統計数理(2004)
第 52 巻 第 2 号 393–405

[研究詳解]

情報幾何学に基づく確率伝搬法の解析

池田 思朗[1] ・田中 利幸[2] ・甘利 俊一[3]

(受付 2004 年 2 月 27 日；改訂 2004 年 6 月 29 日)

要　　旨

1980 年代後半 Pearl が提案した確率伝搬法は，大規模なグラフィカルモデルに対する確率推論のための計算手法である．同等の手法は統計物理学，統計学，誤り訂正符号の復号法などにも存在し，広く用いられている．確率伝搬法は木の構造のグラフに対してはグラフの大きさに比例した計算量で厳密解が得られる．しかしループを持つグラフに対しては繰り返し計算の収束性，および得られた結果の近似精度ともに理論的には十分理解されていなかった．一方で確率伝搬法は実用上有効な手法であり，その性質を理論的に明らかにすることは重要である．本研究では情報幾何学に基づく枠組みにより確率伝搬法を表現し，収束性や近似精度を議論する．

キーワード： 確率伝搬法，情報幾何学，グラフィカルモデル．

1. はじめに

グラフィカルモデル(Lauritzen and Spiegelhalter (1988), Jordan (1999))では，複数の確率変数の同時分布をグラフによって表現する．グラフによって表現された複数の確率変数の一部のみが観測されたとき，その条件付確率分布から観測されていない確率変数の値を推論する問題を考える．この問題は確率推論と呼ばれ，人工知能，統計物理学，情報理論など様々な分野で重要である．

Pearl (1988) はこの問題に対し，確率伝搬法(Belief Propagation)と呼ばれる簡便な繰り返し計算法を提案した．確率伝搬法は条件付き分布から各確率変数の周辺分布を求め，確率推論を得る．木のグラフに対しては収束が保証され，グラフの辺の数に比例した計算量で正しい推論結果が得られることが分かっているが，グラフがループを持つ場合には必ずしも収束しないこと，収束した場合でも得られる解が一般に近似解となることが知られている．

同等の計算手法は他の分野でも広く用いられている．統計物理のベーテ(Bethe)近似(Kabashima and Saad (1998))，また，低密度パリティ検査符号(Gallager (1962), MacKay (1999))，ターボ符号(Berrou et al. (1993), McEliece et al. (1998))といった誤り訂正符号の復号法などは確率伝搬法と等しいことが知られている．統計物理や符号理論で扱う問題に対するグラフィカルモデルは一般にはループを持ち，収束性や近似精度といった問題がある．

確率推論の問題は MCMC などの手法によっても解が得られる．その場合ループがあっても構わないが，精度の良い解を得るために必要な計算量は多い．前に挙げた誤り訂正符号ではグラフの確率変数の数が通常数百から千程度であり，実時間で復号するには計算量の少ない確率

[1] 統計数理研究所：〒106–8569 東京都港区南麻布 4–6–7
[2] 東京都立大学大学院 工学研究科：〒192–0397 東京都八王子市南大沢 1–1
[3] 理化学研究所 脳科学総合研究センター：〒351–0198 埼玉県和光市広沢 2–1

伝搬法が適している．

これまで我々は情報幾何学(甘利・長岡(1993), Amari and Nagaoka (2000))に基づき確率伝搬法を表現し，解の安定性，近似精度といった問題を扱ってきた(Ikeda et al. (2002)， 池田 他 (2002)，Tanaka et al. (2002)，Ikeda et al. (2003, 2004a, 2004b))．本稿ではこれまでの結果をまとめ，新たに得られた 3 次の摂動展開に基づく近似精度の評価について示す．

2. 確率推論の問題と情報幾何的枠組み

2.1 問題の表現

$\boldsymbol{x} = (x_1, \ldots, x_n)^T$ を観測できない確率変数，$\boldsymbol{y} = (y_1, \ldots, y_m)^T$ を観測された確率変数とする．本稿では簡単のため x_i が 2 値変数，特に $x_i \in \{-1, +1\}$ の場合を考える．多値変数への拡張は簡単であり，連続値への拡張も場合によっては可能である(Ikeda et al. (2003, 2004b))．

確率推論の問題は $\boldsymbol{y}$ の条件付きでの $\boldsymbol{x}$ の分布 $q(\boldsymbol{x}|\boldsymbol{y})$ から(簡単のため以下では $q(\boldsymbol{x}|\boldsymbol{y})$ を $q(\boldsymbol{x})$ とかくことにする)，$\boldsymbol{x}$ に関する推論を得ることである．1 つの方法は $q(\boldsymbol{x})$ を最大にする $\boldsymbol{x}$ (MAP 推論: maximum a posteriori)を用いることである．MAP 推論は推論結果が真の $\boldsymbol{x}$ と異なる確率を最小にするが，探索空間が n とともに指数関数的に増える．ここでは別の推論，周辺事後確率分布の最大化(MPM 推論: maximization of the posterior marginals)を考える．$q(\boldsymbol{x})$ の周辺分布を $q(x_i),\ i = 1, \ldots, n$ とするとき MPM 推論では各成分の推論結果を $q(x_i = +1) \geq q(x_i = -1)$ ならば $\hat{x}_i = +1$，それ以外の場合は $\hat{x}_i = -1$ とする．この結果各 x_i が誤って推論される確率は最小となる．η_i を $q(\boldsymbol{x})$ による x_i の期待値とする，

$$\eta_i = E_q[x_i] = \sum_{x_i} x_i q(x_i)\,.$$

MPM 推論は $\hat{x}_i = \mathrm{sgn}\ \eta_i$ とする．これは周辺分布の積 $\prod_{i=1}^n q(x_i)$ あるいは $\boldsymbol{x}$ の期待値

$$\boldsymbol{\eta} = E_q[\boldsymbol{x}]\,,$$

が分かればただちに計算できる．仮に n とともに指数的に増加しない，簡単な計算によって η あるいは $\prod_{i=1}^n q(x_i)$ が計算できれば MPM 推論は有効な推論手法となる．本稿で扱う確率伝搬法は簡単な計算によって $\prod_{i=1}^n q(x_i)$ の近似を得ようというものである．

今 x_i は 2 値なので，全ての $q(\boldsymbol{x})$ に対し $\ln q(\boldsymbol{x})$ は $\{x_i\}$ の高々 n 次関数として表現できる．グラフィカルモデルで表現できる問題の多くは確率変数間の相互関係が限られており，低次の関数で表現できる場合が多い．そこで $\ln q(\boldsymbol{x})$ を次のように表現する

$$\ln q(\boldsymbol{x}) = \boldsymbol{h} \cdot \boldsymbol{x} + \sum_{r=1}^{L} c_r(\boldsymbol{x}) - \psi_q\,, \tag{2.1}$$

ここで $\boldsymbol{h} \cdot \boldsymbol{x} = \sum_i h_i x_i$ は x_i の線形項，$c_r(\boldsymbol{x}), r = 1, \ldots, L$，は単項式でも多項式でもよいが高次項を表し，$\psi_q$ は規格化定数の対数である．$c_r(\boldsymbol{x})$ はクリーク関数とも呼ばれる．ボルツマンマシンや古典的なスピングラスのモデルでは $c_r(\boldsymbol{x})$ は x_i の二次関数である．

$$c_r(\boldsymbol{x}) = w_{ij} x_i x_j\,,$$

ここで r は辺の番号を表すもので，x_i と x_j 間の辺を示す．以下ではこのモデルをボルツマンマシンと呼ぶ(図 1)．

一般のグラフィカルモデル，特に無向グラフでは $q(\boldsymbol{x})$ を以下のようにクリーク関数の積として定義することが多い．

$$q(\boldsymbol{x}) = \frac{1}{Z_q} \prod_{i=1}^{n} \phi_i(x_i) \prod_{r \in \mathcal{C}} \phi_r(\boldsymbol{x}_r)\,,$$

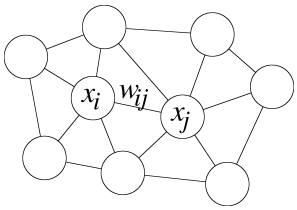

図 1.　ボルツマンマシン.

ここで $\mathcal{C}$ はクリークの集合である. (2.1) **式の表現と** $\phi_i(x_i)$, $\phi_r(\boldsymbol{x}_r)$ **とは次の関係がある.**

$$h_i = \frac{1}{2}\ln\frac{\phi_i(x_i=+1)}{\phi_i(x_i=-1)}\,,\quad c_r(\boldsymbol{x}) = \ln\phi_r(\boldsymbol{x}_r)\,,\quad \psi_q = \ln Z_q\,.$$

ボルツマンマシンでは $\phi_r(\boldsymbol{x}_r)$ **は** $\phi_r(x_i,x_j)$ **となる. 本論文では** $c_r(\boldsymbol{x})$, $r=1,\ldots,L$ **は一次独立であるとする. なお, 本稿では無向グラフのみを扱うが全ての有向グラフは無向グラフで表現できることから一般性は失っていない(**Lauritzen and Spiegelhalter **(**1988**)).**

2.2　情報幾何学と多様体

本節では情報幾何学に基づく議論のための準備を行う.
まず, 確率伝搬法を考えるために必要な確率分布全体の族 S **を考える.**

$$S=\{p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v}) \;\big|\; p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v})=\exp[\boldsymbol{h}\cdot\boldsymbol{x}+\boldsymbol{\theta}\cdot\boldsymbol{x}+\boldsymbol{v}\cdot\boldsymbol{c}(\boldsymbol{x})-\psi(\boldsymbol{\theta},\boldsymbol{v})],\ \boldsymbol{\theta}\in\Re^n,\boldsymbol{v}\in\Re^L\}\,, \tag{2.2}$$

$(\boldsymbol{\theta},\boldsymbol{v})$ **は自然パラメータであり,** $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_n)^T$, $\boldsymbol{v}=(v_1,\ldots,v_L)^T$, $\boldsymbol{c}(\boldsymbol{x})=(c_1(\boldsymbol{x}),\ldots,c_L(\boldsymbol{x}))^T$, $\boldsymbol{v}\cdot\boldsymbol{c}(\boldsymbol{x})=\sum_{r=1}^{L}v_rc_r(\boldsymbol{x})$ **である.** S **は定義より指数型分布族である.** $q(\boldsymbol{x})$ **が** S **に含まれることは** $\boldsymbol{\theta}=\boldsymbol{0}$, $\boldsymbol{v}=\boldsymbol{1}_L=(1,\ldots,1)^T$ **とおけば簡単に確かめられる. 次に** $v=\boldsymbol{0}$ **となる** S **の部分多様体を** M_0 **と呼ぶことにする.**

$$M_0=\{p_0(\boldsymbol{x};\boldsymbol{\theta})=\exp[\boldsymbol{h}\cdot\boldsymbol{x}+\boldsymbol{\theta}\cdot\boldsymbol{x}-\psi_0(\boldsymbol{\theta})]\mid\boldsymbol{\theta}\in\Re^n\}\,.$$

M_0 **の分布は各成分が独立であり, 自然パラメータは** θ **である. さらに, 各成分が独立な分布は全て** M_0 **に含まれることから** $q(\boldsymbol{x})$ **の周辺分布の積** $\prod_{i=1}^{n}q(x_i)$ **も** M_0 **に含まれる. すなわち** $\prod_{i=1}^{n}q(x_i)$ **に対応する** M_0 **の座標** θ **を見つければ** MPM **推論が得られる.**

次に S **の中の** e**-平坦,** m**-平坦な部分多様体を定義する.**

e**-平坦な部分多様体:　部分多様体** $M\subset S$ **は次の** $r(\boldsymbol{x};t)$ **が全ての** $t\in[0,1], q(\boldsymbol{x}), p(\boldsymbol{x})\in M$ **に対して** M **に含まれるとき** e**-平坦である.**

$$\ln r(\boldsymbol{x};t)=(1-t)\ln q(\boldsymbol{x})+t\ln p(\boldsymbol{x})+c(t)\,,$$

$c(t)$ **は規格化定数である.** e**-平坦な部分多様体が** 1 **次元の曲線のとき, 特に** e**-測地線と呼ぶ.** $\{r(\boldsymbol{x};t)\mid t\in[0,1]\}$ **は** $p(\boldsymbol{x})$ **と** $q(\boldsymbol{x})$ **を結ぶ** e**-測地線である.**

m**-平坦な部分多様体:　部分多様体** $M\subset S$ **は次の** $r(\boldsymbol{x};t)$ **が全ての** $t\in[0,1], q(\boldsymbol{x}), p(\boldsymbol{x})\in M$ **に対して** M **に含まれるとき** m**-平坦である.**

$$r(\boldsymbol{x};t)=(1-t)q(\boldsymbol{x})+tp(\boldsymbol{x})\,.$$

m**-平坦な部分多様体が** 1 **次元の曲線のとき, 特に** m**-測地線と呼ぶ.** $\{r(\boldsymbol{x};t)\mid t\in[0,1]\}$ **は** $p(\boldsymbol{x})$ **と** $q(\boldsymbol{x})$ **を結ぶ** m**-測地線である.**

定義よりただちに指数型分布族が e-平坦であることがわかる．したがって S と M_0 は e-平坦である．次に m-射影を定義する．本稿では e-平坦な多様体への m-射影を考えれば十分であるので，次のように定義する．

Definition 1. M を S の e-平坦な部分多様体とし，$q(\boldsymbol{x})\in S$ とする．$q(\boldsymbol{x})$ から M 上の点 $p(\boldsymbol{x})$ への m-測地線が Fisher 情報行列を Riemann 計量として M と直交する点を $q(\boldsymbol{x})$ から M への m-射影と呼ぶ．

Theorem 1. M が S の e-平坦な部分多様体であるとき $q(\boldsymbol{x})\in S$ から M への m-射影は唯一である．今 $q(\boldsymbol{x})\in S$ から M への m-射影を $\Pi_M\circ q(\boldsymbol{x})$ とかくと，$\Pi_M\circ q(\boldsymbol{x})$ は M において $q(\boldsymbol{x})$ からの Kullback-Leibler (KL)divergence を最小にする点である．すなわち

$$\Pi_M\circ q(\boldsymbol{x}) = \operatorname*{argmin}_{p(\boldsymbol{x})\in M} D[q(\boldsymbol{x}); p(\boldsymbol{x})] = \operatorname*{argmin}_{p(\boldsymbol{x})\in M} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} \,. \tag{2.3}$$

ここで $D[\cdot;\cdot]$ は KL-divergence である．

2.3 MPM 推論

本節では MPM 推論が $q(\boldsymbol{x})$ から M_0 への m-射影と等しいことを示す．(2.3) 式の定義から $q(\boldsymbol{x})$ から M_0 への m-射影がパラメータ $\boldsymbol{\theta}^*$ に対応するならば

$$p_0(\boldsymbol{x};\boldsymbol{\theta}^*) = \Pi_{M_0}\circ q(\boldsymbol{x})$$

とかける．これより M_0 への m-射影により求まるパラメータ $\boldsymbol{\theta}$ を次のように書くことにする．

$$\boldsymbol{\theta}^* = \pi_{M_0}\circ q(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{\theta}} D[q(\boldsymbol{x}); p_0(\boldsymbol{x};\boldsymbol{\theta})] \,.$$

$D[q(\boldsymbol{x}); p_0(\boldsymbol{x};\boldsymbol{\theta})]$ を $\boldsymbol{\theta}$ で微分し

$$\sum_{\boldsymbol{x}} \boldsymbol{x} q(\boldsymbol{x}) - \partial_{\boldsymbol{\theta}}\psi_0(\boldsymbol{\theta}^*) = \mathbf{0} \,. \tag{2.4}$$

$\partial_{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ による微分を表す．指数型分布族の定義から

$$\partial_{\boldsymbol{\theta}}\psi_0(\boldsymbol{\theta}) = \partial_{\boldsymbol{\theta}} \ln \sum_{\boldsymbol{x}} \exp(\boldsymbol{h}\cdot\boldsymbol{x} + \boldsymbol{\theta}\cdot\boldsymbol{x}) = \sum_{\boldsymbol{x}} \boldsymbol{x} p_0(\boldsymbol{x};\boldsymbol{\theta}) \,. \tag{2.5}$$

ここで M_0 の期待値パラメータ $\boldsymbol{\eta}_0(\boldsymbol{\theta})$ を次のように定義する．

$$\boldsymbol{\eta}_0(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}} \boldsymbol{x} p_0(\boldsymbol{x};\boldsymbol{\theta}) = \partial_{\boldsymbol{\theta}}\psi_0(\boldsymbol{\theta}) \,. \tag{2.6}$$

M_0 の 2 つの座標系 $\boldsymbol{\theta}$ と $\boldsymbol{\eta}_0$ とは 1 対 1 に対応するので，(2.4), (2.5), (2.6) 式から m-射影が $q(\boldsymbol{x})$ の周辺化と等しく，MPM 推論と等価であることがわかる．

3. 確率伝搬法の情報幾何

3.1 確率伝搬法の情報幾何的表現

本節では確率伝搬法の情報幾何的な表現を与える．グラフィカルモデルにおける確率伝搬法はメッセージと呼ばれる変数の更新規則を定義するのが通常である (Pearl (1988), Weiss (2000))．まず，ボルツマンマシンに対する確率伝搬法を Yedidia et al. (2001) にしたがって定義する．確率伝搬法ではメッセージ $m_{ji}(x_i)$ を次の式で更新し，全てが収束した後，得られたメッセージ

$m^*_{ji}(x_i)$ を用いてビリーフ $b_i(x_i)$ を求める．

$$
\begin{aligned}
m^{t+1}_{ji}(x_i) =& \frac{1}{Z}\sum_{x_j}\phi_j(x_j)\phi_{ij}(x_i,x_j)\prod_{k\in\mathcal{N}(j)\setminus i} m^t_{kj}(x_j)\\
b_i(x_i) =& \frac{1}{Z'}\pi_i(x_i)\sum_{k\in\mathcal{N}(i)} m^*_{ki}(x_i)\,.
\end{aligned}
\tag{3.1}
$$

$\mathcal{N}(j)$ は j 番目の節につながっている節の集合，Z，Z' はそれぞれ $m^{t+1}_{ji}(x_i)$ および $b_i(x_i)$ の x_i に関する和を 1 に規格化する．木のグラフではビリーフ $b_i(x_i)$ は $q(\boldsymbol{x})$ の x_i に関する周辺分布 $q(x_i)$ と一致するが，ループのある場合には一般に一致せず，近似となる．

以下ではこの確率伝搬法の情報幾何学に基づく表現を与える．確率伝搬法では $q(\boldsymbol{x})$ と $p_0(\boldsymbol{x};\boldsymbol{\theta})$ さらに次式の $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$, $r=1,\ldots,L$, を用いる．

$$
p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)=\exp[\boldsymbol{h}\cdot\boldsymbol{x}+c_r(\boldsymbol{x})+\boldsymbol{\zeta}_r\cdot\boldsymbol{x}-\psi_r(\boldsymbol{\zeta}_r)]\,,\quad \boldsymbol{\zeta}_r\in\Re^n\,,\quad r=1,\ldots,L\,.
$$

$p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$ はクリーク関数を 1 つだけ含んだ確率分布である．$c_r(\boldsymbol{x})$ は一般に $\{x_i\}$ の高次の関数であるが，$\{c_r(\boldsymbol{x})\}$ の全てを含む場合に比べて 1 つのみを含む場合は扱い易い．他のクリーク関数は $\boldsymbol{\zeta}_r\cdot\boldsymbol{x}$ で代替されている．$p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$ は指数型分布族であり，以下で定義される M_r は e-平坦である．

$$
M_r=\{p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)\mid\boldsymbol{\zeta}_r\in\Re^n\}\,,\quad r=1,\ldots,L\,.
$$

自然パラメータは $\boldsymbol{\zeta}_r$ である．M_r の期待値パラメータを $\boldsymbol{\eta}_r(\boldsymbol{\zeta}_r)$ と定める．

$$
\boldsymbol{\eta}_r(\boldsymbol{\zeta}_r)=\partial_{\boldsymbol{\zeta}_r}\psi_r(\boldsymbol{\zeta}_r)=\sum_{\boldsymbol{x}}\boldsymbol{x}p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)\,,\quad r=1,\ldots,L\,.
\tag{3.2}
$$

$\boldsymbol{\eta}_r(\boldsymbol{\zeta}_r)$ を求めるための計算量は $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$ の周辺化と同等である．このためには $c_r(\boldsymbol{x})$ に含まれる x_i の全ての組み合わせについて計算しなければいけないが，この計算は可能だとする．全ての r に対して $\boldsymbol{\eta}_r(\boldsymbol{\zeta}_r)$ を求めるほうが $q(\boldsymbol{x})$ の周辺化を直接行うより圧倒的に計算量は少ない．なお，$\boldsymbol{\zeta}_r$ から $\boldsymbol{\eta}_r(\boldsymbol{\zeta}_r)$ を求めるのは簡単だが，一般に逆は簡単でないことを注意しておく．

確率伝搬法では $q(\boldsymbol{x})$ の代わりに $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$ の周辺化を全ての辺の数だけ行い，全体として $q(\boldsymbol{x})$ の近似を行う．つまり $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$, $r=1,\ldots,L$ によって各 $c_r(\boldsymbol{x})$ の影響を表現し，全ての影響を $\boldsymbol{\theta}$ にまとめ，$p_0(\boldsymbol{x};\boldsymbol{\theta})$ によって $\prod_i q(x_i)$ を近似する．その際，繰り返し計算によって $\{\boldsymbol{\zeta}_r\}$ および $\boldsymbol{\theta}$ を更新する．

我々は，誤り訂正符号で用いられる特殊なグラフィカルモデルに対しては池田 他(2002)および Ikeda et al.(2004a)で，一般のグラフに対しては Ikeda et al.(2003, 2004b)で確率伝搬法の情報幾何的な表現を得ている(Richardson (2000)も同様の枠組みを与えている)．ここではその結果を示す．

$p_0(\boldsymbol{x};\boldsymbol{\theta}^t)$ および $p_r(\boldsymbol{x};\boldsymbol{\zeta}^t_r)$ を時刻 t におけるそれぞれ M_0, M_r, $r=1,\ldots,L$ での $q(\boldsymbol{x})$ の近似とする．確率伝搬法は $\boldsymbol{\zeta}^t_r$ および $\boldsymbol{\theta}^t$ を以下のように更新する．

確率伝搬法の情報幾何的表現

(1) 初期値を $t=0$, $\boldsymbol{\xi}^t_r=\boldsymbol{0}$, $\boldsymbol{\zeta}^t_r=\boldsymbol{0}$, $r=1,\ldots,L$ とする．

(2) t を 1 つずつ増加させ $\boldsymbol{\xi}^{t+1}_r$, $r=1,\ldots,L$ を次のように更新する．

$$
\boldsymbol{\xi}^{t+1}_r=\pi_{M_0}\circ p_r(\boldsymbol{x};\boldsymbol{\zeta}^t_r)-\boldsymbol{\zeta}^t_r.
\tag{3.3}
$$

(3) $\boldsymbol{\theta}^{t+1}$ と $\boldsymbol{\zeta}^{t+1}_r$ を以下のように更新する．

$$
\boldsymbol{\zeta}^{t+1}_r=\sum_{r'\neq r}\boldsymbol{\xi}^{t+1}_{r'},\quad \boldsymbol{\theta}^{t+1}=\sum_r\boldsymbol{\xi}^{t+1}_r=\frac{1}{L-1}\sum_r\boldsymbol{\zeta}^{t+1}_r\,.
\tag{3.4}
$$

(4) 2 **と** 3 **を** $\{\xi_r^t\}$ **が収束するまで繰り返す.**

$\boldsymbol{\theta}^t = \sum_r \boldsymbol{\xi}_r^t$, $\boldsymbol{\theta}^t = \boldsymbol{\xi}_r^t + \boldsymbol{\zeta}_r^t$ **は常に成り立つ. 確率伝搬法が収束した時点でのパラメータを** $\boldsymbol{\theta}^*$, $\{\boldsymbol{\zeta}_r^*\}$, $\{\boldsymbol{\xi}_r^*\}$ **とする. 直感的に確率伝搬法を理解するには** $q(\boldsymbol{x})$, $p_0(\boldsymbol{x};\boldsymbol{\theta}^*)$, $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)$, $r = 1,\ldots,L$ **に関する以下の関係を見るのが良い.**

$$
\begin{aligned}
q(\boldsymbol{x}) &= \exp[\boldsymbol{h}\cdot\boldsymbol{x} + c_1(\boldsymbol{x}) + \cdots + c_r(\boldsymbol{x}) + \cdots + c_L(\boldsymbol{x}) - \psi_q] \\
p_0(\boldsymbol{x};\boldsymbol{\theta}^*) &= \exp[\boldsymbol{h}\cdot\boldsymbol{x} + \boldsymbol{\xi}_1^*\cdot\boldsymbol{x} + \cdots + \boldsymbol{\xi}_r^*\cdot\boldsymbol{x} + \cdots + \boldsymbol{\xi}_L^*\cdot\boldsymbol{x} - \psi_0(\boldsymbol{\theta}^*)] \\
p_1(\boldsymbol{x};\boldsymbol{\zeta}_1^*) &= \exp[\boldsymbol{h}\cdot\boldsymbol{x} + c_1(\boldsymbol{x}) + \cdots + \boldsymbol{\xi}_r^*\cdot\boldsymbol{x} + \cdots + \boldsymbol{\xi}_L^*\cdot\boldsymbol{x} - \psi_1(\boldsymbol{\zeta}_1^*)] \\
&\vdots \\
p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*) &= \exp[\boldsymbol{h}\cdot\boldsymbol{x} + \boldsymbol{\xi}_1^*\cdot\boldsymbol{x} + \cdots + c_r(\boldsymbol{x}) + \cdots + \boldsymbol{\xi}_L^*\cdot\boldsymbol{x} - \psi_r(\boldsymbol{\zeta}_r^*)] \\
&\vdots \\
p_L(\boldsymbol{x};\boldsymbol{\zeta}_L^*) &= \exp[\boldsymbol{h}\cdot\boldsymbol{x} + \boldsymbol{\xi}_1^*\cdot\boldsymbol{x} + \cdots + \boldsymbol{\xi}_r^*\cdot\boldsymbol{x} + \cdots + c_L(\boldsymbol{x}) - \psi_L(\boldsymbol{\zeta}_L^*)]\,.
\end{aligned}
$$

この式から確率伝搬法では各 $c_r(\boldsymbol{x})$ **を** $\boldsymbol{\xi}_r^*\cdot\boldsymbol{x}$ **によって表現していることがわかる.**

情報幾何的表現で用いた ξ_r **と** (3.1) **式で用いたメッセージとの関係をボルツマンマシンの場合に示す. 辺** r **が** i **と** j **の節を結ぶものとすると,**

$$
\xi_{r,i} = \frac{1}{2}\ln\frac{m_{ji}(x_i=+1)}{m_{ji}(x_i=-1)}\,,\quad \xi_{r,j} = \frac{1}{2}\ln\frac{m_{ij}(x_j=+1)}{m_{ij}(x_j=-1)}\,,\quad \xi_{r,k} = 0 \text{ for } k\neq i,j\,.
$$

この関係を用いると (3.1) **式のメッセージの更新則と** (3.3) **式が対応すること,** $p_0(\boldsymbol{x};\boldsymbol{\theta}^*)$ **が** $\prod_i b_i(x_i)$ **と等しいことが分る.**

3.2 停留点の持つ性質

Theorem 2. 確率伝搬法の停留点は以下の 2 **つの条件を満たす(**Ikeda et al. **(**2004a**)).**

m-**条件:** $\boldsymbol{\theta}^* = \pi_{M_0}\circ p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)$.

e-**条件:** $\boldsymbol{\theta}^* = \dfrac{1}{L-1}\displaystyle\sum_{r=1}^{L}\boldsymbol{\zeta}_r^*$.

確率伝搬法の停留点で m-**条件が満たされていることは** (3.3) **式と** $\boldsymbol{\theta}^* = \boldsymbol{\zeta}_r^* + \boldsymbol{\xi}_r^*$ **から簡単に確かめられる.** e-**条件は** (3.4) **式からただちに確かめられる.**

情報幾何的な意味を明らかにするため, S **の** 2 **つの部分多様体** M^* **と** E^* **を定義する.**

$$
\begin{aligned}
M^* &= \left\{ p(\boldsymbol{x}) \;\middle|\; p(\boldsymbol{x})\in S,\ \sum_{\boldsymbol{x}}\boldsymbol{x}p(\boldsymbol{x}) = \sum_{\boldsymbol{x}}\boldsymbol{x}p_0(\boldsymbol{x};\boldsymbol{\theta}^*) = \boldsymbol{\eta}_0(\boldsymbol{\theta}^*)\right\}, \\
E^* &= \left\{ p(\boldsymbol{x}) = Cp_0(\boldsymbol{x};\boldsymbol{\theta}^*)^{t_0}\prod_{r=1}^{L} p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)^{t_r} \;\middle|\; \sum_{r=0}^{L} t_r = 1,\ t_r\in\Re\right\}, \\
&\quad C : \textbf{規格化定数}.
\end{aligned}
\tag{3.5}
$$

M^* **は** m-**平坦な部分多様体であり** M_0 **および** M_r, $r = 1,\ldots,L$ **と直交する. 一方,** E^* **は** e-**平坦な部分多様体である. また定義より,** M^* **は** $p_0(\boldsymbol{x};\boldsymbol{\theta}^*)$ **を,** E^* **は** $p_0(\boldsymbol{x};\boldsymbol{\theta}_0^*)$ **と** $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)$, $r = 1,\ldots,L$ **を含む.**

前の 2 **つの条件はこれらの部分多様体を用いて次のようにかける.**

m-**条件:** M^* **が** $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)$, $r = 1,\ldots,L$ **を含む.**

e-条件: E^* が $q(\boldsymbol{x})$ を含む.

m-条件が上のように書き直せることは定義より明らかである．e-条件については (3.5) 式で $t_0 = -(L-1)$, $t_1 = \cdots = t_L = 1$ とおけば $\boldsymbol{\theta}^* = \sum_{r=1}^{L} \boldsymbol{\zeta}_r^*/(L-1)$ と同値なことがわかる．

確率伝搬法が収束した点では e-条件と m-条件が同時に満たされるが，だからといって $p_0(\boldsymbol{x};\boldsymbol{\theta}^*)$ が真の周辺分布 $\prod_{i=1}^{n} q(x_i)$ であるわけではない．これは M^* と E^* との間の差から生じる．M^* と E^* はともに $p_0(\boldsymbol{x};\boldsymbol{\theta}^*)$ と $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)$ を含んでいるが，他の点については明らかではない．本来ならば $q(\boldsymbol{x})$ が M^* に含まれればよいのだが，計算量を減らすためにこれを E^* で置き換えるのが確率伝搬法である．同様の仕組みは平均場近似などでも用いられている．

特殊なのはグラフが木の構造の場合である．木のグラフでは確率伝搬法によって必ず収束し，正しい周辺分布が求まることが分っている（Pearl (1988)）．これより次の結果が得られる．

Proposition 1. $q(\boldsymbol{x})$ が木のグラフで表現できるとき $q(\boldsymbol{x})$, $p_0(\boldsymbol{x};\boldsymbol{\theta}^*)$, $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r^*)$, $r = 1, \ldots, L$ が M^* と E^* に含まれる．

木のグラフの場合には $p_0(\boldsymbol{x};\boldsymbol{\theta}^*) = \prod_i q(x_i)$ であり，確率伝搬法によって真の周辺分布が求まるが，ループのあるグラフの場合には一般には $q(\boldsymbol{x}) \notin M^*$ である．すなわち確率伝搬法の収束点は特殊な場合を除き，真の周辺分布を与えない．

3.3 停留点の安定性

ここでは線形近似に基づき，確率伝搬法の停留点の局所的な安定性を考える．停留点 $\{\boldsymbol{\zeta}_r^*\}$ に摂動 $\Delta\boldsymbol{\zeta}_r$ を加え $\boldsymbol{\zeta}_r = \boldsymbol{\zeta}_r^* + \Delta\boldsymbol{\zeta}_r$ とする．確率伝搬法のステップ(2)の結果を $\boldsymbol{\xi}_r = \boldsymbol{\xi}_r^* + \Delta\boldsymbol{\xi}_r$ とすると $\Delta\boldsymbol{\xi}_r$ は次のように表される．

$$\Delta\boldsymbol{\xi}_r = I_0(\boldsymbol{\theta}^*)^{-1} I_r(\boldsymbol{\zeta}_r^*)\Delta\boldsymbol{\zeta}_r - \Delta\boldsymbol{\zeta}_r = (I_0(\boldsymbol{\theta}^*)^{-1} I_r(\boldsymbol{\zeta}_r^*) - E_n)\Delta\boldsymbol{\zeta}_r \,.$$

ここで $I_0(\boldsymbol{\theta}^*)$ は $p_0(\boldsymbol{x};\boldsymbol{\theta})$ の，$I_r(\boldsymbol{\zeta}_r^*)$ は $p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)$ の Fisher 情報量行列であり E_n は n 次の単位行列である．確率伝搬法の計算を一ステップ行った後のパラメータを $\boldsymbol{\zeta}_r' = \boldsymbol{\zeta}_r^* + \Delta\boldsymbol{\zeta}_r'$ とかくと，

$$\Delta\boldsymbol{\zeta}_r' = \sum_{r' \neq r}^{L} (I_0(\boldsymbol{\theta}^*)^{-1} I_{r'}(\boldsymbol{\zeta}_{r'}^*) - E_n)\Delta\boldsymbol{\zeta}_{r'} \,.$$

この結果から次の条件が成り立つとき，確率伝搬法の停留点が安定であることがわかる．

Theorem 3. (Ikeda et al. (2004a))停留点の周りでの確率伝搬法の性質は次のように近似できる．

$$\begin{pmatrix} \Delta\boldsymbol{\zeta}_1' \\ \vdots \\ \Delta\boldsymbol{\zeta}_L' \end{pmatrix} = T \begin{pmatrix} \Delta\boldsymbol{\zeta}_1 \\ \vdots \\ \Delta\boldsymbol{\zeta}_L \end{pmatrix} ,$$

ここで

$$T = \begin{pmatrix} O & I_0^{-1}I_2 - E_n & \cdots & I_0^{-1}I_L - E_n \\ I_0^{-1}I_1 - E_n & O & & \vdots \\ \vdots & & \ddots & \vdots \\ I_0^{-1}I_1 - E_n & \cdots & \cdots & O \end{pmatrix} ,$$

である．ただし $I_0 = I_0(\boldsymbol{\theta}^*)$, $I_r = I_r(\boldsymbol{\zeta}_r^*)$ とした．T の全ての固有値 λ_i, $i = 1, \ldots, nL$ が $|\lambda_i| < 1$ を満たすとき，停留点は安定である．

上の結果は全てのパラメータを同時に更新する場合の停留点の条件である．実際にはパラメータを一つずつ更新する場合やランダムに更新するなど様々な方法があり，それぞれで安定性は多少異なる．いずれの場合も同様な解析が可能である．

3.4 解の近似精度

本節では摂動展開に基づく近似精度の評価，すなわち確率伝搬法で求まった結果と真の周辺分布との差について述べる．2 次の摂動展開の結果については Ikeda et al.（2004a）にあるが，まずその結果を一般のグラフィカルモデルに対して導き，その後ボルツマンマシンに対する 3 次の摂動展開の結果について述べる．

以下では (2.2) 式の $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v})$ を用いて議論する．定義より明らかに

$$p_0(\boldsymbol{x};\boldsymbol{\theta})=p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{0})\,,\quad p_r(\boldsymbol{x};\boldsymbol{\zeta}_r)=p(\boldsymbol{x};\boldsymbol{\zeta}_r,\boldsymbol{e}_r)\,,\quad q(\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{0},\boldsymbol{1}_L)$$

である．ただし

$$\boldsymbol{e}_r=(0,\cdots,0,\underset{\substack{\uparrow\\ r}}{1},0,\cdots,0)^T\,,\quad \boldsymbol{1}_L=(\underbrace{1,\cdots,1}_{L})^T=\sum_{r=1}^{L}\boldsymbol{e}_r\,.$$

さらに $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v})$ の期待値パラメータを次のように定義する

$$\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{v})=\partial_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta},\boldsymbol{v})=\sum_{\boldsymbol{x}}\boldsymbol{x}p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v})\,.$$

近似精度の解析のため $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v})$ を M^* に拘束する．すなわち

$$\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{v})=\boldsymbol{\eta}(\boldsymbol{\theta}^*,\boldsymbol{0})=\boldsymbol{\eta}(\boldsymbol{\theta}^*)$$

が常に満たされるようにする．このとき $\boldsymbol{\theta}$ は $\boldsymbol{v}$ の関数である．必要があれば $\boldsymbol{\theta}(\boldsymbol{v})$ とかく．以下では $\boldsymbol{v}$ が $\boldsymbol{0}$ から $\boldsymbol{1}_L$ まで変化するにつれて $\boldsymbol{\theta}$ が $\boldsymbol{\theta}^*$ からどのように変化するかを評価する．まず $\boldsymbol{\theta}(\boldsymbol{v})$ の 2 次までの摂動展開の結果を示す．M^* に拘束されていることから

$$\boldsymbol{0}=\frac{d}{d\boldsymbol{v}}\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{v})=\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}}+\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{v}}\,.\tag{3.6}$$

$\partial\boldsymbol{\eta}/\partial\boldsymbol{\theta}$ および $\partial\boldsymbol{\eta}/\partial\boldsymbol{v}$ は $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{v})$ の Fisher 情報量行列の一部である．これ以降 i,j,k は $\boldsymbol{\theta}$ の r,s,t は $\boldsymbol{v}$ の成分を表すことにする．$G_{\boldsymbol{\theta\theta}}=(\partial\boldsymbol{\eta}/\partial\boldsymbol{\theta})$, $G_{\boldsymbol{\theta v}}=(\partial\boldsymbol{\eta}/\partial\boldsymbol{v})$ と定める．ここで任意の $\boldsymbol{v}$ と $\boldsymbol{\theta}=\boldsymbol{\theta}(\boldsymbol{v})$ に対して $G_{\boldsymbol{\theta\theta}}=I_0(\boldsymbol{\theta})$ であることに注意する．(3.6) 式から

$$\boldsymbol{0}=I_0(\boldsymbol{\theta})\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}}+G_{\boldsymbol{\theta v}}(\boldsymbol{\theta})\,,\quad \frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}}=-I_0^{-1}(\boldsymbol{\theta})G_{\boldsymbol{\theta v}}(\boldsymbol{\theta})\,.\tag{3.7}$$

これにより $\boldsymbol{\theta}(\boldsymbol{v})$ の $\boldsymbol{v}$ に関する 1 次の微分 $\partial\boldsymbol{\theta}/\partial\boldsymbol{v}$ が得られた．また $\tilde{G}_{\boldsymbol{\theta v}}$ を $-\partial\boldsymbol{\theta}/\partial\boldsymbol{v}$ と定義する．2 次の微分係数は

$$\frac{d^2}{d\boldsymbol{v}d\boldsymbol{v}}\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{v})=0\,,$$

より

$$I_0(\boldsymbol{\theta})\frac{\partial^2\boldsymbol{\theta}}{\partial\boldsymbol{v}\partial\boldsymbol{v}'}=-T_{\boldsymbol{\theta vv}'}-T_{\boldsymbol{\theta\theta}}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}'}-T_{\boldsymbol{\theta\theta v}}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}'}-T_{\boldsymbol{\theta\theta v}'}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}}\tag{3.8}$$

となる．ここで

$$T_{\boldsymbol{\theta\theta\theta}}=\frac{\partial^3\psi}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}}\,,\quad T_{\boldsymbol{\theta\theta v}}=\frac{\partial^3\psi}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}\partial\boldsymbol{v}}\,,\quad T_{\boldsymbol{\theta vv}'}=\frac{\partial^3\psi}{\partial\boldsymbol{\theta}\partial\boldsymbol{v}\partial\boldsymbol{v}'}$$

である．$\partial\boldsymbol{\theta}/\partial\boldsymbol{v} = -\tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta})$ および $\partial^2\boldsymbol{\theta}/\partial\boldsymbol{v}^2$ を $(\boldsymbol{\theta}, \boldsymbol{v}) = (\boldsymbol{\theta}^*, \mathbf{0})$ で評価し，$\boldsymbol{\theta}(\boldsymbol{v})$ をまずは 2 次のテーラー展開により $\boldsymbol{v}$ で近似する．$(\boldsymbol{\theta}, \boldsymbol{v}) = (\boldsymbol{\theta}^*, \mathbf{0})$ におけるオペレータ $d/d\boldsymbol{v}$ を以下のように定義すれば

$$\frac{d}{d\boldsymbol{v}} = B = \frac{\partial}{\partial \boldsymbol{v}} - \tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta}^*)\frac{\partial}{\partial \boldsymbol{\theta}}\,.$$

$\partial^2\boldsymbol{\theta}/\partial\boldsymbol{v}\partial\boldsymbol{v}$ は

$$\left.\frac{\partial^2\boldsymbol{\theta}}{\partial\boldsymbol{v}\partial\boldsymbol{v}}\right|_{\boldsymbol{v}=\mathbf{0}} = -I_0(\boldsymbol{\theta}^*)^{-1}B^2\boldsymbol{\eta}(\boldsymbol{\theta}^*)\,,$$

となる．簡単のため B^2 の (r, s) 成分を $B_{rs} = B_rB_s$ とかくことにする．なお $(d^2/dvdv)\boldsymbol{\eta}(\boldsymbol{\theta}^*) = \mathbf{0}$ であっても一般に $B^2\boldsymbol{\eta}(\boldsymbol{\theta}^*) \neq \mathbf{0}$ である．

$\boldsymbol{\theta}(\boldsymbol{v})$ を $(\boldsymbol{\theta}^*, \mathbf{0})$ の周りで $\boldsymbol{v}$ の 2 次までで近似した結果は以下の通りである．

$$\begin{aligned}\boldsymbol{\theta}(\boldsymbol{v}) &= \boldsymbol{\theta}^* + \left.\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{v}}\right|_{\boldsymbol{v}=\mathbf{0}}\boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^T\left.\frac{\partial^2\boldsymbol{\theta}}{\partial\boldsymbol{v}\partial\boldsymbol{v}}\right|_{\boldsymbol{v}=\mathbf{0}}\boldsymbol{v} + o(||\boldsymbol{v}||^3) \\ &\simeq \boldsymbol{\theta}^* - \tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta}^*)\boldsymbol{v} - \frac{1}{2}\boldsymbol{v}^T I_0^{-1}(\boldsymbol{\theta}^*)(B^2\boldsymbol{\eta}(\boldsymbol{\theta}^*))\boldsymbol{v}\,.\end{aligned} \tag{3.9}$$

ここまで m-条件を考えていたが，e-条件を考慮する必要がある．e- 条件は

$$\boldsymbol{\theta}^* = -\sum_{r=1}^{L}(\boldsymbol{\zeta}_r^* - \boldsymbol{\theta}^*)\,, \tag{3.10}$$

とかける．次の分布を考える

$$p(\boldsymbol{x}; \boldsymbol{\zeta}_r, \epsilon\boldsymbol{e}_r) = \exp[\boldsymbol{h}\cdot\boldsymbol{x} + \boldsymbol{\zeta}_r\cdot\boldsymbol{x} + \epsilon c_r(\boldsymbol{x}) - \psi(\boldsymbol{\zeta}_r, \epsilon\boldsymbol{e}_r)]\,. \tag{3.11}$$

ここで $p(\boldsymbol{x}; \boldsymbol{\zeta}_r, \epsilon\boldsymbol{e}_r)\big|_{\epsilon=1} = p_r(\boldsymbol{x}; \boldsymbol{\zeta}_r)$ である．$p(\boldsymbol{x}; \boldsymbol{\zeta}_r, \epsilon\boldsymbol{e}_r)$, $r = 1, \ldots, L$ が M^* に含まれるように拘束すると (3.9) 式の結果から $\boldsymbol{\zeta}_r - \boldsymbol{\theta}^*$ は ϵ によって以下のように近似できる．

$$\boldsymbol{\zeta}_r - \boldsymbol{\theta}^* \simeq -\tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta}^*)\boldsymbol{e}_r\epsilon - \frac{1}{2}I_0^{-1}(\boldsymbol{\theta}^*)B_{rr}\boldsymbol{\eta}(\boldsymbol{\theta}^*)\epsilon^2\,.$$

以下の議論では θ に対する $c_r(\boldsymbol{x})$ の影響は小さいと仮定する．この仮定の下では ϵ がある程度大きな値を取っても近似が成り立つことから $\epsilon \to 1$ とした点を考える．この意味での近似を $\approx$ によって表すこととする．$\epsilon \to 1$ ととれば $\boldsymbol{\zeta}_r - \boldsymbol{\theta}^*$ は $\boldsymbol{\zeta}_r^* - \boldsymbol{\theta}^*$ となり

$$\boldsymbol{\zeta}_r^* - \boldsymbol{\theta}^* \approx -\tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta}^*)\boldsymbol{e}_r - \frac{1}{2}I_0^{-1}(\boldsymbol{\theta}^*)B_{rr}\boldsymbol{\eta}(\boldsymbol{\theta}^*)\,.$$

(3.10) 式から $\boldsymbol{\theta}^*$ は次式を満たす．

$$\boldsymbol{\theta}^* = -\sum_{r=1}^{L}(\boldsymbol{\zeta}_r^* - \boldsymbol{\theta}^*) \approx \tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta}^*)\mathbf{1}_L + \frac{1}{2}I_0^{-1}(\boldsymbol{\theta}^*)\sum_r B_{rr}\boldsymbol{\eta}(\boldsymbol{\theta}^*)\,. \tag{3.12}$$

ここで次の分布を考える．

$$p(\boldsymbol{x}; \boldsymbol{u}, \epsilon\mathbf{1}_L) = \exp[\boldsymbol{h}\cdot\boldsymbol{x} + \boldsymbol{u}\cdot\boldsymbol{x} + \epsilon\mathbf{1}_L\cdot\mathbf{c}(\boldsymbol{x}) - \psi(\boldsymbol{u}, \epsilon\mathbf{1}_L)]\,. \tag{3.13}$$

この分布を M^* に拘束し ϵ を 0 から 1 まで増やすと $\boldsymbol{u}$ は $\boldsymbol{\theta}(\mathbf{1}_L)$ となるとする．一般に $\boldsymbol{\theta}(\mathbf{1}_L) \neq \mathbf{0}$ であることから $p(x; \boldsymbol{\theta}(\mathbf{1}_L), \mathbf{1}_L)$ は $q(\boldsymbol{x})$ と等しくない．(3.9) 式の結果から

$$\boldsymbol{\theta}(\mathbf{1}_L) - \boldsymbol{\theta}^* \approx -\tilde{G}_{\boldsymbol{\theta v}}(\boldsymbol{\theta}^*)\mathbf{1}_L - \frac{1}{2}I_0^{-1}(\boldsymbol{\theta}^*)\sum_{r,s}B_{rs}\boldsymbol{\eta}(\boldsymbol{\theta}^*)\,.$$

(3.12) 式から

$$\boldsymbol{\theta}(\mathbf{1}_L) \approx -\frac{1}{2} I_0^{-1}(\boldsymbol{\theta}^*) \sum_{r \neq s} B_{rs} \boldsymbol{\eta}(\boldsymbol{\theta}^*)\,. \tag{3.14}$$

期待値パラメータはテーラー展開に基づき

$$\boldsymbol{\eta}(\mathbf{0}, \mathbf{1}_L) \simeq \boldsymbol{\eta}(\boldsymbol{\theta}(\mathbf{1}_L), \mathbf{1}_L) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\eta}(\boldsymbol{\theta}^*) \boldsymbol{\theta}(\mathbf{1}_L) \approx \boldsymbol{\eta}(\boldsymbol{\theta}^*) + \frac{1}{2} \sum_{r \neq s} B_{rs} \boldsymbol{\eta}(\boldsymbol{\theta}^*) \tag{3.15}$$

と近似できる．$\eta(\mathbf{0}, \mathbf{1}_L)$ は $q(\boldsymbol{x})$ による x の期待値である．したがって (3.15) 式は $q(\boldsymbol{x})$ による x の期待値と確率伝搬法による結果との差を示している．

Theorem 4. 真の条件付確率 $q(\boldsymbol{x})$ による x の期待値を $\boldsymbol{\eta}_{MPM} = \boldsymbol{\eta}(\mathbf{0}, \mathbf{1}_L)$ とし，確率伝搬法によって求まった期待値を $\eta(\theta^*)$ とする．2 次までの摂動展開の結果 $\boldsymbol{\eta}_{MPM}$ と $\eta(\theta^*)$ の差は以下のように近似できる．

$$\boldsymbol{\eta}_{MPM} \approx \boldsymbol{\eta}(\boldsymbol{\theta}^*) + \frac{1}{2} \sum_{r \neq s} B_{rs} \boldsymbol{\eta}(\boldsymbol{\theta}^*)\,. \tag{3.16}$$

この定理に示された近似誤差は M^* の e-曲率と関係している（Ikeda et al.（2004a））．グラフィカルモデルが与えられればこの項を具体的に計算できる．具体的な形はモデルによって異なるが，$c_r(\boldsymbol{x})$ が $\{x_i\}$ の単項式の場合には次の定理が得られている (Tanaka et al.（2002））．

Theorem 5. $c_r(\boldsymbol{x})$, $c_{r'}(\boldsymbol{x})$, $r \neq r'$ が共通な x_i を 2 つ以上含まないとき，摂動展開による 2 次の誤差項 $\sum_{r \neq s} B_{rs} \boldsymbol{\eta}(\boldsymbol{\theta}^*)/2$ は 0 となる．

ボルツマンマシンでは $c_r(\boldsymbol{x}) = w_{ij} x_i x_j$ となるため，$c_r(\boldsymbol{x})$, $c_{r'}(\boldsymbol{x})$, $r \neq r'$ は共通の x_i を持っていたとしても一つのみである．したがってこの近似誤差項は 0 となることがわかる．

ボルツマンマシンの誤差項を評価するためには少なくとも 3 次の摂動展開を行う必要がある．ここでは最近得た結果を示す．3 次までの近似は

$$\boldsymbol{\theta}(\boldsymbol{v}) \simeq \boldsymbol{\theta}^* + \left.\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\mathbf{0}} \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^T \left.\frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{v} \partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\mathbf{0}} \boldsymbol{v} + \frac{1}{6} \sum_{r,s,t} \left.\frac{\partial^3 \boldsymbol{\theta}}{\partial v_r \partial v_s \partial v_t}\right|_{\boldsymbol{v}=\mathbf{0}} v_r v_s v_t \tag{3.17}$$

から求まるわけだが，(3.14) 式の誤差項については 2 次までの項が 0 となることから

$$\boldsymbol{\theta}(\mathbf{1}_L) \simeq \frac{1}{6} \left[\sum_{r,s,t} \left.\frac{\partial^3 \boldsymbol{\theta}}{\partial v_r \partial v_s \partial v_t}\right|_{\boldsymbol{v}=\mathbf{0}} - \sum_r \left.\frac{\partial^3 \boldsymbol{\theta}}{\partial v_r^3}\right|_{\boldsymbol{v}=\mathbf{0}} \right]$$

となる．$\partial^3 \boldsymbol{\theta} / \partial v_r \partial v_s \partial v_t$ は

$$\frac{d^3}{dv dv dv} \boldsymbol{\eta}(\boldsymbol{\theta}, \boldsymbol{v}) = 0\,,$$

から多少の計算をすれば求まる．ボルツマンマシンに対して具体的に計算をすると，次の結果を得る．

$$\theta_i(\mathbf{1}_L) \approx \eta_i^* \sum_{j \neq k} w_{ij} w_{jk} w_{ki} (1 - \eta_j^{*2})(1 - \eta_k^{*2})\,.$$

これは一番短いループのみが $\theta_i(\mathbf{1}_L)$ に寄与することを示している（図 2）．

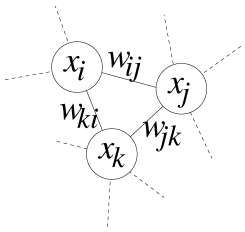

図 2. 誤差項へ寄与する短いループ.

Theorem 6. 3 次までの摂動展開により，ボルツマンマシンに対する η_{MPM} と $\eta(\theta^*)$ の差は以下のように近似できる.

$$\eta_{MPM,i} \approx \eta_i^* - \eta_i^*(1-\eta_i^{*2})\sum_{j\neq k} w_{ij}w_{jk}w_{ki}(1-\eta_j^{*2})(1-\eta_k^{*2})\,. \tag{3.18}$$

現在は 3 次までの項の結果しかないが，さらに高次の摂動展開はより長いループを反映した項が現われるものと考えられる.

(3.16), (3.18) 式の結果を用いると MPM 解の精度が向上できる期待がある．しかし，例えばボルツマンマシンの場合に計算してみると，必ずしも精度が向上するとも限らない．原因としてはいくつか考えられる．摂動展開の仮定としている θ への $c_r(x)$ の影響が小さいとするが，w_{ij} がある程度大きい場合にはこの近似が成りたたない，また，ボルツマンマシンなどではより高次の項の影響があり，3 次の摂動展開の項のみを考えることはあまり意味がない，などである．精度は確率伝搬法の一つの重要な問題である．今後，どうやって 2 次，3 次，あるいはより高次の摂動展開と精度向上を結びつけるかが課題の一つである.

4. まとめ

確率伝搬法は統計物理学や誤り訂正符号など，応用上重要な手法である．本稿では確率伝搬法を理解するための情報幾何学的な枠組みについて述べた.

一般に確率伝搬法で重要な問題は収束性と近似精度である．我々は提案した枠組みに基づき，局所的な安定性の条件を示し，近似誤差についても摂動展開から主要項を示した．近似誤差については 3 次までの漸近展開の結果では最も短いループが主要項を構成しているが，今後より高次の展開についても調べる必要がある.

確率伝搬法は様々な分野で独自の研究がなされており，関連するアルゴリズムや近似についても広く論じられている．情報幾何学による枠組みではそれらを全て表現し，等しく扱うことが可能である．我々は CCCP, TRP, GBP といった関連するアルゴリズムがこの枠組みで表現できることを示した(Ikeda et al. (2004b))．今後は各分野で得られた結果を理解し，新たなアルゴリズムの提案に結び付けたいと考えている.

謝　辞

査読者には有益な御指摘を頂きました．感謝致します.

参 考 文 献

甘利俊一，長岡浩司(1993)．**『情報幾何の方法』，岩波講座 応用数学［対象 12］，岩波書店，東京．**

Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*, AMS and Oxford University Press, Providence, Rhode Island.

Berrou, C., Glavieux, A. and Thitimajshima, P. (1993). Near Shannon limit error-correcting coding and decoding: Turbo-codes, *Proceedings of IEEE International Conference on Communications*, Geneva, Switzerland.

Gallager, R. G. (1962). Low density parity check codes, *IRE Transactions on Information Theory*, **IT-8**, 21–28.

池田思朗，田中利幸，甘利俊一(2002)．**ターボ復号の情報幾何，電子情報通信学会論文誌，J85-D-II**(5), 758–765.

Ikeda, S., Tanaka, T. and Amari, S. (2002). Information geometrical framework for analyzing belief propagation decoder, *Advances in Neural Information Processing Systems*, Vol. 14 (eds. T. G. Dietterich, S. Becker and Z. Ghahramani), 407–414, MIT Press, Cambridge, Massachusetts.

Ikeda, S., Tanaka, T. and Amari, S. (2003). Stochastic reasoning, free energy and information geometry, Research Memorandum, No. 890, The Institute of Statistical Mathematics, Tokyo.

Ikeda, S., Tanaka, T. and Amari, S. (2004a). Information geometry of turbo codes and low-density parity-check codes, *IEEE Transactions on Information Theory*, **50**(6), 1097–1114.

Ikeda, S., Tanaka, T. and Amari, S. (2004b). Stochastic reasoning, free energy and information geometry, *Neural Computation*, **16**(9), 1779–1810.

Jordan, M. I. (1999). *Learning in Graphical Models*, MIT Press, Cambridge, Massachusetts.

Kabashima, Y. and Saad, D. (1998). Belief propagation vs. TAP for decoding corrupted messages, *Europhysics Letters*, **44**(5), 668–674.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society B*, **50**, 157–224.

MacKay, D. J. C. (1999). Good error-correcting codes based on very sparse matrices, *IEEE Transactions on Information Theory*, **45**(2), 399–431.

McEliece, R. J., MacKay, D. J. C. and Cheng, J.-F. (1998). Turbo decoding as an instance of Pearl's "belief propagation" algorithm, *IEEE Journal on Selected Areas in Communications*, **16**(2), 140–152.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, California.

Richardson, T. J. (2000). The geometry of turbo-decoding dynamics, *IEEE Transactions on Information Theory*, **46**(1), 9–23.

Tanaka, T., Ikeda, S. and Amari, S. (2002). Information-geometrical significance of sparsity in Gallager codes, *Advances in Neural Information Processing Systems*, Vol. 14 (eds. T. G. Dietterich, S. Becker and Z. Ghahramani), 527–534, MIT Press, Cambridge, Massachusetts.

Weiss, Y. (2000). Correctness of local probability propagation in graphical models with loops, *Neural Computation*, **12**(1), 1–41.

Yedidia, J. S., Freeman, W. T. and Weiss, Y. (2001). Bethe free energy, Kikuchi approximation, and belief propagation algorithms, Technical Report 2001-16, Mitsubishi Electric Research Laboratories, Cambridge, Massachusetts.

Proceedings of the Institute of Statistical Mathematics Vol. 52, No. 2, 393–405 (2004) 405

Information Geometrical Framework to Analyze Belief Propagation Algorithm

Shiro Ikeda[1], Toshiyuki Tanaka[2] and Shun-ichi Amari[3]

[1]The Institute of Statistical Mathematics
[2]Department of Electronics and Information Engineering, Tokyo Metropolitan University
[3]Brain Science Institute, RIKEN

Belief propagation (BP) is a universal method of stochastic reasoning. It gives exact inference for stochastic models with tree interactions, and works well even if the models have loopy interactions. Its performance has been analyzed separately in many fields, such as, AI, statistical physics, information theory, and information geometry. The present paper provides a unified framework for understanding BP. The stability of BP is analyzed from this framework, and its approximation accuracy is investigated.

Key words: Belief propagation, information geometry, graphical model.