Mathematical structures of loopy belief propagation and cluster variation method

Kazuyuki Tanaka
Graduate School of Information Sciences, Tohoku University, 6-3-09 Aramaki-aza-aoba, Aoba-ku, Sendai 980-8579, Japan
E-mail: kazu@mapip.is.tohoku.ac.jp

Abstract. The mathematical structures of loopy belief propagation are reviewed for graphical models in probabilistic information processing in the standpoint of cluster variation method. An extension of adaptive TAP approaches is given by introducing a generalized scheme of the cluster variation method. Moreover the practical message update rules in loopy belief propagation are summarized also for quantum systems. It is suggested that the loopy belief propagation can be reformulated for quantum electron systems by using density matrices of ideal quantum lattice gas system.

1. Introduction
Advanced mean field methods are powerful applications in probabilistic information processing[1]. Many computer scientists, statistical scientists as well as physicists are interested in the mathematical structures of mean field theory. As one of advanced mean field methods, we have loopy belief propagations (LBP)[2, 3, 4] They have been applied to many problems in computer sciences. One of the successful applications is to probabilistic image processing [5, 6, 7]. Others is to construct algorithms for error correcting codes and other communication technologies[8, 9, 10]. In probabilistic inferences, LBP has been applied also to probabilistic inferences[11, 12, 13].

Recently, it is suggested that LBP can be derived from the cluster variation method (CVM)[4]. The CVM is one of statistical mechanical methods and is an extension of mean field theory[14, 15, 16, 17, 18, 19]. Cycles corrections in loop belief propagations have been discussed[20, 21, 22, 23, 24]. The convergence of algorithms in LBP have been investigated[25]. Some authors have discussed the accuracy of LBP in statistical inferences based on Gaussian graphical models by comparing results obtained through LBP with those based on exact calculation[26, 27, 28]. Averages, variance and covariances in Gaussian graphical model can be calculated by using multi-dimensional Gaussian integral formulas.

Another approach of advanced mean field methods is an adaptive TAP (Thouless-Anderson-Palmer) method[29, 30, 31]. The method is formulated by using the calculations of some statistical quantities in Gaussian graphical models and is applicable to the computations in nontrivial graphical models. Some authors have applied it to some practical problems in computer sciences[32, 33].

Recently, it is expected that the success of advanced mean-field approaches to probabilistic information processing is extended to quantum statistical mechanical approaches[35, 36, 37] The
CVM which can derive some algorithms of LBP has been formulated also in the case of quantum systems[38, 39, 40]. In the present paper, the mathematical structure of LBP and an extension of the adaptive TAP approaches is given. Moreover we give the practical message update rules in LBP. The practical message update rules in loopy belief propagation are summarized also for quantum systems. In section 2, we give brief reviews of conventional LBP. In section 3, we propose an extension of the loopy belief propagation by using CVM for Gaussian graphical models. In section 4, a new interpretation of the adaptive TAP approaches by means of the framework in section 3. In section 5, we derive message passing formulas of the conventional LBP for quantum systems in statistical-mechanical informatics. Section 6 is concluding remarks and we mention the relationship between conventional mean-field approaches to quantum electron systems and LBP. Our suggestions in section 6 is based on Refs.[39, 40].

2. Loopy belief propagation

In this section, we survey the conventional LBP. The message update rules are derived by using the scheme of CVM. Messages of LBP corresponds to effective fields in the statistical mechanics.

In order to explain the framework of LBP, we should define some notations for hyperedges. Hyperedge is a set of nodes. When a node \( i \) belongs to a hyperedge \( \gamma \), we call \( i \) an element of \( \gamma \) and we express it in terms of the notation \( i \in \gamma \). When the node \( i \) belongs to a hyperedge \( \gamma \), \( i \) can be regarded as a proper subset of \( \gamma \) and use the notation \( i < \gamma \). When all the node \( i \) in the hyperedge \( \gamma' \) belong to a hyperedge \( \gamma \), \( \gamma' \) can be regarded as a subset of \( \gamma \) and use the notation \( \gamma' \subseteq \gamma \).

We denote the set of all the nodes by \( V \). First of all, we have to specify a set of hyperedges, \( E \). Every hyperedges must not be a subset of another element in the set of hyperedges, \( E \). Every common set of any two or more hyperedges is only a node. We denote the set of hyperedges by \( E \). We consider such a set \( V_c \) of clusters that a hyperedge or a node is in \( C \) if and only if it is the common node of two or more hyperedges in \( E \).

A random variable \( x_i \) is associated with every node \( i \in V \). Every random variable \( x_i \) takes \(+1\) and \(-1\). A random variable vector defined for the set of all the nodes \( V \) is denoted by \( \vec{x}_V \equiv (x_1, x_2, \cdots, x_{|V|})^T \). Suppose that the nodes \( \gamma_1, \gamma_2, \cdots, \gamma_{|\gamma|} \) belonging to the hyperedge \( \gamma \) are ordered so that \( \gamma_1 < \gamma_2 < \cdots < \gamma_{|\gamma|} \).

Let the \( |\gamma| \)-dimensional vectors \( \vec{x}_\gamma \) be defined by \( \vec{x}_\gamma \equiv (x_{\gamma_1}, x_{\gamma_2}, \cdots, x_{|\gamma|})^T \). We consider the following probability distribution for random variable vector \( \vec{x}_V \):

\[
P_V(\vec{x}_V) \equiv \frac{\prod_{\gamma \in E} f_{\gamma}(\vec{x}_\gamma)}{\sum_{\vec{z}_V} \prod_{\gamma \in E} f_{\gamma}(\vec{z}_\gamma)}. \tag{1}
\]

Here \( \sum_{\vec{z}_V} \equiv \sum_{z_1 = \pm 1} \sum_{z_2 = \pm 1} \cdots \sum_{z_{|V|} = \pm 1} \) is the summation over all the possible configuration for the random variable vector. The purpose of the present paper is to calculate the marginal probabilities

\[
P_i(x_i) \equiv \sum_{\vec{z}_V} \delta_{x_i, z_i} P_V(\vec{z}_V) \quad (i \in V), \quad P_{\gamma}(\vec{x}_\gamma) \equiv \sum_{\vec{z}_V} \delta_{\vec{x}_\gamma, \vec{z}_\gamma} P_V(\vec{z}_V) \quad (\gamma \in E). \tag{2}
\]

Now we consider the Kulback-Leibler divergence \( D[P_V || Q_V] \) defined by

\[
D[P_V || Q_V] \equiv \sum_{\vec{z}_V} Q_V(\vec{z}_V) \ln \left( \frac{Q_V(\vec{z}_V)}{P_V(\vec{z}_V)} \right). \tag{3}
\]
By substituting equation (1) to equation (3), we have

\[ D[P_V||Q_V] = -\sum_{\gamma \in E} \sum_{\vec{z}_\gamma} Q_\gamma(\vec{z}_\gamma) \ln(f_\gamma(\vec{z}_\gamma)) + \sum_{\vec{z}_V} Q_V(\vec{z}_V) \ln Q_V(\vec{z}_V) + \ln\left(\sum_{\vec{z}_V, \gamma \in E} f_\gamma(\vec{z}_\gamma)\right). \]  

(4)

We approximately restrict the trial probability distribution \(Q_V(\vec{z}_V)\) to the following form:

\[ Q_V(x_V) = \left(\prod_{\gamma \in E} Q_\gamma(x_\gamma)\right)^{\prod_{i \in V} Q_i(x_i) - |\partial i| + 1}. \]  

(5)

By using equation (5), equation (4) can be rewritten as

\[ D[P_V||Q_V] = \mathcal{F}[\{Q_\gamma|\gamma \in V_c \cup E\}] + \ln\left(\sum_{\vec{z}_V, \gamma \in E} f_\gamma(\vec{z}_\gamma)\right), \]  

(6)

where

\[ \mathcal{F}[\{Q_\gamma|\gamma \in V_c \cup E\}] \equiv -\sum_{\gamma \in E} \sum_{\vec{z}_\gamma} Q_\gamma(\vec{z}_\gamma) \ln(f_\gamma(\vec{z}_\gamma)) \]

\[-\sum_{i \in V} (|\partial i| - 1) \sum_{\vec{z}_i} Q_i(z_i) \ln Q_V(z_i) + \sum_{\gamma \in E} \sum_{\vec{z}_\gamma} Q_\gamma(\vec{z}_\gamma) \ln Q_V(\vec{z}_\gamma). \]  

(7)

The marginal probability distributions \(Q_i(x_i)\) and \(Q_\gamma(\vec{z}_\gamma)\) satisfy the following consistencies:

\[ \sum_{\vec{z}_\gamma} z_i Q_i(z_i) = \sum_{\vec{z}_\gamma} z_i Q_\gamma(\vec{z}_\gamma) \quad (i \in V_c, \gamma \in \partial i). \]  

(8)

and the normalization conditions:

\[ \sum_{\vec{z}_\gamma} z_i Q_i(z_i) = \sum_{\vec{z}_\gamma} Q_\gamma(\vec{z}_\gamma) = 1 \quad (i \in V_c, \gamma \in E). \]  

(9)

We introduce some Lagrange multipliers \(\lambda_{i,\gamma}, \lambda_i\) and \(\lambda_\gamma\) to ensure the consistency conditions (8) and the normalization conditions (9) as follows:

\[ \mathcal{L}[\{Q_\gamma|\gamma \in V_c \cup E\}] \equiv \mathcal{F}[\{Q_\gamma|\gamma \in V_c \cup E\}] - \sum_{\gamma \in E} \lambda_\gamma \left(\sum_{\vec{z}_\gamma} Q_\gamma(\vec{z}_\gamma) - 1\right) \]

\[ -\sum_{i \in V_c} \lambda_i \left(\sum_{z_i} Q_i(z_i) - 1\right) - \sum_{i \in V_c, \gamma \in \partial i} \lambda_{i,\gamma} \left(\sum_{z_i} z_i Q_i(z_i) - \sum_{\vec{z}_\gamma} z_i Q_\gamma(\vec{z}_\gamma)\right). \]  

(10)

By taking the variational calculations of \(\mathcal{L}[\{Q_\gamma|\gamma \in V_c \cup E\}]\) in equation (10) and by driving the extremum conditions, the marginal probabilities \(\{P_\gamma(\vec{x}_\gamma)|\gamma \in E\}\) and \(\{P_i(x_i)|i \in C \setminus E\}\) are approximately expressed in terms of the Lagrange multipliers as follows:

\[ Q_\gamma(\vec{x}_\gamma) = \frac{f_\gamma(\vec{x}_\gamma) \exp\left(\sum_{\vec{z}_\gamma} \lambda_{i,\gamma} x_i\right)}{\sum_{\vec{z}_\gamma} f_\gamma(\vec{z}_\gamma) \exp\left(\sum_{\vec{z}_\gamma} \lambda_{i,\gamma} z_i\right)} \quad (\gamma \in E), \]  

(11)

\[ Q_i(x_i) = \frac{\exp\left(\lambda_{i,i} x_i\right)}{\sum_{z_i} \exp\left(\lambda_{i,i} z_i\right)} \quad (i \in V_c), \]  

(12)
where $\lambda_{i,i}(x_i)$ are determined so as to satisfy the following equations:

$$
(\partial i| - 1)\lambda_{i,i} = \sum_{\{\gamma|> i, \gamma \in E\}} \lambda_{i,\gamma} \quad (i \in V_c).
$$

(13)

The marginal probability distributions in equations (11) and (12) can be reduced to the representations in LBP by introducing new parameters $h_{\gamma \rightarrow i}$ defined from the Lagrange multipliers $\lambda_{i,\gamma}$ by means of the following linear transformation:

$$
\lambda_{i,\gamma} = \sum_{\{\gamma'|\gamma \in \partial i\}} h_{\gamma \rightarrow i},
$$

(14)

The quantities $\exp(h_{\gamma \rightarrow i},x_i)$ corresponds to the message that hyperedge $\gamma$ sends to node $i$ in LBP. The new parameter $h_{\gamma \rightarrow i}$ is referred to as an effective field in the statistical mechanics.

By using equation (14), the approximate marginal probability distributions $Q_\gamma(\vec{x}_\gamma)$ and $Q_i(x_i)$ are rewritten as follows:

$$
Q_\gamma(\vec{x}_\gamma) = \frac{1}{Z_\gamma} f_\gamma(\vec{x}_\gamma) \exp\left( \sum_{\{j|j \in \gamma\} \{\gamma'|\gamma' \in \partial j\} \gamma} \sum_{j \in j \in j \in j} h_{\gamma' \rightarrow j} x_j \right) \quad (\gamma \in E),
$$

(15)

$$
Q_i(x_i) = \frac{1}{Z_i} \exp\left( \sum_{\{\gamma'|\gamma \in \partial i\}} h_{\gamma' \rightarrow i} x_i \right) \quad (i \in V_c).
$$

(16)

From equation (8), it is valid that

$$
Q_i(x_i) = \sum_{\vec{z}_\gamma} \delta_{z_i,x_i} Q_V(\vec{z}_\gamma).
$$

(17)

By substituting equations (15) and (16) to equation (17), we derive the simultaneous fixed point equations for effective fields as follows:

$$
\exp(h_{\gamma \rightarrow i},x_i) = \frac{Z_i}{Z_\gamma} \sum_{\vec{z}_\gamma} \delta_{z_i,x_i} f_\gamma(\vec{z}_\gamma) \exp\left( \sum_{\{j|j \in \gamma\} \{\gamma'|\gamma' \in \partial j\} \gamma} \sum_{j \in j \in j \in j} h_{\gamma' \rightarrow j} z_j \right) \quad (i \in V_c, \gamma \in \partial i),
$$

(18)

such that

$$
h_{\gamma \rightarrow i} = \sum_{x_i} \sum_{\vec{z}_\gamma} \delta_{z_i,x_i} f_\gamma(\vec{z}_\gamma) \exp\left( \sum_{\{j|j \in \gamma\} \{\gamma'|\gamma' \in \partial j\} \gamma} \sum_{j \in j \in j \in j} h_{\gamma' \rightarrow j} z_j \right) \quad (i \in V_c, \gamma \in \partial i).
$$

(19)

We remark that $h_{\gamma \rightarrow i}$ is referred to as a effective field and $\exp(h_{\gamma \rightarrow i},x_i)$ corresponds to a message from the hyperedge $\gamma$ to the node $i$.

The marginal probability distributions $Q_i(x_i)$ and $Q_\gamma(\vec{x}_\gamma)$ in equation (8) can be expressed in terms of expectation values of products of random variables as follows:

$$
Q_i(x_i) = \frac{1}{2} + \frac{1}{2} \langle x_i \rangle_i x_i \quad (i \in V_c),
$$

(20)

$$
Q_\gamma(\vec{x}_\gamma) = \frac{1}{2\gamma!} + \frac{1}{2\gamma!} \sum_{\gamma' \leq \gamma} \langle \prod_{k \in \gamma'} x_k \rangle_\gamma \prod_{k \in \gamma'} x_k \quad (\gamma \in E),
$$

(21)

where

$$
\langle x_i \rangle_i = \sum_{z_i = \pm 1} z_i Q_i(z_i), \quad \langle \prod_{k \in \gamma'} x_k \rangle_\gamma = \sum_{\vec{z}_\gamma} \prod_{k \in \gamma'} x_k Q_\gamma(\vec{z}_\gamma).
$$

(22)
From equation (21), we have

$$\sum_{\tilde{z}_i} \delta_{x_i,z_i} Q_\gamma(\tilde{z}_\gamma) = \frac{1}{2} + \frac{1}{4} (x_i)_{\gamma} x_i. \quad (23)$$

By comparing equation (23) with equation (20), we see that the consistency condition (8) is equivalent to the following reducibility condition:

$$Q_i(x_i) = \sum_{\tilde{z}_j} \delta_{x_i,z_i} Q_\gamma(\tilde{z}_\gamma) \ (x_i = \pm 1, i \in V, \gamma \in \partial i). \quad (24)$$

After setting $m_i = \langle x_i \rangle_i = \langle x_i \rangle_\gamma$ and $m_{\gamma'} = \langle \prod_{k \in \gamma'} x_k \rangle_{\gamma'}$, we substitute equations (20)-(21) to equation (25) and rewrite the approximate free energy $\mathcal{F}[\{Q_\gamma | \gamma \in V \cup E\}]$ as follows:

$$\mathcal{F}[\{Q_\gamma | \gamma \in V \cup E\}] = -|E| - \sum_{\gamma \in E} \sum_{\gamma' \subseteq \gamma} w_{\gamma'} m_{\gamma'}$$

$$- \sum_{i \in V} (|\partial i| - 1) \sum_{z_i} \left( \frac{1}{2} + \frac{1}{2} m_i z_i \right) \ln \left( \frac{1}{2} + \frac{1}{2} m_i z_i \right)$$

$$+ \sum_{\gamma \in E} \sum_{\tilde{z}_\gamma} \left( \frac{1}{2|\gamma|} + \frac{1}{2|\gamma|} \sum_{\gamma' \subseteq \gamma} m_{\gamma'} \left( \prod_{k \in \gamma'} z_k \right) \right) \ln \left( \frac{1}{2|\gamma|} + \frac{1}{2|\gamma|} \sum_{\gamma' \subseteq \gamma} m_{\gamma'} \left( \prod_{k \in \gamma'} z_k \right) \right), \quad (25)$$

where

$$w_{\gamma'} = \frac{1}{2|\gamma|} \sum_{\tilde{z}_\gamma} \left( \prod_{k \in \gamma'} z_k \right) \ln f_\gamma(\tilde{z}_\gamma) \quad (\gamma' \subseteq \gamma, \gamma \in E). \quad (26)$$

By taking the first deviation with respect to $m_i \ (i \in V)$ and $m_{\gamma'} \ (\gamma' \subseteq \gamma, \gamma \in E)$, we can derive the simultaneous equations. For the case that $E$ is the set of edges, the present framework has been given in Ref.[41]. Moreover, the framework has been extended to higher level approximation methods in Ref.[42].

3. An extension of loopy belief propagation

The conventional LBP has been formulated by means of a set of marginal probability distributions for hyperedges as shown in the previous section. In such formulation, it is hard to calculate some covariances between a pair of nodes which do not belong to a hyperedge. In the present section, we give an extension of LBP in the stand point of CVM. This framework is constructed by combining LBP for graphical models with discrete random variables with the one for Gaussian graphical models.

In the similar way of the conventional LBP, we introduce marginal probability distributions $Q_i(x_i) \equiv \sum_{\tilde{z}_i} \delta_{z_i,x_i} Q_{V}(\tilde{z}_\gamma)$ and $Q_\gamma(\tilde{z}_\gamma) \equiv \sum_{\tilde{z}_\gamma} \delta_{\tilde{z}_\gamma} Q_{V}(\tilde{z}_\gamma)$. Here the random variable $x_i$ at each node $i$ takes $+1$ and $-1$. Moreover we consider marginal probability density functions $\rho_i(\xi_i)$, $\rho_{\gamma}(\tilde{\xi}_\gamma)$ and $\rho_{V}(\tilde{\zeta}_V)$, where the random variable $\xi_i$ at each node $i$ takes any real number in the interval $(-\infty, +\infty)$. These marginal probability density functions have the same average, variance and covariance as the ones of the marginal probability distributions $Q_i(z_i) \ (i \in V)$ and $Q_\gamma(\tilde{z}_\gamma) \ (\gamma = \partial i \cap \partial j \in E)$. Instead of equation (7), we consider the following approximate free energy:

$$\mathcal{F}[\rho_V, \{Q_\gamma, \rho_{\gamma} | \gamma \in V \cup E\}]$$

$$= -\sum_{\gamma \in E} \sum_{\tilde{z}_\gamma} Q_\gamma(\tilde{z}_\gamma) \ln(f_\gamma(\tilde{z}_\gamma)) + \int \rho_{V}(\tilde{\zeta}_V) \ln(\rho_{V}(\tilde{\zeta}_V)) d\tilde{\zeta}_V.$$
The marginal probabilities for hyperedges and nodes are determined so as to minimize it under

\[
+ \sum_{\gamma \in E} \left( \sum_{z_i} Q_{\gamma}(z_i) \ln(Q_{\gamma}(z_i)) - \int \rho_{\gamma}(\vec{z}_i) \ln(\rho_{\gamma}(\vec{z}_i)) d\vec{z}_i \right)
- \sum_{i \in V_c} \left( |\partial i| - 1 \right) \left( \sum_{z_i} Q_i(z_i) \ln(Q_i(z_i)) - \int \rho_i(\vec{z}_i) \ln(\rho_i(\vec{z}_i)) d\vec{z}_i \right).
\]

(27)

The marginal probabilities for hyperedges and nodes are determined so as to satisfy the consistency conditions:

\[
\left\{ \begin{array}{l}
\sum_{z_i} P_i(z_i) = \sum_{\vec{z}_i} P_{\gamma}(\vec{z}_i) = \int \zeta_i \rho_i(\vec{z}_i) d\zeta_i = \int \zeta_i \rho_{\gamma}(\vec{z}_i) d\vec{z}_\gamma = \int \zeta_i \rho_V(\vec{z}_i) d\vec{z}_V \quad (i \in V_c), \\
\int \zeta_i \rho_i(\vec{z}_i) d\zeta_i = \int \zeta_i \rho_{\gamma}(\vec{z}_i) d\vec{z}_\gamma = \int \zeta_i \rho_V(\vec{z}_i) d\vec{z}_V \quad (i \in V \setminus V_c), \\
\int \zeta_i^2 \rho_i(\vec{z}_i) d\zeta_i = \int \zeta_i^2 \rho_{\gamma}(\vec{z}_i) d\vec{z}_\gamma = \int \zeta_i^2 \rho_V(\vec{z}_i) d\vec{z}_V = 1 \quad (i \in V), \\
\sum_{\vec{z}_i} P_i(\vec{z}_i) = \int \zeta_i \rho_{\gamma}(\vec{z}_i) d\vec{z}_\gamma = \int \zeta_i \rho_{\gamma}(\vec{z}_i) d\vec{z}_\gamma = \int \zeta_i \rho_V(\vec{z}_i) d\vec{z}_V \quad (i \in V, j \in V, \gamma = \partial i \cap \partial j \in E), \\
\end{array} \right.
\]

(28)

and the normalization conditions:

\[
\sum_{z_i} P_i(z_i) = \sum_{\vec{z}_i} P_{\gamma}(\vec{z}_i) = \int \rho_i(\vec{z}_i) d\zeta_i = \int \rho_{\gamma}(\vec{z}_i) d\vec{z}_\gamma = \int \rho_V(\vec{z}_i) d\vec{z}_V = 1.
\]

(29)

By taking the variational calculations and by driving the extremum conditions, the marginal probabilities \( Q_{\gamma}(\vec{x}_\gamma) \), \( \rho_{\gamma}(\vec{x}_\gamma) \) \((\gamma \in E), Q_i(x_i), \rho_i(x_i) \quad (i \in V_c) \) and \( \rho_V(\vec{z}_V) \) are approximately expressed in terms of the Lagrange multipliers as follows:

\[
Q_{\gamma}(\vec{x}_\gamma) = \frac{f_{\gamma}(\vec{x}_\gamma) \exp \left( \vec{h}_{\gamma} \vec{x}_\gamma - \frac{1}{2} \vec{x}_\gamma^T D_{\gamma}^T \vec{x}_\gamma \right)}{\sum_{\vec{z}_\gamma} f_{\gamma}(\vec{z}_\gamma) \exp \left( \vec{h}_{\gamma} \vec{z}_\gamma - \frac{1}{2} \vec{z}_\gamma^T D_{\gamma}^T \vec{z}_\gamma \right)} \quad (\gamma \in V_c \cup E),
\]

(30)

\[
\rho_{Vg}(\vec{z}_V) = \frac{\exp \left( \vec{h}_{Vg} \vec{z}_V - \frac{1}{2} \vec{z}_V^T D_{Vg} \vec{z}_V \right)}{\int \exp \left( \vec{h}_{Vg} \vec{z}_V + \frac{1}{2} \vec{z}_V^T D_{Vg} \vec{z}_V \right) d\vec{z}_V},
\]

(31)

\[
\rho_{\gamma g}(\vec{z}_i) = \frac{\exp \left( \vec{h}_{\gamma g} \vec{z}_i - \frac{1}{2} \vec{z}_i^T D_{\gamma g} \vec{z}_i \right)}{\int \exp \left( \vec{h}_{\gamma g} \vec{z}_i + \frac{1}{2} \vec{z}_i^T D_{\gamma g} \vec{z}_i \right) d\vec{z}_i} \quad (\gamma \in V \cup E).
\]

(32)

Here \( h_i, h_{ig}, \vec{h}_{\gamma}, \vec{h}_{\gamma g}, \vec{h}_{Vg}, D_i, D_{ig}, D_{\gamma}, D_{\gamma g} \) and \( D_{Vg} \) are determined so as to satisfy the
The matrices $D_{\gamma}$, $D_{\gamma g}$ and $D_{V g}$ are symmetric matrices.

By using equation (28) and equation (33) and by introducing a new parameter $h_{\gamma - i} = h_i - \vec{h}_\gamma|i|$, we derive the following simultaneous equations for Lagrange multipliers $h_i$, $h_{ig}$, $\vec{h}_\gamma$, $\vec{h}_{\gamma g}$, $\vec{h}_{V g}$, $D_i$, $D_{ig}$, $D_{\gamma}$, $D_{\gamma g}$ and $D_{V g}$ as follows:

\begin{equation}
\begin{aligned}
(m_i - \nu_i) = \sum_{\gamma_i} \sum_{\gamma_j} P_{\gamma_i}(z_i) \ (i \in \mathcal{V}_C),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
m_i = \sum_{\gamma_i} \sum_{\gamma_j} z_i P_{\gamma_j}(z_i) \ (i \in \mathcal{V}_C, \gamma \in \partial i),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{\gamma} | i \rangle = 0 \ (i \in \mathcal{V}, i \gamma E),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{\gamma} | j \rangle = \frac{1}{2} \sum_{\gamma_i} \sum_{\gamma_j} z_i z_j \ln(P_{\gamma}(z_i)) \ (i \in \mathcal{V}, j \in \mathcal{V}, \gamma = \partial i \cap \partial j \in E),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
D_{ig} = (1 - m_i^2)^{-1} \ (i \in \mathcal{V}),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{\gamma g}^{-1} | i \rangle = 1 - m_i^2 \ (\gamma \in \mathcal{E}, i < \gamma),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{\gamma g}^{-1} | j \rangle = \sum_{\gamma_i} \sum_{\gamma_j} z_i z_j P_{\gamma}(z_i) - m_i m_j \ (\gamma \in \mathcal{E}, i < \gamma, j < \gamma),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{V g} | i \rangle = -(|i| - 1) D_{ig} + \sum_{\gamma \in \partial i} \langle i | D_{\gamma g} | i \rangle \ (i \in \mathcal{V}, j \in \mathcal{V}),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{V g} | j \rangle = \langle i | D_{\gamma g} | j \rangle - \langle i | D_{\gamma} | j \rangle \ (i \in \mathcal{V}, j \in \mathcal{V}, \partial i \cap \partial j \in E),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\langle i | D_{V g} | j \rangle = 0 \ (i \in \mathcal{V}, j \in \mathcal{V}, \partial i \cap \partial j = \phi),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\vec{h}_{V g} | i \rangle = \vec{m}_{V g}^T D_{V g} | i \rangle \ (i \in \mathcal{V}),
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\vec{h}_{\gamma g} | i \rangle = \vec{m}_{\gamma g}^T D_{\gamma g} | i \rangle \ (i \in \gamma, \gamma \in \mathcal{E}),
\end{aligned}
\end{equation}
The marginal probability distribution

\[ x \text{ variable} \] is set to

This probability distribution corresponds to the one that

\[ E \] framework given in section 3. Moreover, an extended framework of adaptive TAP are also

4. An interpretation of adaptive TAP approach from cluster variation method

In this section, we give an interpretation of adaptive TAP approaches by means of the

\[ F \] approximate free energy:

\[ -\infty \] in the interval \((\partial i, \partial j = 1)\) and \(ij\) consisting of two

\[ V \] only edges \(ij\) consisting of two nodes and is set to \(f_{ij}(\vec{x}_{ij}) \equiv \exp(\frac{1}{|\partial i|} \theta_i + \frac{1}{|\partial j|} \theta_j + J_{ij} x_i x_j)\) in equation (1).

We introduce marginal probability distributions \(Q_i(x_i) \equiv \sum_{\vec{x}_i} \delta_{x_i} Q_V(\vec{x}_V)\). Here the random

variable \(x_i\) at each node \(i\) takes +1 and −1. Moreover we consider marginal probability density functions \(\rho_i(\xi_i)\) and \(\rho_V(\vec{\xi}_V)\), where the random variable \(\xi_i\) at each node \(i\) takes any real number in the interval \((-\infty, +\infty)\). In the similar way to equation (27), we consider the following approximate free energy:

\[ F[\rho_V, \{Q_i, \rho_i | i \in V\}] \equiv -\sum_{i \in V} \theta_i \int \xi_i \rho_V(\vec{\xi}_V) d\vec{\xi} - \sum_{ij \in E} J_{ij} \int \xi_i \xi_j \rho_V(\vec{\xi}_V) d\vec{\xi} \]

\[ + \int \rho_V(\vec{\xi}_V) \ln(\rho_V(\vec{\xi}_V)) d\vec{\xi}_V + \sum_{i \in V} \left( \sum_{z_i = \pm 1} Q_i(z_i) \ln Q_i(z_i) - \int_{-\infty}^{+\infty} \rho_i(\xi_i) \ln(\rho_i(\xi_i)) d\xi_i \right). \]

The marginal probability distribution \(Q_i(x_i)\) and the marginal probability density functions
\( \rho_i(\xi_i) \) and \( \rho_V(\xi_V) \) should satisfy the following consistencies:

\[
\begin{cases}
\int_{-\infty}^{+\infty} \xi_i \rho_i(\xi_i) d\xi_i = \sum_{\xi_i = \pm 1} z_i Q_i(z_i), \\
\int_{-\infty}^{+\infty} \xi_i^2 \rho_i(\xi_i) d\xi_i = \sum_{\xi_i = \pm 1} z_i^2 Q_i(z_i), \\
\int_{-\infty}^{+\infty} \rho_i(\xi_i) d\xi_i = 1.
\end{cases}
\]  

(54)

\( Q_i(x_i), \rho_i(\xi_i) \) and \( \rho_V(\xi_V) \) are determined so as to minimize the above approximate free energy \( \mathcal{F}[\rho_V, \{Q_i, \rho_i|i\in V\}] \) under the constraint conditions (54).

By introducing Lagrange multipliers

\[
\tilde{h}_{Vg} = \langle 1|\tilde{h}_{Vg}, \langle 2|\tilde{h}_{Vg}, \ldots, \langle |V|\tilde{h}_{Vg} \rangle, \\
D_{Vg} = \begin{pmatrix}
\langle 1|D_{Vg}|1 \rangle & \langle 1|D_{Vg}|2 \rangle & \cdots & \langle 1|D_{Vg}|V \rangle \\
\langle 2|D_{Vg}|1 \rangle & \langle 2|D_{Vg}|2 \rangle & \cdots & \langle 2|D_{Vg}|V \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle |V|D_{Vg}|1 \rangle & \langle |V|D_{Vg}|2 \rangle & \cdots & \langle |V|D_{Vg}|V \rangle
\end{pmatrix}
\]  

(55)

(56)

to ensure the constraint conditions. We remark that all off-diagonal elements of the matrix \( D_{Vg} \) are equal to zero. By taking the first variation of the approximate free energy \( \mathcal{F}[\rho_V, \{Q_i, \rho_i|i\in V\}] \) with respect to the marginal, we can derive the approximate expressions of \( Q_i(x_i), \rho_i(\xi_i) \) and \( \rho_V(\xi_V) \) as follows:

\[
Q_i(x_i) = \frac{\exp(h_i x_i)}{\sum_{\xi_i} \exp(h_i \xi_i)} \quad (i\in V),
\]

(57)

\[
\rho_V(\xi_V) = \frac{\exp((\tilde{h}_{Vg} - \bar{\theta})\xi_V - \frac{1}{2} \xi_V^2 (D_{Vg} - J)\xi_V)}{\int \exp((\tilde{h}_{Vg} - \bar{\theta})\xi_V - \frac{1}{2} \xi_V^2 (D_{Vg} - J)\xi_V) d\xi_V},
\]

(58)

\[
\rho_i(\xi_i) = \frac{\exp((h_i - \tilde{h}_{Vg}|i|)\xi_i - \frac{1}{2} D_{ig}\xi_i^2)}{\int \exp((h_i - \tilde{h}_{Vg}|i|)\xi_i - \frac{1}{2} D_{ig}\xi_i^2) d\xi_i} \quad (i\in V),
\]

(59)

where \( J \) is the \(|V|\times|V|\) matrix whose \((i, j)\)-elements are defined by \( \langle i|J|j \rangle \equiv \begin{cases} J_{ij} & (ij\in E) \\ 0 & (\text{otherwise}) \end{cases} \) for any nodes \( i(\in V) \) and \( j(\in V) \). The effective fields \( h_i, \tilde{h}_{Vg} \) and \( D_{Vg} \) are determined so as to satisfy the consistencies (54). The matrix \( D_{Vg} \) are symmetric matrices. The deterministic equations of \( h_i, \tilde{h}_{Vg} \) and \( D_{Vg} \) are reduced to the following simultaneous equations:

\[
\begin{align*}
\tanh(h_i) &= (h_i - \tilde{h}_{Vg}|i|)D_{ig}^{-1} = (\tilde{h}_{Vg} - \bar{\theta})(D_{Vg} - J)^{-1}|i| \quad (i\in V), \\
1 - \tanh^2(h_i) &= D_{ig}^{-1} = (i\in V), \\
\langle |i|D_{Vg}|j \rangle &= 0 \quad (i\neq j, i\in V, j\in V),
\end{align*}
\]

(60)

equation (60) is equivalent to the deterministic equation in the adaptive TAP approach for the probabilistic model in equation (52). The minimization of the approximate free energy (53) with
respect to the marginal under the constraint conditions (54) is one of an interpretation of the adaptive TAP approach from CVM.

In the standpoint of CVM, we can propose an extension of the adaptive TAP approach by introducing the following approximate free energy:

\[
F[p_V, \{Q_i, \rho_i|i \in V\}, \{Q_{ij}, \rho_{ij}|i, j \in E\}] \equiv - \sum_{i \in V} \theta_i \int \zeta_i \rho_V(\zeta_i)V d\zeta_i - \sum_{ij \in E} J_{ij} \int \zeta_i \zeta_j \rho_V(\zeta_i)V d\zeta_iV + \int \rho_V(\zeta)V \ln(\rho_V(\zeta)V) d\zeta_V + \sum_{ij \in E} \left( \sum_{z_i = \pm 1} \sum_{z_j = \pm 1} Q_{ij}(z_i, z_j) \ln Q_{ij}(z_i, z_j) - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{ij}(\zeta_i, \zeta_j) \ln(\rho_{ij}(\zeta_i, \zeta_j)) d\zeta_i d\zeta_j \right) - \sum_{i \in V} ((\partial i) - 1) \left( \sum_{z_i = \pm 1} Q_i(z_i) \ln Q_i(z_i) - \int_{-\infty}^{+\infty} \rho_i(\zeta_i) \ln(\rho_i(\zeta_i)) d\zeta_i \right).
\]  

(61)

In the present framework, the marginal probability distributions \(Q_i(x_i), Q_{ij}(x_i, x_j)\) and the marginal probability density functions \(\rho_i(\xi_i), \rho_{ij}(\xi_i, \xi_j)\) and \(\rho_V(\xi_V)\) should satisfy the following consistencies:

\[
\begin{align*}
\int_{\zeta_1 = \pm 1} \cdots \int_{\zeta_V = \pm 1} \zeta_i Q_i(z_i) d\zeta_i &= \sum_{z_i = \pm 1} \sum_{z_j = \pm 1} z_i Q_{ij}(z_i, z_j) (i \in V, j \in \partial i), \\
\int_{\zeta_1 = \pm 1} \cdots \int_{\zeta_V = \pm 1} \zeta_i^2 Q_i(z_i) d\zeta_i &= \sum_{z_i = \pm 1} \sum_{z_j = \pm 1} z_i z_j Q_{ij}(z_i, z_j) (i \in V, j \in \partial i), \\
\int_{\zeta_1 = \pm 1} \cdots \int_{\zeta_V = \pm 1} \zeta_i \zeta_j Q_{ij}(z_i, z_j) d\zeta_i d\zeta_j &= \sum_{z_i = \pm 1} \sum_{z_j = \pm 1} z_i z_j Q_{ij}(z_i, z_j) (i \in V, j \in \partial i), \\
\int_{-\infty}^{+\infty} \rho_V(\zeta)V d\zeta_V &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{ij}(\zeta_i, \zeta_j) d\zeta_i d\zeta_j = \int_{-\infty}^{+\infty} \rho_i(\zeta_i) d\zeta_i = 1 (j \in \partial i).
\end{align*}
\]  

(62)

The marginal are determined so as to minimize the above approximate free energy \(F[p_V, \{Q_i, \rho_i|i \in V\}, \{Q_{ij}, \rho_{ij}|i, j \in E\}\) under the constraint conditions (62).

By using the constraint conditions (62), \(F[p_V, \{Q_i, \rho_i|i \in V\}, \{Q_{ij}, \rho_{ij}|i, j \in E\}\) can be rewritten as

\[
F[p_V, \{Q_i, \rho_i|i \in V\}, \{Q_{ij}, \rho_{ij}|i, j \in E\}] \equiv - \sum_{ij \in E} \sum_{z_i = \pm 1} \sum_{z_j = \pm 1} \left( \frac{1}{2} \theta_i z_i + \frac{1}{2} \theta_j z_j + J_{ij} z_i z_j \right) Q_{ij}(z_i, z_j)
\]

\[
+ \int \rho_V(\zeta)V \ln(\rho_V(\zeta)V) d\zeta_V + \sum_{ij \in E} \left( \sum_{z_i = \pm 1} \sum_{z_j = \pm 1} Q_{ij}(z_i, z_j) \ln Q_{ij}(z_i, z_j) - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{ij}(\zeta_i, \zeta_j) \ln(\rho_{ij}(\zeta_i, \zeta_j)) d\zeta_i d\zeta_j \right) - \sum_{i \in V} ((\partial i) - 1) \left( \sum_{z_i = \pm 1} Q_i(z_i) \ln Q_i(z_i) - \int_{-\infty}^{+\infty} \rho_i(\zeta_i) \ln(\rho_i(\zeta_i)) d\zeta_i \right).
\]

(63)

The approximate free energy (63) corresponds to the one given in the case of the probabilistic distribution (52) by equation (27).

5. Quantum belief propagation

In this section, we derive the message passing rule of LBP in quantum graphical model. Quantum graphical models are usually expressed in terms of density matrices on hypergraphs.
We define the vector representations of quantum states and the matrix representation of density matrix and show how to construct the approximate free energy for the density matrix in CVM. The framework in the present section is an extension of the conventional LBP and is based on Ref.[38].

Each random variable \( x_i (i \in V) \) takes two possible states specified by +1 and −1. As two representations which corresponds to the two possible states +1 and −1, we introduce two vectors \( |1\rangle \equiv (1, 0) \) and \( |-1\rangle \equiv (0, 1) \). Moreover we introduce a notation

\[
\langle \bar{x}_V | \equiv \langle x_1, x_2, \cdots, x_{|V|} | \equiv \langle x_1 | \otimes \langle x_2 | \otimes \cdots \otimes \langle x_{|V|} |, \quad |\bar{x}_V \rangle \equiv \langle \bar{x}_V |^T,
\]

so that we have \( |+1,+1\rangle = (1, 0, 0, 0) \), \( |+1,-1\rangle = (0, 1, 0, 0) \), \( |-1,+1\rangle = (0, 0, 1, 0) \) and \( |-1,-1\rangle = (0, 0, 0, 1) \).

We consider the following representation

\[
R(V) \equiv \frac{\exp(-H(V))}{\text{tr} \exp(-H(V))},
\]

where

\[
\begin{align*}
H(V) & \equiv \sum_{\gamma \in E} H(\gamma), \\
\langle \bar{x} | H(\gamma) | \bar{y} \rangle & \equiv \langle \bar{x}_\gamma | E(\gamma) | \bar{y}_\gamma \rangle \left( \prod_{i \in V \setminus \gamma} \delta_{x_i, y_i} \right).
\end{align*}
\]

Though one of the goals in the statistical inference by means of Bayesian networks is to compute the marginal probability at each node or at each hyperedge, the computations correspond to the one of reduced density matrices in the probabilistic information processing by using a density matrix. Reduced density matrices are defined by

\[
\begin{align*}
\langle \bar{x}_\gamma | R_\gamma | \bar{x}_\gamma \rangle & \equiv \sum_{\bar{z}} \sum_{\bar{\zeta}} \langle \bar{z} | R(V) | \bar{\zeta} \rangle \left( \prod_{j \in \gamma} \delta_{z_j, x_j} \delta_{\zeta_j, y_j} \right) \left( \prod_{k \in V \setminus \gamma} \delta_{z_k, c_k} \right) \quad (\gamma \in E), \\
\langle x_i | R_i | y_i \rangle & \equiv \sum_{\bar{z}} \sum_{\bar{\zeta}} \langle \bar{z} | R(V) | \bar{\zeta} \rangle \delta_{x_i, z_i} \delta_{z_i, y_i} \left( \prod_{k \in V \setminus \{i\}} \delta_{z_k, c_k} \right) \quad (i \in V),
\end{align*}
\]

where \( \bar{z} \equiv (z_1, z_2, \cdots, z_{|V|}) \) and \( \sum_{\bar{z}} \equiv \sum_{z_1=\pm 1} \sum_{z_2=\pm 1} \cdots \sum_{z_{|V|}=\pm 1} \). In this paper, the expressions of equation (68) are expressed in terms of the following notations:

\[
\begin{align*}
R(\gamma) & \equiv \text{tr}_\gamma R(V), \\
R(i) & \equiv \text{tr}_i R(V).
\end{align*}
\]

We explain the representation of equation (1) in terms of a density matrix. By using functions \( f_\gamma(\bar{x}_\gamma) \ (\gamma \in E) \) in equation (1), we set the \( 2^{|\gamma|} \times 2^{|\gamma|} \) matrices \( E(\gamma) \) as

\[
\langle \bar{x}_\gamma | E(\gamma) | \bar{y}_\gamma \rangle = -\left( \prod_{i \in \gamma} \delta_{x_i, y_i} \right) \text{ln}(f_\gamma(\bar{x}_\gamma)).
\]

If the energy matrix \( E \) is a diagonal matrix whose all off-diagonal elements \( \langle \bar{x} | E \rangle \bar{y} \) \( (\bar{x} \neq \bar{y}) \) are equal to zero, \( R(V) \) are also diagonal matrix whose diagonal elements \( \langle \bar{x} | R(V) | \bar{x} \rangle \) are given as
as follows:

\[
\begin{align*}
\langle \vec{x}\gamma | R(\gamma) | \vec{y}\gamma \rangle &= \left( \prod_{j \in \gamma} \delta_{x_j, y_j} \right) P_\gamma(\vec{x}\gamma), \\
\langle x_i | R(i) | y_i \rangle &= \delta_{x_i, y_i} P_V(x_i),
\end{align*}
\]

respectively.

We introduce four kinds of $2 \times 2$ matrices:

\[
\begin{align*}
X^{(+1,+1)} &\equiv | +1 \rangle \langle +1 | = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \\
X^{(+1,-1)} &\equiv | +1 \rangle \langle -1 | = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\
X^{(-1,+1)} &\equiv | -1 \rangle \langle +1 | = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\
X^{(-1,-1)} &\equiv | -1 \rangle \langle -1 | = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},
\end{align*}
\]

as well as $2 \times 2$ identity matrix $I$. In terms of these matrices $X^{(xy)}$, $2^{|\gamma|} \times 2^{|\gamma|}$ matrix $X^{(xy)}(\gamma) (i \in \gamma)$, and $2^{|V|} \times 2^{|V|}$ matrix $X^{(xy)}(V) (i \in V)$ are defined as follows:

\[
X_i^{(xy)}(\gamma) \equiv I \otimes I \otimes \cdots \otimes I \otimes X^{(xy)}(\gamma) \otimes I \otimes \cdots \otimes I,
\]

\[
X_i^{(xy)}(V) \equiv I \otimes I \otimes \cdots \otimes I \otimes X^{(xy)}(V) \otimes I \otimes \cdots \otimes I.
\]

Instead of equation (7), we consider the following free energy expressed in terms of the reduced density matrices $Q(i) (i \in V_c)$ and $Q(\gamma) (\gamma \in E)$ of a $2^{|V|} \times 2^{|V|}$ trial density matrix $Q$:

\[
F[\{Q(\gamma) | \gamma \in V_c \cup E\}] \equiv -\sum_{\gamma \in E} \text{tr}[Q(\gamma) E(\gamma)] + \sum_{\gamma \in E} \text{tr}[Q(\gamma) \ln(Q(\gamma))] \\
-\sum_{i \in V_c} (|\partial i| - 1) \text{tr}[Q(i) \ln(R(i))].
\]

The reduced density matrices for hyperedges and nodes are determined so as to minimize it under the consistency conditions:

\[
Q(i) = \text{tr}_\gamma Q(\gamma) \quad (\gamma \in \partial i, i \in V_c),
\]

and the normalizations

\[
\text{tr}[Q(i)] = 1 \quad (i \in V_c), \quad \text{tr}[Q(\gamma)] = 1 \quad (\gamma \in E).
\]

The reduced density matrices $Q(i) (i \in V_c)$ and $Q(\gamma) (\gamma \in E)$ which are determined so as to minimize the free energy $F[\{Q(\gamma) | \gamma \in V_c \cup E\}]$ under the above constraint conditions (79) and (80) are regarded as approximations of reduced density matrices $R(i) \equiv \text{tr}_\gamma R(V)$ and $R(\gamma) \equiv \text{tr}_\gamma R(V)$ of the density matrix $R(V)$. 


In the quantum belief propagation, some reduced density matrices \( \{Q(i)|i\in V_c\} \) and \( \{Q(\gamma)|\gamma\in E\} \) are approximately expressed as follows:

\[
Q(\gamma) = \frac{\exp(\sum_{j\in\gamma} \sum_{x=\pm 1} \sum_{y=\pm 1} \langle x|L(j, \gamma)|y\rangle X_j^{(xy)}(\gamma))}{\text{tr}\left[\exp(\sum_{j\in\gamma} \sum_{x=\pm 1} \sum_{y=\pm 1} \langle x|L(j, \gamma)|y\rangle X_j^{(xy)}(\gamma))\right]} \quad (\gamma\in E),
\]

(81)

\[
Q(i) = \frac{\exp(L(i, i))}{\text{tr}[\exp(L(i, i))]} \quad (i\in V_c),
\]

(82)

where \( L(i, i) \) and \( L(i, \gamma) \) are \( 2\times2 \) matrices and are determined so as to satisfy the following equations:

\[
0 = -(|\partial i| - 1)L(i, i) + \sum_{\{\gamma: \gamma\geq i, \gamma\in E\}} L(i, \gamma) \quad (i\in V_c),
\]

(83)
as well as the reducibility conditions (79).

Now we introduce effective fields defined by the following linear transformations:

\[
L(i, \gamma) = \sum_{\{\gamma': \gamma'\in \partial i \setminus \gamma\}} h_{\gamma\rightarrow \gamma'} \quad (i\in V_c, \gamma\in \partial i).
\]

(84)

By using equation (84), we can rewrite the reduced density matrices \( Q(i) \) and \( Q(\gamma) \) as

\[
Q(\gamma) = \frac{1}{Z_\gamma} \exp\left(-E(\gamma) + \sum_{j\in\gamma} \sum_{\gamma'\in \partial j \setminus \gamma} \sum_{x=\pm 1} \sum_{y=\pm 1} \langle x|h_{\gamma\rightarrow \gamma'}|y\rangle X_j^{(xy)}(\gamma')\right) \quad (\gamma\in E),
\]

(85)

\[
Q(i) = \frac{1}{Z_i} \exp\left(\sum_{\gamma\in \partial i} h_{\gamma\rightarrow -i}\right) \quad (i\in V_c),
\]

(86)

where

\[
Z_\gamma \equiv \text{tr}\left[\exp\left(-E(\gamma) + \sum_{j\in\gamma} \sum_{\gamma'\in \partial j \setminus \gamma} \sum_{x=\pm 1} \sum_{y=\pm 1} \langle x|h_{\gamma\rightarrow \gamma'}|y\rangle X_j^{(xy)}(\gamma')\right)\right] \quad (\gamma\in E),
\]

(87)

\[
Z_i \equiv \text{tr}\left(\sum_{\gamma\in \partial i} h_{\gamma\rightarrow -i}\right) \quad (i\in V_c).
\]

(88)

By substituting equations (85) and (86) to equation (79), we can derive the simultaneous equations to determine the effective fields \( h_{\gamma\rightarrow -i} \) as follows:

\[
h_{\gamma\rightarrow i} = -\sum_{\gamma'\in \partial i \setminus \gamma} h_{\gamma'\rightarrow -i}
\]

\[
+ \ln \left(\text{tr}_\gamma \left[\frac{Z_i}{Z_\gamma} \exp\left(-E(\gamma) + \sum_{j\in\gamma} \sum_{\gamma'\in \partial j \setminus \gamma} \sum_{x'=\pm 1} \sum_{y'=\pm 1} \langle x'|h_{\gamma'\rightarrow -i}|y'\rangle X_j^{(x'y')}(\gamma')\right)\right]\right) \quad (i\in V_c, \gamma\in \partial i),
\]

(89)
Here $h_{\gamma \rightarrow j}$ can be referred to as an effective field in the statistical mechanics. It is known that effective fields correspond to messages in the conventional loopy belief propagation. Equation (89) can be regarded as a message passing rule in the quantum belief propagation.

As one of the examples, we consider the reduced density matrix for the probabilistic inference system given as “Asia”. The energy matrix $H(V)$ can be given as the following $2^8 \times 2^8$ matrix:

$$H(V) = H(13) + H(24) + H(25) + H(346) + H(568) + H(67),$$

(90)

where

$$\begin{align*}
\langle \vec{x} | H(13) | \vec{y} \rangle &= \langle x_1, x_3 | E(13) | y_1, y_3 \rangle \delta_{x_2, y_2} \delta_{x_4, y_4} \delta_{x_5, y_5} \delta_{x_6, y_6} \delta_{x_7, y_7} \delta_{x_8, y_8}, \\
\langle \vec{x} | H(24) | \vec{y} \rangle &= \langle x_2, x_4 | E(24) | y_2, y_4 \rangle \delta_{x_1, y_1} \delta_{x_3, y_3} \delta_{x_5, y_5} \delta_{x_6, y_6} \delta_{x_7, y_7} \delta_{x_8, y_8}, \\
\langle \vec{x} | H(25) | \vec{y} \rangle &= \langle x_2, x_5 | E(25) | y_2, y_5 \rangle \delta_{x_1, y_1} \delta_{x_3, y_3} \delta_{x_4, y_4} \delta_{x_6, y_6} \delta_{x_7, y_7} \delta_{x_8, y_8}, \\
\langle \vec{x} | H(346) | \vec{y} \rangle &= \langle x_3, x_4, x_6 | E(346) | y_3, y_4, y_6 \rangle \delta_{x_1, y_1} \delta_{x_2, y_2} \delta_{x_5, y_5} \delta_{x_7, y_7} \delta_{x_8, y_8}, \\
\langle \vec{x} | H(568) | \vec{y} \rangle &= \langle x_5, x_6, x_8 | E(568) | y_5, y_6, y_8 \rangle \delta_{x_1, y_1} \delta_{x_2, y_2} \delta_{x_3, y_3} \delta_{x_4, y_4} \delta_{x_7, y_7}, \\
\langle \vec{x} | H(67) | \vec{y} \rangle &= \langle x_6, x_7 | E(67) | y_6, y_7 \rangle \delta_{x_1, y_1} \delta_{x_2, y_2} \delta_{x_3, y_3} \delta_{x_4, y_4} \delta_{x_5, y_5} \delta_{x_8, y_8},
\end{align*}$$

(91)

The density matrix $R(V)$ is also $2^8 \times 2^8$ matrix.

By the context of above quantum belief propagation, reduced density matrices are approximately given as follows:

$$\begin{align*}
Q(13) &= \frac{1}{Z_1} \exp(E(13) + I \otimes h_{346 \rightarrow 3}), \\
Q(24) &= \frac{1}{Z_2} \exp(E(24) + h_{25 \rightarrow 2} \otimes I + I \otimes h_{346 \rightarrow 4}), \\
Q(25) &= \frac{1}{Z_3} \exp(E(25) + h_{24 \rightarrow 2} \otimes I + I \otimes h_{568 \rightarrow 5}), \\
Q(346) &= \frac{1}{Z_{346}} \exp(E(346) + h_{13 \rightarrow 3} \otimes I \otimes I + I \otimes h_{24 \rightarrow 4} \otimes I \\
&\quad \quad \quad \quad + I \otimes I \otimes h_{568 \rightarrow 6} + I \otimes I \otimes h_{67 \rightarrow 6}), \\
Q(568) &= \frac{1}{Z_{568}} \exp(E(568) + h_{25 \rightarrow 5} \otimes I \otimes I + I \otimes h_{568 \rightarrow 6} \otimes I + I \otimes h_{67 \rightarrow 6} \otimes I), \\
Q(67) &= \frac{1}{Z_{67}} \exp(E(67) + h_{346 \rightarrow 6} \otimes I + h_{568 \rightarrow 6} \otimes I),
\end{align*}$$

(92)

$$\begin{align*}
Q(2) &= \frac{1}{Z_4} \exp(h_{24 \rightarrow 2} + h_{25 \rightarrow 2}), \\
Q(3) &= \frac{1}{Z_5} \exp(h_{346 \rightarrow 3} + h_{13 \rightarrow 3}), \\
Q(4) &= \frac{1}{Z_6} \exp(h_{346 \rightarrow 4} + h_{24 \rightarrow 4}), \\
Q(5) &= \frac{1}{Z_7} \exp(h_{25 \rightarrow 5} + h_{568 \rightarrow 5}), \\
Q(6) &= \frac{1}{Z_8} \exp(h_{346 \rightarrow 6} + h_{568 \rightarrow 6} + h_{67 \rightarrow 6}),
\end{align*}$$

(93)

where $Z_i$ and $Z_\gamma$ are normalization constants. The above density matrices $Q_i$ and $Q_\gamma$ have the following reducibility conditions:

$$\begin{align*}
Q(2) &= \text{tr}_2 Q(24) = \text{tr}_2 Q(25), \\
Q(3) &= \text{tr}_3 Q(13) = \text{tr}_3 Q(346), \\
Q(4) &= \text{tr}_4 Q(24) = \text{tr}_4 Q(346), \\
Q(5) &= \text{tr}_5 Q(25) = \text{tr}_5 Q(568), \\
Q(6) &= \text{tr}_6 Q(346) = \text{tr}_6 Q(568) = \text{tr}_6 Q(67),
\end{align*}$$

(94)

By substituting equations (92)-(93) to equation (94), we obtain the following recursion formula
to determine the effective fields $h_{\gamma\rightarrow\gamma}$:

$$
\begin{align*}
    h_{24\rightarrow2} &= -h_{25\rightarrow2} + \ln\left(\frac{2}{\mathcal{Z}_{24}} \text{tr}_{\gamma} \exp(E(24) + h_{25\rightarrow2} \otimes I + I \otimes h_{346\rightarrow4})\right), \\
    h_{25\rightarrow2} &= -h_{24\rightarrow2} + \ln\left(\frac{2}{\mathcal{Z}_{24}} \text{tr}_{\gamma} \exp(E(25) + h_{24\rightarrow2} \otimes I + I \otimes h_{568\rightarrow5})\right), \\
    h_{13\rightarrow3} &= -h_{346\rightarrow3} + \ln\left(\frac{2}{\mathcal{Z}_{13}} \text{tr}_{\gamma} \exp(E(13) + I \otimes h_{346\rightarrow3})\right), \\
    h_{346\rightarrow3} &= -h_{13\rightarrow3} + \ln\left(\frac{2}{\mathcal{Z}_{13}} \text{tr}_{\gamma} \exp(E(346) + h_{13\rightarrow3} \otimes I \otimes I + I \otimes h_{24\rightarrow4} \otimes I)\right), \\
    h_{24\rightarrow4} &= -h_{346\rightarrow4} + \ln\left(\frac{2}{\mathcal{Z}_{24}} \text{tr}_{\gamma} \exp(E(24) + h_{25\rightarrow2} \otimes I + I \otimes h_{346\rightarrow4})\right), \\
    h_{346\rightarrow4} &= -h_{24\rightarrow4} \\
        &+ \ln\left(\frac{2}{\mathcal{Z}_{346}} \text{tr}_{\gamma} \exp(E(346) + h_{13\rightarrow3} \otimes I \otimes I \otimes h_{568\rightarrow6} + h_{67\rightarrow6})\right), \\
    h_{25\rightarrow5} &= -h_{568\rightarrow5} + \ln\left(\frac{2}{\mathcal{Z}_{25}} \text{tr}_{\gamma} \exp(E(25) + h_{24\rightarrow2} \otimes I + I \otimes h_{568\rightarrow5})\right), \\
    h_{568\rightarrow5} &= -h_{25\rightarrow5} \\
        &+ \ln\left(\frac{2}{\mathcal{Z}_{568}} \text{tr}_{\gamma} \exp(E(568) + h_{25\rightarrow5} \otimes I \otimes I + I \otimes (h_{568\rightarrow6} + h_{67\rightarrow6}) \otimes I)\right), \\
    h_{346\rightarrow6} &= -h_{568\rightarrow6} - h_{67\rightarrow6} \\
        &+ \ln\left(\frac{2}{\mathcal{Z}_{346}} \text{tr}_{\gamma} \exp(E(346) + h_{13\rightarrow3} \otimes I \otimes I \otimes h_{568\rightarrow6} + h_{67\rightarrow6})\right), \\
    h_{568\rightarrow6} &= -h_{346\rightarrow6} - h_{67\rightarrow6} \\
        &+ \ln\left(\frac{2}{\mathcal{Z}_{568}} \text{tr}_{\gamma} \exp(E(568) + h_{25\rightarrow5} \otimes I \otimes I + I \otimes (h_{568\rightarrow6} + h_{67\rightarrow6}) \otimes I)\right), \\
    h_{67\rightarrow6} &= -h_{346\rightarrow6} - h_{568\rightarrow6} + \ln\left(\frac{2}{\mathcal{Z}_{67}} \text{tr}_{\gamma} \exp(E(67) + h_{346\rightarrow6} + h_{568\rightarrow6}) \otimes I\right).
\end{align*}
$$

(95)

6. Concluding remarks

In the present paper, the fundamental structure of conventional LBP has been reviewed and some new advanced mean field approaches has been proposed. Our proposed approaches in sections 3 and 4 are based on the exact calculations of Gaussian graphical model and are constructed by combining their exact calculation with CVM. They can be applied to some discrete probabilistic models and we have given some interpretations and general extensions in the adaptive TAP approaches.

It is interesting to compare the conventional adaptive TAP approach with our proposed extensions in some numerical experiments for the probability distribution in equation (52). It is left to a separate paper. Cycle corrections in the present method in section 3 may be estimated by using the perturbative method based on Plefka expansion. It is also interesting problem.

Also message propagation rules in LBP for quantum systems have been derived in section 5 and an explicit example for the density matrix of graphical models with eight nodes have been shown. We have to remark that the formal structures of LBP for quantum systems are different from the conventional LBP.

In the quantum systems, we have to consider not only quantum spin systems but also quantum electron systems with Fermion or Boson particles. Moreover, it is known that quantum spin systems including the Heisenberg model can be expressed in terms of the Bose lattice gas. In such electron systems, the tractable models of mean-field approaches are ideal Bose or Fermi gases.

As one of fundamental quantum systems, we have a Bose lattice gas system whose energy matrix is defined by

$$
H(V) = \mu \sum_{i \in V} a_i^\dagger a_i + t \sum_{ij \in E} (a_i^\dagger a_j + a_j^\dagger a_i) + u \sum_{i \in V} a_i^\dagger a_i^\dagger a_i a_i,
$$

(96)
where \( E \) is a square lattice with periodic boundary conditions along \( x \)- and \( y \)-directions. \( a_i^\dagger \) and \( a_i \) are Bose creation and annihilation operators.

We introduce reduced density matrices \( \rho_V, \rho_i \) and \( Q_i \) as follows:

\[
\begin{align*}
\rho_V &= \frac{\exp\left(-\sum_{i,j \in V} (\mu + h_i, V) a_i^\dagger a_i - t \sum_{i,j \in E} (a_i^\dagger a_j + a_j^\dagger a_i)\right)}{\text{tr}\left[\exp\left(-\sum_{i,j \in V} (\mu + h_i, V) a_i^\dagger a_i - t \sum_{i,j \in E} (a_i^\dagger a_j + a_j^\dagger a_i)\right)\right]} \\
\rho_i &= \frac{\exp\left(-h_i, a_i^\dagger a_i\right)}{\text{tr}\left[\exp\left(-h_i, a_i^\dagger a_i\right)\right]} \\
Q_i &= \frac{\exp\left(-h_i, a_i^\dagger a_i - u a_j^\dagger a_i a_i^\dagger a_i\right)}{\text{tr}\left[\exp\left(-h_i, a_i^\dagger a_i - u a_j^\dagger a_i a_i^\dagger a_i\right)\right]}
\end{align*}
\]

Here \( \rho_V \) and \( \rho_i \) are density matrices of the free Bose particles. The energy matrices of density matrices \( \rho_V \) and \( \rho_i \) have quadratic forms of the Bose creation and annihilation operators.

We consider the following approximate free energy for the Bose lattice gas in equation (96), which is defined by

\[
\mathcal{F}[\rho_V, \{Q_i, \rho_i| i \in V\}] = \text{tr}\left[\left(\sum_{i \in V} \mu a_i^\dagger a_i + \sum_{i,j \in E} t (a_i^\dagger a_j + a_j^\dagger a_i)\right) \rho_V\right] + \sum_{i \in V} u \text{tr}\left[a_i^\dagger a_i^\dagger a_i a_i Q_i\right] + \text{tr}\left[\rho_V \ln \rho_V\right] + \sum_{i \in V} \left(\text{tr}\left[Q_i \ln Q_i\right] - \text{tr}\left[\rho_{if} \ln \rho_{if}\right]\right).
\]

In the similar way to the previous section, reduced density matrices \( \rho_V, \rho_i \) and \( Q_i \) are determined so as to minimize \( \mathcal{F}[\rho_V, \{Q_i, \rho_i| i \in V\}] \) under the consistencies

\[
\text{tr}[a_i^\dagger a_i \rho_V] = \text{tr}[a_i^\dagger a_i \rho_i] = \text{tr}[a_i^\dagger a_i Q_i] \quad (i \in V),
\]

and the normalizations

\[
\text{tr}[\rho_V] = \text{tr}[\rho_i] = \text{tr}[Q_i] = 1 \quad (i \in V).
\]

By taking the first deviation of the approximate free energy, reduced density matrices \( \rho_V, \rho_i \) and \( Q_i \) can be derived as the expression in equation (97). \( h_{i,if}, h_{i,i}, h_{i,Vf} \) and \( u_{i,i} \) is determined so as to satisfy the consistencies (99) and the following simultaneous linear equations

\[
h_{i,if} = h_{i,i} + h_{i,Vf} \quad (i \in V).
\]

From the reduced density matrices determined by means of equations (99) and (101), some approximate expectation values of the density matrix \( R(V) \) can be given as follows:

\[
\begin{align*}
\text{tr}[a_i^\dagger a_i R(V)] &\simeq \text{tr}[a_i^\dagger a_i \rho_V] \quad (i \in V), \\
\text{tr}[a_i^\dagger a_j + a_j^\dagger a_i R(V)] &\simeq \text{tr}[a_i^\dagger a_j + a_j^\dagger a_i \rho_V] \quad (i,j \in E), \\
\text{tr}[a_i^\dagger a_i^\dagger a_i a_i R(V)] &\simeq \text{tr}[a_i^\dagger a_i^\dagger a_i a_i Q_i] \quad (i \in V).
\end{align*}
\]

The reduced density matrices \( \rho_V \) and \( \rho_i \) corresponds to the density matrices of the ideal Bose lattice gases with \( V \) nodes and one node, respectively. The expectation value \( \text{tr}[a_i^\dagger a_j + a_j^\dagger a_i \rho_V] \) can be computed by using the discrete Fourier transformations.

This framework was proposed in Ref.[39] and was applied to analyze the low-temperature behaviour of Heisenberg model in Ref.[40]. It corresponds to an extension of spin wave theory[43]. When quantum systems have locally uniform model parameters, for example, interactions and external fields and so on, the corresponding reduced density matrix for ideal Bose gas system can be treated by using the discrete Fourier transformation. However, in the many cases of
probabilistic information processing, it is expected to formulate the algorithms that is applicable
to density matrices with locally non-uniform model parameters, This is left for future research.

Acknowledgements
This work was partly supported by the Grants-In-Aid (No.18079002) and the Global COE (Center of Excellence) Program “Center of Education and Research for Information Electronics Systems” for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References
[1] Opper M and Saad D (eds) 2001 Advanced Mean Field Methods — Theory and Practice — (Cambridge: MIT Press)
[2] Kabashima Y and Saad D 1998 Europhysics Letters 44 668
[3] Kschischang F R, Frey R J and Loeliger H -A 2001 IEEE Transactions on Information Theory 47 498
[4] Yedidia J S, Freeman W T and Weiss Y 2005 IEEE Transactions on Information Theory 51 2252
[5] Freeman W T, Jones T R and Pasztor E C 2002 IEEE Computer Graphics and Applications 22 56
[6] Tanaka K 2002 J. Phys. A: Math. Gen. 35 R81
[7] Willsky A S 2002 Proceedings of IEEE 90 1396
[8] Frey B J 1998 Graphical Models for Machine Learning and Digital Communication (Cambridge: MIT Press)
[9] MacKay D J C 2003 Information Theory, Inference, and Learning Algorithms (Cambridge: Cambridge University Press)
[10] Kabashima Y and Saad D 2004 J. Phys. A: Math. Gen. 37 R1
[11] Pearl J 1988: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (San Francisco: Morgan Kaufmann Publishers Inc.)
[12] Jensen F V 2001 Bayesian Networks and Decision Graphs Springer
[13] Tanaka K 2003 IEICE Transactions on Information and Systems E86-D 1228
[14] Kikuchi R 1951 Phys. Rev. 81 988
[15] Morita T 1957 J. Phys. Soc. Jpn. 12 753
[16] Morita T 1972 J. Math. Phys. 13 115
[17] Morita T 1984 J. Stat. Phys. 34 319
[18] Morita T 1990 J. Stat. Phys. 59 819
[19] Pelizola A 2005 J. Phys. A: Math. Gen. 38 R309
[20] Weiss Y 2000 Neural Computation 12 1
[21] Montanari A and Rizzo T 2005 J. Stat. Mech.: Theory and Experiment P10011
[22] Marinari E and Smerjian G 2006 J. Stat. Mech.: Theory and Experiment P06019
[23] Yasuda M and Tanaka K 2006 J. Phys. Soc. Jpn 75 084006
[24] Yasuda M and Tanaka K 2007 J. Phys. A: Math. Theor. 40 9993
[25] Heskes T 2004 Neural Computation 16 2379
[26] Weiss Y Freeman W T 2001 Neural Computation 13 2173
[27] Tanaka K, Shouno H, Okada M, Titterington D M 2004: J. Phys. A: Math. Gen. 37 8675
[28] Tanaka K and Titterington D M: J. Phys. A: Math. Theor. 40 11285
[29] Opper M and Winther O 2001 Phys. Rev. Letts. 86 3695
[30] Opper M and Winther O 2001 Phys. Rev. E 64 05613186
[31] Opper M and Winther O 2005 Journal of Machine Learning Research 1 1
[32] Hoen-Sorensen P A F R and Winther O 2002 Neural Computation 2002 889
[33] Csato L, Opper M and Winther O 2003 Complexity 8 64
[34] Birrell G and Cugliandolo L F 2001 Phys. Rev. B, 64 014206
[35] Tanaka K and Horiguchi T 1997 IEICE Transactions A, J80-A 2117 (in Japanese); translated in Electronics and Communications in Japan 3: Fundamental Electronic Science 83 84
[36] Suzuki S, Nishimori H and Suzuki M 2007 Phys. Rev. E 75 051112
[37] Hofstetter H 2007 Phys. Rev. B 76 201102(R)
[38] Morita T 1957 J. Phys. Soc. Jpn. 12 1060
[39] Morita T 1994 Prog. Theor. Phys. 92 1081
[40] Morita T 1995 J. Phys. Soc. Jpn. 64 1211
[41] Horiguchi 1981 Physica A 107 360
[42] Yasuda M and Horiguchi T 2006 Physica A 368 83
[43] Kubo R 1952 *Phys. Rev.* **87** 568