximate marginal densities at small covariances.

This paper is organized as follows. In section **2**, we review the scheme of belief propagation. In section **3**, we show the main results of this paper. In section **4**, we give the brief proofs (due to limitations of space). In section **5**, conclusion and future works follow.

2. Belief Propagation

Here, we review the scheme of belief propagation. We review the belief propagation for tree structures in section **3.1**, loopy belief propagation in section **3.2**, and the LBP which is applied to a Gaussian distribution in **3.3**.

2.1 Belief Propagation for Tree Graphs

Let the target distribution to which we apply belief propagation be the pairwise form as

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

where Z is the normalization constant and B is the set which shows the existence of correlation between x_i and x_j. For example, when $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ and correlations exist in pairs $\{x_1, x_2\}$, $\{x_2, x_3\}$, and $\{x_1, x_4\}$ respectively, the set B is expressed as $B = \{\{12\}, \{23\}, \{14\}\}$. When the graph corresponding to the set 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 $\{\mathcal{M}_{i\to j}\}$, $\{\mathcal{M}_{j\to i}\}$, $(\{ij\} \in B)$.

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

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.

2.2 Loopy Belief Propagation

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

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

Because eqs.(3) are composed by $2|B|$ equations in total, 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 $\mathbf{x} \in R^d$:

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

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

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

(5)

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

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

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

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, we put $\{\tilde{\lambda}_{i\to j}\}$ as $\tilde{\lambda}_{i\to j} \equiv \lambda_{i\to j} - \tilde{s}_{j,j}$. 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 approximate marginal densities $\{\tilde{p}_i(x_i)\}$ and $\{\tilde{p}_{ij}(x_i, x_j)\}$ by substituting eqs.(6) into eqs.(2). That is, we obtain marginal densities by substituting $\{\lambda_{i\to j}\}$ which satisfy eqs.(7) into

$$\tilde{p}_i(x_i) \propto \exp\Big\{-\frac{\sum_{k\in\mathcal{N}_i} \lambda_{k\to i}}{2} x_i^2\Big\},$$

$$\tilde{p}_{ij}(x_i, x_j) \propto \exp\Big\{-(x_i, x_j)\frac{\tilde{S}_{i,j}}{2}\begin{pmatrix} x_i \\ x_j \end{pmatrix}\Big\},$$

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

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

3. Main Results

Our aim is to derive messages, approximate marginal densities, and the KL distances at fixed points of LBP. In section **3.1**, we exactly calculate these quantities when the inverse covariance matrix S corresponds to the graph with a single loop. In section **3.2**, we develop the results for a single loop to the graph with a single loop and arbitrary tree structures. In section **3.3**, we give the expansions of inverse covariances of approximate marginal densities when the covariances are small. The proofs for each claim are written in section **4**.

3.1 On the Graphs with a Single Loop

When the inverse covariance matrix S corresponds to the graph which has only a single loop as illustrated in Fig.1, without loss of generality, we can restrict the set B to $B = \{\{12\}, \{23\}, \cdots, \{d-1d\}, \{d1\}\}$. Then, for the messages $\{\mathcal{M}_{i\to j}\}$ at fixed points of LBP, the following theorem holds.

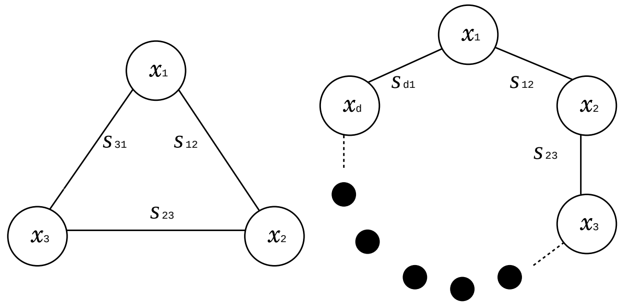

図 1 examples of single loops

[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 $\{\lambda_{i\to j}\}$ at fixed points are given by

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

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

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

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

Theorem 1 gives immediately the following corollary for approximate marginal densities at a fixed point of LBP.

[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

$$\Lambda_i = \frac{\det S}{\Delta_{i,i}}\Big(1 + \frac{(-1)^d 4 s_{1,2}\cdots s_{d,1}}{\det S}\Big)^{\frac{1}{2}},$$

$$\tilde{S}_{i,i+1} = \begin{pmatrix} \frac{E_{i,i+1}}{\Delta_{i,i}} & s_{i,i+1} \\ s_{i,i+1} & \frac{E_{i,i+1}}{\Delta_{i+1,i+1}} \end{pmatrix}, \quad (10)$$

where

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

From corollary 1, we know that, if $s_{i,i+1} \to 0$, $\{\Lambda_i\}$ and $\tilde{S}_{i,i+1}$ hold

$$\Lambda_i \to \frac{\det S}{\Delta_{i,i}}, \quad i \in \{1, \cdots, d\},$$

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

respectively by using $E_{i,i+1} \to \det S$. These results are equivalent to the true inverse variances and covariances of marginal densities respectively. Since $s_{i,i+1} \to 0$ means that a single loop (e.g., Fig.1) turns to a tree graph and LBP changes to BP algorithm for tree structures, eq.(12) agrees with and explains the well-known fact that BP provides true marginal densities but LBP provides approximate marginal densities.

In order to know the difference between true and approximate marginal densities obtained by LBP, if we calculate the Kullback-Leibler (KL) and symmetrized KL (SKL) distances between both marginal densities, the following theorem is obtained. Here, KL and SKL distances are defined as

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

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

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

$$\mathrm{KL}(p_i||\tilde{p}_i) = -\frac{1}{2} + \frac{1}{2}(1+\epsilon)^{\frac{1}{2}} - \frac{1}{4}\log\Big(1+\epsilon\Big),$$

$$\mathrm{SKL}(p_i, \tilde{p}_i) = -1 + \frac{1}{2}(1+\epsilon)^{-\frac{1}{2}} + \frac{1}{2}(1+\epsilon)^{\frac{1}{2}}. \tag{15}$$

where ϵ is defined as

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

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

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

$$\begin{aligned}\mathrm{SKL}(p_{i,i+1}, \tilde{p}_{i,i+1}) &= -2 + \Bigg[1 + \frac{1}{1-\bar{\Delta}_{i,i+1}}\Bigg\{(\frac{1+\sqrt{1+\epsilon}}{2} \\ &- \frac{s_{i,i+1}\Delta_{i,i+1}}{\det S})^2 - \frac{s_{i,i+1}^2\Delta_{ii}\Delta_{i+1,i+1}}{(\det S)^2}\Bigg\}^{-1}\Bigg]\frac{1+\sqrt{1+\epsilon}}{2},\end{aligned} \tag{18}$$

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

From theorem 2, we can know that, if $\epsilon \to 0$, both KL and SKL go to 0. Hence, KL and SKL distances are decided by the parameter ϵ. In other words, the parameter ϵ determines the accuracy of LBP when the inverse covariance matrix S forms the graph with only a single loop.

3.2 A Single Loop and Tree Structures

The results for a single loop in section **3.2** immediately tell us the case that the graph has a single loop and arbitrary tree structures. In practice, we show the inverse variances $\{\Lambda_i\}$ at a fixed point of LBP through an example. We calculate $\{\Lambda_i\}$ when the graph is expressed as Fig.(3). The undirected graph in Fig.(3) can be equivalent to the directed graph in Fig.(4) after replacing inverse variances $\{s_{i,i}\}$ as

$$s_{1,1} \to s'_{1,1}\left(\equiv s_{1,1} - \frac{s_{4,1}^2}{s_{4,4}}\right),$$

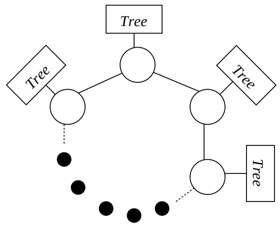

図 2 loop + tree structures

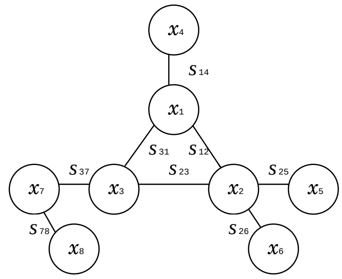

図 3 an example

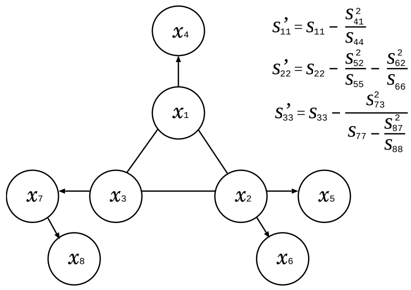

図 4 the same grpah as Fig.3

$$s_{2,2} \to s'_{2,2}\left(\equiv s_{2,2} - \frac{s_{5,2}^2}{s_{5,5}} - \frac{s_{6,2}^2}{s_{6,6}}\right),$$

$$s_{3,3} \to s'_{3,3}\left(\equiv s_{3,3} - \frac{s_{7,3}^2}{s_{7,7} - \frac{s_{8,7}^2}{s_{8,8}}}\right) \tag{19}$$

(These replacements are understandable by solving eqs.(7) directly). In Fig.(4), the direct edges indicate that there exist messages in those directions but not in reverse direction (e.g., message $\mathcal{M}_{1\to 4}(x_4)$ exists but message $\mathcal{M}_{4\to 1}(x_1)$ does not between nodes x_1 and x_4). Then, since the nodes x_1, x_2, and x_3 form a closed single loop, referring to the results of corollary 1, we obtain inverse variances $\{\Lambda'_i\}$, $i \in \{1,2,3\}$ as

$$\Lambda'_i = \frac{\det S'}{\Delta'_{i,i}}\Big(1 + \frac{(-1)^d 4 s_{1,2}s_{2,3}s_{3,1}}{\det S'}\Big)^{\frac{1}{2}}. \tag{20}$$

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.

$$\Lambda'_4 = s_{4,4} - \frac{s_{1,4}^2}{\Lambda'_1 +}\frac{s_{4,1}^2}{s_{4,4}}, \qquad \Lambda'_5 = s_{5,5} - \frac{s_{2,5}^2}{\Lambda'_2 +}\frac{s_{5,2}^2}{s_{55}},$$

$$\Lambda'_6 = s_{6,6} - \frac{s_{2,6}^2}{\Lambda'_2 +}\frac{s_{6,2}^2}{s_{6,6}}, \qquad \Lambda'_8 = s_{8,8} - \frac{s_{7,8}^2}{\Lambda'_7 +}\frac{s_{8,7}^2}{s_{8,8}},$$

$$\Lambda'_7 = s_{77} - \frac{s_{8,7}^2}{s_{8,8}} - \frac{s_{3,7}^2}{\Lambda'_3 +}\frac{s_{7,3}^2}{s_{7,7} -}\frac{s_{8,7}^2}{s_{8,8}} \tag{21}$$

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

In a similar way, we can obtain $\{\mathcal{M}_{i\to j}\}$, $\{p_i(x_i)\}$, $\{p_{ij}(x_i, x_j)\}$, and KL distances in various connected graphs which have at most a single loop.

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

In this section, we aim at obtaining the expansions of $\{\Lambda_i\}$

when the graph has arbitrary structures but covariances are 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 satisfies $0 \leqq s \leqq 1$. 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, we clarify the expansions of $\{\Lambda_i(s)\}$ with respect to s. In this paper we derive the terms up to third-order. Before that, we prepare the following theorem for inverse variances $\{\Lambda_i\}$.

[Theorem 3] The approximate inverse variances $\{\Lambda_i\}$ at fixed points of 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 $\pm$ indicates either of positive or negative for each term in the summation.

Since $\pm$ assigns either of positive or negative for each term, there exist $\sum_{i=1}^{d} 2^{|\mathcal{N}_i|}$ simultaneous equations in eqs.(22) in total. We note that there exist extra simultaneous equations which has no solution $\{\Lambda_i\}$. For example, if $|\mathcal{N}_i| \geqq 2$ for $^{\exists}i$ and all the signs are negative, there is no solution.

At $s = 0$, the approximate inverse variances should be equal to true inverse variances (i.e.,$\Lambda_i(0) = s_{i,i}$ for $i \in \{1, \cdots, d\}$) since the matrix S corresponds to a tree graph. After 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) have the expansions as

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

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

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

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 $\{\Lambda_i\}$ give exact coefficients up to second-order term but differ from third-order term. Then, the KL distances between both marginal densities have the expansions as follows.

[Theorem 5] KL distances between true marginal densities $p_i(x_i)$ and approximate marginal densities $\tilde{p}_i(x_i)$ have no term up to third-order with respect to s. That is, $\mathrm{KL}(p_i||\tilde{p}_i) = O(s^4)$ for $i \in \{1, \cdots, d\}$ hold.

Theorem 5 tells us that approximate marginal densities calculated by LBP are so close to the true marginal densities when covariances in S are small (i.e., $s \ll 1$).

4. Proofs

We give the brief proof for each claim.

Proof of Theorem 1

When $B = \{\{12\}, \{23\}, \cdots, \{d-1d\}, \{d1\}\}$, from decision equations (7), $\tilde{\lambda}_{d\to 1}$ satisfies the following equation by using the notation of continued fractions.

$$\tilde{\lambda}_{d\to 1} = -\frac{s_{d,1}^2}{s_{d,d} -} \frac{s_{d-1,d}^2}{s_{d-1,d-1} -} \cdots \frac{s_{2,3}^2}{s_{2,2} -} \frac{s_{1,2}^2}{s_{1,1} + \tilde{\lambda}_{d\to 1}} \tag{26}$$

For the right-hand side of eq.(26),

$$\frac{1}{s_{k,k} -} \frac{s_{k-1,k}^2}{s_{k-1,k-1} -} \cdots \frac{s_{1,2}^2}{s_{1,1} + \tilde{\lambda}_{d\to 1}} = \frac{\Delta_{1,1}^{(1,\cdots,k-1)} \tilde{\lambda}_{d\to 1} + \Delta_{k,k}}{\Delta_{1,1} \tilde{\lambda}_{d\to 1} + \det S - 2s_{k,1}\Delta_{k,1} - s_{k,1}^2 \Delta_{1,1}^{(1,\cdots,k-1)}} \tag{27}$$

holds when $k \in \{2, \cdots, d\}$ by using induction. Here, $\Delta_{11}^{(1,\cdots,k-1)}$ is the cofactor which variables $x_1, \cdots, x_{k-1}$ create. From eq.(27), eq.(26) can be rewritten as a quadratic equation

$$\Delta_{1,1} \tilde{\lambda}_{d\to 1}^2 + (\det S - 2s_{d,1}\Delta_{d,1}) \tilde{\lambda}_{d\to 1} + s_{d,1}^2 \Delta_{d,d} = 0 \tag{28}$$

Then, the solution $\tilde{\lambda}_{d\to 1}$ is given by

$$\tilde{\lambda}_{d\to 1} = \frac{-\det S + 2S_{d,1}\Delta_{d,1} \pm \sqrt{\mathcal{D}}}{2\Delta_{1,1}},$$
$$\mathcal{D} = (\det S)^2 - 4s_{d1}(\Delta_{d1} \det S - s_{d1}\Delta_{d1}^2 + s_{d1}\Delta_{dd}\Delta_{11}) = (\det S)^2 + (-1)^d 4 s_{1,2} s_{2,3} \cdots s_{d-1,d} s_{d,1} \det S. \tag{29}$$

For the second equal sign in $\mathcal{D}$, we use the equation

$$\Delta_{d,d}\Delta_{1,1} - \Delta_{d,1}^2 = \Delta_{1,1}^{(1,\cdots,d-1)} \det S. \tag{30}$$

Eq.(30) is proved by induction using tridiagonal matrices. After substituting eq.(29) into $\lambda_{d\to 1} = \tilde{\lambda}_{d\to 1} + \tilde{s}_{1,1}$, we obtain $\lambda_{d\to 1}$. We obtain the other values $\{\lambda_{i-1\to i}\}$ by applying cyclic permutations $(1, \cdots, d)$ to $\lambda_{d\to 1}$, the value $\lambda_{1\to d}$ by applying the permutation $\binom{1,2,\cdots,d}{d,d-1,\cdots,1}$ to $\lambda_{d\to 1}$, and the other values $\{\lambda_{i\to i-1}\}$ by applying cyclic permutations $(1, \cdots, d)$

to $\lambda_{1\to d}$. (Q.E.D.)

Proof of Corollary 1

By substituting eqs.(9) into eqs.(8) and imposing $\Lambda_i > 0$, we obtain eqs.(10). (Q.E.D.)

Proof of Theorem 2

After substituting eqs.(10) into eqs.(8), we calculate KL and SKL distances (14). (Q.E.D.)

Proof of Theorem 3

The simultaneous equations (7) are rewritten as the $2|B|$ simultaneous equations

$$\tilde{\lambda}_{i\to j} = -\frac{s_{i,j}^2}{\Lambda_i - \tilde{\lambda}_{j\to i}} \quad (31)$$

by using the variable transformation of $\Lambda_i = \sum_{k\in\mathcal{N}_k} \lambda_{k\to i}$. Then, the variable $\tilde{\lambda}_{i\to j}$ satisfies

$$\tilde{\lambda}_{i\to j} = -\frac{s_{i,j}^2}{\Lambda_i +} \frac{s_{j,i}^2}{\Lambda_j - \tilde{\lambda}_{i\to j}} \quad (32)$$

and can be written as

$$\tilde{\lambda}_{i\to j} = \frac{\Lambda_j}{2} \pm \frac{\Lambda_j}{2}\sqrt{1 + \frac{4s_{i,j}^2}{\Lambda_i\Lambda_j}}. \quad (33)$$

Since $\tilde{\lambda}_{i\to j}$ satisfies $\Lambda_j = \sum_{k\in\mathcal{N}_j} \tilde{\lambda}_{k\to j} + s_{j,j}$ and we substitute eq.(33) into this equation, we obtain the decision equations for $\{\Lambda_j\}$ as follows.

$$\Lambda_j = s_{j,j} + \sum_{k\in\mathcal{N}_j}\left(\frac{\Lambda_j}{2} \pm \frac{\Lambda_j}{2}\sqrt{1 + \frac{4s_{k,j}^2}{\Lambda_k\Lambda_j}}\right). \quad (34)$$

(Q.E.D.)

Proof of theorem 4

By substituting $s = 0$, $\Lambda_i = s_{i,i}$ into eqs.(22), we know that all the signs have to be positive. When we substitute the expansions

$$\Lambda_i = s_{i,i} - a_{i1}s - a_{i2}s^2 - a_{i3}s^3 + \cdots \quad (35)$$

into eqs.(23) and compare the coefficients up to third-order term, we obtain

$$\frac{a_{i1}}{s_{i,i}} = 0, \qquad \frac{a_{i2}}{s_{i,i}} + \frac{a_{i1}^2}{s_{i,i}} = \sum_{k=1(\neq i)}^{s} \bar{s}_{k,i},$$

$$\frac{a_{i3}}{s_{i,i}} + \frac{2a_{i1}a_{i2}}{s_{i,i}^2} + \frac{a_{i1}^3}{s_{i,i}^3} = \sum_{k=1(\neq i)}^{d} \bar{s}_{k,i}\left(\frac{a_{k1}}{s_{k,k}} + \frac{a_{i1}}{s_{i,i}}\right), \quad (36)$$

where we put $\frac{s_{i,j}^2}{s_{i,i}s_{j,j}} \equiv \bar{s}_{i,j}$. By solving these equations, we obtain

$$a_{i1} = a_{i3} = 0, \quad a_{i2} = \sum_{k=1(\neq i)}^{d} \frac{s_{k,i}^2}{s_{k,k}}. \quad (37)$$

(Q.E.D.)

Proof of Theorem 5

By substituting eqs.(24)(25) into KL distance

$$\mathrm{KL}(p_i||\tilde{p}_i) = -\frac{1}{2} + \frac{1}{2}\frac{\Lambda_i}{\det S/\Delta_{ii}} + \frac{1}{2}\log\frac{\det S/\Delta_{ii}}{\Lambda_i}, \quad (38)$$

we expand $\mathrm{KL}(p_i||\tilde{p}_i)$ up to third-order. Then, the second and third terms of eq.(38) are expanded as

$$\frac{1}{2}\frac{\Lambda_i}{\det S/\Delta_{ii}} = \frac{1}{2} - \frac{1}{6}\mathrm{tr}[(S_d^{-1}S_o)^3 - \{(\Delta_{ii})_d^{-1}(\Delta_{ii})_o\}^3]s^3 + O(s^4),$$

$$\frac{1}{2}\log\frac{\det S/\Delta_{ii}}{\Lambda_i} = \frac{1}{6}\mathrm{tr}[(S_d^{-1}S_o)^3 - \{(\Delta_{ii})_d^{-1}(\Delta_{ii})_o\}^3]s^3 + O(s^4) \quad (39)$$

respectively and $\mathrm{KL}(p_i||\tilde{p}_i) = O(s^4)$ are obtained. (Q.E.D.)

5. Conclusion and Future Works

In this paper, we focus on applying LBP to a multi-dimensional Gaussian distribution and analytically clarify the messages, approximate marginal densities and KL distances at fixed points of LBP under some conditions. We know that the accuracy of LBP are decided by a simple parameter ϵ when the inverse covariance matrix is described as a graph with a single loop. For the arbitrary graph, we derive the expansions of inverse variances of approximate marginal densities at small covariances. We have several future works, some of which are to derive convergence rate to fixed points of LBP, to compare with numerical experiments, and to propose a correction algorithm for LBP.

6. Acknowledgements

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, 2006.

文　献

[1] T. Heskes, "On the Uniqueness of Loopy Belief Propagation Fixed Points", *Neural computation* **16**(11), 2379-2414, 2004.

[2] Y. Weiss, "Belief propagation and revision in networks with loops", Technical Report 1616, MIT AI lab, 1997.

[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.(Japanese)

[9] 田中和之, "確率モデルによる画像処理技術入門", 森北出版, 2006.