

## ガウス分布におけるベータ近似の理論解析

# Theoretical Analysis of Bethe Approximation in Gaussian Distribution

西山 悠\*

Yu Nishiyama

渡辺澄夫†

Sumio Watanabe

**Abstract:** For the calculation of marginal distributions, belief propagation (BP) is known to be a efficient algorithm. Loopy belief propagation (LBP), which is the BP applied to the probabilistic distribution that has loops, has the problems of under what conditions LBP converges or how accurate the solution is. Meanwhile, the fixed points of LBP update rule are known to accord with the extrema of Bethe free energy. In this paper, we focus on applying Bethe approximation to the standard model of Gaussian distribution and analytically evaluate the accuracy of the approximation. As a result, we reveal that what quantity determines the accuracy and the convergence condition when Gaussian distribution forms a simple single loop. For the arbitrary graphs, we obtain the expansions at small covariances. Based on the analytical solution, we do the numerical experiments.

**Keywords:** Bethe approximation, Loopy belief propagation, Single loop, Gaussian Distribution

## 1 まえがき

計算量が指数的に増大する周辺分布の計算に確率伝搬法 (Belief Propagation;BP) のアルゴリズムの有効性が知られている。確率伝搬法は、周辺分布を計算するのに、他の確率変数の実現値による和 (積分) を直接的に行うのではなく、メッセージと呼ばれる確率変数ノード間を伝搬する量を用いて、メッセージに対する更新式を施した後に、メッセージの積によって周辺分布を構成するものである。確率伝搬法のアルゴリズムは、元々、確率分布のグラフ構造が木構造であるという条件の下で導出されるものであるが、導出されるアルゴリズム自体は、グラフ構造がループ構造を持つ確率分布の場合にも適用可能なものになっている。ループ構造を持つ確率分布に適用した確率伝搬法 (Loopy Belief Propagation;LBP) は、理論的な保証がないために、アルゴリズムが収束しないといった問題点を抱え、収束した場合には近似的な周辺

分布を与えるアルゴリズムとなっている。LBP が収束するための条件について、LBP 更新式の固定点の一意性に注目して、一意的な固定点を持つための条件について研究されている [1]。LBP の近似精度についても、理論的に広く研究され [2] [3] [4]、情報幾何学に基づく確率伝搬法の解析 [5][6] も行われている。また確率伝搬法の固定点は、統計物理から由来されるベータ自由エネルギーと呼ばれる評価関数の極値で与えられることが知られている [7][8]。

本稿では、確率伝搬法を多次元正規分布に適用した場合に、確率伝搬法の収束性、収束値を解析的、数値的に考察する。多次元正規分布のときのベータ近似の近似精度については、確率的画像処理において、劣化画像から修復画像を構成する際に、周辺尤度最大化を与えるハイパーパラメータの値の精度の点から数値的に研究されている [9][10]。またガウシアンマルコフ確率場 (GMRF) のときに理論的に研究されている [4]。

本稿では、確率伝搬法を多次元正規分布に適用した結果、1重ループのグラフ構造で表わされる正規分布の場合にはLBP解の解析解を与え、LBPが収束するための(必要)条件を明らかにする。任意のグラフ構造で表わされる正規分布の場合には、相関が小さいときのLBP解の展開式を与える。その後、確率伝搬法の数値実験を行うことで得られた理論的結果と比較する。

\*東京工業大学大学院総合理工学研究科知能システム科学専攻, 226-8503 横浜市緑区長津田町 4259 メールボックス:R2-5, tel.045-924-5018, e-mail nishiyudesu@cs.pi.titech.ac.jp, Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Mailbox:R2-5, 4259, Nagatsuta-chou, Midori-ku, Yokohama, 226-8503, Japan.

†東京工業大学精密工学研究所, 226-8503 横浜市緑区長津田町 4259 メールボックス:R2-5, tel.045-924-5018, e-mail swatanab@pi.titech.ac.jp, P&I Lab., Tokyo Institute of Technology, Mailbox:R2-5, 4259, Nagatsuta-chou, Midori-ku, Yokohama, 226-8503, Japan.

次節以降の構成は次のようになっている．2でベータ近似の数学的枠組みを説明し，3でガウス分布における確率伝搬法の解析解を与える．4でメッセージを使わない確率伝搬法を述べた後に，5で確率伝搬法の数値解を解析解と比較する．最後に結論と今後の展開を与える．

## 2 ベータ近似

ここではベータ自由エネルギーの数学的枠組みを説明する [10] ．

### 2.1 ベータ自由エネルギー

周辺分布を求めたい確率分布を  $\mathbf{x} \in R^d$ ， $Z$  を正規化定数，集合  $B$  を変数間の相互作用の存在を表す変数対の集合とした下で

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

の形で表わされるものとする．今，確率分布  $p(\mathbf{x})$  を試験分布  $r(\mathbf{x})$  で近似することを考えるのに， $r(\mathbf{x})$  から  $p(\mathbf{x})$  へのカルバック距離を考えれば，カルバック距離を小さくすることは，汎関数

$$F(r) = -S(r) - \int r(\mathbf{x}) \log \prod_{\{ij\} \in B} W_{ij}(x_i, x_j) d\mathbf{x},$$

$$S(r) \equiv - \int r(\mathbf{x}) \log r(\mathbf{x}) d\mathbf{x}$$

を，試験分布  $r(\mathbf{x})$  について小さくすることと等しい．このときベータ自由エネルギーは，エントロピー項  $S(r)$  を

$$\begin{aligned} S(r) &\simeq S_{Bethe}(\{r_i\}, \{r_{ij}\}) \\ &\equiv \sum_{i=1}^d S_i(r_i) + \sum_{\{ij\} \in B} \{S_{ij}(r_{ij}) - S_i(r_i) - S_j(r_j)\} \\ &= \sum_{\{ij\} \in B} S_{ij}(r_{ij}) - \sum_{i=1}^d (|\mathcal{N}_i| - 1) S_i(r_i) \end{aligned}$$

と近似した下で評価関数

$$F_{Bethe}(\{r_i\}, \{r_{ij}\}) = -S_{Bethe} - \sum_{\{ij\} \in B} \int r_{ij} \log W_{ij} dx_i dx_j$$

で与えられる (ただし引数は適宜省略している)．このベータ自由エネルギーを確率密度関数  $\{r_i\}, \{r_{ij}\}$  について最小化する．エントロピー項の近似は，確率分布 (1) が木構造で表わされる場合には，厳密に正しく等号となる．ここで  $|\mathcal{N}_i|$  は， $i$  の含まれている元  $\{ij\} \in B$  の個数であり，変数ノード  $x_i$  につながっている最近接の変数ノードの個数である．ベータ自由エネルギーを

$\{r_i\}, \{r_{ij}\}$  に対する拘束条件，互いに周辺分布の関係にある確率密度関数であるという条件の下で極値条件を与えると，関数  $F_{Bethe}$  から構成されるラグランジュ関数

$$\begin{aligned} L_{Bethe}(\{r_i\}, \{r_{ij}\}) &= F_{Bethe}(\{r_i\}, \{r_{ij}\}) \\ &+ \sum_{i=1}^d \nu_i \left( \int r_i dx_i - 1 \right) + \sum_{\{ij\} \in B} \left\{ \nu_{ij} \left( \int r_{ij} dx_i dx_j - 1 \right) \right. \\ &\left. + \int \eta_{ij,i}(x_i) \left( \int r_{ij} dx_j - r_i \right) dx_i + \int \eta_{ij,j}(x_j) \left( \int r_{ij} dx_i - r_j \right) dx_j \right\} \end{aligned}$$

を考え， $L_{Bethe}$  の  $r_i, r_{ij}$  における変分を取ることで，ベータ自由エネルギーの極値を与える確率分布  $r_i^*, r_{ij}^*$  は

$$r_i^* = \frac{1}{Z_i} \prod_{k \in \mathcal{N}_i} \exp \left\{ -\frac{\eta_{ik,i}(x_i)}{|\mathcal{N}_i| - 1} \right\}$$

$$r_{ij}^* = \frac{1}{Z_{ij}} W_{ij} \exp \{ -\eta_{ij,i}(x_i) - \eta_{ij,j}(x_j) \}$$

で与えられる．ラグランジュ未定乗数を  $\exp\{-\eta_{ik,i}(x_i)\} = \prod_{l \in \mathcal{N}_i \setminus \{k\}} \mathcal{M}_{l \rightarrow i}(x_i)$  と変数変換すると，これらの確率分布は，

$$r_i^* = \frac{1}{Z_i} \prod_{k \in \mathcal{N}_i} \mathcal{M}_{k \rightarrow i}(x_i),$$

$$r_{ij}^* = \frac{W_{ij}}{Z_{ij}} \left( \prod_{k \in \mathcal{N}_i \setminus \{j\}} \mathcal{M}_{k \rightarrow i}(x_i) \right) \left( \prod_{k \in \mathcal{N}_j \setminus \{i\}} \mathcal{M}_{k \rightarrow j}(x_j) \right).$$

で表される．これらの式は確率伝搬法のメッセージを用いた周辺分布の計算式と一致している．式中のメッセージ  $\mathcal{M}_{i \rightarrow j}(x_i)$ ， $\mathcal{M}_{i \rightarrow j}(x_i)$ ， $(\{ij\} \in B)$  は， $r_i, r_{ij}$  の拘束条件から導かれる更新式

$$\mathcal{M}_{i \rightarrow j}^{(t+1)}(x_j) = \frac{1}{Z_{ij}} \int_{-\infty}^{\infty} W_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}_i \setminus \{j\}} \mathcal{M}_{k \rightarrow i}^{(t)}(x_i) dx_i,$$

$$\mathcal{M}_{j \rightarrow i}^{(t+1)}(x_i) = \frac{1}{Z_{ji}} \int_{-\infty}^{\infty} W_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}_j \setminus \{i\}} \mathcal{M}_{k \rightarrow j}^{(t)}(x_j) dx_j$$

の固定点  $\mathcal{M}_{i \rightarrow j}^*(x_j)$ ， $\mathcal{M}_{j \rightarrow i}^*(x_i)$ ， $(\{ij\} \in B)$  によって構成される．

### 2.2 ガウス分布におけるベータ自由エネルギー

ターゲットの確率分布 (1) を，平均ベクトル  $\mathbf{0}$ ，精度行列  $S$  の多次元ガウス分布である

$$p(\mathbf{x}) = \sqrt{\frac{\det S}{(2\pi)^d}} \exp \left\{ -\frac{1}{2} \mathbf{x}^T S \mathbf{x} \right\}$$

とする．以後，このガウス分布を  $N(\mathbf{0}, S)$  と書き，行列  $S$  の各成分を  $(S)_{ij} = s_{ij}$  と書くことにする．このとき確率分布 (1) の  $W_{ij}(x_i, x_j)$  は 1 つに，

$$W_{ij}(x_i, x_j) = \exp\left\{-\frac{1}{2}\left(\frac{s_{ii}}{|\mathcal{N}_i|}x_i^2 + 2s_{ij}x_ix_j + \frac{s_{jj}}{|\mathcal{N}_j|}x_j^2\right)\right\}$$

と表すことができる．ただし任意の変数  $x_i$  について  $|\mathcal{N}_i| > 0$  である．以降では  $\frac{s_{ii}}{|\mathcal{N}_i|} = \tilde{s}_{ii}$  と書くことにする．メッセージ  $\mathcal{M}_{i \rightarrow j}(x_j)$  を  $N(0, \lambda_{i \rightarrow j})$  の正規分布とする．このとき，メッセージの更新式は，メッセージの精度（分散の逆数） $\{\lambda_{i \rightarrow j}\}$  の更新式

$$\begin{aligned}\bar{\lambda}_{i \rightarrow j}^{(t+1)} &= -\frac{\tilde{s}_{ij}^2}{1 + \sum_{k \in \mathcal{N}_i \setminus \{j\}} \bar{\lambda}_{k \rightarrow i}^{(t)}}, \\ \bar{\lambda}_{j \rightarrow i}^{(t+1)} &= -\frac{\tilde{s}_{ij}^2}{1 + \sum_{k \in \mathcal{N}_j \setminus \{i\}} \bar{\lambda}_{k \rightarrow j}^{(t)}}, \quad \{ij\} \in B\end{aligned}\quad (2)$$

に帰着される．ただし  $\bar{\lambda}_{i \rightarrow j} \equiv \frac{\lambda_{i \rightarrow j}}{s_{jj}} - \frac{1}{|\mathcal{N}_j|}$  と置き， $\tilde{s}_{ij}$  は  $\tilde{s}_{ij} \equiv s_{ij}/\sqrt{s_{ii}\sqrt{s_{jj}}}$  である．精度の更新式より，多次元ガウス分布のときのベータ自由エネルギーについて，その極値条件を与える確率分布  $r_i^*, r_{ij}^*$  は，精度の更新式の固定点  $\lambda_{i \rightarrow j}^*, \lambda_{j \rightarrow i}^*, \{ij\} \in B$  を用いて

$$\begin{aligned}r_i^* &\sim N(0, \Lambda_i^*), \quad \Lambda_i^* = \sum_{k \in \mathcal{N}_i} \lambda_{k \rightarrow i}^*, \\ r_{ij}^* &\sim N(0, \tilde{S}_{ij}), \\ \tilde{S}_{ij} &\equiv \begin{pmatrix} \tilde{s}_{ii} + \sum_{k \in \mathcal{N}_i \setminus \{j\}} \lambda_{k \rightarrow i}^* & s_{ij} \\ s_{ij} & \tilde{s}_{jj} + \sum_{k \in \mathcal{N}_j \setminus \{i\}} \lambda_{k \rightarrow j}^* \end{pmatrix}\end{aligned}$$

で与えられる．

### 3 解析解

ここでは，正規分布にベータ近似を適用したときのベータ自由エネルギーについて，その極値を与える周辺密度関数（ベータ近似解）を解析的に評価する．ベータ近似解は LBP によって計算される周辺密度関数（LBP 解）と等しい．3.1, 3.2 の中の定理の証明は [11] を参照．

#### 3.1 1 重ループのときの解析解

まず，正規分布の表わすグラフ構造として，最も単純なループ構造である 1 重ループの場合に，ベータ近似解の解析解を与える．1 重ループを考えると，一般性を失うことなく，ボンド集合  $B$  を  $B \in \{\{12\}, \{23\}, \dots, \{d-1d\}, \{d1\}\}$  に制限することができる．このときベータ近似解を与える固定点のメッセージについて次が成り立つ．

定理 1 ガウス分布が 1 重ループのグラフ構造で表わされるとき，ベータ近似解を与えるメッセージの固定点

$\mathcal{M}_{i \rightarrow j}^*(x_j)$  は，以下の密度関数で与えられる．

$$\begin{aligned}\mathcal{M}_{i \rightarrow i+1}^*(x_{i+1}) &\sim N(0, \lambda_{i \rightarrow i+1}^*), \\ \lambda_{i \rightarrow i+1}^* &\equiv \frac{s_{i,i+1}\Delta_{i,i+1} - s_{i+1,i+2}\Delta_{i+1,i+2} \pm \sqrt{D}}{2\Delta_{i+1,i+1}}, \\ \mathcal{M}_{i \rightarrow i-1}^*(x_{i-1}) &\sim N(0, \lambda_{i \rightarrow i-1}^*), \\ \lambda_{i \rightarrow i-1}^* &\equiv \frac{s_{i,i-1}\Delta_{i,i-1} - s_{i-1,i-2}\Delta_{i-1,i-2} \pm \sqrt{D}}{2\Delta_{i-1,i-1}}, \\ D &\equiv (\det S)^2 + (-1)^d 4s_{1,2}s_{2,3}\cdots s_{d-1,d}s_{d,1} \det S.\end{aligned}$$

ただし  $i \in \{1, 2, \dots, d\}$  であり，周期的境界条件である  $s_{d,d+1} \equiv s_{d,1}$ ,  $s_{1,0} \equiv s_{1,d}$  が成り立っている． $\Delta_{i,j}$  は行列  $S$  の  $(i, j)$  余因子である．

定理 1 の固定点のメッセージについて， $2\lambda_{j-1 \rightarrow j}^* \Delta_{j,j}$  の量を考えると，ノード上に一様に存在する  $\sqrt{D}$  の値に，ノード  $x_{j-1}$  から  $x_j$  に流れる  $s_{j-1,j} \Delta_{j-1,j}$  の量を加えノード  $x_j$  から  $x_{j+1}$  に流れ出る  $s_{j,j+1} \Delta_{j,j+1}$  を引くことで  $2\lambda_{j-1 \rightarrow j}^* \Delta_{j,j}$  の値が構成されることは興味深い．

定理 1 より，ベータ近似解を与える周辺密度関数について以下が成り立つ．

定理 2 1 重ループのガウス分布において，ベータ近似解（等しく LBP 解）を与える周辺密度関数  $r_i^*, r_{ij}^*$  はそれぞれ以下の密度関数で与えられる．

$$\begin{aligned}r_i^* &\sim N(0, \Lambda_i^*), \quad \Lambda_i^* = \frac{\det S}{\Delta_{i,i}} \sqrt{1 + \frac{(-1)^d 4s_{1,2}\cdots s_{d,1}}{\det S}}, \\ r_{i,i+1}^* &\sim N(0, \tilde{S}_{i,i+1}^*), \quad \tilde{S}_{i,i+1}^* = \begin{pmatrix} \frac{E_{i,i+1}}{\Delta_{i,i}} & s_{i,i+1} \\ s_{i,i+1} & \frac{E_{i,i+1}}{\Delta_{i+1,i+1}} \end{pmatrix}.\end{aligned}$$

ただし  $E_{i,i+1}$  は

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

で与えられる．

定理 2 で与えられた精度の解析解  $\Lambda_i^*, \tilde{S}_{i,i+1}^*$  は， $s_{i,i+1} \rightarrow 0$  の極限で，それぞれ

$$\Lambda_i^* \rightarrow \frac{\det S}{\Delta_{i,i}}, \quad \tilde{S}_{i,i+1}^* \rightarrow \begin{pmatrix} \frac{\det S}{\Delta_{i,i}} & 0 \\ 0 & \frac{\det S}{\Delta_{i+1,i+1}} \end{pmatrix}$$

の値に近づくが，これらの値は精度行列  $S$  を持つ正規分布の，真の周辺密度関数の精度となっている．このことは，1 重ループにおいて  $s_{i,i+1} \rightarrow 0$  とすることは，ループ構造は木構造へと移り変わり，LBP が木構造に対する BP となることから，BP においては確率分布の周辺分布が厳密に計算されることを説明している．定理 2 の

結果より，1重ループのときにLBPの近似精度を決定づけている量は

$$\epsilon \equiv \frac{(-1)^{d_4} s_{1,2} \cdots s_{d,1}}{\det S}$$

の値であることがわかる．精度行列  $S$  から計算される  $|\epsilon|$  の値が大きければLBPの近似精度は悪くなり， $|\epsilon|$  の値が小さければLBPの近似精度が良いものとなる． $\epsilon = 0$  のときに真の周辺分布の精度と一致する．さらに，精度  $\Lambda_i^*$  が実数解となるためには，平方根の中身が非負より  $\epsilon \geq -1$  が必要条件であることがわかる．このことから，1重ループで表わされるガウス分布においては， $\epsilon \geq -1$  が，LBPが収束するための条件と言える（十分条件にもなっていないと思われるが証明はまだ不明）．実際，5節で，正定値行列ではあるが  $\epsilon \geq -1$  を満たさない行列で，数値実験の結果，LBPが収束しない行列の例を挙げる．

### 3.2 相関が小さいときのLBP解の展開式

3.1節では，1重ループで表わされる精度行列のベータ近似解析解を与えたが，ここでは多重ループを含む任意のグラフ構造を表す精度行列で，ただし変数間の相関が小さいとしたときの，LBP解の展開式を与える．目的のために，精度行列  $S$  を対角成分を保持する行列  $S_d$  と，非対角成分を保持する行列  $S_o$  とに分けた後に，新しいパラメータ  $s$  を導入して  $S = S_d + sS_o$  としておく．このときLBP解をパラメータ  $s$  について3次のオーダーまで求める．その結果，次が成り立つ．

**定理 3** 任意のグラフ構造をとる精度行列  $S$  において，正規分布におけるベータ自由エネルギーのベータ近似解（等しくLBP解） $r_i^*$  は  $s$  について次式で展開される．

$$r_i^* \sim N(0, \Lambda_i^*(s)),$$

$$\Lambda_i^*(s) = s_{i,i} - \sum_{j=1(\neq i)}^d \frac{s_{j,i}^2}{s_{j,j}} s^2 + O(s^4). \quad (3)$$

定理 3 で与えられるベータ近似解の  $s$  による展開式の一方で，ガウス分布の真の周辺密度関数は  $s$  について次式で展開される．

$$p_i \sim N(0, \det S / \Delta_{i,i}),$$

$$\frac{\det S}{\Delta_{i,i}} = s_{i,i} - \sum_{j=1(\neq i)}^d \frac{s_{j,i}^2}{s_{j,j}} s^2 + \delta_i s^3 + O(s^4),$$

$$\delta_i \equiv \frac{s_{i,i}}{3} [\text{tr}(S_d^{-1} S_o)^3 - \text{tr}\{(S_d)_{i,i}^{-1} (S_o)_{i,i}\}^3]. \quad (4)$$

ここで  $(S_d)_{i,i}, (S_o)_{i,i}$  は，それぞれ行列  $S_d, S_o$  から  $i$  行  $i$  列を除いてできる小行列である．展開式 (3) と (4) を比較することで，相関が小さいときにLBPの近似精度を決定する量は， $\delta_i$  の値であることがわかる．

## 4 メッセージを使わない確率伝搬法

### 4.1 周辺分布の精度 $\Lambda_i$ の更新式

通常確率伝搬法では，周辺分布を計算するのに，メッセージを更新していき，そのメッセージの固定点を周辺分布構成式に代入することで得られるものとなっているが，メッセージを介さない，周辺分布の直接的な更新式を考えることで，次の更新式が得られる [11] ．

1. **Step1:** 精度  $\Lambda_i^{(0)}$ ， $i \in \{1, \dots, d\}$  の初期値を設定する．

2. **Step2:** 精度  $\Lambda_i^{(t)}$ ， $i \in \{1, \dots, d\}$  を更新式

$$\Lambda_i^{(t+1)} \leftarrow \frac{2s_{i,i}}{2 - |\mathcal{N}_i| + \sum_{j \in \mathcal{N}_i} \sqrt{1 + \frac{4s_{i,j}^2}{\Lambda_i^{(t)} \Lambda_j^{(t)}}}} \quad (5)$$

にしたがって更新する．

3. **Step3:** 十分小さい  $\epsilon_0$  について， $\sum_{i=1}^d |\Lambda_i^{(t+1)} - \Lambda_i^{(t)}| < \epsilon_0$  を満たすとき，計算を終了する．さもなければ  $\Lambda_i^{(t)} \leftarrow \Lambda_i^{(t+1)}$  として Step2 に戻る．

この周辺分布を直接更新していく更新式は，エッジ ( $\{ij\} \in B$ ) の個数の2倍あるメッセージの更新式とは異なり，ノード数の個数である  $d$  個の更新式で済むことから，その意味で計算量が少ない更新式となっている．

### 4.2 更新式 (5) の導出

ここでは，メッセージを介さない周辺分布の精度の更新式 (5) を導出する．更新式 (5) は，メッセージの精度の固定点方程式 (2) において，周辺分布の精度  $\Lambda_i = \sum_{k \in \mathcal{N}_i} \lambda_{k \rightarrow i}$ ， $(i \in \{1, \dots, d\})$  の式を使い，メッセージの精度  $\lambda_{i \rightarrow j}$ ， $\lambda_{j \rightarrow i}$ ， $(\{ij\} \in B)$  を消去することで得られる．具体的には， $\Lambda_i = \sum_{k \in \mathcal{N}_i} \lambda_{k \rightarrow i}$ ， $(i \in \{1, \dots, d\})$  を使えば，固定点方程式 (2) は， $\bar{\Lambda}_i \equiv \Lambda_i / s_{i,i}$  とおいた下で，

$$\bar{\lambda}_{i \rightarrow j} = -\frac{\bar{s}_{i,j}^2}{\Lambda_i - \bar{\lambda}_{j \rightarrow i}},$$

$$\bar{\lambda}_{j \rightarrow i} = -\frac{\bar{s}_{j,i}^2}{\Lambda_j - \bar{\lambda}_{i \rightarrow j}}, \quad \{ij\} \in B$$

と変形される．これらの式より，連分数の表記を使って

$$\bar{\lambda}_{i \rightarrow j} = -\frac{\bar{s}_{i,j}^2}{\Lambda_i + \frac{\bar{s}_{j,i}^2}{\Lambda_j - \bar{\lambda}_{i \rightarrow j}}}$$

$$\bar{\lambda}_{j \rightarrow i} = -\frac{\bar{s}_{i,j}^2}{\Lambda_j + \frac{\bar{s}_{i,j}^2}{\Lambda_i - \bar{\lambda}_{j \rightarrow i}}}, \quad \{ij\} \in B$$

と書くことができ、これら  $\bar{\lambda}_{i \rightarrow j}, \bar{\lambda}_{j \rightarrow i}$  の2方程式を解くことで、ただちに、

$$\begin{aligned}\bar{\lambda}_{i \rightarrow j} &= \frac{\bar{\Lambda}_j}{2} \pm \frac{\bar{\Lambda}_j}{2} \sqrt{1 + \frac{4s_{i,j}^2}{\Lambda_i \Lambda_j}}, \\ \bar{\lambda}_{j \rightarrow i} &= \frac{\bar{\Lambda}_i}{2} \pm \frac{\bar{\Lambda}_i}{2} \sqrt{1 + \frac{4s_{j,i}^2}{\Lambda_j \Lambda_i}}, \quad \{ij\} \in B\end{aligned}\quad (6)$$

が得られる。周辺分布の精度について  $\Lambda_i = \sum_{k \in \mathcal{N}_i} \lambda_{k \rightarrow i}$ , ( $i \in \{1, \dots, d\}$ ) が成り立っていたことに留意すれば、この式に (6) を代入することで最終的に、

$$\Lambda_j = s_{j,j} + \sum_{k \in \mathcal{N}_j} \left( \frac{\Lambda_j}{2} \pm \frac{\Lambda_j}{2} \sqrt{1 + \frac{4s_{k,j}^2}{\Lambda_k \Lambda_j}} \right) \quad (7)$$

が  $j \in \{1, \dots, d\}$  で得られる。ただし符号  $\pm$  はそれぞれの項で任意にとる。このようにして、メッセージの精度を使わない、周辺分布の精度についての固定点方程式が得られた。符号  $\pm$  をすべて  $+$  として、固定点方程式 (7) を変形することで、目的の更新式 (5) が得られる。

## 5 数値実験

ここではガウス分布に対する確率伝搬法の数値実験を行い、3節で得られた理論的結果との比較を行う。

### 5.1 1重ループのときの確率伝搬法

まずはじめに、精度行列が1重ループで表される場合について考察する。精度行列  $S$  が  $3 \times 3, 4 \times 4$  などの小さい行列のときに、LBP 解が真の周辺分布と比較的大きくずれる行列の例を示す。精度行列  $S$  が

$$S_1 = \begin{pmatrix} 10 & -9.9 & 1 \\ -9.9 & 10 & 0.4 \\ 1 & 0.4 & 10 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 10 & 9 & 0 & 1 \\ 9 & 10 & 4 & 0 \\ 0 & 4 & 10 & 1.9 \\ 1 & 0 & 1.9 & 10 \end{pmatrix}$$

で表される場合 (相関の程度が大きい行列であるが)、確率伝搬法の結果は、それぞれ図1, 2となる。図は、プロット点が更新式の反復回数 (横軸) に対する精度  $\Lambda_i$  (縦軸) の収束の様子を示しており、破線が定理2のLBP解析解、直線が真の精度の値を表している。図1, 2から、確率伝搬法の更新式の収束先は、定理2で与えたLBP解析解と一致し、LBP解と真値はいくらもずれることがわかる。実際、これらの行列  $S_1, S_2$  についての  $\epsilon$  の値はそれぞれ  $\epsilon = 41.684211, \epsilon = 25.786993$  となり近似の程度が悪いことがわかる。

次に、このLBPの近似精度を決める  $\epsilon$  の量について、次元  $d$  に対する  $\epsilon$  の値の振る舞いを考察する。精度行列  $S$  として  $d \times d$  行列

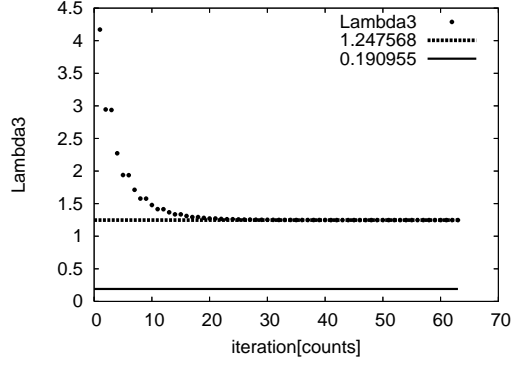


図1: 行列  $S_1$  における  $\Lambda_3$  の収束の様子

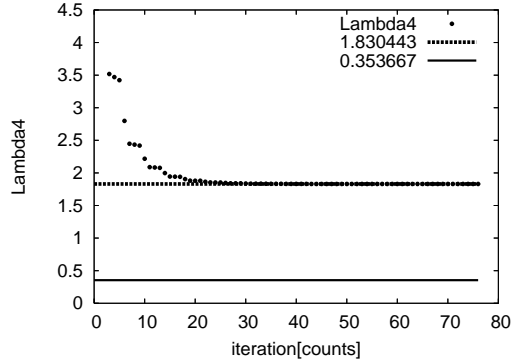


図2: 行列  $S_2$  における  $\Lambda_4$  の収束の様子

$$S_3(d) = \begin{pmatrix} 10 & 4.9 & 0 & \dots & 0 & 4.9 \\ 4.9 & 10 & 4.9 & 0 & \dots & 0 \\ 0 & 4.9 & 10 & 4.9 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 4.9 & 10 & 4.9 \\ 4.9 & 0 & \dots & 0 & 4.9 & 10 \end{pmatrix}$$

を考えるとすれば、次元  $d$  に対する  $\epsilon$  のグラフは図3となる。もう一つ、行列  $S_3(d)$  の代わりに、 $4.9 \rightarrow -4.9$  に変更した行列  $S_4(d)$  についてのグラフは図4となる。図3, 4からLBPの近似精度は次元  $d$  の増加に対して指数的に良くなることがわかる。このことは1重ループにおいては、次元  $d$  の増加によってローカルには木構造に近づくという直観的事実と合っている。指数的に良くなることは、 $\epsilon$  の値に対して、

$$\epsilon(d) = \frac{4(-s)^d}{\det S} \simeq 4 \left( -\frac{s}{s_0} \right)^d$$

の評価から説明づけられる。図3において、次元  $d$  の偶数奇数によってLBPの近似精度が振動することは興味深い。

### 5.2 相関が小さいときのLBP解

ここでは相関が小さいときのLBP解について数値的に考察する。表1は、いくつかの精度行列について、以下の量を並べた表になっている。

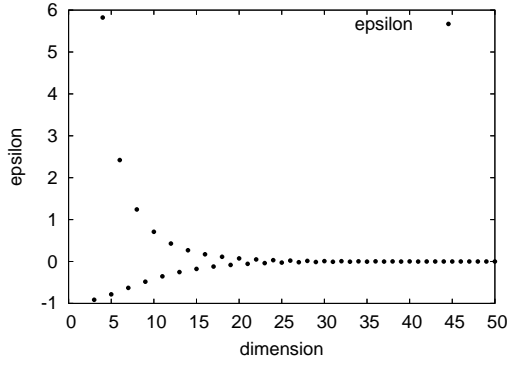


図 3: 行列  $S_3(d)$  における LBP 近似精度の変化

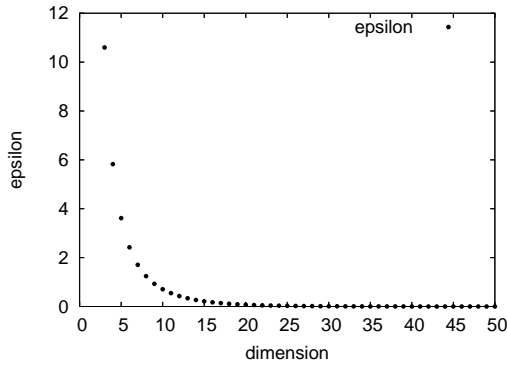


図 4: 行列  $S_4(d)$  における LBP 近似精度の変化

1. (i) メッセージの更新式の後，周辺分布構成式によって得られる通常の LBP 解．
2. (ii) メッセージを用いない周辺分布による直接的な更新式の LBP 解．
3. (iii) LBP 解析解の  $s^3$  までの展開式の値 (3)．
4. (iv) (iii) の LBP 解に，式 (4) における  $\delta_i s^3$  の値を加えて，解を補正した値．
5. (v) 1 重ループのときの LBP 解析解
6. (vi) 真の周辺密度関数の精度の値．

表中の行列は以下で与えられる精度行列である．

$$S_5 = \begin{pmatrix} 10 & 1 & 0 & 0 \\ 1 & 10 & 1 & 0 \\ 0 & 1 & 10 & 1 \\ 0 & 0 & 1 & 10 \end{pmatrix}, S_6 = \begin{pmatrix} 10 & 1 & 0 & 1 \\ 1 & 10 & 1 & 0 \\ 0 & 1 & 10 & 1 \\ 1 & 0 & 1 & 10 \end{pmatrix},$$

$$S_7 = \begin{pmatrix} 10 & 1 & 1 & 1 \\ 1 & 10 & 1 & 1 \\ 1 & 1 & 10 & 1 \\ 1 & 1 & 1 & 10 \end{pmatrix}, S_8 = \begin{pmatrix} 10 & 3 & 3 & 3 \\ 3 & 10 & 3 & 3 \\ 3 & 3 & 10 & 3 \\ 3 & 3 & 3 & 10 \end{pmatrix},$$

$$S_9 = \begin{pmatrix} 10 & 1 & 1 & 7 \\ 1 & 10 & 1 & 1 \\ 1 & 1 & 10 & 1 \\ 7 & 1 & 1 & 10 \end{pmatrix}, S_{10} = \begin{pmatrix} 10 & 9 & 1 \\ 9 & 10 & 5 \\ 1 & 5 & 10 \end{pmatrix}.$$

表 1: LBP 解の数値結果

|                              | (i)      | (ii)     | (iii)     |
|------------------------------|----------|----------|-----------|
| $S_5 (\Lambda_1)$            | 9.898980 | 9.898980 | 9.900000  |
| $S_6$                        | 9.797959 | 9.797959 | 9.800000  |
| $S_7$                        | 9.693747 | 9.693747 | 9.700000  |
| $S_8$                        | 6.468627 | 6.468627 | 7.300000  |
| $S_9 (\Lambda_1, \Lambda_4)$ | 4.790380 | 4.790380 | 4.900000  |
| $S_9 (\Lambda_2, \Lambda_3)$ | 9.487039 | 9.487039 | 9.700000  |
| $S_{10} (\Lambda_2)$         | —        | —        | -0.600000 |

|                              | (iv)     | (v)      | (vi)     |
|------------------------------|----------|----------|----------|
| $S_5 (\Lambda_1)$            | 9.900000 | 9.898980 | 9.898980 |
| $S_6$                        | 9.800000 | 9.797959 | 9.795918 |
| $S_7$                        | 9.760000 | 9.752057 | 9.750000 |
| $S_8$                        | 8.920000 | 8.516626 | 8.312500 |
| $S_9 (\Lambda_1, \Lambda_4)$ | 5.200000 | 5.097716 | 5.083333 |
| $S_9 (\Lambda_2, \Lambda_3)$ | 9.880000 | 9.831310 | 9.803571 |
| $S_{10} (\Lambda_2)$         | 0.300000 | —        | 0.202020 |

$S_5$  は木構造のグラフ構造を持つ精度行列であり， $S_6$  は 1 重ループのグラフ構造を持つ行列， $S_7$  は一様全結合のグラフを表し， $S_8$  は一様全結合のグラフで，行列  $S_7$  よりも相関の程度が大きい行列である． $S_9, S_{10}$  は全結合のグラフを表すが一様ではない行列である．

表 1 の数値結果 (i),(ii) にあるように，周辺分布の更新式 (5) は，(i) と同じ LBP 解に収束する．精度行列  $S_5$  は木構造であることから，LBP 解 (i),(ii)，1 重ループ解 (v) は共に，真の値 (vi) に一致している．精度行列  $S_6$  は 1 重ループ構造であることから，LBP 解 (i),(ii) は 1 重ループ解 (v) と一致しているが，真値 (vi) からはずれてきている．精度行列  $S_7$  は全結合であることから LBP 解 (i),(ii) は 1 重ループ解 (v)，真値 (vi) からともはずれた値となっている．また，LBP 解 (i),(ii) は (iii) とほぼ一致していることから，相関の値が小さいときには，定理 3 の結果が正しいことを数値的に支持している．(iii) を (iv) に補正することで，真値 (vi) に近づいていることも見てとれる．精度行列  $S_8$  は行列  $S_7$  より相関の程度が高いことから，LBP 解 (i),(ii) は真値 (vi) より比較的大きくずれ，展開式 (iii) による LBP 解の近似も荒いものとなっている．精度行列  $S_{10}$  は確率伝搬法が収束しない行列となっており，実際， $\epsilon$  の値は， $\epsilon = -9$  となり  $\epsilon \geq -1$  の収束条件を満たさない．行列  $S_7, S_8, S_9$  は 1 重ループ構造でないことから，1 重ループ解 (v) を計算することに何ら理論的保証はないが，(v) の値が LBP 解 (i),(ii) よりも真値 (vi) に近づいていることは興味深い．このことは，定理 2 の解析解は 1 重ループという条件下で成立していたものだが，1 重ループ以外のグラフ構造においても良好な解が得られる可能性を数値的に示

峻している .

## 6 結論と今後の展開

本稿では, 正規分布に対するベータ自由エネルギーを解析し, ベータ近似解の解析解を 1 重ループのグラフ構造の場合に与えた. その結果, LBP の近似精度を決定する量を明らかにし, LBP が収束する (必要) 条件を与えた. 任意のグラフ構造については相関が小さいときのベータ近似解の展開式を与えた. LBP の更新式として, メッセージを使わない, 周辺分布を直接更新する更新式を提案した. 理論的結果と数値実験を比較することで, 数値計算が理論的結果を支持していることを確認し, 特に, 1 重ループ構造においては, LBP の近似精度が振動する場合があることや, 1 重ループの解析解を 1 重ループでないグラフ構造のときにも適用した結果, LBP 解よりも良好な結果が得られる場合があることを数値的に示した.

今後の展開として, 確率画像処理において, LBP, 一般化された信念伝搬法 (GBP), TAP 自由エネルギーを使った方法のそれぞれの近似精度が, ガウシアングラフィカルモデルのときに数値的に比較研究されているが [12], それらの解析解を評価し, 本稿との比較が考えられる.

## 7 謝辞

本研究の一部は, 科学研究費補助金 18079007 および特別研究員 19005165 の援助を受けている. 査読者の方々には有意義なコメントを頂きました. 感謝致します.

## 参考文献

- [1] T. Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, Vol. 16, No. 11, pp. 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*, Vol. 12, No. 1, pp. 1–41, 2000.
- [4] Y. Weiss and W. Freeman. Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Computation*, Vol. 13, No. 10, pp. 2173–2200, 2001.
- [5] S. Ikeda, T. Tanaka, and S. Amari. Stochastic reasoning, free energy, and information geometry. *Neural Computation*, Vol. 16, No. 9, pp. 1779–1810, 2004.
- [6] 池田思朗, 田中利幸, 甘利俊一. 情報幾何学に基づく確率伝搬法の解析. *統計数理*, Vol. 43, No. 500, pp. 393–405, 2005.
- [7] Y. Kabashima and D. Saad. The TAP approach to intensive and extensive connectivity systems. In M. Oppor and D. Saad, editors, *Advanced Mean Field Methods - Theory and Practice*, pp. 65–84. MIT Press, 2001.
- [8] J. Yedidia, W. Freeman, and Y. Weiss. Bethe free energy, kikuchi approximations, and belief propagation algorithms. Technical Report TR2001-16, Mitsubishi Electric Research Laboratories, 2001.
- [9] K. Tanaka, H. Shouno, and 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, 2004.
- [10] 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.
- [11] Y. Nishiyama and S. Watanabe. Theoretical Analysis of Accuracy of Belief Propagation in Gaussian Models. *IEICE Technical Report*, Vol. 107, No. 50, pp. 23–28, 2007.
- [12] 田中和之. ガウシアングラフィカルモデルに基づく確率的情報処理における一般化された信念伝搬法. *電子情報通信学会誌 D-II*, Vol. J88-D-II, No. 12, pp. 2368–2379, 2005.