In this paper, we review some recent developments of computational techniques in the probabilistic information processing and the statistical machine learning theory based on the statistical-mechanical informatics. In the first part, we explain the expectation-maximization algorithm for the estimations of hyperparameters based on maximum likelihood framework and show an expectation-maximization algorithm for Gaussian graphical model in noise reduction problems. Next, we explain some mathematical frameworks of a loopy belief propagation for probabilistic graphical models. The loopy belief propagations is familiar techniques of approximate computations of marginal probability distributions. We show an application of the loopy belief propagation to an approximate expectation-maximization algorithm for generalized sparse Gaussian graphical model in noise reduction problems. Moreover, we explain an approximate computational techniques of marginal probability distributions between every pairs of nodes by combining the loopy belief propagation with linear response formulas. In the second part, we review some fundamental aspects of the loopy belief propagation in the statistical-mechanical stand point of view. We first review an mean field theory of Ising model, which is one of the most fundamental probabilistic graphical models in the statistical mechanical informatics.

KEYWORDS: Probabilistic Information Processing, Statistical-Mechanical Informatics, Statistical Machine Learning, Statistical Inference, Belief Propagation, Advanced Mean-Field Method, Markov Random Field
1 Probability Theory

We consider a set of random variables \( \{A_i| i \in V\} \), where \( V=\{1,2,\cdots,|V|\} \) and \(|V|\) denotes the number of elements of the set \( V \). Each random variable \( A_i \) takes any integer numbers in \( \Omega_i=\{0,1,2,\cdots,|\Omega_i|-1\} \). We introduce a vector representation \( A = (A_1,A_2,\cdots,A_{|V|}) \) in terms of \(|V|\) random variables \( A_1,A_2,\cdots,A_{|V|} \). The vector representation \( A \) is referred to as Probability Vector. If the probabilities \( Pr\{A = a \} \) for the events \( A = a \) are expressed in terms of a function \( P(a) \) as follows:

\[
Pr\{A = a \} = P(a) (a \in \Omega_1 \times \Omega_2 \times \cdots \times \Omega_{|V|}),
\]

such that

\[
Pr\{A_1 = a_1, A_2 = a_2, \cdots, A_{|V|} = a_{|V|} \} = P(a_1, a_2, \cdots, a_{|V|}) (a_1 \in \Omega_1, a_2 \in \Omega_2, \cdots, a_{|V|} \in \Omega_{|V|}).
\]

the function \( P(a) = P(a_1, a_2, \cdots, a_{|V|}) \) and the \(|V|\)-dimensional vector \( a = (a_1, a_2, \cdots, a_{|V|}) \) are referred to as Joint Probability Distribution and State Vector, respectively. Marginal Probabilities \( P_i(a_i) \) and \( P_{i,j}(a_i, a_j) \) are defined by

\[
P_i(a_i) = \sum_{z_{i+1} \in \Omega_{i+1}} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{a_i,z_i} P(z_1, z_2, \cdots, z_{|V|}) (a_i \in \Omega_i, i \in V),
\]

\[
P_{i,j}(a_i, a_j) = \sum_{z_{i+1} \in \Omega_{i+1}} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{a_i,z_i} \delta_{a_j,z_j} P(z_1, z_2, \cdots, z_{|V|}) (a_i \in \Omega_i, a_j \in \Omega_j, i \in V, j \in V, i < j)
\]

\[
P_{i,j,k}(a_i, a_j, a_k) = \sum_{z_{i+1} \in \Omega_{i+1}} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{a_i,z_i} \delta_{a_j,z_j} \delta_{a_k,z_k} P(z_1, z_2, \cdots, z_{|V|}) (a_i \in \Omega_i, a_j \in \Omega_j, a_k \in \Omega_k, i \in V, j \in V, k \in V, i < j < k)
\]

\[
P_{i,j,k,l}(a_i, a_j, a_k, a_l) = \sum_{z_{i+1} \in \Omega_{i+1}} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{a_i,z_i} \delta_{a_j,z_j} \delta_{a_k,z_k} \delta_{a_l,z_l} P(z_1, z_2, \cdots, z_{|V|}) (a_i \in \Omega_i, a_j \in \Omega_j, a_k \in \Omega_k, a_l \in \Omega_l, i \in V, j \in V, k \in V, l \in V, i < j < k < l)
\]

where \( \delta_{a,b} = \begin{cases} 1 & (a = b) \\ 0 & (a \neq b) \end{cases} \) is a Kronecker’s delta. and \( a_y \in \Omega_{|V|} \) means \((\{a_k \in \Omega_k|k \in \gamma\})\). For \( \gamma = \{n_1, n_2, \cdots, n_{|\gamma|}\} \), the state vector \( a_y \) is defined by

\[
a_y = (a_i|i \in \gamma) \equiv (a_{n_1}, a_{n_2}, \cdots, a_{n_{|\gamma|}}) \quad (\gamma = \{n_1, n_2, \cdots, n_{|\gamma|}\} \subset V, n_1 < n_2 < \cdots < n_{|\gamma|}).
\]

These definitions of Eqs.(1.3)-(1.4) can be reduced to

\[
P_i(a_i) = \sum_{a_{i+1} \in \Omega_{i+1}} \cdots \sum_{a_{|V|} \in \Omega_{|V|}} P(a) (a \in \Omega_i)
\]

\[
P_{i,j}(a_i, a_j) = \sum_{a_{i+1} \in \Omega_{i+1}} \cdots \sum_{a_{i+j} \in \Omega_{i+j}} \sum_{a_{i+j+1} \in \Omega_{i+j+1}} \cdots \sum_{a_{|V|} \in \Omega_{|V|}} P(a)
\]

and Eq.(1.7) is generally expressed by

\[
P_y(a_y) = \sum_{a_{|V| \setminus \gamma} \in \Omega_{|V| \setminus \gamma}} P(a) (a_y \in \Omega_y)
\]

where

\[
[a_{|V| \setminus \gamma} \in \Omega_{|V| \setminus \gamma}] \equiv \bigcap_{i \in V \setminus \gamma} [a_i \in \Omega_i]
\]

Moreover, by using Eqs.(1.3) and (1.4), we have the following relationship between the above marginal probabilities as follows:

\[
P_i(a_i) = \sum_{a_{j \neq i} \in \Omega_{j \neq i}} P_{i,j}(a_i, a_j) (a_i \in \Omega_i, i \in V, j \in V)
\]
\[ P_j(a_j) = \sum_{a_i \in \Omega} P_{i,j}(a_i, a_j) \quad (a_j \in \Omega_j, \ i \in \mathcal{V}, \ j \in \mathcal{V}). \quad (1.14) \]

The proof of Eq.(1.13) is as follows:

\[
\sum_{a_j \in \Omega_j} P_{i,j}(a_i, a_j) = \sum_{a_j \in \Omega_j} \sum_{z_1, z_2, \cdots, z_{|\mathcal{V}|} \in \Omega_{|\mathcal{V}|}} \delta_{a_i, z_i} \delta_{a_j, z_j} P(z_1, z_2, \cdots, z_{|\mathcal{V}|}) \\
= \sum_{z_1, z_2, \cdots, z_{|\mathcal{V}|} \in \Omega_{|\mathcal{V}|}} \delta_{a_i, z_i} \left( \sum_{a_j \in \Omega_j} \delta_{a_j, z_j} \right) P(z_1, z_2, \cdots, z_{|\mathcal{V}|}) \\
= \sum_{z_1, z_2, \cdots, z_{|\mathcal{V}|} \in \Omega_{|\mathcal{V}|}} \delta_{a_i, z_i} P(z_1, z_2, \cdots, z_{|\mathcal{V}|}) = P_i(a_i). \quad (1.15) \]

In the similar arguments, we have also Eq.(1.13).

In the case that each random variable \( a_i \) takes any real numbers in the interval \((-\infty, +\infty)\) and a finite interval in \((-\infty, +\infty)\) is defined by \((b_i, c_i)\) for every \(i = 1, 2, \cdots, |\mathcal{V}|\), we should consider the probability for an event \([A_1 \in (b_1, c_1)] \cap [A_2 \in (b_2, c_2)] \cap \cdots \cap [A_{|\mathcal{V}|} \in (b_{|\mathcal{V}|}, c_{|\mathcal{V}|})]\). Let us suppose that the probability for a event defined by the region \((b_1, c_1) \times (b_2, c_2) \times \cdots \times \{b_{|\mathcal{V}|}, c_{|\mathcal{V}|}\}\) in a \(|\mathcal{V}|\)-dimensional space \((-\infty, +\infty)^{|\mathcal{V}|}\) can be expressed in terms of a function \(f(a) = f(a_1, a_2, \cdots, a_{|\mathcal{V}|})\) as follows:

\[
\Pr\{A_1 \in (b_1, c_1), A_2 \in (b_2, c_2), \cdots, A_{|\mathcal{V}|} \in (b_{|\mathcal{V}|}, c_{|\mathcal{V}|})\} = \Pr\{A_1 \in (b_1, c_1), A_2 \in (b_2, c_2), \cdots, A_{|\mathcal{V}|} \in (b_{|\mathcal{V}|}, c_{|\mathcal{V}|})\} = \int_{b_1}^{c_1} \cdots \int_{b_{|\mathcal{V}|}}^{c_{|\mathcal{V}|}} f(z_1, z_2, \cdots, z_{|\mathcal{V}|}) dz_1 dz_2 \cdots dz_{|\mathcal{V}|}. \quad (1.16)
\]

If we introduce

\[
F(a_1, a_2, \cdots, a_{|\mathcal{V}|}) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_{|\mathcal{V}|}} f(z_1, z_2, \cdots, z_{|\mathcal{V}|}) dz_1 dz_2 \cdots dz_{|\mathcal{V}|},
\]

\[
f(a_1, a_2, \cdots, a_{|\mathcal{V}|}) = \frac{\partial}{\partial a_1} \frac{\partial}{\partial a_2} \cdots \frac{\partial}{\partial a_{|\mathcal{V}|}} F(a_1, a_2, \cdots, a_{|\mathcal{V}|}).
\]

In such cases of continuous random variables, the functions \(f(a) = f(a_1, a_2, \cdots, a_{|\mathcal{V}|})\) are referred to as a Joint Probability Density Function and a Distribution Function. For the probability density function \(f(a) = f(a_1, a_2, \cdots, a_{|\mathcal{V}|})\), we can define marginal probability density functions as follows:

\[
f_i(a_i) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \delta(a_i - z_i) f(z_1, z_2, \cdots, z_{|\mathcal{V}|}) dz_1 dz_2 \cdots dz_{|\mathcal{V}|}, \quad (1.19)
\]

\[
f_{i,j}(a_i, a_j) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \delta(a_i - z_i) \delta(a_j - z_j) f(z_1, z_2, \cdots, z_{|\mathcal{V}|}) dz_1 dz_2 \cdots dz_{|\mathcal{V}|}, \quad (1.20)
\]

where \(\delta(a)\) is the Dirac’s delta function. Moreover, by using Eqs.(1.19) and (1.20), we have the following relationship between the above marginal probability density functions as follows:

\[
f_i(a_i) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(a_i - z_i) f_{i,j}(z_i, z_j) dz_i dz_j = \int_{-\infty}^{+\infty} f_{i,j}(a_i, a_j) da_j \quad (a_i \in (-\infty, +\infty), \ i \in \mathcal{V}, \ j \in \mathcal{V}),
\]

\[
f_j(a_j) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(a_j - z_j) f_{i,j}(z_i, z_j) dz_i dz_j = \int_{-\infty}^{+\infty} P_{i,j}(a_i, a_j) da_i \quad (a_j \in (-\infty, +\infty), \ i \in \mathcal{V}, \ j \in \mathcal{V}).
\]

Hereafter, when we consider a continuous state space \((-\infty, +\infty)\) for the random variable \(a_i\), we remark that the notation \(\sum_{a_i \in \Omega_i}\) means the integral \(\int_{-\infty}^{+\infty} da_i\).

We introduce two kinds of probability vectors \(A = (A_1, A_2, \cdots, A_{|\mathcal{V}|})\) and \(D = (D_1, D_2, \cdots, D_{|\mathcal{V}|})\) and their corresponding state vectors \(a = (a_1, a_2, \cdots, a_{|\mathcal{V}|})\) and \(d = (d_1, d_2, \cdots, d_{|\mathcal{V}|})\) and consider a joint probability distribution \(P(a, d) = P(a_1, a_2, \cdots, a_{|\mathcal{V}|}, d_1, d_2, \cdots, d_{|\mathcal{V}|})\) as follows:

\[
\Pr\{A = a, D = d\} = P(a, d). \quad (1.23)
\]
When an event \( D = d \) is given, \textit{Conditional Probability Distribution} \( P(a|d) \) of \( A \) is defined by
\[
P(a|d) = \frac{P(a,d)}{P_D(d)},
\]
where
\[
P_D(d) = \sum_{a \in \Omega} P(a,d).
\]
When an event \( A = a \) is given, the conditional probability distribution \( P(d|a) \) of \( D \) is expressed as
\[
P(d|a) = \frac{P(a,d)}{P_A(a)},
\]
where
\[
P_A(a) = \sum_d P(a,d).
\]
Eqs.(1.24) and (1.26) can be rewritten to
\[
P(a,d) = P(a|d)P_D(d)
\]
\[
P(a,d) = P(d|a)P_A(a)
\]
From Eqs.(1.28) and (1.29), we have
\[
P(a|d)P_D(d) = P(d|a)P_A(a)
\]
and then derive so-called \textit{Bayes Formulas} as follows:
\[
P(a|d) = \frac{P(a|d)P_A(a)}{P_D(d)} = \frac{P(a|d)P_A(a)}{\sum_a P(a|d)P_A(a)}.
\]
We remark that the second equality in Eq.(1.31) can be verified by substituting Eq.(1.29) to the right-hand side of Eq.(1.25). Now we consider the meaning of the Bayes formulas by regarding \( a \) and \( d \) to state vectors of information sources and data. A data set \( d \) is assumed to be generated by according to the conditional probability distribution \( P(d|a) \) when an event for \( A = a \) for information sources has occurred. And an event for \( A = a \) for information sources has occurred by according to the probability distribution \( P_A(a) = P_A(a_1,a_2,\cdots,a_{|V|}) \). In such a situation, \( P(a|d) = P(a_1,a_2,\cdots,a_{|V|}|d) \) can be regarded as the conditional probability distribution to of the information source \( a \) after the data set \( d \) has been observed. \( P(a|d) = P(a_1,a_2,\cdots,a_{|V|}|d) \) is referred to as \textit{Posterior Probability Distribution} and \( P_A(a) = P_A(a_1,a_2,\cdots,a_{|V|}) \) is called \textit{Prior Probability Distribution}. We regard the information sources \( a \) as \textit{parameter} and can estimate it by means of the posterior probability distribution. One of the choiceto estimate the parameter \( a = (a_1,a_2,\cdots,a_{|V|}) \) is a \textit{Maximum A Posteriori} (MAP) estimation as follows:
\[
\hat{a}(d) = \left(\hat{a}_1(d),\hat{a}_2(d),\cdots,\hat{a}_{|V|}(d)\right)
\equiv \arg \max_{z_1 \in \Omega_1,z_2 \in \Omega_2,\cdots,z_{|V|} \in \Omega_{|V|}} P(z_1,z_2,\cdots,z_{|V|}|d).
\]
As one of familiar probability distributions, we have the \textit{Boltzmann Distribution}:
\[
P(a_1,a_2,\cdots,a_{|V|}|d) = \frac{1}{Z(d,T)} \exp \left( \frac{-1}{T} H(a_1,a_2,\cdots,a_{|V|}|d) \right) (T > 0).
\]
\[
Z(d,T) \equiv \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \exp \left( \frac{-1}{T} H(z_1,z_2,\cdots,z_{|V|}|d) \right).
\]
Here \( T \) and \( Z(d,T) \) is referred to as a \textit{Temperature} and a \textit{Partition Function} and
\[
F(d,T) \equiv -\ln(Z(d,T))
\]
corresponds to a \textit{Free Energy} of \( P(a|d) \). For the Boltzmann distribution, the MAP criteria is equivalent to the following optimization problem:
\[
\hat{a}(d) = \left(\hat{a}_1(d),\hat{a}_2(d),\cdots,\hat{a}_{|V|}(d)\right) \equiv \arg \min_{z_1 \in \Omega_1,z_2 \in \Omega_2,\cdots,z_{|V|} \in \Omega_{|V|}} H(z_1,z_2,\cdots,z_{|V|}|d).
\]
Here \( H(a|d) = H(a_1, a_2, \ldots, a_V|d) \) is referred to as the Hamiltonian or Energy Function in the statistical mechanical informatics. Another choice is a Maximum Posterior Marginal (MPM) estimation defined by

\[
\hat{a}_i(d) = \arg\max_{z_i \in \Omega_i} P_i(z_i|d) \quad (i \in V),
\]

where

\[
P_i(a_i|d) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} \delta_{a_i,z_i} P(z_1, z_2, \ldots, z_i, \ldots, z_V|d) \quad (i \in V).
\]

We remark that the estimates obtained by MAP and MPM estimations are usually different from each other except for some special cases. In this paper, we mainly adopt the MPM estimation to determine the estimates of parameter vector \( a \) for a given data point \( d \).

As the proximity between two probability distributions \( P(a) \) and \( Q(a) \), we introduce a Kullback-Leibler Divergence (KL divergence) as follows: such that

\[
\text{KL}[P|Q] = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} P(z_1, z_2, \ldots, z_V|d) \ln \left( \frac{Q(z_1, z_2, \ldots, z_V|d)}{P(z_1, z_2, \ldots, z_V|d)} \right).
\]

\[
\text{KL}[P|Q] \equiv \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} Q(z_1, z_2, \ldots, z_V|d) \ln \left( \frac{Q(z_1, z_2, \ldots, z_V|d)}{P(z_1, z_2, \ldots, z_V|d)} \right). 
\]

\[
\text{KL}[P|Q] = - \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} Q(z_1, z_2, \ldots, z_V|d) \left( \frac{P(z_1, z_2, \ldots, z_V|d)}{Q(z_1, z_2, \ldots, z_V|d)} - 1 \right)
\]

\[
= - \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} P(z_1, z_2, \ldots, z_V|d) + \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} Q(z_1, z_2, \ldots, z_V|d) = 0. 
\]

Let us suppose that a joint probability distribution \( P(a) \) is expressed as a function which is parameterized by an \( K \)-dimensional vector \( \theta = (\theta_1, \theta_2, \ldots, \theta_K) \) with \( \theta_i \) taking any real value in the interval \(( -\infty, +\infty) \). In this case, \( P(a) \) can be rewritten as \( P(a|\theta) \). We have \( D \) configurations \( a^{(1)}, a^{(2)}, \ldots, a^{(D)} \) for the state vector \( a \) as data points.

First of all, we have \( D \) configurations \( d^{(1)}, d^{(2)}, \ldots, d^{(D)} \) for the state vector \( d \) as data points. These data points are assumed to be generated as muntual independent events by according to the same identical probability distribution. We introduce a probability distribution \( P^*(d) = P^*(d_1, d_2, \ldots, d_V|d) \) in terms of the above data points as follows:

\[
P^*(d) = \frac{1}{D} \prod_{n=1}^{D} \prod_{i \in V} \delta_{d_i^{(n)}, d_i},
\]

where is referred to as an Empirical Joint Probability Distribution. Now we consider a joint probability distribution \( P(d|\theta) \) which is parameterized by an \( K \)-dimensional vector \( \theta = (\theta_1, \theta_2, \ldots, \theta_K) \) with \( \theta_i \) taking any real value in the interval \(( -\infty, +\infty) \). We introduce the KL divergence between \( P^*(d) \) and \( P(d|\theta) \) as follows:

\[
\text{KL}[P|P^*] = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} P^*(z_1, z_2, \ldots, z_V|d) \ln \left( \frac{P^*(z_1, z_2, \ldots, z_V|d)}{P(z_1, z_2, \ldots, z_V|\theta)} \right). 
\]

and the estimate \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K) \) of the parameter vector \( \theta \) is determined so as to minimize \( \text{KL}[P|P^*] \) as follows:

\[
\hat{\theta} = \arg\min_{\theta} \text{KL}[P|P^*]. 
\]

Because \( \text{KL}[P|P^*] \) can be approximately rewritten as

\[
\text{KL}[P|P^*] \simeq \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} P^*(z_1, z_2, \ldots, z_V|d) \ln \left( P^*(z_1, z_2, \ldots, z_V|d) \right)
\]

\[
- \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} P^*(z_1, z_2, \ldots, z_V|d) \ln \left( P(z_1, z_2, \ldots, z_V|\theta) \right)
\]

\[
= \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} \left( \frac{1}{D} \sum_{n=1}^{D} \prod_{i \in V} \delta_{d_i^{(n)}, z_i} \right) \ln \left( \frac{1}{D} \sum_{n=1}^{D} \prod_{i \in V} \delta_{d_i^{(n)}, z_i} \right). 
\]
\[
\begin{align*}
&- \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \left( \frac{1}{D} \sum_{d=1}^{D} \prod_{i \in V} \delta_{i}^{(n)}(z_i) \right) \ln(P(z_1, z_2, \cdots, z_{|V|} | \theta)) \\
&= \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \left( \frac{1}{D} \sum_{d=1}^{D} \prod_{i \in V} \delta_{i}^{(n)}(d_i) \right) \ln \left( \frac{1}{D} \sum_{d=1}^{D} \prod_{i \in V} \delta_{i}^{(n)}(z_i) \right) \\
&= -\frac{1}{D} \sum_{d=1}^{D} \ln \left( \prod_{i \in V} \delta_{i}^{(n)}(d_i) \right) \\
&\quad \text{for any data points for the state vector } P
\end{align*}
\]

The deterministic equation of the estimate \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_K) \) can be reduced to
\[
\hat{\theta} = \arg\max_{\theta} \prod_{n=1}^{D} P(d^{(n)} | \theta).
\]

It means that the minimization (1.43) of the KL divergence between the empirical joint probability distribution \( P^*(a) = P^*(d_1, d_2, \cdots, d_{|V|}) \) and the parameterized joint probability distribution \( P(a | \theta) \) can be reduced to a Maximum Likelihood Estimation for the parameter vector \( \theta \) by regarding the joint probability distribution \( P(d^{(1)}, d^{(2)}, \cdots, d^{(n)}) \equiv \prod_{n=1}^{D} P(d^{(n)} | \theta) \) to realize the data points \( d^{(1)}, d^{(2)}, \cdots, d^{(n)} \) as a Likelihood Function of the parameter vector \( \theta \) when the data points \( d^{(1)}, d^{(2)}, \cdots, d^{(n)} \) are given.

Next, we devise the state vector \( a \) to two state vectors \( (a_1, a_2, \cdots, a_{|V|}) \) and \( (d_1, d_2, \cdots, d_{|V|}) \) and let us suppose that we have \( D \) configurations \( \left\{ (d_1^{(n)}, d_2^{(n)}, \cdots, d_{|V|}^{(n)}) \mid n = 1, 2, \cdots, D \right\} \) for the state vector \( (d_1, d_2, \cdots, d_{|V|}) \) as data points and any data points for the state vector \( (a_1, a_2, \cdots, a_{|V|}) \) are missing. In this case, we consider the empirical joint probability distribution \( P_D(d_1, d_2, \cdots, d_{|V|} | \theta) \) and the marginal probability \( P(d_1, d_2, \cdots, d_{|V|} | \theta) \) as follows:
\[
P_D(d_1, d_2, \cdots, d_{|V|} | \theta) = \sum_{a_1 \in \Omega_1} \sum_{a_2 \in \Omega_2} \cdots \sum_{a_{|V|} \in \Omega_{|V|}} P(a_1, a_2, \cdots, a_{|V|}, d_1, d_2, \cdots, d_{|V|} | \theta).
\]

We introduce the KL divergence between \( P^*(d_1, d_2, \cdots, d_{|V|}) \) and \( P(d_1, d_2, \cdots, d_{|V|} | \theta) \) as follows:
\[
\text{KL}[P_D || P_D^*] \equiv \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} P_D(z_1, z_2, \cdots, z_{|V|} | \theta) \ln \left( \frac{P_D(z_1, z_2, \cdots, z_{|V|} | \theta)}{P^*(d_1, d_2, \cdots, d_{|V|})} \right),
\]
and the estimate \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_K) \) of the parameter vector \( \theta \) is determined so as to minimize \( \text{KL}[P_D || P_D^*] \) as follows:
\[
\hat{\theta} = \arg\min_{\theta} \text{KL}[P_D || P_D^*].
\]

The deterministic equation of the estimate \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_K) \) can be reduced to
\[
\hat{\theta} = \arg\max_{\theta} \prod_{n=1}^{D} P_D(d_1^{(n)}, d_2^{(n)}, \cdots, d_{|V|}^{(n)} | \theta).
\]
consisting of pairs of nodes $i$ and $j$. In the conventional EM algorithm, after setting the hyperparameter vector $\theta$ at fixed true values $\theta^*$ of hyperparameters, the marginal likelihood $P_{1,2,\ldots,|V|}(a_{i+1}, a_{i+2}, \ldots, a_{|V|})$ is statistically maximized at $\theta = \theta^*$. The estimates $\hat{\theta}$ in Eq. (1.50) are expected to be close to the true values $\theta = \theta^*$ of hyperparameters when our observed data vector $(a_{i+1}, a_{i+2}, \ldots, a_{|V|})$ has been produced.

In order to realize the maximization of marginal likelihood in Eq. (1.50), an Expectation-Maximization (EM) algorithm are often introduced in the statistical machine learning theory. In the EM algorithm, we first introduce

$$
Q(\theta \mid \theta^*, \{(d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)}) | n = 1, 2, \ldots, D\})
$$

in terms of the marginal likelihood in Eq. (1.51) and the posterior probability distribution

$$
P(a_1, a_2, \ldots, a_{|V|}|d_1, d_2, \ldots, d_{|V|}, \theta) \equiv \frac{P(a_1, a_2, \ldots, a_{|V|}, d_1, d_2, \ldots, d_{|V|}|\theta)}{P_D(d_1, d_2, \ldots, d_{|V|}|\theta)},
$$

where

$$
P_D(d_1, d_2, \ldots, d_{|V|}|\theta) \equiv \sum_{z_1, z_2, \ldots, z_{|V|} \in \Omega_{z_{|V|}}} \sum_{z_1, z_2, \ldots, z_{|V|} \in \Omega_{z_{|V|}}} P(z_1, z_2, \ldots, z_{|V|}, d_1, d_2, \ldots, d_{|V|}|\theta).
$$

In the conventional EM algorithm, after setting the hyperparameter vector $\hat{\theta}$ to an initial vector, the following procedures are repeated until $\hat{\theta}$ converges:

E-step: Compute $Q(\theta \mid \theta^*, \{(a_1^{(n)}, a_2^{(n)}, \ldots, a_{|V|}^{(n)}) | n = 1, 2, \ldots, D\})$ for any $K$-dimensional real vectors $\theta \in (-\infty, +\infty)^K$.

M-step: Update $\hat{\theta}$ as follows:

$$
\hat{\theta} := \arg\max_{\theta \in (-\infty, +\infty)^K} Q(\theta \mid \theta^*, \{(a_1^{(n)}, a_2^{(n)}, \ldots, a_{|V|}^{(n)}) | n = 1, 2, \ldots, D\}).
$$

The extremum conditions $Q(\theta \mid \theta^*, \{(d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)}) | n = 1, 2, \ldots, D\})$ with respect to $\theta$

$$
\left[ \frac{\partial}{\partial \theta_k} Q(\theta \mid \theta^*, \{(d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)}) | n = 1, 2, \ldots, D\}) \right]_{\theta = \hat{\theta}} = 0 \quad (n = 1, 2, \ldots, K),
$$

can be reduced to the ones of the marginal likelihood $P((d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)}) | n = 1, 2, \ldots, D) | \theta)$ as follows:

$$
\sum_{n=1}^{D} \frac{1}{P(d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)} | \theta)} \frac{\partial}{\partial \theta_k} P(d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)} | \theta) \bigg|_{\theta = \hat{\theta}} = 0,\quad (n = 1, 2, \ldots, K),
$$

such that

$$
\left[ \frac{\partial}{\partial \theta_k} \prod_{n=1}^{D} P_D(d_1^{(n)}, d_2^{(n)}, \ldots, d_{|V|}^{(n)} | \theta) \right]_{\theta = \hat{\theta}} = 0 \quad (n = 1, 2, \ldots, K).
$$

2 Expectation-Maximization Algorithm for Maximum Marginal Likelihood Estimation in Gaussian Graphical Models

We introduce a graph $(V, E)$ where $V = \{1, 2, \ldots, |V|\}$ is a set of all the nodes and $E$ is a set of all the edges $\{i, j\}$ consisting of pairs of nodes $i$ and $j$. We define the state variables $a_i$ and $d_i$ ($i \in V$), which take any real values in the interval $(-\infty, +\infty)$. We consider a probability density function $P(a)$ and a conditional probability density function...
\( P(d|α, β, γ) \) defined by

\[
f(a|α, γ) \equiv \frac{1}{Z(α, γ)} \exp\left( -\frac{1}{2} α \sum_{(i,j)\in E} (a_i - a_j)^2 - \frac{1}{2} γ \sum_{i\in V} a_i^2 \right),
\]
\[
f(d|α, β) \equiv \sqrt{\frac{β}{2π}} |V| \exp\left( -\frac{1}{2} β \sum_{i\in V} (d_i - d_i)^2 \right),
\]
on the state space \( Ω = (-∞, +∞)^|V| \), \( Z(α, γ) \) is a normalization constant of \( f(a|α, γ) \), which is defined by

\[
Z(α, γ) \equiv \int_{-∞}^{+∞} \int_{-∞}^{+∞} \cdots \int_{-∞}^{+∞} \exp\left( -\frac{1}{2} α \sum_{(i,j)\in E} (z_i - z_j)^2 - \frac{1}{2} γ \sum_{i\in V} z_i^2 \right) dz_1 dz_2 \cdots dz_{|V|}.
\]

The posterior probability density function \( f(a|d, α, β, γ) \) in Eq.(2.13), is defined by

\[
f(a|d, α, β, γ) \equiv \frac{f(d|a, β) f(a|α, γ)}{f(d|α, β, γ)},
\]
where the denominator \( f(d|α, β, γ) \) in the right-hand side is the probability density function of \( d \) defined by

\[
f(d|α, β, γ) \equiv \int_{-∞}^{+∞} \int_{-∞}^{+∞} \cdots \int_{-∞}^{+∞} f(z, d|α, β, γ) dz_1 dz_2 \cdots dz_{|V|} = \int_{-∞}^{+∞} \int_{-∞}^{+∞} \cdots \int_{-∞}^{+∞} f(d|z, β) f(z|α, γ) dz_1 dz_2 \cdots dz_{|V|}.
\]

Now we regard the probability \( f(d|α, β, γ) \) of observed data \( d \) with the hyperparameters \( α, β \) and \( γ \) given as a likelihood function of \( (α, β, γ) \) with the data \( d \) given and estimate the hyperparameters \( α, β \) and \( γ \) so as to maximize the marginal likelihood \( P(d|α, β, γ) \) as follows:

\[
(\hat{α}, \hat{β}, \hat{γ}) = \arg \max_{α, β, γ} f(d|α, β, γ).
\]

After determining the hyperparameters, estimates \( \hat{a}_i(\hat{α}, \hat{β}, \hat{γ})d \) for all the nodes \( i\in V \) are determined so as to maximize the posterior marginal probability distribution \( P_j(a_i|d, \hat{α}, \hat{β}, \hat{γ}) \) at each node as follows:

\[
\hat{a}_i(\hat{α}, \hat{β}, \hat{γ}) = \arg \max_{z_i\in (-∞, +∞)} f_i(z_i|d, \hat{α}, \hat{β}, \hat{γ}),
\]
where

\[
f_i(a_i|d, \hat{α}, \hat{β}, \hat{γ}) \equiv \int_{-∞}^{+∞} \int_{-∞}^{+∞} \cdots \int_{-∞}^{+∞} δ(z_i - a_i) P(z|d, \hat{α}, \hat{β}, \hat{γ}) dz_1 dz_2 \cdots dz_{|V|}.
\]

By substituting Eqs.(2.1) and (2.2), the posterior probability density function of Eq.(2.4) and the marginal likelihood of Eq.(2.5) are expressed as follows:

\[
f(a|d, α, β, γ) = \frac{1}{Z(d, α, β, γ)} \exp\left( -\frac{1}{2} α \sum_{(i,j)\in E} (a_i - a_j)^2 - \frac{1}{2} γ \sum_{i\in V} a_i^2 - \frac{1}{2} β \sum_{i\in V} (d_i - d_i)^2 \right),
\]
\[
f(a|d, α, β, γ) = \frac{1}{Z(d, α, β, γ)} \sqrt{\frac{β}{2π}} |V| \exp\left( -\frac{1}{2} α \sum_{(i,j)\in E} (a_i - a_j)^2 - \frac{1}{2} γ \sum_{i\in V} a_i^2 - \frac{1}{2} β \sum_{i\in V} (d_i - d_i)^2 \right),
\]
where \( Z(d, α, β, γ) \) is a normalization constant of the posterior probability density function and is defined by

\[
Z(d, α, β, γ) \equiv \int_{-∞}^{+∞} \int_{-∞}^{+∞} \cdots \int_{-∞}^{+∞} \exp\left( -\frac{1}{2} α \sum_{(i,j)\in E} (z_i - z_j)^2 - \frac{1}{2} γ \sum_{i\in V} z_i^2 - \frac{1}{2} β \sum_{i\in V} (z_i - z_i)^2 \right) dz_1 dz_2 \cdots dz_{|V|}.
\]

The posterior probability density function \( P(z|d, α, β, γ) \) in Eq.(2.13), is defined by

\[
f(a|d, α, β, γ) = \frac{f(d|a, β) f(a|α, γ)}{f(d|α, β, γ)},
\]
where the denominator \( f(d|α, β, γ) \) in the right-hand side is the probability density function of \( d \) defined by Eq.(2.5).
Now we regard the probability \( f(d|\alpha, \beta, \gamma) \) of observed data \( d \) with the hyperparameters \( \alpha, \beta \) and \( \gamma \) given as a likelihood function of \((\alpha, \beta, \gamma)\) with the data \( d \) given and estimate the hyperparameters \( \alpha, \beta \) and \( \gamma \) so as to maximize the marginal likelihood \( f(d|\alpha, \beta, \gamma) \) in Eq.\((2.6)\). After determining the hyperparameters, estimates \( \hat{\alpha}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \) for all the nodes \((i \in V)\) are determined so as to maximize the posterior marginal probability distribution \( f_i(a_i|d, \hat{\alpha}, \hat{\beta}, \hat{\gamma}) \) at each node by Eqs.\((2.7)-(2.8)\).

We introduce the Q-function defined by

\[
Q(\alpha, \beta, \gamma|\alpha', \beta', \gamma', d) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(z|d, \alpha', \beta', \gamma') \ln \left( P(z, d|\alpha, \beta, \gamma) \right) dz_1dz_2 \cdots dz_{|V|},
\]

(2.13)

for the probability density functions \( f(a|\alpha, \gamma) \) and \( f(a|d, \beta) \). By applying these expressions to the update rule of the EM algorithm:

\[
(\alpha(t+1), \beta(t+1), \gamma(t+1)) \Leftarrow \arg \max_{(a, b, c) \in (0, 1, \infty)^3} Q(a, b, c|\alpha(t), \beta(t), \gamma(t), d),
\]

(2.14)

and by considering the extremum conditions of \( Q(\alpha, \beta, \gamma|\alpha', \beta', \gamma', d) \) with respect to \( a, b \) and \( c \), the update rule (2.14) from \((\alpha(t), \beta(t), \gamma(t))\) to \((\alpha(t+1), \beta(t+1), \gamma(t+1))\) can be reduced to the simultaneous deterministic equations: for \((\alpha(t+1), \beta(t+1), \gamma(t+1))\), with \((\alpha(t), \beta(t), \gamma(t))\) being given, as follows:

\[
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|E|} \sum_{i,j \in E} (z_i - z_j)^2 \right) f(z|\alpha(t+1), \beta(t+1), \gamma(t+1)) dz_1dz_2 \cdots dz_{|V|}
\]

(2.15)

\[
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} z_i^2 \right) f(z|\alpha(t+1), \beta(t+1), \gamma(t+1)) dz_1dz_2 \cdots dz_{|V|}
\]

(2.16)

\[
\beta(t+1)^{-1} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} (z_i - d_i)^2 \right) f(z|\alpha(t), \beta(t), \gamma(t)) dz_1dz_2 \cdots dz_{|V|}
\]

(2.17)

The practical procedures to solve Eqs.\((2.15), (2.16)\) and \((2.17)\) are summarized as follows:

**EM Algorithm for Gaussian Graphical Model (1)**

**Step 1** Input a given data point \(d\) and \(t=0\). Set \(\alpha(0), \beta(t)\) and \(\gamma(t)\) as initial values.

**Step 2** Update \(t\) by \(t \Leftarrow t+1\) and compute \(u, v\) and \(\beta(t)\) from \(\alpha(t-1), \beta(t-1)\) and \(\gamma(t-1)\) as follows:

\[
u \Leftarrow \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} z_i^2 \right) f(z|\alpha(t-1), \beta(t-1), \gamma(t-1)) dz_1dz_2 \cdots dz_{|V|},
\]

(2.19)

\[
\beta(t) \Leftarrow \left( \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} (z_i - d_i)^2 \right) f(z|\alpha(t-1), \beta(t-1), \gamma(t-1)) dz_1dz_2 \cdots dz_{|V|} \right)^{-1}.
\]

(2.20)

**Step 3** Repeat the following updates of \(\alpha(i)\) and \(\gamma(i)\) until they converge:

\[
\alpha(i) \Leftarrow \alpha(i) \left( \frac{1}{u} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|E|} \sum_{i,j \in E} (z_i - z_j)^2 \right) f(z|\alpha(i), \gamma(i)) dz_1dz_2 \cdots dz_{|V|} \right)^{1/2},
\]

(2.21)

\[
\gamma(i) \Leftarrow \gamma(i) \left( \frac{1}{v} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} z_i^2 \right) f(z|\alpha(i), \gamma(i)) dz_1dz_2 \cdots dz_{|V|} \right)^{1/2}.
\]

(2.22)
Step 4 Set $\tilde{\alpha} \leftarrow \alpha(t)$, $\tilde{\beta} \leftarrow \beta(t)$, $\tilde{\gamma} \leftarrow \gamma(t)$ and

$$\hat{a}_i(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} | d) \leftarrow \arg \max_{z_i \in (-\infty, +\infty]} f_i(z_i | d, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \ (i \in V).$$  \hfill (2.23)

Stop if $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\gamma}$ converge, and go to Step 2 otherwise.

The right-hand side of Eqs. (2.22) - (2.22) are rewritten in terms of the normalization constants $Z(d, \alpha, \beta, \gamma)$ and $Z(\alpha, \gamma)$ as follows:

$$
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|E|} \sum_{(i,j) \in E} (z_i - z_j)^2 \right) f(z | d, \alpha, \beta, \gamma) dz_1 dz_2 \cdots dz_{|V|} = -\frac{2}{|V|} \frac{\partial}{\partial \alpha} \ln Z(d, \alpha, \beta, \gamma),
$$

$$
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} z_i^2 \right) f(z | d, \alpha, \beta, \gamma) dz_1 dz_2 \cdots dz_{|V|} = -\frac{2}{|V|} \frac{\partial}{\partial \beta} \ln Z(d, \alpha, \beta, \gamma),
$$

$$
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{(i,j) \in E} (z_i - z_j)^2 \right) f(z | \alpha, \beta, \gamma) dz_1 dz_2 \cdots dz_{|V|} = -\frac{2}{|V|} \frac{\partial}{\partial \gamma} \ln Z(\alpha, \gamma),
$$

$$
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \frac{1}{|V|} \sum_{i \in V} z_i^2 \right) f(z | \alpha, \gamma) dz_1 dz_2 \cdots dz_{|V|} = -\frac{2}{|V|} \frac{\partial}{\partial \gamma} \ln Z(\alpha, \gamma)
$$

By using the $|V|$-dimensional Gaussian integral formulas for an $|V| \times |V|$ positive definite real symmetric matrix $A$ and an $|V|$-dimensional real vector $m$:

$$
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left( -\frac{1}{2} (z - m)^T A (z - m) \right) dz_1 dz_2 \cdots dz_{|V|} = \sqrt{\frac{(2\pi)^{|V|}}{\det(A)}}
$$

the computations of $Z(d, \alpha, \beta, \gamma)$ and $Z(\alpha, \gamma)$ can be reduced to ones of determinants of $|V| \times |V|$ matrices as follows:

$$
Z(\alpha, \gamma) = \sqrt{\frac{(2\pi)^{|V|}}{\det(VI + \alpha C)}}
$$

$$
Z(\alpha, \beta, \gamma) = \sqrt{\frac{(2\pi)^{|V|}}{\det((\beta + \gamma)I + \alpha C)}}
$$

$$
\times \exp \left( -\frac{1}{2} \beta d (\gamma I + \alpha C) \left( (\beta + \gamma)I + \alpha C \right)^{-1} d^T \right)
$$

$C$ is an $|V| \times |V|$ matrix whose $(i, j)$-component $\langle i | C | j \rangle$ $(i \in V, i \in V)$ is defined by

$$
\langle i | C | j \rangle \equiv \begin{cases} 
| \partial i | & (i = j), \\
-1 & (i, j) \in E), \\
0 & \text{otherwise},
\end{cases}
$$

where $\partial i \equiv \{ (i, j) | (i, j) \in E \}$ is the set of all the nodes connected with the node $i$ by an edge, such that all the neighbouring nodes of $i$. $I$ is the $|V| \times |V|$ identity matrix.

The posterior probability density function in Eq.(2.4) is rewritten as

$$
f(a | d, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) = \sqrt{\frac{(2\pi)^{|V|}}{\det((\beta + \gamma)I + \alpha C)}}
$$

$$
\times \exp \left( \frac{1}{2} \left( a - ((\beta + \gamma)I + \alpha C)^{-1} d \right)^T ((\beta + \gamma)I + \alpha C) \left( a - ((\beta + \gamma)I + \alpha C)^{-1} d \right) \right),
$$

\hfill (2.33)

and then the maximization of the posterior marginal in Eq.(2.7) ac be reduced to

$$
\mathbf{\overline{a}}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} | d) = ((\beta U) + \tilde{\gamma} I + \tilde{\alpha} C)^{-1} d.
$$

\hfill (2.34)
By substituting Eqs. (2.1), (2.2), (2.30) and (2.31) to Eq. (2.5), we have
\[
f(d|\alpha, \beta, \gamma) = \sqrt{\frac{\beta}{2\pi}} |\mathcal{V}| \frac{Z(d, \alpha, \beta, \gamma)}{Z(\alpha, \gamma)}
\]
\[
= \sqrt{\frac{\det(\beta(\gamma I + \alpha C)}{(2\pi)^{|\mathcal{V}|} \det((\beta + \gamma)I + \alpha C)}
\]
\[
\times \exp\left(-\frac{1}{2}\beta d(\gamma I + \alpha C)((\beta + \gamma)I + \alpha C)^{-1} d^T\right).
\]
(2.35)

By substituting Eqs. (2.24)-(2.28) and Eqs. (2.30)-(2.31) to Eqs. (2.18), (2.19), (2.20), (2.21) and (2.22), the above EM algorithm can be rewritten as the following practical expressions:

**EM Algorithm in Gaussian Graphical Model (2)**

**Step 1** Input a given data point \(d\) and \(t \leftarrow 0\). Set \(\alpha(t), \beta(t)\) and \(\gamma(t)\) as initial values.

**Step 2** Update \(t\) by \(t \leftarrow t + 1\) and compute \(u, v\) and \(\beta(t)\) from \(\alpha(t-1), \beta(t-1)\) and \(\gamma(t-1)\) as follows:
\[
u \leftarrow \left(\frac{1}{|\mathcal{E}|} \text{Tr} \left(C((\beta(t-1) + \gamma(t-1))I + \alpha(t-1)C)^{-1}\right)\right)^{-1} \beta(t-1)^2 d C((\beta(t-1) + \gamma(t-1))I + \alpha(t-1)C)^{-1} d^T
\]
(2.36)
\[
u \leftarrow \left(\frac{1}{|\mathcal{E}|} \text{Tr} \left(C((\beta(t-1) + \gamma(t-1))I + \alpha(t-1)C)^{-1}\right)\right)^{-1} \beta(t-1)^2 d C((\beta(t-1) + \gamma(t-1))I + \alpha(t-1)C)^{-1} d^T
\]
(2.37)
\[
\beta(t) \leftarrow \left(\frac{1}{|\mathcal{V}|} \text{Tr} \left(C((\beta(t-1) + \gamma(t-1))I + \alpha(t-1)C)^{-1}\right)\right)^{-1} + \frac{1}{|\mathcal{V}|} \beta(t-1)^2 d C((\beta(t-1) + \gamma(t-1))I + \alpha(t-1)C)^{-1} d^T
\]
(2.38)

**Step 3** Repeat the following updates of \(\alpha(t)\) and \(\gamma(t)\) until they converge:
\[
\alpha(t) \leftarrow \alpha(t) \left(\frac{1}{|\mathcal{E}|} \text{Tr} \left(C((\gamma(t)I + \alpha(t)C)^{-1}\right)\right)^{1/2}
\]
(2.39)
\[
\gamma(t) \leftarrow \gamma(t) \left(\frac{1}{|\mathcal{V}|} \text{Tr} \left(C((\gamma(t)I + \alpha(t)C)^{-1}\right)\right)^{1/2}
\]
(2.40)

**Step 4** Set \(\tilde{\alpha} \leftarrow \alpha(t), \tilde{\beta} \leftarrow \beta(t), \tilde{\gamma} \leftarrow \gamma(t)\), and
\[
\tilde{\alpha}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \leftarrow ((\beta(U) + \gamma)I + \alpha C)^{-1} d.
\]
(2.41)

Stop if \(\tilde{\alpha}, \tilde{\beta}\) and \(\tilde{\gamma}\) converge, and go to Step 2 otherwise.

Now we consider a \(M \times N\) square grid graph with periodic boundary conditions for the abscissa and the ordinate as \(G = (V, E)\) as shown in Fig.1. In our square grid graph, the position vector \(r_i\) of each node \(i\) is assigned by
\[
r_i \equiv (i - 1) \text{mod}(M), -\left\lfloor \frac{i - 1}{M} \right\rfloor.
\]
(2.42)

where \(|V| = MN\) and \(\lfloor \cdot \rfloor\) is a floor function. When the abscissa and the ordinate of \(r_i\) are denoted by \(m\) and \(n\), such that \(r_i = (m, n)\), the periodic boundary conditions for the abscissa and the ordinate means that \(m = M\) and \(n = N\) are
interpreted as $m = 0$ and $n = 0$, respectively. We introduce $MN \times MN$ unitary matrix $U$ and its conjugate matrix $U^\dagger$ as follows:

\[
\langle m, n | U | k, l \rangle \equiv \frac{1}{\sqrt{MN}} \exp \left( -i \frac{2\pi km}{M} - i \frac{2\pi ln}{N} \right) \tag{2.43}
\]

\[
\langle k, l | U^\dagger | m, n \rangle \equiv \frac{1}{\sqrt{MN}} \exp \left( i \frac{2\pi km}{M} + i \frac{2\pi ln}{N} \right) \tag{2.44}
\]

We can confirm the following relations for $U$ and $U^\dagger$:

\[
\sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \langle m, n | U | k, l \rangle \langle k, l | U^\dagger | m', n' \rangle = \delta_{m,m'} \delta_{n,n'} \tag{2.45}
\]

\[
\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle k, l | U^\dagger | m, n \rangle \langle m, n | U | k', l' \rangle = \delta_{k,k'} \delta_{l,l'} \tag{2.46}
\]

In order to derive Eqs. (2.46) and (2.46) we use the following calculation:

\[
\sum_{k=0}^{M-1} \exp \left( -i \frac{2\pi km}{M} \right) = \frac{1 - \exp \left( -i \frac{2\pi (m - m')}{M} \right)}{1 - \exp \left( -i \frac{2\pi}{M} \right)} = 0 \tag{2.47}
\]

Eqs. (2.45) and (2.45) means that $UU^\dagger = U^\dagger U = I$, where $I$ is an $MN \times MN$ unit matrix, such that we have $C^\dagger = C^{-1}$. Now we denote the components of the matrices $C$ and $I$ by two-dimensional representations as

\[
\langle m, n | C | m', n' \rangle \equiv 4 \delta_{m', m} \delta_{n', n} - \delta_{m', m-1} \delta_{n', n} - \delta_{m', m+1} \delta_{n', n} - \delta_{m', m} \delta_{n', n-1} - \delta_{m', m} \delta_{n', n+1} \tag{2.48}
\]

\[
\langle m, n | I | m', n' \rangle \equiv \delta_{m', m} \delta_{n', n} (m, m' = 0, 1, \ldots, M - 1; n, n' = 0, 1, \ldots, N - 1) \tag{2.49}
\]

We can confirm the following equality:

\[
\langle k, l | U^\dagger C U | k', l' \rangle = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \langle k, l | U^\dagger | m, n \rangle \langle m, n | C | m', n' \rangle \langle m', n' | U | k', l' \rangle \]

\[
= \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} 4 \delta_{m', m} \delta_{n', n} - \delta_{m', m-1} \delta_{n', n} - \delta_{m', m+1} \delta_{n', n} - \delta_{m', m} \delta_{n', n-1} - \delta_{m', m} \delta_{n', n+1} \tag{2.50}
\]
\[
\begin{align*}
- \delta_{m',m} \delta_{n',n-1} - \delta_{m',m} \delta_{n',n+1}
\times \exp \left( i \frac{2 \pi k m}{M} + i \frac{2 \pi l n}{N} \right) 
\times \exp \left( - i \frac{2 \pi k' m'}{M} - i \frac{2 \pi l' n'}{N} \right) 
= \left( 4 - \exp \left( i \frac{2 \pi k}{M} \right) \right) 
\times \left( \frac{1}{N} \sum_{n'=0}^{N-1} \exp \left( \frac{2 \pi (l-l') n'}{N} \right) \right) 
\times \frac{1}{M} \sum_{m'=0}^{M-1} \exp \left( i \frac{2 \pi (k-k') m'}{M} \right) 
\times \left( \frac{1}{N} \sum_{n'=0}^{N-1} \exp \left( \frac{2 \pi (l-l') n'}{N} \right) \right) 
= \delta_{k,k'} \delta_{l,l'} \left( 4 - 2 \cos \left( \frac{2 \pi k}{M} \right) - 2 \cos \left( \frac{2 \pi l}{N} \right) \right).
\end{align*}
\]

It means that \( C \) has been diagonalized by the unitary matrix \( U \) as follows:

\[
C = U \Lambda U^T,
\]

where

\[
\langle k,l|A|k',l' \rangle \equiv \delta_{k,k'} \delta_{l,l'} \lambda(k,l),
\]

\[
\lambda(k,l) \equiv \delta_{k,k'} \delta_{l,l'} \left( 4 - 2 \cos \left( \frac{2 \pi k}{M} \right) - 2 \cos \left( \frac{2 \pi l}{N} \right) \right).
\]

The marginal likelihood in Eq. (2.35) is reduced by

\[
\frac{1}{MN} \ln \left( P(d|\alpha,\beta,\gamma) \right) = \frac{1}{2} \ln (2\pi) + \frac{1}{2MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \ln \left( \gamma + \alpha \lambda(k,l) \right) 
- \frac{1}{2MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \ln \left( \beta + \gamma + \alpha \lambda(k,l) \right) 
- \frac{1}{2MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \frac{\beta(t-1)^2 \lambda(k,l)}{\left( \beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l) \right)^2} 
- \frac{1}{2MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \frac{\beta(t-1)^2 \lambda(k,l)}{\left( \beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l) \right)^2}.
\]

Moreover, Eqs. (2.36)-(2.40) and Eq. (2.41) are also replaced to the following useful representations for practical computations:

\[
u \left( \begin{array}{c}
\lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l)
\end{array} \right)
\]

\[
u \left( \begin{array}{c}
\lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l)
\end{array} \right)
\]

\[
\beta(t) \left( \begin{array}{c}
\lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l)
\end{array} \right)
\]

\[
\alpha(t) \left( \begin{array}{c}
\lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l) \\
\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda(k,l)
\end{array} \right)
\]
Eqs. (2.55)-(2.59) are only for the discrete Fourier transformation of data vector \( \mathbf{d} \). Such reductions of computational time come from the spatially translational symmetries of square grid graphs with periodic boundary conditions.

Standard images which are used as original images in our numerical experiments of probabilistic image processing are shown in Fig. 2.

Fig. 3. Degraded images \( \mathbf{d} \) generated by according to an additive white Gaussian noise with the average 0 and the variance \( \sigma^2 = 40^2 \) from the standard images of Fig. 2. (a) Lena. (b) Mandrill. (c) Pepper.

\[
\begin{align*}
\gamma(t) &= \gamma(t) \left( \frac{1}{2MN} \sum_{k=0}^{M-1N-1} \sum_{l=0}^{N-1} \gamma(i-1) + \alpha(i-1) \lambda(k,l) \right)^{1/2}, \\
\hat{u}(\alpha, \beta, \gamma \mathbf{d}) &= \frac{1}{\sqrt{MN}} \sum_{k=0}^{M-1N-1} \sum_{l=0}^{N-1} \left( \alpha + \gamma \lambda(k,l) \right) \\
&\times \left( \cos \left( \frac{2\pi k}{M} + \frac{2\pi l}{N} \right) \Re\left( \gamma(k,l) \right) \sin \left( \frac{2\pi k}{M} + \frac{2\pi l}{N} \right) \Im\left( \gamma(k,l) \right) \right) \\
&\quad \left( m, n \rightarrow \left( i-1 \mod(M), \left\lfloor \frac{i-1}{M} \right\rfloor \right) \right).
\end{align*}
\] (2.60)

These representations are very useful for implementing the procedure as a practical computational program. Actually, although Eqs. (2.36)-(2.40) needs computations of inverse matrices of \( |V| \times |V| \), the massive parts of computations in Eqs. (2.55)-(2.59) are only for the discrete Fourier transformation of data vector \( \mathbf{d} \). Such reductions of computational time come from the spatially translational symmetries of square grid graphs with periodic boundary conditions.

Fig. 3 は本節で用いる劣化画像である。Degraded images \( \mathbf{d} \) which are generated by according to an additive white Gaussian noise with the average 0 and the variance \( \sigma^2 = 40^2 \) are shown in Fig. 2. The additive white Gaussian noise \( n_i \) is according to the following probability density function:

\[
g(n_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{1}{2\sigma^2} n_i^2 \right) \quad (i \in V, n_i \in (-\infty, +\infty), \sigma > 0)
\] (2.61)

and \( n_i \) is generated independently at each pixel \( i \). By adding the generated random number vector \( \mathbf{n} = (n_1, n_2, \cdots, n_{|V|}) \) to the original image \( \mathbf{a} \), such that \( \mathbf{d} = \mathbf{a} + \mathbf{n} \), degraded images \( \mathbf{d} \) are generated. We know that the uniform distribution \( h(a) \equiv \frac{1}{\{a \in [0, 1]\}} \) has an average \( \frac{1}{2} \) and a variance \( \frac{1}{12} \). If we generate \( N \) random numbers \( \rho_1, \rho_2, \cdots, \rho_N \) from the uniform distribution \( h(a) \), the central limit theorem mention that \( \frac{1}{N} (\rho_1 + \rho_2 + \cdots + \rho_N) - \frac{1}{2} \) follows the gauss distribution with an average \( \frac{1}{2} \) and a variance \( \frac{1}{12N} \), and then \( \sigma \sqrt{12N} \left( \frac{1}{N} (\rho_1 + \rho_2 + \cdots + \rho_N) - \frac{1}{2} \right) \) follows the gauss distribution with an average \( 0 \) and a variance \( \sigma^2 \) for the limit \( N \to +\infty \). By using these statistical properties, we generate Gaussian random numbers \( n_i (i \in V) \) independently at each pixel. In Table 1, we show the mean square error:

\[
\text{MSE}(\mathbf{a}, \mathbf{b}) \equiv \frac{1}{|V|} \sum_{i \in V} (a_i - d_i),
\] (2.62)
and the signal to noise ratio

$$\text{SNR}(a,b) \equiv 10\log_{10} \left( \frac{\frac{1}{|V|} \sum_{i \in V} (a_i - \overline{a})^2}{\frac{1}{|V|} \sum_{i \in V} (d_i - \hat{d}_i)^2} \right) \text{ (dB)},$$

(2.63)

between the original image $a$ and the degraded image $d$. Here $\overline{a}$ is defined by

$$\overline{a} \equiv \frac{1}{|V|} \sum_{i \in V} (a_i).$$

(2.64)

Restored images $\hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}|d)$ obtained by applying the expectation-maximization algorithm for the Gaussian graphical model on the square grid graph to the degraded images $d$ in Fig.3 are shown in Fig.4. In Table 2, we show estimates $\hat{\alpha}$, $\hat{\beta} = 1/\sqrt{\hat{\beta}}$, $\hat{\gamma}$ and the logarithm of marginal likelihood per pixel $\ln f(d, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$, the mean square error between the original image and the estimated image $\text{MSE}(a, \hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}|d))$, and the signal to noise ratio $\text{SNR}(a, \hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}|d))$.

Table 1. Mean square error $\text{MSE}(a,d)$ and the signal to noise ratio $\text{SNR}(a,d)$ between the standard image $a$ in Fig. 2 and the degraded image $d$ in Fig. 3.

<table>
<thead>
<tr>
<th></th>
<th>Lena</th>
<th>Mandrill</th>
<th>Pepper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{MSE}(a,b)$</td>
<td>1406.90</td>
<td>1514.35</td>
<td>1446.55</td>
</tr>
<tr>
<td>$\text{SNR}(a,b)$</td>
<td>2.887 (dB)</td>
<td>−0.075 (dB)</td>
<td>2.937 (dB)</td>
</tr>
</tbody>
</table>

Table 2. Estimates $\hat{\alpha}$, $\hat{\beta} = 1/\sqrt{\hat{\beta}}$, $\hat{\gamma}$ and the logarithm of marginal likelihood per pixel $\ln f(d, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$, the mean square error between the original image and the estimated image $\text{MSE}(a, \hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}|d))$, and the signal to noise ratio $\text{SNR}(a, \hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}|d))$ obtained by applying the expectation-maximization algorithm for the Gaussian graphical model on the square grid graph to the degraded image $d$ in Fig.3. (a) Lena. (b) Mandrill. (c) Pepper.

<table>
<thead>
<tr>
<th></th>
<th>Lena</th>
<th>Mandrill</th>
<th>Pepper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>0.000522</td>
<td>0.000765</td>
<td>0.000504</td>
</tr>
<tr>
<td>$\hat{\beta} = 1/\sqrt{\hat{\beta}}$</td>
<td>31.985</td>
<td>37.982</td>
<td>31.901</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>0.0000000135</td>
<td>0.0000000470</td>
<td>0.0000000119</td>
</tr>
<tr>
<td>$\ln f(d, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$</td>
<td>−10.3045</td>
<td>−10.4294</td>
<td>−10.3140</td>
</tr>
<tr>
<td>$\text{MSE}(a, \hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}</td>
<td>d))$</td>
<td>305.49</td>
<td>315.00</td>
</tr>
<tr>
<td>$\text{SNR}(a, \hat{a}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}</td>
<td>d))$</td>
<td>9.520 (dB)</td>
<td>6.743 (dB)</td>
</tr>
</tbody>
</table>

Before closing this subsection, we consider the limit of $\varepsilon \to 0$ after setting $\gamma = \alpha \varepsilon$. The extremum conditions $\frac{\partial}{\partial \alpha} P(d|\alpha, \beta, \alpha \varepsilon) = 0$ and $\frac{\partial}{\partial \beta} P(d|\alpha, \beta, \alpha \varepsilon) = 0$ in the limit of $\varepsilon \to 0$ are derived as

$$\alpha = \left( \frac{1}{|E|} \text{Tr} \left( \beta C^{-1} \right) + \frac{1}{|E|} \langle \beta I + \alpha C \rangle^{-1} \beta I \rangle^2 d^2 \right)^{-1},$$

(2.65)
\[
\beta = \left( \frac{1}{|V|} \text{Tr} \left( (\beta I + \alpha C)^{-1} \right) \right) + \frac{1}{|V|} \beta d \alpha^2 (\beta I + \alpha C)^{-1} d t^{-1}. \tag{2.66}
\]

3 Belief Propagation

We consider a probability distribution \( P(a_1, a_2, a_3, a_4) \) defined by

\[
P(a_1, a_2, a_3, a_4) = \frac{1}{Z} W_{(1,2)}(a_1, a_2) W_{(1,3)}(a_1, a_3) W_{(1,4)}(a_1, a_4), \tag{3.1}
\]

and explain how to construct an explicit belief propagation procedure to compute marginal probability distributions:

\[
P_i(a_i) = \sum_{z_i \in \Omega_i} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \delta_{a_i, z_i} P(z_1, z_2, z_3, z_4) \quad (i \in \{1, 2, 3, 4\}), \tag{3.3}
\]

and

\[
P_{(1,i)}(a_1, a_i) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \delta_{a_1, z_1} \delta_{a_i, z_i} P(z_1, z_2, z_3, z_4) \quad (i \in \{1, 2, 3, 4\}). \tag{3.4}
\]

It is valid that

\[
\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_{(1,2)}(a_1, z_2) W_{(1,3)}(a_1, z_3) W_{(1,4)}(a_1, z_4) = \prod_{k \in \{2, 3, 4\} \setminus \{i\}} \left( \sum_{z_k \in \Omega_k} W_{(1,k)}(a_1, z_k) \right), \tag{3.5}
\]

and

\[
\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \delta_{a_1, z_1} W_{(1,2)}(z_1, z_2) W_{(1,3)}(z_1, z_3) W_{(1,4)}(z_1, z_4) = \sum_{z_1 \in \Omega_1} W_{(1,1)}(z_1, a_i) \prod_{k \in \{2, 3, 4\} \setminus \{i\}} \left( \sum_{z_k \in \Omega_k} W_{(1,k)}(z_1, z_k) \right). \tag{3.6}
\]

From these equalities, the marginal probability distributions in Eqs.(3.3) and (3.4) are reduced to the following representations for the belief propagation:

\[
P_i(a_i) = \frac{M_{2-i}(a_i) M_{3-i}(a_i) M_{4-i}(a_i)}{\sum_{z_i \in \Omega_i} M_{1-i}(z_i)}, \tag{3.7}
\]

\[
P_{(1,i)}(a_1, a_i) = \frac{\left( \prod_{k \in \{2, 3, 4\} \setminus \{i\}} M_{k-i}(a_i) \right) W_{(1,i)}(a_1, a_i)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \left( \prod_{k \in \{2, 3, 4\} \setminus \{i\}} M_{k-i}(z_1) \right) W_{(1,i)}(z_1, z_i)} \tag{3.9}
\]

where

\[
M_{1-i}(a_1) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \delta_{a_1, z_1} W_{(1,i)}(z_1, z_i), \tag{3.10}
\]

\[
M_{1-i}(a_i) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \delta_{a_i, z_i} W_{(1,2)}(z_1, z_2)
\times W_{(1,3)}(z_1, z_3) W_{(1,4)}(z_1, z_4). \tag{3.11}
\]
In the belief propagation, $M_{i ightarrow j}(a_j)$ is referred to as a message from a node $i$ to a node $j$. The message in Eq. (3.11) is rewritten as

$$ M_{i ightarrow j}(a_j) = \sum_{z_i \in \Omega_i} \left( \prod_{k \in \{2,3,4\} \setminus \{i\}} M_{k ightarrow i}(z_i) \right) W_{i,j}(z_i, a_j). \quad (3.12) $$

In Eq. (3.12), we see that the out-going message $M_{i ightarrow j}(a_j)$ from the node 1 to the node $i$ is represented in terms of all the in-coming messages $M_{k ightarrow i}(z_i)$ ($k \in \{2,3,4\} \setminus \{i\}$) to the node 1 except the message $M_{j ightarrow i}(a_1)$ from the node 1 to the node 1. Eq. (3.12) can be regarded as one of recursion formulas for messages in the probability distribution (3.1). The practical procedures to compute the marginal probability distributions in Eqs. (3.3) and (3.4) are given as follows:

### Procedures for Computations of Marginals by Belief Propagation on Tree Graph

**Step 1:** Compute messages from the node 1 to the $i$ ($i = 2, 3, 4$):

$$ M_{1,i}(a_i) \equiv \sum_{z_1 \in \Omega_1} \sum_{i \in \Omega_i} \delta_{a_i,z_i} W_{1,i}(z_1, a_i) \quad (i \in \{2,3,4\}). \quad (3.13) $$

**Step 2:** Compute messages from the node 1 to the $i$ ($i = 2, 3, 4$):

$$ M_{i,j}(a_j) \equiv \sum_{z_i \in \Omega_i} \left( \prod_{k \in \{2,3,4\} \setminus \{i\}} M_{k ightarrow i}(z_i) \right) W_{i,j}(z_i, a_i) \quad (i \in \{2,3,4\}). \quad (3.14) $$

**Step 3:** Compute marginal probability distributions:

$$ P_i(a_i) \equiv \frac{M_{1,i}(a_i) M_{i,j}(a_j) M_{j,1}(a_1)}{\sum_{z_j \in \Omega_j} \sum_{z_i \in \Omega_i} \left( \prod_{k \in \{2,3,4\} \setminus \{i\}} M_{k ightarrow i}(z_i) \right) W_{i,j}(z_i, a_i) }. \quad (3.15) $$

$$ P_i(a_i) \equiv \frac{M_{i,j}(a_j)}{\sum_{z_j \in \Omega_j} \sum_{z_i \in \Omega_i} \left( \prod_{k \in \{2,3,4\} \setminus \{i\}} M_{k ightarrow i}(z_i) \right) W_{i,j}(z_i, a_i) }. \quad (3.16) $$

$$ P_{1,i}(a_1, a_i) \equiv \frac{M_{1,i}(a_i)}{\sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} \left( \prod_{k \in \{2,3,4\} \setminus \{i\}} M_{k ightarrow i}(z_i) \right) W_{i,j}(z_i, a_i) }. \quad (3.17) $$

We can regard the probability distribution (3.1) as a graphical representation in terms of the graph consisting of the set of nodes, $V = \{2,3,4\}$, and the set of edges, $E = \{\{1,2\}, \{1,3\}, \{1,4\}\}$.

We consider a probability distribution $P(a)$ defined by

$$ P(a) \equiv \frac{1}{Z} \left( \prod_{i \in V} W_i(a_i) \right) \left( \prod_{(i,j) \in E} W_{i,j}(a_i, a_j) \right), \quad (3.18) $$

$$ Z \equiv \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} \left( \prod_{i \in V} W_i(z_i) \right) \left( \prod_{(i,j) \in E} W_{i,j}(z_i, z_j) \right). \quad (3.19) $$

$V$ is the set of all the nodes and $E$ is the set of all the edges. In the statistical-mechanical informatics, $Z$ is referred to as a **Partition Function** and

$$ F \equiv -\ln(Z), \quad (3.20) $$

corresponds to a **Free Energy** of the graphical model in Eq. (3.18).

Let us suppose that the graph $G = (V, E)$ is a connected tree. and consider a probability distribution $P(a)$ defined by Eqs. (3.18) and (3.19) on the connected tree. For every node $i \in V$, we can divide the tree graph $G$ to $\partial i$ subgraphs in which the node $i$ is a leaf. The subgraph of $G$ which satisfy the following properties is denoted by the notation $G(i, \{i,j\})$ ($i, j \in \partial i, i \in V$):

1. $G(i, \{i,j\})$ is a connected tree.
2. $G(i, \{i,j\}) \cap G(i, \{k,l\}) = i \in V$ ($i, j \in \partial i \setminus \{k\}, i, k \in \partial i \setminus \{j\}$).
3. $G(i, \{i,j\}) \cap G(j, \{i,j\}) = \{i,j\} \in E$. 

17
In the subgraph $G(\{i, j\})$, the node $i$ is a leaf and the edge $\{i, j\}$ is a leaf edge. The set of all the nodes belonging to the subgraph $G(\{i, j\})$ is denoted by $V(\{i, j\})$. We introduce messages $M_{(i,j) \rightarrow (a_i)}$ and $M_{(i,j) \rightarrow (a_j)}$ for every edge $\{i, j\} \in E$:

$$M_{(i,j) \rightarrow (a_i)} = \sum_{z_{V(\{i,j\})} \in \Omega_{V(\{i,j\})}} \delta_{a_i, z_i} \left( \prod_{k \in V(\{i,j\}) \setminus \{i\}} W_k(z_k) \right) \times \left( \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i\}} W_{k,l}(z_k, z_l) \right) \quad \forall \{i, j\} \in E, a_i \in \Omega_i, \quad (3.21)$$

$$M_{(i,j) \rightarrow (a_j)} = \sum_{z_{V(\{i,j\})} \in \Omega_{V(\{i,j\})}} \delta_{a_j, z_j} \left( \prod_{k \in V(\{j\}) \setminus \{j\}} W_k(z_k) \right) \times \left( \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{j\}} W_{k,l}(z_k, z_l) \right) \quad \forall \{i, j\} \in E, a_j \in \Omega_j, \quad (3.22)$$

where $\sum_{z_{V(\{i,j\})}}$ is the summation with respect to the state variables $z_k$ for all $k \in V(\{i, j\})$ over all the possible configurations. For the probability distribution $P(a)$ in Eq. (5.1), the marginal probability distributions of the state variable $a_i$ and the state vector $(a_i, a_j)$ are expressed in terms of the messages $\{M_{j \rightarrow i}(a_i) \mid \{i, j\} \in \partial i, i \in V\}$ as

$$P_i(a_i) = \frac{1}{Z_i} W_i(a_i) \prod_{\{k,l\} \in \partial i \setminus \{i\}} M_{k \rightarrow i}(a_i) \quad \forall i \in V, \quad (3.23)$$

$$P_{\{i,j\}}(a_i, a_j) = \frac{1}{Z_{\{i,j\}}} \left( \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i,j\}} M_{\{k,l\} \rightarrow \{a_i, a_j\}} \right) W_i(a_i) W_{\{i,j\}}(a_i, a_j) \times W_{\{i,j\}}(a_j) \left( \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i,j\}} M_{\{k,l\} \rightarrow \{a_i, a_j\}} \right) \quad \forall \{i, j\} \in E, \quad (3.24)$$

where

$$Z_i = \sum_{z_{V(\{i,j\})}} W_i(a_i) \prod_{\{k,l\} \in \partial i \setminus \{i\}} M_{k \rightarrow i}(z_i) \quad \forall i \in V, \quad (3.25)$$

$$Z_{\{i,j\}} = \sum_{z_{V(\{i,j\})}} \left( \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i,j\}} M_{\{k,l\} \rightarrow \{z_i, z_j\}} \right) W_i(a_i) W_{\{i,j\}}(z_i, z_j) \times W_{\{i,j\}}(a_j) \left( \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i,j\}} M_{\{k,l\} \rightarrow \{z_i, z_j\}} \right) \quad \forall \{i, j\} \in E. \quad (3.26)$$

From the definitions (3.21) and (3.22), we can derive the following recursion formulas for messages $\{M_{(i,j) \rightarrow (a_i)} \mid \{i, j\} \in \partial i, i \in V\}$ and $\{M_{(i,j) \rightarrow (a_j)} \mid \{i, j\} \in \partial j, i \in V\}$:

$$M_{(i,j) \rightarrow (a_i)} = \sum_{z_{V(\{i,j\})}} W_{\{i,j\}}(a_i, z_j) \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i,j\}} M_{\{k,l\} \rightarrow \{z_j\}} \quad \forall \{i, j\} \in E, \{i, j\} \in \partial j, \quad (3.27)$$

$$M_{(i,j) \rightarrow (a_j)} = \sum_{z_{V(\{i,j\})}} W_{\{i,j\}}(z_i, a_j) \prod_{\{k,l\} \in \partial V(\{i,j\}) \setminus \{i,j\}} M_{\{k,l\} \rightarrow \{z_i\}} \quad \forall \{i, j\} \in E, \{i, j\} \in \partial i. \quad (3.28)$$

The set of recursion formulas in Eqs. (3.27)-(3.28) is referred to as a message passing rule in the belief propagation. By using the expression of marginals in terms of the messages, the probability distribution $P(a)$ is expressed in terms of the marginals as follows:

$$P(a) = \left( \prod_{i \in V} P_i(a_i) \right) \left( \prod_{\{i,j\} \in E} \frac{P_{\{i,j\}}(a_i, a_j)}{P_i(a_i) P_j(a_j)} \right). \quad (3.29)$$

Next, we consider the probability distribution of Eqs. (3.18) and (3.19) on a graph $G = (V, E)$ with some cycles. In
such situations, Eq. (3.29) is not always valid. We introduce the following trial probability distribution

$$Q(a) = \left( \prod_{i \in V} Q_i(a_i) \right) \left( \prod_{\{i,j\} \in E} \frac{Q_{\{i,j\}}(a_i, a_j)}{Q_i(a_i)Q_j(a_j)} \right),$$

(3.30)

where

$$Q_i(a_i) = \sum_z \delta_{a_i,z} Q(z),$$

(3.31)

$$Q_{\{i,j\}}(a_i, a_j) = \sum_z \delta_{a_i,z} Q(z),$$

(3.32)

By using Eqs. (3.30), (3.31) and (3.32), the KL divergence $KL[P||Q]$ for $P(a)$ in Eqs. (3.18)-(3.19) on a graph $G = (V, E)$ with cycles can be written down to

$$KL[P||Q] = -\sum_{i \in V} \sum_{z \in \Omega_i} Q_i(z_i) \ln(W_i(z_i))$$

$$- \sum_{\{i,j\} \in E} \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} Q_{\{i,j\}}(z_i, z_j) \ln(W_{\{i,j\}}(z_i, z_j))$$

$$+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i))$$

$$+ \sum_{\{i,j\} \in E} \sum_{\{z_i \in \Omega_i, z_j \in \Omega_j\}} Q_{\{i,j\}}(z_i, z_j) \ln(Q_{\{i,j\}}(z_i, z_j))$$

$$+ \ln \left( \sum_z \left( \prod_{i \in V} W_i(z_i) \right) \left( \prod_{\{i,j\} \in E} W_{\{i,j\}}(z_i, z_j) \right) \right),$$

(3.33)

so that we have

$$KL[P||Q] = \mathcal{F}_{\text{Bethe}}[\{Q_i|_{i \in V}\}, \{Q_{\{i,j\}}|_{\{i,j\} \in E}\}] - F,$$

(3.34)

where

$$\mathcal{F}_{\text{Bethe}}[\{Q_i|_{i \in V}\}, \{Q_{\{i,j\}}|_{\{i,j\} \in E}\}] = \sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i))$$

$$- \sum_{\{i,j\} \in E} \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} Q_{\{i,j\}}(z_i, z_j) \ln(W_{\{i,j\}}(z_i, z_j))$$

$$+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i))$$

$$+ \sum_{\{i,j\} \in E} \sum_{\{z_i \in \Omega_i, z_j \in \Omega_j\}} Q_{\{i,j\}}(z_i, z_j) \ln(Q_{\{i,j\}}(z_i, z_j)).$$

(3.35)

We remark that $\mathcal{F}_{\text{Bethe}}[\{Q_i|_{i \in V}\}, \{Q_{\{i,j\}}|_{\{i,j\} \in E}\}]$ is often referred to as Bethe Free Energy Functional.

We consider to determine the marginal probability distributions $\{Q_i(a_i)|_{a_i \in \Omega_i, i \in V}\}$ and $\{Q_{\{i,j\}}(a_i, a_j)|_{a_i \in \Omega_i, a_j \in \Omega_j, {\{i,j\} \in E}}$ so as to minimize the KL divergence $KL[P||Q]$ under the following constraint conditions

$$Q_i(a_i) = \sum_{a_j \in \Omega_j} Q_{\{i,j\}}(a_i, a_j) \ (\{i, j\} \in E),$$

(3.36)

$$Q_j(a_j) = \sum_{a_i \in \Omega_i} Q_{\{i,j\}}(a_i, a_j) \ (\{i, j\} \in E),$$

(3.37)

$$\sum_{a_i \in \Omega_i} Q_{\{i,j\}}(a_i, a_j) = 1 \ (\{i, j\} \in E),$$

(3.38)

$$\sum_{a_i \in \Omega_i} Q(a_i) = 1 \ (i \in V).$$

(3.39)
The set of approximate marginal probability distributions, \( \{ \hat{Q}_i(a_i) \}_{i \in V, a_i \in \Omega_i} \), and 
\( \{ \hat{Q}_{(i,j)}(a_i, a_j) \}_{\{i,j\} \in E, a_i \in \Omega_i, a_j \in \Omega_j} \), in the loopy belief propagation is determined so as to minimize the KL divergence KL\[P||Q\] such that the Bethe free energy functional \( \mathcal{F}_{\text{Bethe}}\left[ \{ Q_i \}_{i \in V}, \{ Q_{(i,j)} \}_{\{i,j\} \in E} \right] \) under the normalization conditions and the reducibility conditions of each trial marginal probability distributions. To ensure the reducibilities (3.36)-(3.37) and the normalizations (3.38)-(3.39), we introduce the Lagrange multipliers \( \hat{\lambda}_{i,j}(a_i), \hat{\lambda}_{j,i}(a_i), \lambda_{i,j} \) as follows:

\[
\mathcal{L}\left[ \{ Q_i \}_{i \in V}, \{ Q_{(i,j)} \}_{\{i,j\} \in E} \right] = - \sum_{i \in V} \sum_{a_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i)) \\
+ \sum_{\{i,j\} \in E} \sum_{a_i \in \Omega_i} Q_{(i,j)}(z_i, z_j) \ln(W_{(i,j)}(z_i, z_j)) \\
+ \sum_{i \in V} \left( 1 - |\partial_i| \right) \sum_{a_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i)) \\
+ \sum_{\{i,j\} \in E} \sum_{a_i \in \Omega_i} \lambda_{i,j}(z_i) \left( Q_i(z_i) - \sum_{z_j \in \Omega_j} Q_{(i,j)}(z_i, z_j) \right) \\
+ \sum_{\{i,j\} \in E} \left( \lambda_{i,j} + 1 \right) \left( \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} Q_{(i,j)}(z_i, z_j) - 1 \right) \\
+ \sum_{i \in V} \left( \lambda_i + 1 \right) \left( \sum_{z_i \in \Omega_i} Q_i(z_i) - 1 \right)
\]

(3.40)

By taking the first derivatives of \( \mathcal{L}\left[ \{ Q_i \}_{i \in V}, \{ Q_{(i,j)} \}_{\{i,j\} \in E} \right] \) with respect to \( Q(a_i) \) and \( Q_{(i,j)}(a_i, a_j) \), the extremum conditions

\[
\left[ \frac{\partial}{\partial Q_k(a_k)} \mathcal{L}\left[ \{ Q_i \}_{i \in V}, \{ Q_{(i,j)} \}_{\{i,j\} \in E} \right] \right]_{\{Q_i = \hat{Q}_i \}_{i \in V}, \{Q_{(i,j)} = \hat{Q}_{(i,j)} \}_{\{i,j\} \in E}} = 0 \\
(k \in V),
\]

(3.41)

\[
\left[ \frac{\partial}{\partial Q_{(k,l)}(a_k, a_l)} \mathcal{L}\left[ \{ Q_i \}_{i \in V}, \{ Q_{(i,j)} \}_{\{i,j\} \in E} \right] \right]_{\{Q_i = \hat{Q}_i \}_{i \in V}, \{Q_{(i,j)} = \hat{Q}_{(i,j)} \}_{\{i,j\} \in E}} = 0 \\
(\{k,l\} \in E),
\]

(3.42)

are reduce to

\[
\hat{Q}_i(a_i) = W_i(a_i) \exp\left( \lambda_i + \hat{\lambda}_{i,j}(a_i) \right) (i \in V),
\]

\[
\hat{Q}_{(i,j)}(a_i, a_j) = W_{(i,j)}(a_i, a_j) \times \exp\left( \lambda_{i,j} + \hat{\lambda}_{i,j}(a_i) + \hat{\lambda}_{j,i}(a_j) \right) (\{i,j\} \in E),
\]

(3.43)

\[
\lambda_{i,j}(a_i) = \frac{1}{|\partial_i| - 1} \sum_{\{k,l\} \in \partial i, |\partial i| \geq 2} \lambda_{i,j}(a_i) (i \in V, |\partial i| \geq 2),
\]

(3.44)

\[
\lambda_{i,j}(a_i) = \lambda_{i,(k,l)}(a_i) = 0 (i \in V, \{k,l\} \in \partial_i, |\partial i| = 1),
\]

(3.45)

\[
\lambda_q = -\ln\left( \sum_{z_i \in \Omega_i} \exp\left( \hat{\lambda}_{i,j}(z_i) \right) \right) (\{i,j\} \in E),
\]

(3.46)

\[
\lambda_{i,j} = -\ln\left( \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} W_{(i,j)}(z_i, z_j) \right)
\]

(3.47)
We introduce messages $\{\mathcal{M}_{(i,j)\rightarrow}(a_i)\}$ and $\{\mathcal{M}_{(i,j)\rightarrow}(a_j)\}$ \((\{i,j\}\in E)\) by

$$\exp\left(\lambda_{i,(i,j)}(a_i)\right) = W_i(a_i) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(a_i) \quad (\{i,j\}\in E, a_i\in \Omega_i).$$

By substituting Eqs. (3.49) and (3.50) to Eqs. (3.43) and (3.44), approximate expressions of marginal probabilities $\mathcal{Q}(a_i)$ and $\mathcal{Q}_{(i,j)}(a_i, a_j)$ are given as

$$\mathcal{Q}(a_i) = \sum_{z_i\in \Omega_i} W_i(z_i) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(z_i) \quad (i\in V),$$

$$\mathcal{Q}_{(i,j)}(a_i, a_j) = \frac{W_i(a_i) W_j(a_j) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(a_i) \prod_{(k,l)\in \partial(j,j)} \mathcal{M}_{(k,l)\rightarrow}(a_j)}{\sum_{z_i, z_j \in \Omega_i} W_i(z_i) W_j(z_j) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(z_i) \prod_{(k,l)\in \partial(j,j)} \mathcal{M}_{(k,l)\rightarrow}(z_j)} \quad (\{i,j\}\in E).$$

By substituting Eqs. (3.51) and (3.52) to

$$\mathcal{Q}(a_i) = \sum_{a_i \in \Omega_i} \mathcal{Q}_{(i,j)}(a_i, a_j) \quad (\{i,j\}\in E),$$

$$\mathcal{Q}(a_j) = \sum_{a_j \in \Omega_j} \mathcal{Q}_{(i,j)}(a_i, a_j) \quad (\{i,j\}\in E),$$

we can derive the following simultaneous recursion formulas to determine $\{\mathcal{M}_{(i,j)\rightarrow}(a_i)\}|\{i,j\}\in E, a_i\in \Omega_i|$ and $\{\mathcal{M}_{(i,j)\rightarrow}(a_j)\}|\{i,j\}\in E, a_j\in \Omega_j|$ as follows:

$$\mathcal{M}_{(i,j)\rightarrow}(a_i) = \left( \sum_{z_i\in \Omega_i} W_i(z_i) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(z_i) \right)^{-1} \times \left( \sum_{z_i, z_j \in \Omega_i} W_i(z_i) W_j(z_j) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(z_i) \prod_{(k,l)\in \partial(j,j)} \mathcal{M}_{(k,l)\rightarrow}(z_j) \right) \times \left( \sum_{z_j \in \Omega_j} W_j(z_j) \prod_{(k,l)\in \partial(j,j)} \mathcal{M}_{(k,l)\rightarrow}(z_j) \right) \quad (\{i,j\}\in E, i\in V).$$

$$\mathcal{M}_{(i,j)\rightarrow}(a_j) = \left( \sum_{z_j\in \Omega_j} W_j(z_j) \prod_{(k,l)\in \partial(j,j)} \mathcal{M}_{(k,l)\rightarrow}(z_j) \right)^{-1} \times \left( \sum_{z_i, z_j \in \Omega_i} W_i(z_i) W_j(z_j) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(z_i) \prod_{(k,l)\in \partial(j,j)} \mathcal{M}_{(k,l)\rightarrow}(z_j) \right) \times \left( \sum_{z_i \in \Omega_i} W_i(z_i) \prod_{(k,l)\in \partial(i,i)} \mathcal{M}_{(k,l)\rightarrow}(z_i) \right) \quad (\{i,j\}\in E).$$
Graphical representations in Fig. 5 are assigned to the messages $\mu_{(i,j)\rightarrow (a_j)}$ and $\mu_{(i,j)\rightarrow i(a_i)}$.

$$\times \sum_{z_i \in \Omega_i} \left( W_{(i,j)}(a_i, z_j) W_i(z_i) \prod_{(k,l) \in \partial_i \setminus \{i,j\}} \mathcal{M}_{(k,l)\rightarrow i(z_i)} \right) \left( \{i,j\} \in E \right).$$  \hspace{1cm} (3.56)

By introducing

$$\mu_{(i,j)\rightarrow i(a_i)} = \frac{\mathcal{M}_{(i,j)\rightarrow (a_i)}}{\sum_{\xi \in \Omega_i} \mathcal{M}_{(i,j)\rightarrow i(\xi)}} \left( \{i,j\} \in E, a_i \in \Omega_i \right),$$  \hspace{1cm} (3.57)

$$\mu_{(i,j)\rightarrow j(a_j)} = \frac{\mathcal{M}_{(i,j)\rightarrow j(a_j)}}{\sum_{\xi \in \Omega_j} \mathcal{M}_{(i,j)\rightarrow j(\xi)}} \left( \{i,j\} \in E, a_j \in \Omega_j \right),$$  \hspace{1cm} (3.58)

Eqs. (3.51)-(3.52) and Eqs. (3.55)-(3.56), can be rewritten as

$$\hat{Q}(a_i) = \frac{1}{Z_i} W_i(a_i) \prod_{\{k,l\} \in \partial i} \mu_{(k,l)\rightarrow i(a_i)} \left( \{i\} \in V, a_i \in \Omega_i \right),$$  \hspace{1cm} (3.59)

$$\hat{Q}_{(i,j)}(a_i, a_j) = \frac{1}{Z_{(i,j)}} W_{(i,j)}(a_i, a_j) W_j(a_j) \left( \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow i(a_i)} \right) \left( \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow j(a_j)} \right) \left( \{i,j\} \in E, a_i \in \Omega_i, a_j \in \Omega_j \right),$$  \hspace{1cm} (3.60)

$$\mu_{(i,j)\rightarrow i(a_i)} = \frac{\sum_{z_i \in \Omega_i} W_{(i,j)}(a_i, z_j) W_j(z_j) \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow j(z_j)}}{\sum_{z_i \in \Omega_i, z_j \in \Omega_j} W_{(i,j)}(z_i, z_j) W_j(z_j) \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow j(z_j)}} \left( \{i,j\} \in \partial i, i \in V \right),$$  \hspace{1cm} (3.61)

$$\mu_{(i,j)\rightarrow j(a_j)} = \frac{\sum_{z_j \in \Omega_j} W_{(i,j)}(a_i, z_j) W_j(z_j) \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow j(z_j)}}{\sum_{z_i \in \Omega_i, z_j \in \Omega_j} W_{(i,j)}(z_i, z_j) W_j(z_j) \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow j(z_j)}} \left( \{i,j\} \in \partial i, i \in V \right).$$  \hspace{1cm} (3.62)

$Z_i$ and $Z_{(i,j)}$ are the normalization constants of the marginal probability distributions $\hat{Q}(a_i)$ and $\hat{Q}_{(i,j)}(a_i, a_j)$ are are given as

$$Z_i = \sum_{z_i \in \Omega_i} W_i(z_i) \prod_{\{k,l\} \in \partial i} \mu_{(k,l)\rightarrow i(z_i)} \left( \{i\} \in V \right),$$  \hspace{1cm} (3.63)

$$Z_{(i,j)} = \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} W_{(i,j)}(z_i, z_j) W_j(z_j) \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow i(z_i)} \left( \prod_{\{k,l\} \in \partial j \setminus \{i,j\}} \mu_{(k,l)\rightarrow j(z_j)} \right) \left( \{i,j\} \in E \right),$$  \hspace{1cm} (3.64)

Graphical representations in Fig. 5 are assigned to the messages $\mu_{(i,j)\rightarrow j(a_j)}$ and $\mu_{(i,j)\rightarrow i(a_i)}$ in Eqs. (3.18) and (3.19). And graphical Representations of Eqs. (3.59)-(3.62) for a probabilistic graphical model of Eqs. (3.18) and (3.19) on a square grid graph $G = (V,E)$ in Fig. 1 are shown in Fig. 6.

By substituting Eqs. (3.59) and (3.60) and using Eqs. (3.53) and (3.54), the KL divergence between the probability distribution (3.19) and

$$\hat{Q}(a) = \left( \prod_{i \in V} \hat{Q}(a_i) \right) \left( \prod_{\{i,j\} \in E} \hat{Q}_{(i,j)}(a_i, a_j) \right),$$  \hspace{1cm} (3.65)
3.6. We consider a prior probability distribution consisting of pairs of nodes $i, j$, and any real values in the interval $(-\infty, +\infty)$, respectively. We define the state variables $a_i$ and $d_i (i \in \mathcal{V})$, which take any integer numbers in $\{0, 1, 2, \cdots, q - 1\}$ and any real values in the interval $(-\infty, +\infty)$, respectively. We consider a prior probability distribution $P(a|\alpha)$ and a conditional probability density function $f(d|a, \beta)$ defined by

$$P(a|\alpha) \equiv \frac{1}{Z(\alpha)} \exp\left( -\frac{1}{2} \alpha \sum_{(i,j) \in E} |a_i - a_j|^p \right) \quad (p > 0),$$  

$$f(d|a, \beta) \equiv \sqrt{\left( \frac{\beta}{2\pi}\right)^{|\mathcal{V}|}} \exp\left( -\frac{1}{2} \beta \sum_{i \in \mathcal{V}} (a_i - d_i)^2 \right).$$

4 Approximate Expectation-Maximization Algorithm by Loopy Belief Propagation for Generalized Sparse Markov Random Field Modeling

We introduce a square grid graph $(\mathcal{V}, E)$ with periodic boundary conditions for the abscissa and the ordinate as $G = (\mathcal{V}, E)$ as shown in Fig.1. Here $\mathcal{V} = \{1, 2, \cdots, |\mathcal{V}|\}$ is a set of all the nodes and $E$ is a set of all the edges $\{i, j\}$ consisting of pairs of nodes $i$ and $j$. We remark that $\partial i$ denote a set of all the neighbouring pairs of nodes $\{i, j\}$ and $|\partial i| = 4$ for every node $i (\in \mathcal{V})$. In Fig.1, we see that $\partial 15 = \{\{9, 15\}, \{14, 15\}, \{16, 15\}, \{15, 21\}\}$ and $\partial 16 = \{\{10, 15\}, \{15, 15\}, \{17, 15\}, \{15, 22\}\}$. Moreover, we should understand that $\partial 1 = \{\{1, 2\}, \{1, 7\}, \{1, 6\}, \{1, 25\}\}$ because of the periodic boundary conditions. We define the state variables $a_i$ and $d_i (i \in \mathcal{V})$, which take any integer numbers in $\{0, 1, 2, \cdots, q - 1\}$ and any real values in the interval $(-\infty, +\infty)$, respectively. We consider a prior probability distribution $P(a|\alpha)$ and a conditional probability density function $f(d|a, \beta)$ defined by

$$P(a|\alpha) \equiv \frac{1}{Z(\alpha)} \exp\left( -\frac{1}{2} \alpha \sum_{(i,j) \in E} |a_i - a_j|^p \right) \quad (p > 0),$$  

$$f(d|a, \beta) \equiv \sqrt{\left( \frac{\beta}{2\pi}\right)^{|\mathcal{V}|}} \exp\left( -\frac{1}{2} \beta \sum_{i \in \mathcal{V}} (a_i - d_i)^2 \right).$$

It means that the loopy belief propagation give an approximate value of $F = -\ln(Z)$ as

$$F = -\ln(Z) \approx -\sum_{i \in \mathcal{V}} \sum_{i \in \mathcal{V}} (1 - |\partial i|) \ln(Z_i) - \sum_{(i,j) \in E} \ln(Z_{i,j}).$$

Fig. 6. Graphical Representations of Eqs.(3.60)-(3.61) for a probabilistic graphical model of Eqs.(3.18) and (3.19) on a square grid graph $G = (\mathcal{V}, E)$ in Fig. 1. (a) Eq.(3.60). (b) Eq.(3.59). (c) Eq.(3.62). (d) Eq.(3.61).
on the state spaces $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_{|\mathcal{V}|}$ and $(-\infty, +\infty)^{|\mathcal{V}|}$ for the state vectors $\bm{a}$ and $d$, respectively. $Z(\alpha)$ is a normalization constant of $P(\bm{a}|\alpha)$, which is defined by

$$Z(\alpha) \equiv \sum_{\alpha_1=0}^{q-1} \sum_{\alpha_2=0}^{q-1} \cdots \sum_{\alpha_{|\mathcal{V}|}=0}^{q-1} \exp \left( \frac{1}{2} \alpha \sum_{i,j \in \mathcal{E}} |z_i - z_j|^p \right).$$

(4.3)

The posterior probability distribution $P(\mathcal{z}|d, \alpha, \beta)$ in Eq.(4.11), is defined by

$$P(\mathcal{z}|d, \alpha, \beta) = \frac{f(d|\alpha, \beta)P(\alpha|\alpha)}{f(d|\alpha, \beta)},$$

(4.4)

where the denominator $f(d|\alpha, \beta)$ in the right-hand side is the probability density function of $d$ defined by

$$f(d|\alpha, \beta) \equiv \sum_{\alpha_1=0}^{q-1} \sum_{\alpha_2=0}^{q-1} \cdots \sum_{\alpha_{|\mathcal{V}|}=0}^{q-1} P(\mathcal{z}, d|\alpha, \beta) = \sum_{\alpha_1=0}^{q-1} \sum_{\alpha_2=0}^{q-1} \cdots \sum_{\alpha_{|\mathcal{V}|}=0}^{q-1} \sum_{\mathcal{z} \in \Omega} f(d|\beta)P(\mathcal{z}|\alpha).$$

(4.5)

By substituting Eqs. (4.1) and (4.2), the posterior probability density function of Eq.(4.4) and the marginal likelihood of Eq.(4.5) are expressed as follows:

$$P(\alpha|d, \alpha, \beta) = \frac{1}{Z(\alpha, \beta)} \exp \left( -\frac{1}{2} \alpha \sum_{i,j \in \mathcal{E}} |a_i - a_j|^p - \frac{1}{2} \beta \sum_{i \in \mathcal{V}} (a_i - d_i)^2 \right),$$

(4.6)

$$P(\alpha, d|\alpha, \beta) = \frac{1}{Z(\alpha, \beta)} \left( \frac{\beta}{2\pi} \right)^{|\mathcal{V}|} \exp \left( -\frac{1}{2} \alpha \sum_{i,j \in \mathcal{E}} |a_i - a_j|^p - \frac{1}{2} \beta \sum_{i \in \mathcal{V}} (a_i - d_i)^2 \right),$$

(4.7)

where

$$Z(\alpha, \beta) \equiv \sum_{\alpha_1=0}^{q-1} \sum_{\alpha_2=0}^{q-1} \cdots \sum_{\alpha_{|\mathcal{V}|}=0}^{q-1} \exp \left( -\frac{1}{2} \alpha \sum_{i,j \in \mathcal{E}} |z_i - z_j|^p - \frac{1}{2} \beta \sum_{i \in \mathcal{V}} (z_i - d_i)^2 \right).$$

(4.8)

Now we remark that

$$f(d|\alpha, \beta) = \left( \frac{\beta}{2\pi} \right)^{|\mathcal{V}|} \frac{Z(d, \alpha, \beta)}{Z(\alpha)}.$$

(4.9)

By substituting Eqs.(4.1) and (4.2) to Eq.(4.5). Now we regard the probability $f(d|\alpha, \beta)$ of observed data $d$ with the hyperparameters $\alpha$ and $\beta$ given as a likelihood function of $(\alpha, \beta)$ with the data $d$ given and estimate the hyperparameters $\alpha$ and $\beta$ so as to maximize the marginal likelihood $f(d|\alpha, \beta)$ as follows:

$$\hat{(\alpha, \beta)} = \arg \max_{(\alpha, \beta)} f(d|\alpha, \beta).$$

(4.10)

The maximization of marginal likelihood (4.10) can be achieved by using the EM algorithm. For the present case, we introduce the $Q$-function defined by

$$Q(\alpha, \beta|\alpha', \beta', d) = \sum_{\mathcal{z} \in \Omega} P(\mathcal{z}|d, \alpha', \beta') \ln \left( P(\mathcal{z}|d, \alpha, \beta) \right),$$

(4.11)

for the probability distributions $P(\alpha|\alpha)$ and $P(\alpha|d, \beta)$. By applying these expressions to the update rule of the EM algorithm:

$$\alpha(t + 1), \beta(t + 1) = \arg \max_{\alpha \in (0, +\infty), \beta \in (0, +\infty)} Q(\alpha(t), \beta(t), d),$$

(4.12)
and by considering the extremum conditions of $Q(a, b | \alpha(t), \beta(t), d)$ with respect to $a$ and $b$, the update rule (4.12) from $(\alpha(t), \beta(t))$ to $(\alpha(t+1), \beta(t+1))$ can be reduced to the simultaneous deterministic equations: for $(\alpha(t+1), \beta(t+1))$, with $(\alpha(t), \beta(t))$ being given, as follows:

$$
\sum_{z \in \Omega} \left( \frac{1}{|E|} \sum_{(i,j) \in E} |z_i - z_j|^p \right) P(z | \alpha(t+1)) = \sum_{z_1=0 \ldots z_{q-1}=0} \left( \frac{1}{|E|} \sum_{(i,j) \in E} |z_i - z_j|^p \right) P(z | d, \alpha(t), \beta(t)), \tag{4.13}
$$

$$
\beta(t+1)^{-1} = \sum_{z_1=0 \ldots z_{q-1}=0} \left( \frac{1}{|V|} \sum_{i \in V} (z_i - d_i)^2 \right) P(z | d, \alpha(t), \beta(t)). \tag{4.14}
$$

By introducing the following marginal probability distributions of the prior probability distribution $P(a | \alpha)$ in Eq.(4.1) and the posterior probability distribution $P(a | d, \alpha, \beta)$ in Eq.(4.6):

$$
P_{i} (a_i | \alpha) \equiv \sum_{z_1=0 \ldots z_{q-1}=0} \delta_{a_i \beta_i} P(z | \alpha), \tag{4.15}
$$

$$
P_{(i,j)} (a_i, a_j | \alpha) \equiv \sum_{z_1=0 \ldots z_{q-1}=0} \delta_{a_i \beta_i} \delta_{a_j \beta_j} P(z | \alpha), \tag{4.16}
$$

$$
P_{(i)} (a_i | d, \alpha, \beta) \equiv \sum_{z_1=0 \ldots z_{q-1}=0} \delta_{a_i \beta_i} P(z | d, \alpha, \beta), \tag{4.17}
$$

$$
P_{(i,j)} (a_i, a_j | d, \alpha, \beta) \equiv \sum_{z_1=0 \ldots z_{q-1}=0} \delta_{a_i \beta_i} \delta_{a_j \beta_j} P(z | d, \alpha, \beta), \tag{4.18}
$$

the above update rules (4.13)-(4.14) in the expectation-maximization algorithm for the present generalized sparse Markov random field model can be reduced to

$$
\frac{1}{|E|} \sum_{(i,j) \in E} \sum_{z = 0}^{q-1} |z_i - z_j|^p P_{(i,j)} (z_i, z_j | \alpha(t+1))
= \frac{1}{|E|} \sum_{(i,j) \in E} \sum_{z = 0}^{q-1} |z_i - z_j|^p P_{(i,j)} (z_i, z_j | d, \alpha(t), \beta(t)), \tag{4.19}
$$

$$
\beta(t+1)^{-1} = \frac{1}{|E|} \sum_{(i,j) \in E} (z_i - d_i)^2 P(z_i | d, \alpha(t), \beta(t)). \tag{4.20}
$$

By setting

$$
W_i (a_i) = \exp \left( - \frac{1}{2} \beta (a_i - d_i)^2 \right), \tag{4.21}
$$

$$
W_{(i,j)} (a_i, a_j) = \exp \left( - \frac{1}{2} \alpha |a_i - a_j|^p \right), \tag{4.22}
$$

in Eqs.(3.59)-(3.62), we can derive approximate marginal probability distributions and message passing rules for the posterior probability distribution $P(a | d, \alpha, \beta)$ in Eq.(4.6) by means of the LBP as follows:

$$
P_i (a_i | d, \alpha, \beta) \approx \frac{1}{Z_i (d, \alpha, \beta)} \exp \left( - \frac{1}{2} \beta (a_i - d_i)^2 \right) \prod_{(k,j) \in e_i} \mu_{(k,j) \rightarrow (i)} (a_i | d, \alpha, \beta), \tag{4.23}
$$

$$
P_{(i,j)} (a_i, a_j | d, \alpha, \beta) \approx \frac{1}{Z_{(i,j)} (d, \alpha, \beta)} \exp \left( - \frac{1}{2} \beta (a_i - d_i)^2 - \frac{1}{2} \alpha |a_i - a_j|^p - \frac{1}{2} \beta (a_j - d_j)^2 \right).
$$
The prior probability distribution \( P(\alpha) \) in Eq.(4.1) has the spatial uniformness and the prior marginal probability distributions \( P_i(a_i|\alpha) \) and \( P_{ij}(a_i,a_j|\alpha) \) defined by Eqs.(4.15) and (4.16) are functions of state variables and hyperparameter \( \alpha \) and all the state variables have the same state space \( \{0,1,2,\cdots,q-1\} \). then the prior marginal probability distributions \( P_i(a_i|\alpha) \) and \( P_{ij}(a_i,a_j|\alpha) \) do not depend on \( i \) and \( \{i,j\} \) and can be expressed in terms of \( P_{\text{node}}(a_i|\alpha) \) and \( P_{\text{edge}}(a_i,a_j|\alpha) \), respectively. Moreover, all the messages \( \mu_{i,j}^{(ai)} \) and \( \mu_{i,j}^{(aj)} \) in Eqs.(3.59)-(3.62) also do not depend on \( \{i,j\} \rightarrow j \) and \( \{i,j\} \rightarrow i \) and can be rewritten to \( \mu(a_i) \) and \( \mu(a_j) \), respectively. In the similar arguments to Eqs.(4.21)-(4.28), we set

\[
W_i(a_i) = 1, \tag{4.29}
\]

\[
W_{i,j}(a_i,a_j) = \exp\left( -\frac{1}{2} \alpha |a_i - a_j|^p \right), \tag{4.30}
\]

in Eqs.(3.59)-(3.62) and derive approximate marginal probability distributions and message passing rules for the prior probability distribution \( P(a|\alpha) \) in Eq.(4.1) by means of the LBP as follows:

\[
P_i(a_i|\alpha) \approx P_{\text{node}}(a_i|\alpha) \equiv \frac{1}{Z_{\text{node}}(\alpha)} \mu(a_i|\alpha)^4 \{\{i,j\} \in E\}, \tag{4.31}
\]

\[
P_{ij}(a_i,a_j|\alpha) \approx P_{\text{edge}}(a_i,a_j|\alpha) \equiv \frac{1}{Z_{\text{edge}}(\alpha)} \exp\left( -\frac{1}{2} \alpha |a_i - a_j|^p \right) \mu(a_i|\alpha)^3 \mu(a_j|\alpha)^3 \{\{i,j\} \in E\}, \tag{4.32}
\]

\[
\mu(\xi|\alpha) = \frac{\sum_{\xi=0}^{q-1} \exp\left( -\frac{1}{2} \alpha |\xi - \xi'|^p \right) \mu(\xi|\alpha)^3}{\sum_{\xi=0}^{q-1} \sum_{\xi'=0}^{q-1} \exp\left( -\frac{1}{2} \alpha |\xi' - \xi'|^p \right) \mu(\xi|\alpha)^3} \{\{i,j\} \in E, \xi \in \{0,1,\cdots,q-1\}\}, \tag{4.33}
\]

\[
Z_{\text{node}}(\alpha) \equiv \sum_{\xi=0}^{q-1} \mu(\xi|\alpha)^3 \{i \in V\}. \tag{4.34}
\]
\[ Z_{\text{edge}}(\alpha) = \sum_{\zeta=0}^{q-1} \sum_{\zeta'=0}^{q-1} \exp \left( -\frac{1}{2} \alpha \left| \zeta - \zeta' \right|^2 \right) \mu(\zeta|\alpha)^3 \mu(\zeta'|\alpha)^3 \left( \{i,j\} \in E \right). \] (4.35)

Moreover, from Eq. (4.9), the logarithm of marginal likelihood per node is written as

\[ \frac{1}{|V|} \ln(f(d|\alpha, \beta)) \simeq \ln \left( \sqrt{\frac{\beta}{2\pi}} \right) + \frac{1}{|V|} \sum_{\{i,j\} \in E} \ln(Z_{\{i,j\}}(d, \alpha, \beta)) - \frac{3}{|V|} \sum_{i \in V} \ln(Z_i(d, \alpha, \beta)) \]

\[ + 2\ln(Z_{\text{edge}}(\alpha)) - 3\ln(Z_{\text{node}}(\alpha)), \] (4.36)
in terms of the \( Z_{\{i,j\}}(d, \alpha, \beta), Z_i(d, \alpha, \beta), Z_{\text{edge}}(\alpha) \) and \( Z_{\text{node}}(\alpha) \).

The practical procedures to solve Eqs. (4.13) and (4.14) are summarized as follows:

**Approximate EM Algorithm in Generalized Sparse Gaussian Graphical Model based on Loopy Belief Propagation**

**Step 1** Input a given data point \( d \) and \( r \equiv 0 \). Set \( \alpha(t) \) and \( \beta(t) \) as initial values.

**Step 2** Update \( r \) by \( r \leftarrow r + 1 \) and compute \( u \) and \( \beta(t) \) from \( \alpha(t-1) \) and \( \beta(t-1) \) as follows:

\[ \mu_{\{i,j\} \to a_i} \leftarrow \frac{\sum_{i \in \Omega_i} \sum_{j \in \Omega_j} \exp \left( -\frac{1}{2} \alpha(t-1)(a_i - z_i)^2 - \frac{1}{2} \beta(t-1)(z_j - d_j)^2 \right) \mu_{\{k,l\} \to j}(z_j)}{\sum_{i \in \Omega_i} \sum_{j \in \Omega_j} \exp \left( -\frac{1}{2} \alpha(t-1)(a_i - z_i)^2 - \frac{1}{2} \beta(t-1)(z_j - d_j)^2 \right) \mu_{\{k,l\} \to j}(z_j)} \] (4.37)

\[ \mu_{a_i \to \{i,j\}} \leftarrow \frac{\sum_{i \in \Omega_i} \sum_{j \in \Omega_j} \exp \left( -\frac{1}{2} \alpha(t-1)(a_i - z_i)^2 - \frac{1}{2} \beta(t-1)(z_j - d_j)^2 \right) \mu_{\{k,l\} \to \{i,j\}}(z_j)}{\sum_{i \in \Omega_i} \sum_{j \in \Omega_j} \exp \left( -\frac{1}{2} \alpha(t-1)(a_i - z_i)^2 - \frac{1}{2} \beta(t-1)(z_j - d_j)^2 \right) \mu_{\{k,l\} \to \{i,j\}}(z_j)} \] (4.38)

\[ B_i \leftarrow \sum_{z_i \in \Omega_i} \exp \left( -\frac{1}{2} \beta(t-1)(z_i - d_i)^2 \right) \left( \prod_{\{i,k\} \in \partial_i} \mu_{\{k,l\} \to \{i,j\}}(z_j) \right) \] (4.39)

\[ B_{\{i,j\}} \leftarrow \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} \exp \left( -\frac{1}{2} \beta(t-1)(z_i - d_i)^2 - \frac{1}{2} \alpha(t-1)(z_i - z_j)^2 - \frac{1}{2} \beta(t-1)(z_j - d_j)^2 \right) \]

\[ \times \left( \prod_{\{k,l\} \in \partial_i \setminus \{i,j\}} \mu_{\{k,l\} \to j}(z_j) \right) \left( \prod_{\{k,l\} \in \partial_j \setminus \{i,j\}} \mu_{\{k,l\} \to \{i,j\}}(z_j) \right) \] (4.40)

\[ u \leftarrow \frac{1}{|E|} \sum_{\{i,j\} \in E} \frac{1}{|B_{\{i,j\}}|} \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} |z_i - z_j|^p \]

\[ \times \exp \left( -\frac{1}{2} \beta(t-1)(z_i - d_i)^2 - \frac{1}{2} \alpha(t-1)(z_i - z_j)^2 - \frac{1}{2} \beta(t-1)(z_j - d_j)^2 \right) \]

\[ \times \left( \prod_{\{k,l\} \in \partial_i \setminus \{i,j\}} \mu_{\{k,l\} \to j}(z_j) \right) \left( \prod_{\{k,l\} \in \partial_j \setminus \{i,j\}} \mu_{\{k,l\} \to \{i,j\}}(z_j) \right). \] (4.41)

\[ \beta(t) \leftarrow \left( \frac{1}{|V|} \sum_{i \in V} \frac{1}{|B_i|} \sum_{z_i \in \Omega_i} (z_i - d_i)^2 \exp \left( -\frac{1}{2} \beta(t-1)(z_i - d_i)^2 \right) \left( \prod_{\{i,k\} \in \partial_i} \mu_{\{k,l\} \to \{i,j\}}(z_j) \right) \right)^{-1}. \] (4.42)

**Step 3** After setting \( \alpha(t) \leftarrow \alpha(t-1) \) and \( \tilde{\mu}(0) \leftarrow 2/q + 1 \), \( \tilde{\mu}(\hat{\xi}) \leftarrow 1/q + 1 \) \( (\hat{\xi} \in \{1, 2, \cdots, q-1\}) \) as initial values in this step, repeat the following updates of \( \alpha(t) \) as well as \( \tilde{\lambda}_{\{i,j\} \to a_i} \) \( (a_i \in \Omega_i, \{i,j\} \in \partial_i, i \in V) \) until they converge:

\[ \mu(\xi) \leftarrow \sum_{\xi=0}^{q-1} \sum_{\xi'=0}^{q-1} \exp \left( -\frac{1}{2} \alpha(t)(\xi - \xi')^2 \right) \tilde{\mu}(\xi)^3 \] (4.43)
5 Probabilistic Inference and Loopy Belief Propagation for Probabilistic Graphical Models on Hypergraphs

We introduce a hypergraph \((V,E)\) where \(V = \{1,2,\ldots,|V|\}\) is a set of all the nodes and \(E\) is a set of all the hyperedges \(\gamma\) consisting of sets of nodes.

\[
P(a) = \frac{1}{Z} \left( \prod_{i \in V} W_i(a_i) \right) \left( \prod_{\gamma \in E} W_{\gamma}(a_{\gamma}) \right),
\]

\[
Z = \sum_{z} \left( \prod_{i \in V} W_i(z_i) \right) \left( \prod_{\gamma \in E} W_{\gamma}(z_{\gamma}) \right),
\]

where \(a_{\gamma} = (a_{n_1}, a_{n_2}, \ldots, a_{n_{|\gamma|}})^T\) is a state vector for every hyperedge \(\gamma = \{n_1, n_2, \ldots, n_{|\gamma|} | n_1 < n_2 < \cdots < n_{|\gamma|}\} \in E\).

Let us suppose that the graph \(G = (V,E)\) is a connected cactus hypertree. For every node \(i \in V\), we can divide the cactus hypertree \(G\) to \(|\partial i|\) subgraphs in which the node \(i\) is a leaf. The subgraph of \(G\) which satisfies the following properties is denoted by the notation \(G(i,\gamma)\) (\(\gamma \in \partial i, i \in V\):

1. \(G(i,\gamma)\) is a connected hypertree.
2. \(G(i,\gamma) \cap G(i,\gamma') = i \in V\) (\(\gamma \in \partial i \backslash \gamma', \gamma' \in \partial i \backslash \gamma\)).
3. \(\bigcap_{i \in \gamma} G(i,\gamma) = \gamma \in E\).
4. \(\bigcup_{\gamma \in \partial i} G(i,\gamma) = G\).

In the subgraph \(G(i,\gamma)\), the node \(i\) is a leaf and the hyperedge \(\gamma\) is a leaf hyperedge. The set of all the nodes belonging to the subgraph \(G(i,\gamma)\) is denoted by \(V(i,\gamma)\). We introduce messages \(M_{\gamma \rightarrow i}(a_i)\) and \(M_{\gamma \rightarrow j}(a_j)\) for every edge \([i,j] \in E\):

\[
M_{\gamma \rightarrow i}(a_i) = \sum_{z_{\gamma \backslash i}} \delta_{a_i, z_i} \left( \prod_{\gamma' \in V(i,\gamma) \setminus \{i\}} W_{\gamma'}(z_{\gamma'}) \right)
\]
Fig. 7. Restored images $\widehat{\alpha}, \widehat{\beta} | \mathbf{d}$ obtained by applying the expectation-maximization algorithm for the generalized sparse Gaussian graphical model on the square grid graph to degraded images $\mathbf{d}$ in Fig. 3. (a) Lena, $p = 2.0$. (b) Lena, $p = 1.0$. (c) Lena, $p = 0.5$. (d) Mandrill, $p = 2.0$. (e) Mandrill, $p = 1.0$. (f) Mandrill, $p = 0.5$. (g) Pepper, $p = 2.0$. (h) Pepper, $p = 1.0$. (i) Pepper, $p = 0.5$.

$$\times \left( \prod_{\gamma \in \mathcal{G}(i, \gamma)} W_\gamma(z_\gamma) \right) (i \in V, \gamma \in \partial i), \quad (5.3)$$

where $\sum_{z_i \in \Omega}$ is the summation with respect to the state variables $z_i$ for all $i \in V(i, \gamma)$ over all the possible configurations.

For the probability distribution $P(a)$ in Eq. (5.1), the marginal probability distributions of the state variable $a_i$ and the state vector $(a_i, a_j)$ are expressed in terms of the messages $\{M_{\gamma \rightarrow i}(a_i) | i \in \partial i, i \in V\}$ as

$$P_i(a_i) = \frac{1}{Z_i} W_i(a_i) \prod_{\gamma \in \partial i} M_{\gamma \rightarrow i}(a_i) (i \in V), \quad (5.4)$$

$$P_\gamma(a_\gamma) = \frac{1}{Z_\gamma} W_\gamma(a_\gamma) \left( \prod_{k \in \mathcal{G}(\gamma)} W_k(a_k) \left( \prod_{\gamma' \in \partial \gamma \setminus \gamma} M_{\gamma' \rightarrow k}(a_k) \right) \right) (\gamma \in E), \quad (5.5)$$

where

$$Z_i \equiv \sum_{z_i \in \Omega_i} \prod_{\gamma \in \partial i} M_{\gamma \rightarrow i}(z_i) (i \in V), \quad (5.6)$$

$$Z_\gamma \equiv \sum_{z_\gamma \in \Omega_\gamma} W_\gamma(z_\gamma) \left( \prod_{k \in \mathcal{G}(\gamma) \setminus \gamma} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} M_{\gamma' \rightarrow k}(z_k) \right) \right) (\gamma \in E). \quad (5.7)$$

From the definition (5.3), we can derive the following recursion formulas for messages $\{M_{\gamma \rightarrow i}(a_i) | \gamma \in \partial i, i \in V\}$:

$$M_{\gamma \rightarrow i}(a_i) = \sum_{z_\gamma \in \Omega_\gamma} \delta_{a_i, z_i} W_\gamma(z_\gamma) \times \left( \prod_{k \in \mathcal{G}(\gamma) \setminus \{i\}} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} M_{\gamma' \rightarrow k}(z_k) \right) \right) \left( \gamma \in \partial i, i \in V \right). \quad (5.8)$$
We remark that every cycle consists of some hyperedges. In such situations, Eq. (5.2) expresses the mean square error between the original image and the estimated image \( \text{MSE} \{a, \tilde{\alpha}, \tilde{\beta}, d\} \), and the signal to noise ratio \( \text{SNR} \{a, \tilde{\alpha}, \tilde{\beta}, d\} \) by means of the expectation-maximization algorithm for the Gaussian graphical model on the square grid graph for the degraded image \( d \) in Fig.3.

(a) Lena. (b) Mandrill. (c) Pepper.

<table>
<thead>
<tr>
<th>Table 3. Estimates ( \tilde{\alpha} ) and ( \tilde{\sigma} = 1/\sqrt{\beta} ), and the logarithm of marginal likelihood per pixel ( \frac{1}{</th>
<th>V</th>
<th>} \ln \left( f \left( d, \tilde{\alpha}, \tilde{\beta} \right) \right) ).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p = 2.0 )</td>
<td>( p = 1.0 )</td>
</tr>
<tr>
<td>( \tilde{\alpha} )</td>
<td>0.000642</td>
<td>0.047460</td>
</tr>
<tr>
<td>( \tilde{\sigma} = 1/\sqrt{\beta} )</td>
<td>33.270</td>
<td>32.860</td>
</tr>
<tr>
<td>( \frac{1}{</td>
<td>V</td>
<td>} \ln \left( f \left( d, \tilde{\alpha}, \tilde{\beta} \right) \right) )</td>
</tr>
<tr>
<td>( \text{MSE} {a, \tilde{\alpha}, \tilde{\beta}, d} )</td>
<td>276.12</td>
<td>234.84</td>
</tr>
<tr>
<td>( \text{SNR} {a, \tilde{\alpha}, \tilde{\beta}, d} )</td>
<td>9.959 (dB)</td>
<td>10.662 (dB)</td>
</tr>
</tbody>
</table>

The set of recursion formulas in Eq.(5.8) is referred to as a message passing rule in the belief propagation. By using the expression of marginals in terms of the messages, the probability distribution \( P(a) \) is expressed in terms of the marginals as follows:

\[
P(a) = \left( \prod_{i \in V} P_i(a_i) \right) \left( \prod_{(\gamma, \delta) \in E} \frac{P_\gamma(a_\gamma)}{P_i(a_i)} \right)
\]

\[
= \left( \prod_{i \in V} P_i(a_i)^{1-|\delta_i|} \right) \left( \prod_{(\gamma, \delta) \in E} P_\gamma(a_\gamma) \right),
\]  

Next, we consider the probability distribution of Eqs.(5.1) and (5.2) on a hypergraph \( G = (V, E) \) with some cycles. We remark that every cycle consists of some hyperedges. In such situations, Eq.(5.9) is not always valid. We introduce the following trial probability distribution

\[
Q(a) = \left( \prod_{i \in V} Q_i(a_i) \right) \left( \prod_{(\gamma, \delta) \in E} \frac{Q_\gamma(a_\gamma)}{Q_i(a_i)} \right)
\]

\[
= \left( \prod_{i \in V} Q_i(a_i)^{1-|\delta_i|} \right) \left( \prod_{(\gamma, \delta) \in E} Q_\gamma(a_\gamma) \right),
\]

where

\[
Q_i(a_i) \equiv \sum_{z} \delta_{a_i, z} Q(z),
\]

\[
Q_\gamma(a_\gamma) \equiv \sum_{z} \delta_{a_\gamma, z} Q(z).
\]

By using Eqs.(5.10), (5.11) and (5.12), the KL divergence \( \text{KL}[P || Q] \) for \( P(a) \) in Eqs.(5.1)-(5.2) on a graph \( G = (V, E) \)
with cycles can be written down to

$$
\text{KL}[P||Q] = - \sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i)) \\
- \sum_{\gamma \in V} \sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \ln(W_\gamma(z_{\gamma})) \\
+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i)) \\
+ \sum_{\gamma \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \ln(Q_\gamma(z_{\gamma})) \\
+ \ln \left( \sum_{z} \left( \prod_{i \in V} W_i(z_i) \right) \left( \prod_{(i,j) \in E} W_j(z_j) \right) \right).
$$

(5.13)

so that we have

$$
\text{KL}[P||Q] = \mathcal{F}_{\text{Bethe}}[\{Q_i|i\in V\}, \{Q_\gamma|\gamma\in E\}] - F,
$$

(5.14)

where

$$
\mathcal{F}_{\text{Bethe}}[\{Q_i|i\in V\}, \{Q_\gamma|\gamma\in E\}] \\
= - \sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i)) - \sum_{\gamma \in \Omega_{\gamma}} \sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \ln(W_\gamma(z_{\gamma})) \\
+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i)) + \sum_{\gamma \in \Omega_{\gamma}} \sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \ln(Q_\gamma(z_{\gamma})).
$$

(5.15)

The partition function $Z$ is given in Eq.(5.2) and $F \equiv - \ln(Z)$ is the free energy of the probabilistic graphical model (5.1). We remark that $\mathcal{F}_{\text{Bethe}}[\{Q_i|i\in V\}, \{Q_\gamma|\gamma\in E\}]$ corresponds to so-called Bethe Free Energy Functional on a cactus hypergraph.

We consider to determine the marginal probability distributions $\{Q_i(a_i)|a_i \in \Omega_i, i \in V\}$ and $\{Q_\gamma(a_\gamma)|a_\gamma \in \Omega_\gamma, \gamma \in E\}$ so as to minimize the KL divergence $\text{KL}[P||Q]$ under the following constraint conditions

$$
Q_i(a_i) = \sum_{z_i \in \Omega_i} \delta_{a_i,z_i} Q_i(z_i) \ (a_i \in \Omega_i, i \in V, \gamma \in E),
$$

(5.16)

$$
\sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) = 1 \ (\gamma \in E),
$$

(5.17)

$$
\sum_{z_i \in \Omega_i} Q_i(z_i) = 1 \ (i \in V).
$$

(5.18)

The set of approximate marginal probability distributions, $\{\hat{Q}_i(a_i)|i \in V, a_i \in \Omega_i\}$, and $\{\hat{Q}_\gamma(a_\gamma)|\gamma \in E, a_\gamma \in \Omega_\gamma\}$, in the loopy belief propagation on a cactus hypergraph is determined so as to minimize the KL divergence $\text{KL}[P||Q]$ such that the Bethe free energy functional $\mathcal{F}_{\text{Bethe}}[\{Q_i|i\in V\}, \{Q_\gamma|\gamma\in E\}]$ under the normalization conditions and the reducibility conditions of each trial marginal probability distributions. To ensure the reducibility (5.17) and the normalizations (5.17)-(5.18), we introduce the Lagrange multipliers $\lambda_i(a_i)$, $\lambda_\gamma$ and $\lambda_\gamma$ as follows:

$$
\mathcal{L}[\{Q_i|i\in V\}, \{Q_\gamma|\gamma\in E\}] = - \sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i)) \\
- \sum_{\gamma \in V} \sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \ln(W_\gamma(z_{\gamma})) \\
+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i)) \\
+ \sum_{\gamma \in \Omega_{\gamma}} \sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \ln(Q_\gamma(z_{\gamma})) \\
+ \sum_{\gamma \in \Omega_{\gamma}} \sum_{z_{\gamma} \in \Omega_{\gamma}} \lambda_i(a_i) \left( Q_i(z_i) - \sum_{z_{\gamma} \in \Omega_{\gamma}} Q_\gamma(z_{\gamma}) \right)
$$

31
We introduce messages $\{M_{\gamma}^{-\rho} (a_i)\}$ by
\[
\exp \left( \lambda_{\gamma} (a_i) \right) = W_i (a_i) \prod_{\gamma \in \partial i, \gamma 
\neq \rho} M_{\gamma^{-\rho}} (a_i) \quad (i \in V, \gamma \in \partial i). 
\] (5.28)

By substituting Eq. (5.28) to Eqs. (5.22) and (5.23), approximate expressions of marginal probabilities $\tilde{Q}(a_i)$ and $\tilde{Q}_{\gamma} (a_{\gamma})$ are given as
\[
\tilde{Q}(a_i) = \sum_{z_i \in \Omega_i} W_i (a_i) \prod_{\gamma \in \partial i} M_{\gamma^{-\rho}} (a_i) \quad (i \in V, |\partial i| \geq 2), 
\] (5.29)
\[
\tilde{Q}_{\gamma} (a_{\gamma}) = \frac{W_{\gamma} (a_{\gamma}) \prod_{k \in \gamma} W_k (a_k) \left( \prod_{\gamma \in \partial k, \gamma \neq \rho} M_{\gamma^{-\rho}} (a_k) \right)}{W_{\gamma} (z_{\gamma}) \prod_{k \in \gamma} W_k (z_k) \left( \prod_{\gamma \in \partial k, \gamma \neq \rho} M_{\gamma^{-\rho}} (z_k) \right)} \quad (\gamma \in E). 
\] (5.30)

By substituting Eqs. (5.29) and (5.30) to
\[
\tilde{Q}(a_i) = \sum_{z_i \in \Omega_i} \delta_{a_i,z_i} \tilde{Q}_{\gamma} (z_{\gamma}) \quad (a_i \in \Omega_i, i \in \gamma, \gamma \in \partial i, |\partial i| \geq 2), 
\] (5.31)
we can derive the following simultaneous recursion formulas to determine \( \{ \mathcal{M}_{y \to i}(a_i) \}_{y \in \partial i, i \in V, |\partial i| \geq 2} \) as follows:

\[
\mathcal{M}_{y \to i}(a_i) = \left( \sum_{z_i \in \Omega_i} W_i(z_i) \prod_{\gamma' \in \partial_i} \mathcal{M}_{y' \to i}(z_i) \right)^{-1} \\
\times \left( \sum_{z_y \in \Omega_y} W_y(z_y) \prod_{k \in \gamma} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mathcal{M}_{y' \to k}(z_k) \right) \right) \\
\times \sum_{z_y \in \Omega_y} \delta_{a_i, z_i} W_y(z_y) \prod_{k \in \gamma(i)} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mathcal{M}_{y' \to k}(z_k) \right) \\
(\gamma \in \partial i, i \in V, |\partial i| \geq 2).
\] (5.32)

They are the simultaneous fixted point equations to determine the messages to nodes \( i \) which are not leafs. Moreover, we define the set of messages to leaves \( \{ \mathcal{M}_{y \to i}(a_i) \}_{y \in \partial i, i \in V, |\partial i| = 1} \) as follows:

\[
\mathcal{M}_{y \to i}(a_i) = \left( \sum_{z_i \in \Omega_i} W_i(z_i) \mathcal{M}_{y \to i}(z_i) \right)^{-1} \\
\times \left( \sum_{z_y \in \Omega_y} W_y(z_y) \prod_{k \in \gamma} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mathcal{M}_{y' \to k}(z_k) \right) \right) \\
\times \sum_{z_y \in \Omega_y} \delta_{a_i, z_i} W_y(z_y) \prod_{k \in \gamma(i)} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mathcal{M}_{y' \to k}(z_k) \right) \\
(i \in \gamma, \gamma \in E, |\partial i| = 1).
\] (5.33)

Eq.(5.33) means that all the messages to leaves are defined in terms of messages determined by Eq.(5.32). In the LBP for cactus hypergraphs, the approximate marginal probability distributions of leaves \( i \) are expressed in terms of the following reducibility conditions:

\[
\hat{\mu}_i(a_i) = \sum_{z_i \in \Omega_i} \delta_{a_i, z_i} \hat{\mu}_i(z_i) \quad (a_i \in \Omega_i, i \in \gamma, \gamma \in E, |\partial i| = 1).
\] (5.34)

By using Eqs. (5.30), (5.33) and (5.34), the approximate marginal probability of leaves \( i \) also can be expressed in terms of the messages determined by Eq.(5.33) as follows:

\[
\hat{\mu}_i(a_i) = \frac{W_i(a_i) \mathcal{M}_{y \to i}(a_i)}{\sum_{z_i \in \Omega_i} W_i(z_i) \mathcal{M}_{y \to i}(z_i)} \\
(i \in \gamma, \gamma \in E, |\partial i| = 1).
\] (5.35)

By introducing

\[
\mu_{y \to i}(a_i) = \frac{\mathcal{M}_{y \to i}(a_i)}{\sum_{z_i \in \Omega_i} \mathcal{M}_{y \to i}(z_i)} \quad (y \in \partial i, i \in V),
\] (5.36)

Eqs. (5.29)-(5.30), (5.32), (5.33) and (5.35) can be rewritten as

\[
\hat{\mu}_i(a_i) = \frac{W_i(a_i) \prod_{\gamma \in \partial i} \mu_{y \to i}(a_i)}{\sum_{z_i \in \Omega_i} W_i(z_i) \prod_{\gamma \in \partial i} \mu_{y \to i}(z_i)} \quad (i \in V),
\] (5.37)

\[
\hat{\mu}_y(a_y) = \frac{W_y(a_y) \prod_{k \in \gamma} W_k(a_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mu_{y' \to k}(a_k) \right)}{\sum_{z_y \in \Omega_y} W_y(z_y) \prod_{k \in \gamma(i)} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mu_{y' \to k}(z_k) \right)} \quad (y \in E),
\] (5.38)

\[
\mu_{y \to i}(a_i) = \frac{\sum_{z_y \in \Omega_y} \delta_{a_i, z_i} W_y(z_y) \prod_{k \in \gamma(i)} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mu_{y' \to k}(z_k) \right)}{\sum_{z_y \in \Omega_y} W_y(z_y) \prod_{k \in \gamma(i)} W_k(z_k) \left( \prod_{\gamma' \in \partial k \setminus \gamma} \mu_{y' \to k}(z_k) \right)}
\]
Explicit procedures to compute some approximate marginals in Eq.(5.37) and (5.38) with simultaneous fixed point equations (5.39) for messages are given as follows:

** Procedures for Computation of Approximate Marginals by Loopy Belief Propagation on Hypergraph**

**Step 1:** Set the messages $\mu_{\gamma \rightarrow i}(a_i)$ ($i \in V, \gamma \in \partial i, a_i \in \Omega_i$) as initial values.

**Step 2:** Repeat the following updates until all messages $\mu_{\gamma \rightarrow i}(a_i)$ ($i \in V, \gamma \in \partial i, a_i \in \Omega_i$) converge

$$
\mu_{\gamma \rightarrow i}(a_i) \Leftarrow \frac{\sum_{z_y \in \Omega_y} \delta_{a_i,z_i} W_{\gamma}(z_y) \prod_{k \in \gamma, \gamma \in \partial \gamma} W_k(z_k) \left( \prod_{\gamma' \in \partial \gamma \setminus \gamma} \mu_{\gamma' \rightarrow k}(z_k) \right)}{\sum_{z_y \in \Omega_y} W_{\gamma}(z_y) \prod_{k \in \gamma, \gamma \in \partial \gamma} W_k(z_k) \left( \prod_{\gamma' \in \partial \gamma \setminus \gamma} \mu_{\gamma' \rightarrow k}(z_k) \right)} (\gamma \in \partial i, i \in V).
$$

**Step 3:** Stop after computing the marginals as follows:

$$
\tilde{Q}_i(a_i) \Leftarrow \frac{W_i(a_i) \prod_{\gamma \in \partial i} \tilde{\mu}_{\gamma \rightarrow i}(a_i)}{\sum_{z_i \in \Omega_i} W_i(z_i) \prod_{\gamma \in \partial i} \tilde{\mu}_{\gamma \rightarrow i}(z_i)} (i \in V),
$$

$$
\tilde{Q}_\gamma(a) \Leftarrow \frac{W_\gamma(a) \prod_{k \in \gamma} W_k(a_k) \left( \prod_{\gamma' \in \partial \gamma \setminus \gamma} \tilde{\mu}_{\gamma' \rightarrow k}(a_k) \right)}{\sum_{z_y \in \Omega_y} W_\gamma(z_y) \prod_{k \in \gamma} W_k(z_k) \left( \prod_{\gamma' \in \partial \gamma \setminus \gamma} \tilde{\mu}_{\gamma' \rightarrow k}(z_k) \right)} (\gamma \in E).
$$

We consider a probabilistic model defined by the following joint probability distribution

$$
Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} = Pr(a_1, a_2, a_3, a_4) (a_1 \in \Omega_1, a_2 \in \Omega_2, a_3 \in \Omega_3, a_4 \in \Omega_4).
$$

In terms of the joint probability $Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\}$ and its marginal probabilities

$$
Pr\{A_1 = a_1\} = \sum_{a_2} \sum_{a_3} \sum_{a_4} Pr\{A_1 = a_1, A_2 = z_2, A_3 = z_3, A_4 = z_4\},
$$

$$
Pr\{A_1 = a_1, A_2 = a_2\} = \sum_{a_3} \sum_{a_4} Pr\{A_1 = a_1, A_2 = a_2, A_3 = z_3, A_4 = z_4\},
$$

$$
Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \sum_{a_4} Pr\{A_1 = a_1, A_2 = a_2, A_3 = z_3, A_4 = z_4\},
$$

some conditional probabilities are defined by

$$
Pr\{A_4 = a_4|A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \frac{Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\}}{Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\}},
$$

$$
Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\} = \frac{Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\}}{Pr\{A_1 = a_1, A_2 = a_2\}},
$$

$$
Pr\{A_2 = a_2|A_1 = a_1\} = \frac{Pr\{A_1 = a_1, A_2 = a_2\}}{Pr\{A_1 = a_1\}}.
$$

If the conditional probability $Pr\{A_4 = a_4|A_1 = a_1, A_2 = a_2, A_3 = a_3\}$ is independend of $a_1$ and does depend only on $a_2$ and $a_3$ as well as $a_4$, it is *conditional independent* of the random variable $a_1$. This conditional independence is represented by

$$
Pr\{A_4 = a_4|A_1 = a_1, A_2 = a_2, A_3 = a_3\} = Pr\{A_4 = a_4|A_2 = a_2, A_3 = a_3\}.
$$
We explain the conditional independency in this example more explicitly. When the state space of $a_i$ is set to $\Omega_i = \{0, 1\}$, the conditional independency (5.51)

$$\Pr\{A_4 = a_4|A_1 = 0, A_2 = a_2, A_3 = a_3\}$$

$$= \Pr\{A_4 = a_4|A_1 = 1, A_2 = a_2, A_3 = a_3\}$$

$$(a_2 \in \{0, 1\}, a_3 \in \{0, 1\}, a_4 \in \{0, 1\}).$$  \hspace{1cm} (5.52)

In the similar arguments, the conditional independency

$$\Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2, A_4 = a_4\} = \Pr\{A_3 = a_3|A_1 = a_1\},$$  \hspace{1cm} (5.53)

means that

$$\Pr\{A_3 = a_3|A_1 = a_1, A_2 = 0, A_3 = 0\}$$

$$= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = 0, A_3 = 1\}$$

$$= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = 1, A_3 = 0\}$$

$$= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = 1, A_3 = 1\} \quad \{a_1 \in \{0, 1\}, a_3 \in \{0, 1\}\},$$  \hspace{1cm} (5.54)

respectively. Under these conditional independencies, the joint probability can be expressed in terms of

$$\Pr\{A_4 = a_4|A_1 = a_1, A_2 = a_2, A_3 = a_3\}, \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}, \Pr\{A_2 = a_2|A_1 = a_1\} \text{ and } \Pr\{A_1 = a_1\}$$

as follows:

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\}$$

$$= \Pr\{A_4 = a_4|A_1 = a_1, A_2 = a_2, A_3 = a_3\}$$

$$\times \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\} \Pr\{A_2 = a_2|A_1 = a_1\} \Pr\{A_1 = a_1\}$$

$$= \Pr\{A_4 = a_4|A_2 = a_2, A_3 = a_3\}$$

$$\times \Pr\{A_3 = a_3|A_1 = a_1\} \Pr\{A_2 = a_2|A_1 = a_1\} \Pr\{A_1 = a_1\}$$

$$(a_1 \in \Omega_1, a_2 \in \Omega_2, a_3 \in \Omega_3, a_4 \in \Omega_4),$$  \hspace{1cm} (5.55)

and can be represented as

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} = P(a_1, a_2, a_3, a_4)$$

$$= \frac{W_{(1, 2)}(a_1, a_2)W_{(1, 3)}(a_1, a_3)W_{(2, 3, 4)}(a_2, a_3, a_4)}{\sum_z W_{(1, 2)}(z_1, z_2)W_{(1, 3)}(z_1, z_3)W_{(2, 3, 4)}(z_2, z_3, z_4)}.$$  \hspace{1cm} (5.56)

where the weights $W_{(2, 3, 4)}(a_2, a_3, a_4)$, $W_{(1, 3)}(a_1, a_3)$ and $W_{(1, 2)}(a_1, a_2)$ are defined by

$$W_{(2, 3, 4)}(a_2, a_3, a_4) \equiv \Pr\{A_4 = a_4|A_2 = a_2, A_3 = a_3\},$$

$$W_{(1, 3)}(a_1, a_3) \equiv \Pr\{A_3 = a_3|A_1 = a_1\},$$

$$W_{(1, 2)}(a_1, a_2) \equiv \Pr\{A_2 = a_2|A_1 = a_1\} \Pr\{A_1 = a_1\}.$$  \hspace{1cm} (5.57)

By using the joint probability $\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\}$, the conditional probability $\Pr\{A_2 = a_2|A_4 = a_4\}$ and $\Pr\{A_3 = a_3|A_4 = a_4\}$ are computed as follows:

$$\Pr\{A_2 = a_2|A_4 = a_4\} = \frac{\Pr\{A_2 = a_2, A_4 = a_4\}}{\Pr\{A_4 = a_4\}}$$  \hspace{1cm} (5.58)

$$\Pr\{A_3 = a_3|A_4 = a_4\} = \frac{\Pr\{A_3 = a_3, A_4 = a_4\}}{\Pr\{A_4 = a_4\}}.$$  \hspace{1cm} (5.59)

In the left-hand side of these equations, $\Pr\{A_2 = a_2, A_4 = a_4\}$, $\Pr\{A_3 = a_3, A_4 = a_4\}$ and $\Pr\{A_4 = a_4\}$ are marginal probabilities of the joint probability $\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\}$ as follows:

$$\Pr\{A_2 = a_2, A_4 = a_4\} = \sum_{z_1 \in \Omega_1} \sum_{z_3 \in \Omega_3} \Pr\{A_1 = z_1, A_2 = a_2, A_3 = z_3, A_4 = a_4\},$$  \hspace{1cm} (5.60)

$$\Pr\{A_3 = a_3, A_4 = a_4\} = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \Pr\{A_1 = z_1, A_2 = a_2, A_3 = z_3, A_4 = a_4\},$$  \hspace{1cm} (5.61)
\[ \Pr\{A_4 = a_4\} = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \Pr\{A_1 = z_1, A_2 = a_2, A_3 = z_3, A_4 = a_4\}. \quad (5.62) \]

By setting \( V = \{1, 2, 3, 4\} \) and \( E = \{\{1, 2\}, \{1, 3\}, \{2, 3, 4\}\} \), the probability distribution \((5.56)\) can be represented by Eq.(5.1) with Eq.(5.2). For the present example, approximate expressions of marginals in Eqs.(5.37) and (5.38) are explicitly written down as follows:

\[ P_1(a_1) = \sum_{a_2 \in \Omega_2} \sum_{a_3 \in \Omega_3} \sum_{a_4 \in \Omega_4} P(a_1, a_2, a_3, a_4) \approx \sum_{z_1 \in \Omega_1} W_1(a_1) \mu_{(1,2)\rightarrow 1}(a_1) \mu_{(1,3)\rightarrow 1}(a_1) \mu_{(1,4)\rightarrow 1}(a_1) \mu_{(2,3,4)\rightarrow 2}(a_2) \mu_{(2,3,4)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.63) \]

\[ P_2(a_2) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} P(z_1, a_2, z_3, z_4) \approx \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \mu_{(1,2)\rightarrow 2}(a_2) \mu_{(2,3,4)\rightarrow 2}(a_2) \mu_{(2,3,4)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.64) \]

\[ P_3(a_3) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} P(z_1, z_2, a_3, z_4) \approx \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \mu_{(1,3)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.65) \]

\[ P_4(a_4) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} P(z_1, z_2, z_3, a_4) \approx \sum_{z_4 \in \Omega_4} \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.66) \]

\[ P_{1,2}(a_1, a_2) = \sum_{a_3 \in \Omega_3} \sum_{a_4 \in \Omega_4} P(a_1, a_2, a_3, a_4) \approx \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_1(a_1) W_1(a_2) \mu_{(1,3)\rightarrow 1}(a_1) \mu_{(2,3,4)\rightarrow 2}(a_2) \mu_{(2,3,4)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.67) \]

\[ P_{1,3}(a_1, a_3) = \sum_{a_2 \in \Omega_2} \sum_{a_4 \in \Omega_4} P(a_1, a_2, a_3, a_4) \approx \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_1(a_1) W_1(a_3) \mu_{(1,2)\rightarrow 1}(a_1) \mu_{(2,3,4)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.68) \]

\[ P_{2,3,4}(a_2, a_3, a_4) = \sum_{a_1 \in \Omega_1} P(a_1, a_2, a_3, a_4) \approx \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_2(a_2, a_3, a_4) \mu_{(2,3,4)\rightarrow 2}(a_2) \mu_{(2,3,4)\rightarrow 3}(a_3) \mu_{(2,3,4)\rightarrow 4}(a_4) \quad (5.69) \]

Simultaneous fixed point equations of messages in Eqs.(5.39) are given as follows:

\[ \mu_{(1,2)\rightarrow 1}(a_1) = \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_1(a_1, z_2) \mu_{(2,3,4)\rightarrow 2}(z_2) \quad (5.70) \]

\[ \mu_{(1,2)\rightarrow 2}(a_2) = \sum_{z_1 \in \Omega_1} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_1(z_1, a_2) \mu_{(1,3)\rightarrow 1}(z_1) \quad (5.71) \]

\[ \mu_{(1,3)\rightarrow 1}(a_1) = \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_1(a_1, z_3) \mu_{(2,3,4)\rightarrow 3}(z_3) \quad (5.72) \]
\[ \mu_{(1,3)\to3}(a_3) = \frac{\sum_{z_1 \in \Omega_1} W_{(1,3)}(z_1, a_3) \mu_{(1,2)\to1}(z_1)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_{(1,3)}(z_1, z_2) \mu_{(1,2)\to1}(z_1)} \]  
(5.73)

\[ \mu_{(2,3,4)\to3}(a_3) = \frac{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} W_{(2,3,4)}(z_1, z_2, z_3) \mu_{(1,2)\to2}(z_2)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_{(2,3,4)}(z_1, z_2, z_3, z_4) \mu_{(1,2)\to2}(z_2)} \]  
(5.74)

\[ \mu_{(3,4)\to2}(a_2) = \frac{\sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} W_{(2,3,4)}(z_2, z_3, a_2) \mu_{(1,3)\to3}(z_3)}{\sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_{(2,3,4)}(z_2, z_3, z_4) \mu_{(1,3)\to3}(z_3)} \]  
(5.75)

\[ \mu_{(2,3)\to4}(a_2) = \frac{\sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_{(2,3,4)}(z_2, z_3, a_2, z_4) \mu_{(1,2)\to2}(z_2) \mu_{(1,3)\to3}(z_3)}{\sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} \sum_{z_5 \in \Omega_5} W_{(2,3,4)}(z_2, z_3, z_4, z_5) \mu_{(1,2)\to2}(z_2) \mu_{(1,3)\to3}(z_3)} \]  
(5.76)

6 Factor Graph Representation of Loopy Belief Propagation

We set a new node at each centroid of hyperedge \( \gamma \) and connect the centroid with all nodes belonging to the hyperedge \( \gamma \) by \(|\gamma|\) line segments. After doing this transformation for every hyperedge \( \gamma \) in the given hypergraph \( G \), we delete all the hyperedges. This procedure generates a new bipartite graph \( G' \) with edges and two kinds of nodes. In the bipartite graph \( G' \), each node which is generated by setting to the centroid of hyperedge \( \gamma \) is denoted by \( \gamma \). The set of all the nodes of \( V' \) consists of \( V \cup E \) and the set of all the edges in the bipartite graph is denoted by \( E' \). In the bipartite graph \( G' \), \( \partial^i \gamma \) is the set of all the neighbouring nodes \( \gamma \in E \) of the node \( i \in V \), while \( \partial^i \gamma \) is the set of all the neighbouring nodes \( i \in V \) of the node \( \gamma \in E \). Such bipartite graphs \( G' \) are referred to as Factor Graphs in the Bayesian network. We introduce new message from original messages in the loopy belief propagation as follows:

\[ \kappa_{\gamma\to\gamma}(a_i) \equiv \mu_{\gamma\to\gamma}(a_i) \quad (\gamma \in \partial^i \gamma, i \in V), \]  
(6.1)

\[ \kappa_{i\to\gamma}(a_i) \equiv \prod_{\gamma' \in \partial^i \gamma \setminus \gamma} \kappa_{\gamma'\to\gamma}(a_i) \quad (i \in \partial^i \gamma, \gamma \in E), \]  
(6.2)

In terms of the new messages, Eqs.(5.37), (5.38) and (5.39) can be rewritten as follows:

\[ \hat{Q}_i(a_i) = \frac{W_i(a_i)}{\sum_{z_1 \in \Omega_1} W_i(a_i) \prod_{\gamma \in \partial^i \gamma} \kappa_{\gamma\to\gamma}(z_i)} (i \in V), \]  
(6.3)

\[ \hat{Q}_\gamma(a_\gamma) = \frac{W_\gamma(a_\gamma) \prod_{j \in \partial^i \gamma} W_j(a_j) \kappa_{\gamma\to\gamma}(a_j)}{\sum_{z_\gamma \in \Omega_\gamma} W_\gamma(z_\gamma) \prod_{j \in \partial^i \gamma} W_j(z_j) \kappa_{\gamma\to\gamma}(z_j)} (\gamma \in E), \]  
(6.4)

\[ \kappa_{i\to\gamma}(a_i) = \prod_{\gamma' \in \partial^i \gamma \setminus \gamma} \kappa_{\gamma'\to\gamma}(a_i) \quad (i \in \partial^i \gamma, \gamma \in E), \]  
(6.5)

\[ \kappa_{\gamma\to\gamma}(a_i) = \frac{\sum_{z_\gamma \in \Omega_\gamma} \delta_{a_i, z_\gamma} W_\gamma(z_\gamma) \prod_{j \in \partial^i \gamma \setminus \gamma} W_j(z_j) \kappa_{\gamma\to\gamma}(z_j)}{\sum_{z_\gamma \in \Omega_\gamma} W_\gamma(z_\gamma) \prod_{j \in \partial^i \gamma \setminus \gamma} W_j(z_j) \kappa_{\gamma\to\gamma}(z_j)} (\gamma \in \partial^i \gamma, i \in V). \]  
(6.6)

We remark that \( \kappa_{i\to\gamma}(a_i) = 1 \) if \( \partial^i \gamma \) is an empty set in Eq. (6.5). They are referred to as Factor Graph Representations of the loopy belief propagation. Explicit procedures to compute some approximate marginals in Eq. (5.37) and (5.38) with simultaneous fixed point equations (5.39) for messages are given as follows:

[Procedures for Loopy Belief Propagation by Factor Graph Representations]
**Step 1:** Set the messages $\tilde{\kappa}_{\gamma\rightarrow i}(a_i) \equiv 1$ ($i \in V$, $\gamma \in \partial_i \setminus i$, $a_i \in \Omega_i$) and $\tilde{\kappa}_{\gamma\rightarrow \gamma}(a_i) \equiv 1$ ($\gamma \in E$, $i \in \partial^3 \gamma$, $a_i \in \Omega_i$) as initial values.

**Step 2:** Repeat the following updates until all messages $\tilde{\kappa}_{\gamma\rightarrow i}(a_i)$ ($i \in V$, $\gamma \in \partial^1 i$, $a_i \in \Omega_i$) and $\tilde{\kappa}_{\gamma\rightarrow \gamma}(a_i)$ ($\gamma \in E$, $i \in \partial^3 \gamma$, $a_i \in \Omega_i$) converge

\[
\tilde{\kappa}_{\gamma\rightarrow i}(a_i) \leftarrow \prod_{\gamma' \in \partial^1 i \setminus \gamma} \tilde{\kappa}_{\gamma'\rightarrow i}(a_i) \quad (i \in \partial^3 \gamma, \gamma \in E),
\]

\[
\tilde{\kappa}_{\gamma\rightarrow \gamma}(a_i) \leftarrow \sum_{z_i \in \Omega_i} \delta_{i,z_i} W_i(z_i) \prod_{j \in \partial^1 \gamma} W_j(z_j) \tilde{\kappa}_{j\rightarrow \gamma}(z_j)
\prod_{z_j \in \partial^3 \gamma} W_j(z_j) \tilde{\kappa}_{j\rightarrow \gamma}(z_j) \quad (\gamma \in \partial^1 i, i \in V).
\]

**Step 3:** Stop after computing the marginals as follows:

\[
\hat{Q}_i(a_i) \leftarrow \sum_{z_i \in \Omega_i} W_i(z_i) \prod_{\gamma \in \partial^1 i} \tilde{\kappa}_{\gamma\rightarrow i}(z_i) \quad (i \in V),
\]

\[
\hat{Q}_\gamma(a_\gamma) \leftarrow \sum_{z_\gamma \in \Omega_\gamma} W_\gamma(z_\gamma) \prod_{j \in \partial^1 \gamma} W_j(z_j) \tilde{\kappa}_{j\rightarrow \gamma}(z_j)
\prod_{z_j \in \partial^3 \gamma} W_j(z_j) \tilde{\kappa}_{j\rightarrow \gamma}(z_j) \quad (\gamma \in E),
\]

In terms of the factor graph representations, the approximate marginals (5.63)-(5.69) in LBP for the probability distribution (5.56) are rewritten as

\[
P_1(a_1) \simeq \frac{W_1(a_1) \kappa_{[1,2] \rightarrow 1}(a_1) \kappa_{[1,3] \rightarrow 1}(a_1)}{\sum_{z_1 \in \Omega_1} W_1(z_1) \kappa_{[1,2] \rightarrow 1}(z_1) \kappa_{[1,3] \rightarrow 1}(z_1)},
\]

\[
P_2(a_2) \simeq \sum_{z_2 \in \Omega_2} \kappa_{[1,2] \rightarrow 2}(a_2) \kappa_{[2,3,4] \rightarrow 2}(a_2),
\]

\[
P_3(a_3) \simeq \sum_{z_3 \in \Omega_3} \kappa_{[1,3] \rightarrow 3}(a_3) \kappa_{[2,3,4] \rightarrow 3}(a_3),
\]

\[
P_4(a_4) \simeq \sum_{z_4 \in \Omega_4} \kappa_{[2,3,4] \rightarrow 4}(a_4)
\]

\[
P_{[1,2]}(a_1, a_2) \simeq \frac{W_1(a_1) W_{[1,2]}(a_1, a_2) \kappa_{[1,2] \rightarrow 1}(a_1) \kappa_{[2,1] \rightarrow 2}(a_2)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_1(z_1) W_{[2,1]}(z_1, z_2) \kappa_{[1,2] \rightarrow 1}(z_1) \kappa_{[2,1] \rightarrow 2}(z_2)},
\]

\[
P_{[1,3]}(a_1, a_3) \simeq \frac{W_1(a_1) W_{[1,3]}(a_1, a_3) \kappa_{[1,3] \rightarrow 1}(a_1) \kappa_{[1,3] \rightarrow 1}(a_3)}{\sum_{z_1 \in \Omega_1} \sum_{z_3 \in \Omega_3} W_1(z_1) W_{[1,3]}(z_1, z_3) \kappa_{[1,3] \rightarrow 1}(z_1) \kappa_{[1,3] \rightarrow 1}(z_3)},
\]

\[
P_{[2,3,4]}(a_2, a_3, a_4) \simeq \frac{W_{[2,3,4]}(a_2, a_3, a_4) \kappa_{[2,3,4] \rightarrow 2}(a_2) \kappa_{[2,3,4] \rightarrow 3}(a_3) \kappa_{[2,3,4] \rightarrow 4}(a_4)}{\sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} \sum_{z_4 \in \Omega_4} W_{[2,3,4]}(a_2, a_3, a_4) \kappa_{[2,3,4] \rightarrow 2}(z_2) \kappa_{[2,3,4] \rightarrow 3}(z_3) \kappa_{[2,3,4] \rightarrow 4}(z_4)}
\]

By using Eqs. (6.5) and (6.6), the simultaneous fixed point equations of messages (5.63)-(5.69) are also written as follows:

\[
\kappa_{[1,2] \rightarrow 1}(a_1) = \kappa_{[1,3] \rightarrow 1}(a_1),
\]

\[
\kappa_{[2,3,4] \rightarrow 2}(a_2) \kappa_{[2,3,4] \rightarrow 3}(a_3) \kappa_{[2,3,4] \rightarrow 4}(a_4)
\]

\[
\kappa_{[1,3] \rightarrow 1}(a_1) \end{equation}
\[ \kappa_{2\rightarrow(1,2)}(a_2) = \kappa_{2\rightarrow(2,3,4)}\kappa_{2\rightarrow(1,2)}(a_2), \]  
\[ \kappa_{1\rightarrow(1,3)}(a_1) = \kappa_{1\rightarrow(1,2)}(a_1), \]  
\[ \kappa_{3\rightarrow(1,3)}(a_3) = \kappa_{2\rightarrow(2,3,4)}(a_3), \]  
\[ \kappa_{2\rightarrow(2,3,4)}(a_2) = \kappa_{1\rightarrow(2,3,4)}(a_2), \]  
\[ \kappa_{3\rightarrow(2,3,4)}(a_3) = \kappa_{1\rightarrow(3,3)}(a_3), \]  
\[ \kappa_{4\rightarrow(2,3,4)}(a_4) = 1, \]  
\[ \kappa_{1\rightarrow(2)}(a_1) = \frac{\sum_{z_1 \in \Omega_1} W_{1\rightarrow(1,2)}(a_1, z_1) \kappa_{2\rightarrow(1,2)}(z_2)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_{1\rightarrow(1,2)}(z_1, z_2) \kappa_{2\rightarrow(1,2)}(z_2)}, \]  
\[ \kappa_{1\rightarrow(2)}(a_2) = \frac{\sum_{z_1 \in \Omega_1} W_{1\rightarrow(2,3,4)}(a_2, z_1, z_4) \kappa_{3\rightarrow(1,3)}(z_3) \kappa_{4\rightarrow(1,2)}(z_4)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \sum_{z_3 \in \Omega_3} W_{1\rightarrow(2,3,4)}(z_2, z_3, z_4) \kappa_{3\rightarrow(1,3)}(z_3) \kappa_{4\rightarrow(1,2)}(z_4)}, \]  
\[ \kappa_{1\rightarrow(2)}(a_3) = \frac{\sum_{z_1 \in \Omega_1} W_{1\rightarrow(1,2)}(z_1, a_3) \kappa_{1\rightarrow(1,3)}(z_1)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_{1\rightarrow(1,2)}(z_1, z_2) \kappa_{1\rightarrow(1,3)}(z_1)}, \]  
\[ \kappa_{1\rightarrow(2)}(a_4) = \frac{\sum_{z_1 \in \Omega_1} W_{1\rightarrow(1,2)}(z_1, a_4) \kappa_{1\rightarrow(1,3)}(z_1)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_{1\rightarrow(1,2)}(z_1, z_2) \kappa_{1\rightarrow(1,3)}(z_1)}, \]  
\[ \mu_{1\rightarrow(2)}(a_4) = \frac{\sum_{z_1 \in \Omega_1} W_{1\rightarrow(1,2)}(z_1, a_4) \kappa_{1\rightarrow(1,3)}(z_1)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_{1\rightarrow(1,2)}(z_1, z_2) \kappa_{1\rightarrow(1,3)}(z_1)}, \]  
\[ \mu_{1\rightarrow(2)}(a_4) = \frac{\sum_{z_1 \in \Omega_1} W_{1\rightarrow(1,2)}(z_1, a_4) \kappa_{1\rightarrow(1,3)}(z_1)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} W_{1\rightarrow(1,2)}(z_1, z_2) \kappa_{1\rightarrow(1,3)}(z_1)}, \]

7 Moment Representation of Loopy Belief Propagation

We introduce a set of orthonormal polynomials \{\Phi_i(\xi, a_i) | \xi \in \Omega_i\} for each node \(i \in V\), where \(\Omega_i = \{0, 1, \cdots, |\Omega_i|\}\) and \(\Phi_i(0, a_i) = 1/\sqrt{|\Omega_i|}\) (\(a_i \in \Omega_i\)). The orthonormal polynomials satisfy the following relations:

\[ \sum_{a_i \in \Omega_i} \Phi_i(\xi_i, a_i) \Phi_i(\xi_j, a_i) = \delta_{\xi_i, \xi_j} \delta_{a_i, a_i} (\xi_i \in \Omega_i, \xi_j \in \Omega_j, i \in V). \]  

(7.1)

In terms of the orthonormal polynomials, \(\ln W_i(a_i)\), \(\ln W_{ij}(a_{ij})\), \(Q_i(a_i)\) and \(Q_{ij}(a_{ij})\) are expanded as Fourier series to

\[ \ln W_i(a_i) = \sum_{\xi_i \in \Omega_i} \kappa_i(\xi_i) \Phi_i(\xi_i, a_i), \]  

(7.2)
\[ \ln W_\gamma(a_\gamma) = \sum_{\zeta \in \mathcal{M}_\gamma} K_\gamma(\zeta_\gamma) \prod_{j \in \gamma} \Phi_j(\zeta_j, a_j), \]  

(7.3)

\[ Q_\gamma(a_\gamma) = \left( \prod_{k \in \gamma} \frac{1}{|\Omega_k|} \right) + \sum_{j \in \gamma} \left( \prod_{k \in \gamma \setminus \{j\}} \frac{1}{|\Omega_k|} \right) \sum_{\zeta \in \mathcal{M}_j(\zeta_\gamma)} \Phi_j(\zeta_\gamma, a_j), \]  

(7.4)

\[ Q_\gamma(a_\gamma) = \left( \prod_{k \in \gamma} \frac{1}{|\Omega_k|} \right) \sum_{\zeta \in \mathcal{M}_j(\zeta_\gamma)} \Phi_j(\zeta_\gamma, a_j), \]  

(7.5)

where

\[ K_i(\xi_i) = \sum_{\gamma \in \Omega_i} \Phi_i(\xi_i, z_i) \ln W_i(z_i), \]  

(7.6)

\[ K_j(\xi_j) = \sum_{z_j \in \Omega_j} \left( \prod_{\xi \in \mathcal{M}_j(\zeta_\gamma)} \Phi_j(\zeta_\gamma, z_j) \right) \ln W_j(z_\gamma), \]  

(7.7)

\[ M_i(\xi_i) = \sum_{\zeta \in \mathcal{M}_i} \Phi_i(\zeta_i, z_i) Q_i(z_i), \]  

(7.8)

\[ M_j(\xi_j) = \sum_{z_j \in \Omega_j} \left( \prod_{\xi \in \mathcal{M}_j(\zeta_\gamma)} \Phi_j(\zeta_\gamma, z_j) \right) Q_j(z_\gamma), \]  

(7.9)

and

\[ \Omega_\gamma \equiv \{ \xi_\gamma \in \mathcal{M}_\gamma \setminus \{0\}, j \in \gamma \}, \]  

(7.10)

\[ \Omega_{\gamma j} \equiv \{ \xi_\gamma \xi_j \in \mathcal{M}_\gamma \setminus \{0\}, \xi_j = 0 (j \in \gamma \setminus \{i\}) \}. \]  

(7.11)

By substituting them to Eq.(5.15), we have

\[ \mathcal{F}_{\text{Bethe}}[\{ Q_i \mid i \in V \}, \{ Q_\gamma \mid \gamma \in E \}] \]

\[ = \mathcal{F}_{\text{Bethe}}[\{ M_i(\xi_i) \mid i \in V, \xi_i \in \Omega_i \setminus \{0\} \}, \{ M_\gamma(\xi_\gamma) \mid \gamma \in E, \xi_\gamma \in \Omega_\gamma \setminus \{0\} \}]] \]

\[ - \sum_{i \in V} \sum_{\xi_i \in \Omega_i \setminus \{0\}} K_i(\xi_i) M_i(\xi_i) - \sum_{\gamma \in E} \sum_{\xi_\gamma \in \Omega_\gamma \setminus \{0\}} K_\gamma(\xi_\gamma) M_\gamma(\xi_\gamma) \]

\[ + \sum_{i \in V} (1 - |\Omega_i|) \sum_{\xi_i \in \Omega_i \setminus \{0\}} \frac{1}{|\Omega_i|} \sum_{\zeta \in \mathcal{M}_i} \Phi_i(\zeta_i, a_i) \]

\[ \times \ln \left( \frac{1}{|\Omega_i|} \right) + \sum_{\zeta \in \mathcal{M}_i} \Phi_i(\zeta_i, a_i) \]

\[ + \sum_{\gamma \in E} \sum_{\xi_\gamma \in \Omega_\gamma \setminus \{0\}} \left( \prod_{k \in \gamma} \frac{1}{|\Omega_k|} \right) + \sum_{j \in \gamma} \left( \prod_{k \in \gamma \setminus \{j\}} \frac{1}{|\Omega_k|} \right) \sum_{\zeta_j \in \Omega_j \setminus \{0\}} M_j(\zeta_j) \Phi_j(\zeta_j, a_j) \]

\[ + \sum_{\xi_\gamma \in \Omega_\gamma \setminus \{0\}} M_\gamma(\xi_\gamma) \prod_{j \in \gamma} \Phi_j(\zeta_j, a_j), \]  

(7.12)

By taking the first derivative of \( \mathcal{F}_{\text{Bethe}}[\{ M_i(\xi_i) \mid i \in V, \xi_i \in \Omega_i \setminus \{0\} \}, \{ M_\gamma(\xi_\gamma) \mid \gamma \in E, \xi_\gamma \in \Omega_\gamma \setminus \{0\} \}]] \) with respect to \( M_i(\xi_i) \) and \( M_\gamma(\xi_\gamma) \), the simultaneous deterministic equations for the extremum point \( M_i(\xi_i) = \hat{M}_i(\xi_i) \) and
\[ \mathcal{M}_i(\xi) = \hat{\mathcal{M}}_i(\xi) \] are reduced to

\[
K_i(\xi) = \left( 1 - |\partial i| \right) \sum_{z_i \in \Omega_i} \Phi_i(\xi_i, z_i) \hat{Q}_i(z_i) \right) + \sum_{\gamma \in \Omega} \sum_{i \in \Omega_i} \Phi_i(\xi_i, z_i) \hat{Q}_j(z_j) \right) \left( i \in V, \xi_i \in \Omega_i \setminus \{0\}, \right), \quad (7.13)
\]

\[
K_j(\xi) = \sum_{z_j \in \Omega_j} \left( \prod_{i \in \Omega_i} \Phi_i(\xi_i, z_i) \right) \hat{Q}_j(z_j) \left( \gamma \in E, \xi_j \in \Omega_j \setminus \{0\}, \gamma \in \bigcup_{j \in \gamma} \Omega_j \cup \{0\} \right), \quad (7.14)
\]

\[
\hat{Q}_i(a_i) = \frac{1}{|\Omega_i|} + \sum_{\xi \in \Omega_i \setminus \{0\}} \hat{M}_i(\xi) \Phi_i(\xi_i, a_i) (i \in V), \quad (7.15)
\]

\[
\hat{Q}_j(\xi_j) = \left( \prod_{i \in \Omega_i} \right) + \sum_{j \in \gamma} \left( \prod_{i \in \Omega_i \setminus \{0\}} \frac{1}{\Omega_i} \right) \sum_{\xi_i \in \Omega_i \setminus \{0\}} \hat{M}_j(\xi_j) \Phi_j(\xi_j, a_j)
\]

By solving Eqs.(7.13) and (7.14), \( \hat{Q}_i(a_i) \) and \( \hat{Q}_j(\xi_j) \) can be computed.

If we have \( D \) configurations \( a^{(1)}, a^{(2)}, \ldots, a^{(D)} \) for the state vector \( a \) as data points. These data points are assumed to be generated as muntual independent events by according to the joint probability distribution \( P(a) \) in Eq.(5.1). From all the data points, the marginal probability distributions of every node \( i \) and of every hyperedge \( \gamma \) can be estimated as follows:

\[
P_i(a_i) \equiv \sum_{z_i \in \Omega_i} \delta_{a_i, z_i} P(z) \approx \frac{1}{D} \sum_{d=1}^{D} \delta_{a_i, z_i} \quad (7.17)
\]

\[
P_j(\xi_j) \equiv \sum_{z_j \in \Omega_j} \delta_{\xi_j, z_j} P(z) \approx \frac{1}{D} \sum_{d=1}^{D} \prod_{j \in \gamma} \delta_{a_j, z_j} \quad (7.18)
\]

We remark that \( \hat{Q}_i(a_i) \) and \( \hat{Q}_j(\xi_j) \) in Eqs.(7.14) and (7.14) are approximate expressions of \( P_i(a_i) \) and \( P_j(\xi_j) \) in the LBP, respectively. Then by setting \( \hat{Q}_i(a_i) \) and \( \hat{Q}_j(\xi_j) \) in Eqs.(7.14) and (7.14) to

\[
\hat{Q}_i(a_i) \leftarrow \frac{1}{D} \sum_{d=1}^{D} \delta_{a_i, z_i} \quad (7.19)
\]

\[
\hat{Q}_j(\xi_j) \leftarrow \frac{1}{D} \sum_{d=1}^{D} \prod_{j \in \gamma} \delta_{a_j, z_j} \quad (7.20)
\]

we can estimate \( K_i(\xi_i) \left( i \in V, \xi_i \in \Omega_i \setminus \{0\} \right) \) \( K_j(\xi_j) \left( \gamma \in E, \xi_j \in \Omega_j \setminus \{0\}, \gamma \in \bigcup_{j \in \gamma} \Omega_j \cup \{0\} \right) \), such that \( W_i(a_i) \left( a_i \in \Omega_i \right) \) and \( W_j(\xi_j) \left( \xi_j \in \Omega_j \right) \) in Eqs.(7.2) and (7.3) up to contant terms. \( K_i(\xi_i) \Phi_i(0, a_i) \) and \( K_j(\xi_j) \prod_{j \in \gamma} \Phi_j(0, a_j) \), respectively. This scheme can be regarded as one of approximate supervised learnings in the LBP from compleata data points.

8 Linear Response in Loopy Belief Propagation

In terms of a set of orthonormal polynomials \{ \Phi_i(\xi_i, z_i) | z_i \in \Omega_i, \xi_i \in \Omega_i \} \equiv 1/\sqrt{|\Omega_i|} \}, the marginal probabilities \( P_{i,j}(a_i, a_j) \equiv \sum_{z_i \in \Omega_i} \sum_{z_j \in \Omega_j} P(z) \) for all the pairs of nodes \( i \) and \( j \) are expressed as the following Fourier series:

\[
P_{i,j}(a_i, a_j) = \frac{1}{|\Omega_i||\Omega_j|} + \frac{1}{|\Omega_i|} \sum_{\xi_j \in \Omega_j \setminus \{0\}} \Phi_j(a_j, \xi_j) \mathcal{M}_j(\xi_j)
\]
\[
\begin{align*}
+ \frac{1}{|\Omega_j|} \sum_{\zeta_j \in \Omega_j \setminus \{0\}} \Phi_j(a_i, \zeta_j) \mathcal{M}_{i,j}^*(\zeta_j) \\
+ \sum_{\zeta_j \in \Omega_j \setminus \{0\}} \sum_{\zeta_j \in \Omega_j} \Phi_j(a_i, \zeta_j) \mathcal{M}_{i,j}^*(\zeta_j),
\end{align*}
\]
where
\[
\mathcal{M}_{i,j}^*(\xi_j, \xi_j) \equiv \sum_{z_1 \in \Omega_1} \cdots \sum_{z_2 \in \Omega_2} \Phi_i(z_i, \xi_i) \Phi_j(z_j, \xi_j) P(z_1, z_2, \ldots, z_{|V|}),
\]
(8.2)
\[
\mathcal{M}_j^*(\xi_j) \equiv \sum_{z_1 \in \Omega_1} \cdots \sum_{z_j \in \Omega_j} \Phi_i(z_i, \xi_i) P(z_1, z_2, \ldots, z_{|V|}).
\]
(8.3)

In terms of the probability distribution \( P(a_1, a_2, \ldots, a_{|V|} | \Delta \beta_j(\xi_j)) \), we have the following linear response formulas:
\[
\mathcal{M}_{i,j}^*(\xi_i, \xi_j) - \mathcal{M}_j^*(\xi_j) \mathcal{M}_j^*(\xi_j) = \lim_{\Delta \beta_j(\xi_j) \to 0} \frac{\langle i, \xi_i | \Delta \Pi (\Delta \beta_j(\xi_j)) \rangle \mathcal{M}_j^*(\xi_j)}{\Delta \beta_j(\xi_j)} (i \in V, j \in V),
\]
(8.4)
where
\[
P(a_1, a_2, \ldots, a_{|V|} | \Delta \beta_j(\xi_j)) = \sum_{z_1 \in \Omega_1} \cdots \sum_{z_j \in \Omega_j} \exp \left( \frac{\Delta \beta_j(\xi_j) \Phi_i(z_i, \xi_i)}{\Delta \beta_j(\xi_j)} \right) P(z_1, z_2, \ldots, z_{|V|}).
\]
(8.5)
\[
\langle i, \xi_i | \Delta \Pi (\Delta \beta_j(\xi_j)) \rangle = \sum_{z_j \in \Omega_j} \Phi_i(z_i, \xi_i) P(z_j | \Delta \beta_j(\xi_j)) - \sum_{z_j \in \Omega_j} \Phi_i(z_i, \xi_i) P(z_j).
\]
(8.6)

This linear response formula means that we can approximately compute expectation values \( \mathcal{M}_{i,j}^*(\xi_i, \xi_j) \) for all the pairs of nodes \( i \in V \) and \( j \in V \) by considering the first derivatives of the expectation values \( \sum_{z \in \Omega} \Phi_i(z_i, \xi_i) P(z | \Delta \beta_j(\xi_j)) \) with respect of \( \Delta \beta_j(\xi_j) \) for all the nodes \( i \in V \) in the LBP. They can lead us to the computation of the marginal probability distribution \( P_i(a_i, a_j) \) for every pair of nodes \( i \) and \( j \) which do not belong to the same hyperedge of the other.

We introduce Fourier series of \( \lambda_{i,j}(z_i) \) and \( \lambda_{i,j}(z_i) \) in Eqs.(5.22)-(5.27) as
\[
\lambda_{i,j}(a_i) = \sum_{\zeta_i \in \Omega_i} \Phi_i(z_i, \zeta_i) \Psi_{i,j}(\zeta_i) (a_i \in \Omega_i),
\]
(8.7)
\[
\lambda_{i,j}(z_i) = \sum_{\zeta_i \in \Omega_i} \Phi_i(z_i, \zeta_i) \Psi_{i,j}(\zeta_i) (a_i \in \Omega_i),
\]
(8.8)
where
\[
\Psi_{i,j}(\zeta_i) \equiv \sum_{z_i \in \Omega_i} \lambda_{i,j}(z_i) \Phi_i(z_i, \zeta_i) (\zeta_i \in \Omega_i \setminus \{0\}),
\]
(8.9)
\[
\Psi_{i,j}(\zeta_i) \equiv \sum_{z_i \in \Omega_i} \lambda_{i,j}(z_i) \Phi_i(z_i, \zeta_i) (\zeta_i \in \Omega_i \setminus \{0\}).
\]
(8.10)
Approximate marginal probability distributions in Eqs.(5.22) and (5.23) are rewritten as
\[
\hat{Q}_i(a_i) = \sum_{z_i \in \Omega_i} W_i(a_i) \exp \left( \sum_{\zeta_i \in \Omega_i} \Phi_i(a_i, \zeta_i) \Psi_{i,i}(\zeta_i) \right),
\]
(8.11)
\[
\hat{Q}_i(a_i) = \sum_{z_i \in \Omega_i \setminus \gamma} W_i(a_i) \exp \left( \sum_{\zeta_i \in \Omega_i \setminus \gamma} \Phi_i(a_i, \zeta_i) \Psi_{i,j}(\zeta_i) \right),
\]
(8.12)
where

\[ (1 - |\partial i|)\Psi_{i,j}(\xi_i) + \sum_{j\in Y} \Psi_{i,j}\gamma(\xi_i) = 0 \quad (\xi_i \in \Omega_i, i \in V), \]  

\[ (8.13) \]

\[ \sum_{z_i \in \Omega_i} \Phi_i(z_i, \xi_i) \hat{Q}_i(z_i) = \sum_{z_i \in \Omega_i} \Phi_i(z_i, \xi_i) \hat{Q}_i(z_i) \quad (\xi_i \in \Omega_i, \gamma \in E). \]  

\[ (8.14) \]

On the other hand, approximate marginal probability distributions for \( P(a|\Delta \beta_j(\xi_j)) \) in the LBP are given as follows:

\[ \hat{Q}_i(a_i|\Delta \beta_j(\xi_j)) = \frac{W_i(a_i) \exp(\lambda_{i,j}(a_i|\Delta \beta_j(\xi_j)))}{\sum_{z_i \in \Omega_i} W_i(z_i) \exp(\lambda_{i,j}(z_i|\Delta \beta_j(\xi_j)))} \quad (i \in V), \]  

\[ (8.15) \]

\[ \hat{Q}_j(a_j|\Delta \beta_j(\xi_j)) = \frac{W_j(a_j) \exp(\sum_{k \in Y} \lambda_{k,j}(a_k|\Delta \beta_j(\xi_j)))}{\sum_{z_j \in \Omega_j} W_j(z_j) \exp(\sum_{k \in Y} \lambda_{k,j}(z_k|\Delta \beta_j(\xi_j)))} \quad (\gamma \in E), \]  

\[ (8.16) \]

where

\[ (1 - |\partial i|)\lambda_{i,j}(a_i|\Delta \beta_j(\xi_j)) + \sum_{\gamma \in \partial i} \lambda_{i,j}(a_i|\Delta \beta_j(\xi_j)) = \delta_{i,j}(1 - \delta_{\beta_i,1}) \sum_{\xi_i \in \Omega_i} \Phi_i(a_i, \xi_i) \Delta \beta_j(\xi_j) \quad (i \in V, |\partial i| = 1), \]  

\[ (8.17) \]

\[ \lambda_{i,j}(a_i|\Delta \beta_j(\xi_j)) = \lambda_{i,j}(a_i|\Delta \beta_j(\xi_j)) = 0 \quad (i \in V, |\partial i| = 1), \]  

\[ (8.18) \]

We introduce Fourier transformations of \( \lambda_{i,j}(z_i|\Delta \beta_j(\xi_j)) \) and \( \lambda_{i,j}(z_i|\Delta \beta_j(\xi_j)) \) as

\[ \Psi_{i,j}(\xi_i|\Delta \beta_j(\xi_j)) = \sum_{z_i \in \Omega_i} \Phi_i(z_i, \xi_i) \lambda_{i,j}(z_i|\Delta \beta_j(\xi_j)) \quad (\xi_i \in \Omega_i), \]  

\[ (8.19) \]

\[ \Psi_{i,j}(\xi_i|\Delta \beta_j(\xi_j)) = \sum_{z_i \in \Omega_i} \Phi_i(z_i, \xi_i) \lambda_{i,j}(z_i|\Delta \beta_j(\xi_j)) \quad (\xi_i \in \Omega_i). \]  

\[ (8.20) \]

Eqs.\( (8.15) \)-\( (8.17) \) can be rewritten as follows:

\[ \hat{Q}_j(a_j|\Delta \beta_j(\xi_j)) = \frac{W_j(a_j) \exp\left(\sum_{k \in Y} \Phi_k(a_k, \xi_k) \Psi_{k,j}(\xi_k|\Delta \beta_j(\xi_j))\right)}{\sum_{z_j \in \Omega_j} W_j(z_j) \exp\left(\sum_{k \in Y} \Phi_k(z_k, \xi_k) \Psi_{k,j}(\xi_k|\Delta \beta_j(\xi_j))\right)} \quad (\gamma \in E, j \in V, \xi_j \in \Omega_j \setminus \{0\}), \]  

\[ (8.21) \]

\[ \hat{Q}_i(a_i|\Delta \beta_j(\xi_j)) = \frac{W_i(a_i) \exp\left(\sum_{\xi_i \in \Omega_i \setminus \{0\}} \Phi_i(a_i, \xi_i) \Psi_{i,j}(\xi_i|\Delta \beta_j(\xi_j))\right)}{\sum_{z_i \in \Omega_i} W_i(z_i) \exp\left(\sum_{\xi_i \in \Omega_i \setminus \{0\}} \Phi_i(z_i, \xi_i) \Psi_{i,j}(\xi_i|\Delta \beta_j(\xi_j))\right)} \quad (i \in V, j \in V, \xi_j \in \Omega_j \setminus \{0\}), \]  

\[ (8.22) \]

where

\[ (1 - |\partial i|)\Psi_{i,j}(\xi_i|\Delta \beta_j(\xi_j)) + \sum_{\gamma \in \partial i} \Psi_{i,j}(\xi_i|\Delta \beta_j(\xi_j)) = \delta_{i,j} \delta_{\xi_i,0} \Delta \beta_j(\xi_j) \quad (i \in V, j \in V, \xi_i \in \Omega_i \setminus \{0\}, \xi_j \in \Omega_j \setminus \{0\}), \]  

\[ (8.23) \]

\[ \sum_{z_i \in \Omega_i} \Phi_i(z_i, \xi_i) \hat{Q}_i(z_i|\Delta \beta_j(\xi_j)) = \sum_{z_j \in \Omega_j} \Phi_i(z_i, \xi_i) \hat{Q}_j(z_j|\Delta \beta_j(\xi_j)) \]

\[ \sum_{z_j \in \Omega_j} \Phi_i(z_i, \xi_i) \hat{Q}_j(z_j|\Delta \beta_j(\xi_j)) \]
Now we introduce the derivative of \( \Phi_i(a_i, \xi) \) at node \( i \) with respect to the infinitesimal \( \Delta \beta_j(\xi_j) \) as follows:

\[
\langle \xi_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle = \sum_{z_i \in \Omega} \Phi_i(z_i, \xi_j) \widehat{O}_i(z_i) = \sum_{z_i \in \Omega} \Phi_i(z_i, \xi_j) \widehat{O_i}(z_i)
\]

\[
(i, \xi_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle = \sum_{z_i \in \Omega} \Phi_i(z_i, \xi_j) \widehat{O}_i(z_i) = \sum_{z_i \in \Omega} \Phi_i(z_i, \xi_j) \widehat{O_i}(z_i)
\]

(8.25)

By expanding \( \langle \xi_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle \) and \( \langle i, \xi_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle \) in powers of \( \Delta \beta_j(\xi_j) \), \( \Psi_{i,j}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{i,j}(\xi) \) and \( \Psi_{k,y}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{k,y}(\xi) \) and by retaining only the first order terms, we derive

\[
\langle \xi_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle = \sum_{\xi_j \in \Omega} \langle \xi_i | \mathbf{C}_j(\xi_j) \rangle (\Psi_{i,j}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{i,j}(\xi))
\]

(8.27)

\[
(i, \xi_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle = \sum_{\xi_j \in \Omega} \langle i, \xi_i | \mathbf{C}_j(\xi_j) \rangle (\psi_{i,j}(\xi|\Delta \beta_j(\xi_j)) - \psi_{i,j}(\xi))
\]

(8.28)

(8.29)

where

\[
\langle \xi_i | \mathbf{C}_i(\xi'_i) \rangle = \langle \xi'_i | \mathbf{C}_i(\xi'_i) \rangle = \sum_{z_i \in \Omega} \Phi_i(z_i, \xi'_i) \Phi_i(z_i, \xi'_i) \widehat{O}_i(z_i)
\]

\[
\langle i, \xi_i | \mathbf{C}_j(\xi_j) \rangle = \sum_{z_i \in \Omega} \Phi_i(z_i, \xi_j) \Phi_j(z_j, \xi'_j) \widehat{O}_j(z_j)
\]

(8.30)

(8.31)

For a fixed \( \gamma \), we regard Eqs. (8.27) and (8.28) as simultaneous linear equations for unknowns \( \Psi_{i,j}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{i,j}(\xi) \) and \( \Psi_{k,y}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{k,y}(\xi) \), respectively, and express the solutions as follows:

\[
\Psi_{i,j}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{i,j}(\xi) = \sum_{\xi_j \in \Omega} \langle \xi'_i | \mathbf{X}^{-1}_i(\theta) \rangle \langle \xi'_i | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle
\]

(8.32)

\[
\Psi_{k,y}(\xi|\Delta \beta_j(\xi_j)) - \Psi_{k,y}(\xi) = \sum_{\xi_j \in \Omega} \langle i, \xi_i | \mathbf{X}^{-1}_j(\theta) \rangle \langle k, \xi_k | \Delta \mathbf{m}(\Delta \beta_j(\xi_j)) \rangle
\]

(8.33)

where

\[
\langle \xi_i | \mathbf{X}_i(\xi'_i) \rangle = \langle \xi_i | \mathbf{X}_i(\xi'_i) \rangle = \sum_{\xi_i \in \Omega} \langle i, \xi_i | \mathbf{X}_i(\xi_i') \rangle (\theta \in \Omega, |\omega| \geq 0, \xi_i \in \Omega_i \Omega_i \{0\}, \theta \in \Omega, |\omega| \geq 0, \xi_i \in \Omega_i \Omega_i \{0\})
\]

(8.34)

\[
\langle i, \xi_i | \mathbf{X}_j(\xi_j) \rangle = \langle j, \xi_j | \mathbf{X}_j(\xi_j) \rangle = \sum_{\xi_j \in \Omega} \langle i, \xi_i | \mathbf{C}_j(\xi_j) \rangle (\theta \in \Omega, |\omega| \geq 0, \xi_j \in \Omega_j \Omega_j \{0\}, \theta \in \Omega, |\omega| \geq 0, \xi_j \in \Omega_j \Omega_j \{0\})
\]
By substituting Eqs. (8.32) and (8.33) to Eqs. (8.13) and (8.23) and rewriting \( \langle \xi | \Delta m_t(\Delta \beta_j(\xi)) \rangle \) and \( \langle i, \xi | \Delta m_t(\Delta \beta_j(\xi)) \rangle \) as \( \langle i, \xi | \Delta m_t(\Delta \beta_j(\xi)) \rangle \), we can express the result as
\[
(1 - |\partial i|) \sum_{\xi_i \in \mathbb{R} \setminus \{0\}} \langle \xi_i | X_{i-1}^{-1} | \xi_j \rangle \langle \xi_j | \Delta m_t(\Delta \beta_i(\xi_j)) \rangle
+ \sum_{y \in \partial i} \sum_{\{k \in \mathbb{R} | |\partial k| \geq 2\}} \sum_{\gamma \in \mathbb{R} \setminus \{0\}} \langle i, \xi_j | X_{y^{-1}}^{-1} | k, \xi_k \rangle \langle k, \xi_k | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle = \delta_{i,j} \delta_{\partial i, \partial j} \Delta \beta_j(\xi_j)
\]
(8.37)
such that
\[
\sum_{\{k \in \mathbb{R} | |\partial k| \geq 2\}} \sum_{\gamma \in \mathbb{R}} \sum_{y \in \partial i} \langle i, \xi_j | G[k, \xi_k] \rangle \langle k, \xi_k | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle = \delta_{i,j} \delta_{\partial i, \partial j} \Delta \beta_j(\xi_j)
\]
(8.38)
where
\[
\langle i, \xi_j | G[j, \xi_j] \rangle = \delta_{i,j} (1 - |\partial i|)/\langle \xi_j | X_{j-1}^{-1} | \xi_j \rangle + \sum_{\{y \in \partial i, y \neq \gamma \} \setminus \{\partial i\}} \langle i, \xi_j | X_{y^{-1}}^{-1} | j, \xi_j \rangle
\]
(8.39)
From Eqs. (8.4) and (8.38), we obtain approximate values of \( \mathcal{M}^t_{i,j}(\xi_i, \xi_j) - \mathcal{M}^t_{i,j}(\xi_i, \xi_j) \) between pairs of nodes \( i \) and \( j \) with \(|\partial i| \geq 2\) and \(|\partial j| \geq 2\) as follows:
\[
\mathcal{M}^t_{i,j}(\xi_i, \xi_j) - \mathcal{M}^t_{i,j}(\xi_i, \xi_j) \simeq \frac{\langle i, \xi_i | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle}{\Delta \beta_j(\xi_j)} = \langle i, \xi_i | G^{-1}[j, \xi_j] \rangle
\]
(8.40)
in combining the LBP with the linear response formula (8.4).

Next, we consider the computations of \( \mathcal{M}^t_{i,j}(\xi_i, \xi_j) - \mathcal{M}^t_{i,j}(\xi_i, \xi_j) \) for pairs of nodes \( i \) and \( j \) with \(|\partial i| \geq 2\) and \(|\partial j| = 1\). In such cases, we use Eq. (8.29) and solve it with respect to \( \Psi_{i,j}(\xi_i | \Delta \beta_j(\xi_j)) \) as follows:
\[
\Psi_{i,j}(\xi_i | \Delta \beta_j(\xi_j)) = \sum_{\gamma \in \mathbb{R}} \sum_{k \in \mathbb{R} \setminus \{0\}} \sum_{\{\gamma \neq \partial i, \gamma \neq \partial j\}} \langle \xi_i | X_{y^{-1}}^{-1} | k, \xi_k \rangle \langle k, \xi_k | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle - \delta_{i,j} \delta_{\partial i, \partial j} \Delta \beta_j(\xi_j)
\]
(8.41)
By substituting Eq. (8.37) to Eqs. (8.13) and (8.23) and rewriting \( \langle \xi_i | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle \) and \( \langle i, \xi_i | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle \) as \( \langle i, \xi_i | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle \), we can express the result as
\[
\sum_{\{k \in \mathbb{R} | |\partial k| \geq 2\}} \sum_{\gamma \in \mathbb{R} \setminus \{0\}} \langle i, \xi_i | G[k, \xi_k] \rangle \langle k, \xi_k | \Delta m_t(\Delta \beta_j(\xi_j)) \rangle = -\langle i, \xi_i | R[j, \xi_j] \rangle \Delta \beta_j(\xi_j),
\]
(8.42)
where
\[
\langle i, \xi_i | R[j, \xi_j] \rangle = \sum_{\{y \in \mathbb{R} | \gamma \neq \partial i, \gamma \neq \partial j\}} \sum_{\{\gamma \neq \partial i, \gamma \neq \partial j\}} \langle i, \xi_i | X_{y^{-1}}^{-1} | k, \xi_k \rangle \langle k, \xi_k | \Delta \beta_j(\xi_j) \rangle
\]
(8.43)
\[\langle j, \xi \mid R \mid i, \zeta \rangle = \sum_{\{y \in \mathcal{Y} \mid y \in E \}} \sum_{\{|k \in \mathcal{V} \mid \partial k \geq 2\}} \langle j, \xi \mid C_j \mid k, \zeta \rangle \langle k, \zeta \mid X_k^{-1} \mid i, \zeta \rangle \]

(8.44)

We remark that \(\langle i, \zeta \mid R \mid j, \zeta' \rangle\) and \(\langle j, \xi \mid R \mid i, \zeta \rangle\) are always zero for every pair of nodes, \(\{i, j\}\), which does not belong to \(E\), such that \(\{i, j\} \notin E\). From Eqs.(8.4) and (8.42), we obtain approximate values of \(\mathcal{M}^*_\{i,j\}(\xi_i, \zeta_j) - \mathcal{M}^*_\{i,j\}(\xi_i, \zeta_j)\) between pairs of nodes \(i\) and \(j\) with \(|\partial i| \geq 2\) and \(|\partial j| = 1\) as follows:

\[\mathcal{M}^*_\{i,j\}(\xi_i, \zeta_j) - \mathcal{M}^*_\{i,j\}(\xi_i, \zeta_j) \simeq \lim_{\Delta \beta_{j}(\zeta_{j}) \to 0} \frac{\langle i, \zeta \mid \Delta \overline{\mathcal{M}}_{\{j\}}(\Delta \beta_{j}(\zeta_{j})) \rangle}{\Delta \overline{\mathcal{M}}_{\{j\}}(\Delta \beta_{j}(\zeta_{j}))} = \sum_{\{|k \in \mathcal{V} \mid \partial k \geq 2\}} \langle i, \zeta \mid G^{-1} \mid k, \zeta \rangle \langle k, \zeta \mid R \mid j, \zeta' \rangle \]

(8.45)

In the similar way as Eqs.(8.28) and (8.29), we expand \(\langle i, \zeta \mid \Delta \overline{\mathcal{M}}_{\{j\}}(\Delta \beta_{j}(\zeta_{j})) \rangle\) for \(|\partial i| = 1\) in powers of \(\Delta \beta_{j}(\zeta_{j})\) and retain only the first order terms as follows:

\[\langle i, \zeta \mid \Delta \overline{\mathcal{M}}_{\{j\}}(\Delta \beta_{j}(\zeta_{j})) \rangle = \sum_{\{k \in \mathcal{Y} \mid \partial k \geq 2\}} \sum_{\xi \in \mathcal{X}_k} \langle i, \zeta \mid X_k \mid k, \zeta \rangle \langle \Psi_k(y, k, \Delta \beta_{j}(\zeta_{j})) - \Psi_k(y, k, \zeta_{j}) \rangle + (1 - \delta_{\partial i \partial j})(1 \Delta \beta_{j}(\zeta_{j}))
\]

(8.46)

By substituting Eqs.(8.32) and (8.33) to Eq.(8.46) and by using Eqs.(8.4) and (8.40), we have

\[\mathcal{M}^*_\{i,j\}(\xi_i, \zeta_j) - \mathcal{M}^*_\{i,j\}(\xi_i, \zeta_j) \simeq \lim_{\Delta \beta_{j}(\zeta_{j}) \to 0} \frac{\langle i, \zeta \mid \Delta \overline{\mathcal{M}}_{\{j\}}(\Delta \beta_{j}(\zeta_{j})) \rangle}{\Delta \overline{\mathcal{M}}_{\{j\}}(\Delta \beta_{j}(\zeta_{j}))} = \sum_{\{k \in \mathcal{Y} \mid \partial k \geq 2\}} \sum_{\{|l \in \mathcal{V} \mid \partial l \geq 2\}} \sum_{\xi \in \mathcal{X}_k} \sum_{\zeta \in \mathcal{X}_l} \langle i, \zeta \mid R \mid k, \zeta \rangle \langle k, \zeta \mid G^{-1} \mid l, \zeta \rangle \langle l, \zeta \mid R \mid j, \zeta' \rangle \]

(8.47)

9 Statistical Mechanical Informatics

In this section, we show some applications and extensions of the loopy belief propagation to some probabilistic models in th statistical-mechanical informatcs. The loopy belief propagation is basically equivalent to the Bethe approximation.

9.1 Energy, Entropy and Free Energy in Canonical Distribution

For a given energy function \(H(a) = H(a_1, a_2, \cdots, a_{|V|})\), we consider a distribution in which all state vectors \(a = (a_1, a_2, \cdots, a_{|V|})\) with their constraints \(H(a) = U\) satisfied give the same probability and the probability for the other state vectors are zero. This probability is expressed as follows:

\[\Pr\{A_1 = a_1, A_2 = a_2, \cdots, A_{|V|} = a_{|V|} \mid U\} = \frac{\delta_{H(a_1, a_2, \cdots, a_{|V|}) U}}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{H(z_1, z_2, \cdots, z_{|V|}) U}} = \int_{-\infty}^{\infty} \delta(H(a_1, a_2, \cdots, a_{|V|}) - U - \xi) \, d\xi. \]

(9.1)

Without loss of generality, we can choose that \(H(a)\) is always positive. The number of states which takes \(U\) as an energy is expressed as

\[\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{H(z_1, z_2, \cdots, z_{|V|}) U} = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \int_{-\infty}^{\infty} \delta(H(z_1, z_2, \cdots, z_{|V|}) - U - \xi) \, d\xi. \]

(9.2)
In the statistical mechanics, Boltzmann introduced an entropy \( S(U) \) by using the logarithm of the number of states as follows:

\[
S(U) \equiv k_B \ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{H(z_1, z_2, \ldots, z_{|V|}), U} \right)
\]

\[
= k_B \ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \int_{-\infty}^{+\infty} \delta (H(z_1, z_2, \ldots, z_{|V|}) - U - \xi) \, d\xi \right).
\]

\[(9.3)\]

The entropy means a certainty of states. When probabilities of all the possible states are equivalent of each other, increasing of the number of all the possible states leads growing of uncertainty, such that of the entropy. On the other hand, it is known that the entropy of an isolated system always increase and the probability of decreasing the entropy is almost equal to zero in the second law of thermodynamics. \( \Pr\{A_1 = a_1, A_2 = a_2, \ldots, A_{|V|} = a_{|V|}, U\} \) of Eq.(9.1) is referred to as a Micro canonical distribution in the statistical mechanics. Here \( k_B \) is referred to as a Boltzmann constant, and for material sciences,

\[
k_B = \frac{R}{N_A}
\]

(9.4)

where \( R \) is a gas constant and \( N_A \) is an Avogadro constant. However without loss of generality, we set \( k_B = 1 \) in the statistical-mechanical informatics. Here we introduce

\[
T(U) \equiv \left( \frac{\partial}{\partial U} S(U) \right)^{-1},
\]

(9.5)

where \( \lim_{|V| \to +\infty} T(U) \) is referred to as a Temperature in the statistical mechanics.

We introduce a Laplace transformation of the number of states with respect to \( U \) as follows:

\[
n(\beta) \equiv \int_{0}^{+\infty} e^{-\beta U} \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \int_{-\infty}^{+\infty} \delta (H(z_1, z_2, \ldots, z_{|V|}) - U - \xi) \, d\xi \, dU
\]

\[
= \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \exp (-\beta H(z_1, z_2, \ldots, z_{|V|}))
\]

(9.6)

By taking an inverse Laplace transformation of \( n(\beta) \), we can express the number of states in terms of the following complex integral representation:

\[
\sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{H(z_1, z_2, \ldots, z_{|V|}), U} = \lim_{\lambda \to +\infty} \lim_{-\lambda \to +\infty} \int_{-\infty}^{+\infty} e^{\beta U} n(\beta) \, d\beta = \lim_{\lambda \to +\infty} \int_{-\lambda}^{+\lambda} e^{\beta U} g(\beta) \, d\beta,
\]

(9.7)

where

\[
g(\beta) \equiv \beta + \frac{1}{U} \ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \exp (-\beta H(z_1, z_2, \ldots, z_{|V|})) \right).
\]

(9.8)

When we choose \( U \) and \( H(z_1, z_2, \ldots, z_{|V|}) \) so as to satisfy

\[
\ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \exp (-\beta H(z_1, z_2, \ldots, z_{|V|})) \right) = \mathcal{O}(|V|) \ (|V| \to +\infty),
\]

(9.9)

\[
\frac{1}{U} \ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \exp (-\beta H(z_1, z_2, \ldots, z_{|V|})) \right) = \mathcal{O}(1) \ (|V| \to +\infty),
\]

(9.10)

we obtain

\[
\frac{1}{|V|} \ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \delta_{H(z_1, z_2, \ldots, z_{|V|}), U} \right) = \frac{1}{|V|} \left( \hat{\beta}(U) U + \ln \left( \sum_{z_1 \in \Omega_1, z_2 \in \Omega_2, \ldots} \sum_{z_{|V|} \in \Omega_{|V|}} \exp (-\hat{\beta}(U) H(z_1, z_2, \ldots, z_{|V|})) \right) \right)
\]

\[
+ \mathcal{O}(1) \ (|V| \to +\infty),
\]

(9.11)
where

$$
\frac{\sum \sum \cdots \sum \sum H(z_1, z_2, \ldots, z_{|V|}) \exp \left( -\tilde{\beta}(U)H(z_1, z_2, \ldots, z_{|V|}) \right)}{\sum \sum \cdots \sum \sum \exp \left( -\tilde{\beta}(U)H(z_1, z_2, \ldots, z_{|V|}) \right)} = U. \quad (9.12)
$$

We can express the entropy in terms of $\tilde{\beta}(U)$ as

$$
\lim_{|V| \to +\infty} S(U) = \lim_{|V| \to +\infty} k_B \left( \tilde{\beta}(U)U + \ln \left( \sum \sum \cdots \sum \exp \left( -\tilde{\beta}(U)H(z_1, z_2, \ldots, z_{|V|}) \right) \right) \right), \quad (9.13)
$$

and can derive the following equality:

$$
\frac{\partial}{\partial U} \left( \lim_{|V| \to +\infty} S(U) \right) = \lim_{|V| \to +\infty} \frac{\partial}{\partial U} \left( k_B \tilde{\beta}(U)U + k_B \ln \left( \sum \sum \cdots \sum \exp \left( -\tilde{\beta}(U)H(z_1, z_2, \ldots, z_{|V|}) \right) \right) \right)
\quad = \lim_{|V| \to +\infty} \left( k_B \tilde{\beta}(U) + k_B U \left( \frac{\partial}{\partial U} \tilde{\beta}(U) \right) \right)
\quad - k_B \left( \frac{\partial}{\partial U} \tilde{\beta}(U) \right) \sum \sum \cdots \sum \sum H(z_1, z_2, \ldots, z_{|V|}) \exp \left( -\tilde{\beta}(U)H(z_1, z_2, \ldots, z_{|V|}) \right)
\quad = \lim_{|V| \to +\infty} \left( k_B \tilde{\beta}(U) + k_B U \left( \frac{\partial}{\partial U} \tilde{\beta}(U) \right) \right) - k_B \left( \frac{\partial}{\partial U} \tilde{\beta}(U) \right) U = k_B \lim_{|V| \to +\infty} \tilde{\beta}(U). \quad (9.14)
$$

By comparing Eq.(9.14) with Eq.(9.5), the relationship between $T(U)$ and $\tilde{\beta}(U)$ is expressed as

$$
\lim_{|V| \to +\infty} \tilde{\beta}(U) = \lim_{|V| \to +\infty} \frac{1}{k_B T(U)}. \quad (9.15)
$$

By substituting Eq.(9.12) to Eq.(9.11), we derive the following equality in the limit of $|V| \to +\infty$:

$$
\frac{1}{|V|} \ln \left( \sum \sum \cdots \sum \sum \delta_{H(z_1, z_2, \ldots, z_{|V|}) U} \right) = -\frac{1}{|V|} \sum \sum \cdots \sum \sum P(z_1, z_2, \ldots, z_{|V|} | \tilde{T}(U))
\quad \times \ln \left( P(z_1, z_2, \ldots, z_{|V|} | \tilde{T}(U)) \right)
\quad + O(|V|^{-1}) \quad (|V| \to +\infty) \quad (9.16)
$$

where

$$
P(a_1, a_2, \ldots, a_{|V|} | T) \equiv \frac{\exp \left( -\frac{1}{k_B} H(a_1, a_2, \ldots, a_{|V|}) \right)}{\sum \sum \cdots \sum \sum \exp \left( -\frac{1}{k_B} H(z_1, z_2, \ldots, z_{|V|}) \right)} \quad (9.17)
$$

The new probability distribution $P(a_1, a_2, \ldots, a_{|V|} | T)$ is referred to as a Canonical Distribution or a Boltzmann Distribution in the statistical mechanics. We remark that this distribution has been appeared in the present review as one of representations in posterior probability distributions also in Eqs.(1.33)-(1.34).

From Eq.(9.12) and (9.17), the energy $U$ in the canonical distribution is expressed in terms of $T$ as follows:

$$
U(T) = \sum \sum \cdots \sum H(z_1, z_2, \ldots, z_{|V|}) P(z_1, z_2, \ldots, z_{|V|} | T). \quad (9.18)
$$

By comparing Eq.(9.3) with Eq.(9.16), we can define the entropy of the canonical distribution by

$$
S(T) \equiv -k_B \sum \sum \cdots \sum P(z_1, z_2, \ldots, z_{|V|} | T) \ln \left( P(z_1, z_2, \ldots, z_{|V|} | T) \right), \quad (9.19)
$$

as a function of $T$. Now we regard $T$ as an independent variable in the probability distribution $P(a_1, a_2, \ldots, a_{|V|} | T)$.
although \( U \) is an independent variable in the micro-canonical probability distribution in Eq. (9.3). In the thermodynamics, the Helmholtz free energy \( F(T) \) is given by
\[
F(T) = U(T) - TS(T),
\]
and by using Eqs. (9.18) and (9.19) to Eq. (9.20), we derive the representation of the Helmholtz Free Energy in terms of the canonical distribution as follows:
\[
F(T) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} H(z_1, z_2, \cdots, z_{|V|}) P(z_1, z_2, \cdots, z_{|V|}|T)
+ k_B T \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} P(z_1, z_2, \cdots, z_{|V|}|T) \ln \left( P(z_1, z_2, \cdots, z_{|V|}|T) \right).
\]
(9.21)

By substituting Eq. (9.17) to Eq. (9.21), we have
\[
F(T) = - \ln \left( \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \exp \left( - \frac{1}{k_B T} H(z_1, z_2, \cdots, z_{|V|}) \right) \right).
\]
(9.22)

From Eq. (9.21), we introduce a functional of every trial probability distribution \( Q(a) = Q(a_1, a_2, \cdots, a_{|V|}) \) as
\[
\mathcal{F}[Q] = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} H(z_1, z_2, \cdots, z_{|V|}) Q(z_1, z_2, \cdots, z_{|V|})
+ k_B T \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} Q(z_1, z_2, \cdots, z_{|V|}) \ln \left( Q(z_1, z_2, \cdots, z_{|V|}) \right),
\]
(9.23)
and we consider the minimization of the free energy functional under the normalization condition
\[
\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} Q(z_1, z_2, \cdots, z_{|V|}) = 1
\]
as follows:
\[
\mathcal{Q}(a_1, a_2, \cdots, a_{|V|}) = \min_Q \left\{ \mathcal{F}[Q] \left| \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} Q(z_1, z_2, \cdots, z_{|V|}) = 1 \right. \right\}
\]
(9.24)
We introduce a Langrange multiplier \( \lambda \) for
\[
\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} Q(z_1, z_2, \cdots, z_{|V|}) = 1
\]
as follows:
\[
\mathcal{L}[Q] = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} H(z_1, z_2, \cdots, z_{|V|}) Q(z_1, z_2, \cdots, z_{|V|})
+ k_B T \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} Q(z_1, z_2, \cdots, z_{|V|}) \ln \left( Q(z_1, z_2, \cdots, z_{|V|}) \right)
+ \lambda \left( \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} Q(z_1, z_2, \cdots, z_{|V|}) - 1 \right),
\]
(9.25)
The extremum condition of \( \mathcal{L}[Q] \) with respect to \( Q(a_1, a_2, \cdots, a_{|V|}) \)
\[
\frac{\partial \mathcal{L}[Q]}{\partial Q(a_1, a_2, \cdots, a_{|V|})} \bigg|_{Q(a_1, a_2, \cdots, a_{|V|}) = \mathcal{Q}(a_1, a_2, \cdots, a_{|V|})} = 0
\]
leads to
\[
H(a_1, a_2, \cdots, a_{|V|}) + k_B T \left( 1 + \ln \left( \mathcal{Q}(a_1, a_2, \cdots, a_{|V|}) \right) \right) + \lambda = 0
\]
(9.26)
such that
\[
\mathcal{Q}(a_1, a_2, \cdots, a_{|V|}) = \exp \left( - 1 - \frac{1}{k_B T} \lambda - \frac{1}{k_B T} H(a_1, a_2, \cdots, a_{|V|}) \right).
\]
(9.27)
The Langrange multiplier \( \lambda \) is determined so as to satisfy the normalization condition
\[
\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \mathcal{Q}(z_1, z_2, \cdots, z_{|V|}) = 1
\]
and then the solution of the constrained minimization problem (9.24) is given
by
\[
\tilde{Q}(a_1, a_2, \cdots, a_V) = \frac{\exp\left(-\frac{1}{\beta} H(a_1, a_2, \cdots, a_V)\right)}{\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} \exp\left(-\frac{1}{\beta} H(z_1, z_2, \cdots, z_V)\right)} \quad (9.29)
\]

It means that the constrained minimization problem (9.24) give us the canonical distribution (9.17).

9.2 Mean-Field Method

We consider a probability distribution \( P(a) \) defined by Eqs.(3.18) and (3.19), where \( V \) is the set of all the nodes and \( E \) is the set of all the edges. We introduce the following trial probability distribution:
\[
Q(a) = Q(a_1, a_2, \cdots, a_V) = \prod_{i \in V} Q_i(a_i),
\]
where
\[
Q_i(a_i) = \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} \delta_{z_i, z_i} Q(z_1, z_2, \cdots, z_V) (i \in V).
\]

We can derive the KL divergence KL\([P||Q]\) between \( P(a) \) in Eqs.(3.18)-(3.19) and \( Q(a) \) in Eq.(9.30) as
\[
KL[P||Q] = -\sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i))
\]
\[
- \sum_{i,j \in E} \sum_{z_i \in \Omega_i} Q_i(z_i) Q_j(z_j) \ln(W_{i,j}(z_i, z_j))
\]
\[
+ \sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i))
\]
\[
+ \ln\left( \sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_V \in \Omega_V} \prod_{i \in V} W_i(z_i) \right) \left( \prod_{(i,j) \in E} W_{i,j}(z_i, z_j) \right),
\]
so that we have
\[
KL[P||Q] = \mathcal{F}_{MF}[\{Q_i|i \in V\}] - F,
\]
where
\[
\mathcal{F}_{MF}[\{Q_i|i \in V\}] = -\sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(W_i(z_i))
\]
\[
- \sum_{(i,j) \in E} \sum_{z_i \in \Omega_i} Q_i(z_i) Q_j(z_j) \ln(W_{i,j}(z_i, z_j))
\]
\[
+ \sum_{i \in V} \sum_{z_i \in \Omega_i} Q_i(z_i) \ln(Q_i(z_i)).
\]

Here \( F \equiv -\ln(Z) \) corresponds to the free energy of the system defined by Eqs.(3.18) and (3.19). We remark that \( \mathcal{F}_{MF}[\{Q_i|i \in V\}] \) in Eq.(9.34) is referred to as a Mean-Field Free Energy Functional. The set of trial marginal probability distribution, \( \{\tilde{Q}(a_i)|i \in V, a_i \in \Omega_i\} \), is determined so as to minimize the KL divergence KL\([P||Q]\) under the normalization conditions of each marginal probability distribution \( Q_i(a_i) \) as
\[
\{\tilde{Q}(a_i)|i \in V, a_i \in \Omega_i\} = \arg \min_{\{Q_i|i \in V, a_i \in \Omega_i\}} \left\{KL[P||Q] \sum_{z_i \in \Omega_i} Q_i(z_i) = 1 \right\}
\]
\[
= \arg \min_{\{Q_i|i \in V, a_i \in \Omega_i\}} \left\{\mathcal{F}_{MF}[\{Q_i|i \in V\}] \sum_{z_i \in \Omega_i} Q_i(z_i) = 1 \right\}.
\]

We introduce the Lagrange multiplier \( \lambda_i \) for the normalization condition for \( Q_i(a_i) \) for every \( i \in V \) as follows:
\[
\mathcal{L}[\{Q_i|i \in V\}] = \mathcal{F}_{MF}[\{Q_i|i \in V\}] - \sum_{i \in V} (\lambda_i + 1) \left( \sum_{z_i \in \Omega_i} Q_i(z_i) - 1 \right),
\]
and take the extremum condition:

\[
\left[ \frac{\partial}{\partial \hat{Q}_i(a_i)} \mathcal{L}[\{\hat{Q}_i|i \in \Omega\}] \right]_{\hat{Q}_i(a_i)=\hat{Q}_i(a_1),\hat{Q}_i(a_2)=\hat{Q}_i(a_2),\ldots,\hat{Q}_i(a_V)=\hat{Q}_i(a_V)} = 0
\]

\[
(i \in \Omega, x = (a_1, a_2, \ldots, a_V) \in \Omega), \tag{9.37}
\]

\[
\sum_{\{i,j\} \in \partial i} \sum_{\Omega} \hat{Q}_i(z_i) \ln (W_{i,j}(a_i, z_j)) + \ln (W_i(a_i)) - \ln (\hat{Q}_i(a_i)) + \hat{\lambda}_i = 0, \tag{9.38}
\]

\[
\hat{Q}_i(a_i) = \exp \left( \hat{\lambda}_i + \sum_{\{i,j\} \in \partial i} \sum_{\Omega} \hat{Q}_j(z_j) \ln (W_{i,j}(a_i, z_j)) \right). \tag{9.39}
\]

Each Lagrange multiplier \(\hat{\lambda}_i\) is determined so as to satisfy the normalization condition \(\sum_{z_i \in \Omega} \hat{Q}_i(z_i) = 1\) and the trial probability distribution \(\hat{Q}(a)\) is given as

\[
\hat{Q}(a) = \hat{Q}(a_1, a_2, \ldots, a_V) = \prod_{i \in \Omega} \hat{Q}_i(a_i), \tag{9.40}
\]

where

\[
\hat{Q}_i(a_i) = \exp \left( \hat{\lambda}_i + \ln (W_i(a_i)) + \sum_{\{i,j\} \in \partial i} \sum_{\Omega} \hat{Q}_j(z_j) \ln (W_{i,j}(a_i, z_j)) \right), \tag{9.41}
\]

\[
\hat{\lambda}_i = -\ln \left( \sum_{z_i \in \Omega} \exp \left( \ln (W_i(z_i)) + \sum_{\{i,j\} \in \partial i} \sum_{\Omega} \hat{Q}_j(z_j) \ln (W_{i,j}(z_i, z_j)) \right) \right). \tag{9.42}
\]

### Procedures for Computation of Approximate Marginals \(\hat{Q}_i(a_i)\)

**Step 1:** Set the marginal probability distributions \(\hat{Q}_i(a_i)\) \((i \in V, a_i \in \Omega)\) as initial values.

**Step 2:** Repeat the following updates until all marginals \(\hat{Q}_i(a_i)\) \((i \in V, a_i \in \Omega)\) converge:

\[
\hat{\lambda}_i \leftarrow -\ln \left( \sum_{z_i \in \Omega} W_i(z_i) \exp \left( \sum_{\{i,j\} \in \partial i} \sum_{\Omega} \hat{Q}_j(z_j) \ln (W_{i,j}(z_i, z_j)) \right) \right) \quad (i \in V), \tag{9.43}
\]

\[
\hat{Q}_i(a_i) \leftarrow W_i(a_i) \exp \left( \hat{\lambda}_i + \sum_{\{i,j\} \in \partial i} \sum_{\Omega} \hat{Q}_j(z_j) \ln (W_{i,j}(a_i, z_j)) \right) \quad (i \in V, a_i \in \Omega). \tag{9.44}
\]

### 9.3 Mean Field Approximation and Loopy Belief Propagation for Ising Model

Our probabilistic model in the present section is given in terms of the probability distribution in Eqs. (3.18) and (3.19) on a graph \(\Gamma(V,E)\) which includes some cycles. However, we restrict the state space \(\Omega_i\) at each node \(i\) to \(\{+1, -1\}\). By introducing the following Fourier expansions:

\[
\ln (W_i(a_i)) = \frac{1}{2} \bar{\nu}_i(0) + \frac{1}{2} \bar{\nu}_i(1) a_i, \tag{9.45}
\]

\[
\ln (W_{i,j}(a_i, a_j)) = \frac{1}{4} \bar{\nu}_{i,j}(0,0) + \frac{1}{4} \bar{\nu}_{i,j}(1,0) a_i + \frac{1}{4} \bar{\nu}_{i,j}(0,1) a_j + \frac{1}{4} \bar{\nu}_{i,j}(1,1) a_i a_j, \tag{9.46}
\]

where

\[
\bar{\nu}_i(\xi) \equiv \sum_{\xi \in \{0,1\}} \xi^i \ln (W_i(\xi)) \quad (\xi \in \{0,1\}) \tag{9.47}
\]

\[
\bar{\nu}_{i,j}(\xi, \xi') = \sum_{\xi \in \{0,1\}} \sum_{\xi' \in \{0,1\}} \xi^i \xi'^j \ln (W_{i,j}(\xi, \xi')) \quad (\xi \in \{0,1\}, \xi' \in \{0,1\}), \tag{9.48}
\]
the probability distribution (3.18) with (3.19) can be rewritten as

\[
P(a) = P(a_1, a_2, \cdots, a_{|V|}) = \frac{\exp\left(-H(a_1, a_2, \cdots, a_{|V|})\right)}{\sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} \exp\left(-H(a_1, a_2, \cdots, a_{|V|})\right)}
\]  

(9.49)

where

\[
H(a) = H(a_1, a_2, \cdots, a_{|V|}) = -\sum_{i \in V} \beta_i a_i - \sum_{(i, j) \in E} \alpha_{(i, j)} a_i a_j,
\]

(9.50)

\[
\beta_i = \frac{1}{2} \bar{\eta}_i(1) + \frac{1}{4} \sum_{\{k, j\} \in \Delta i, \zeta \in \{0, 1\}} \bar{\eta}_{k,j}(\zeta, \zeta) \delta_{\zeta, 1} \delta_{\zeta(\{k, j\}, 0)}.
\]

(9.51)

\[
\alpha_{(i, j)} = \frac{1}{4} \bar{\eta}_{(i, j)}(1, 1).
\]

(9.52)

The probabilistic model in Eq.(9.49) is referred to as Ising model in the statistical-mechanical informatics. For the probabilistic model in Eq.(9.49), \(F(\alpha, \beta)\) is introduced as follows:

\[
F(\alpha, \beta) \equiv -\ln \left( \frac{\sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} \exp(-H(z_1, z_2, \cdots, z_{|V|}))}{\sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} \exp(-H(z_1, z_2, \cdots, z_{|V|}))} \right). 
\]

(9.53)

In terms of derivatives \(F(\alpha, \beta)\) in Eq.(9.53) with respect to \(\beta_i\) and \(\alpha_{(i, j)}\), the expectation values of \(\bar{m}_i(\alpha, \beta)\) and \(\bar{m}_{(i, j)}(\alpha, \beta)\) are expressed as

\[
\bar{m}_i(\alpha, \beta) \equiv \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} z_i P(z_1, z_2, \cdots, z_{|V|}) = -\frac{\partial}{\partial \beta_i} F(\alpha, \beta) \quad (i \in V),
\]

(9.54)

\[
\bar{m}_{(i, j)}(\alpha, \beta) \equiv \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} z_i z_j P(z_1, z_2, \cdots, z_{|V|}) = -\frac{\partial}{\partial \alpha_{(i, j)}} F(\alpha, \beta) \quad \{(i, j) \in E\}.
\]

(9.55)

The probability distribution \(P(a|\alpha, \beta)\) minimizes the free energy functional \(\mathcal{F}[Q]\) defined by

\[
\mathcal{F}[Q] \equiv \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} H(z_1, z_2, \cdots, z_{|V|}) Q(z_1, z_2, \cdots, z_{|V|})
\]

+ \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} Q(z_1, z_2, \cdots, z_{|V|}) \ln Q(z_1, z_2, \cdots, z_{|V|}),

(9.56)

under the normalization condition \(\sum_{z_1 \in \Omega_1} \sum_{z_2 \in \Omega_2} \cdots \sum_{z_{|V|} \in \Omega_{|V|}} \hat{Q}(z_1, z_2, \cdots, z_{|V|}) = 1\) is imposed as a constraint condition, such that

\[
P(a|\alpha, \beta) = \arg \min_{\hat{Q}(\alpha)} \left\{ \mathcal{F}[Q] \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} Q(z_1, z_2, \cdots, z_{|V|}) = 1 \right\}.
\]

(9.57)

The free energy functional \(\mathcal{F}[Q] = F(\alpha, \beta)\) for \(Q(a_1, a_2, \cdots, a_{|V|}) = P(a_1, a_2, \cdots, a_{|V|})\)

\((a_1 \in \{+1, -1\}, a_2 \in \{+1, -1\}, \cdots, a_{|V|} \in \{+1, -1\})\) in the statistical-mechanical informatics, \(F(\alpha, \beta)\) is referred to as the free energy of the system defined by the probability distribution \(P(a|\alpha, \beta)\) in Eqs.(9.49)-(9.50)

We introduce a probability distribution

\[
P(a|\Delta \beta_j) = P(a_1, a_2, \cdots, a_{|V|}|\Delta \beta_j) \equiv \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} \frac{\exp(\beta_j z_j) P(a_1, a_2, \cdots, a_{|V|})}{\sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} \exp(\beta_j z_j) P(z_1, z_2, \cdots, z_{|V|})}.
\]

(9.58)

by replacing \(\beta_j\) by \(\beta_j + \Delta \beta_j\) at a fixed node \(j\) in the probability distribution \(P(a|\alpha, \beta)\). For an infinitesimal derivative \(\Delta \beta_j\) in the probability distribution \(P(a|\alpha, \beta + \Delta \beta_j)\), the average \(\bar{m}_i(\alpha, \beta + \Delta \beta_j)\) is derived as follows:

\[
\bar{m}_i(\alpha, \beta + \Delta \beta_j) \equiv \sum_{z_1 \in \{+1, -1\}, z_2 \in \{+1, -1\}, \cdots, z_{|V|} \in \{+1, -1\}} z_i P(z_1, z_2, \cdots, z_{|V|}|\Delta \beta_j).
\]

52
By substituting Eq. (9.58),

\[
\begin{align*}
\sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i \left( 1 + \Delta \beta_j z_j + O(\Delta \beta_j^2) \right) P(z_1, z_2, \ldots, z_{|V|}) \\
= m_i^1(\alpha, \beta) + m_i^{(i,j)}(\alpha, \beta) \Delta \beta_j + O(\Delta \beta_j^2) \\
= m_i^1(\alpha, \beta) + m_i^{(i,j)}(\alpha, \beta) \Delta \beta_j + O(\Delta \beta_j^2) \\
= m_i^1(\alpha, \beta) + m_i^{(i,j)}(\alpha, \beta) \Delta \beta_j + O(\Delta \beta_j^2) (\Delta \beta_j \rightarrow 0).
\end{align*}
\]

Eq. (9.59), can be reduced to the following linear response formulas:

\[
\begin{align*}
\bar{m}^{(i,j)}(\alpha, \beta) - m_i^1(\alpha, \beta) m_j^1(\alpha, \beta) &= \lim_{\Delta \beta_j \to 0} \frac{1}{\Delta \beta_j} \left( \sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i P(z_1, z_2, \ldots, z_{|V|}) \right) \\
&- \sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i P(z_1, z_2, \ldots, z_{|V|}) \right) (i \in V, j \in V). \quad (9.60)
\end{align*}
\]

This linear response formula means that we can compute expectation values \( \bar{m}^{(i,j)}(\alpha, \beta) - m_i^1(\alpha, \beta) m_j^1(\alpha, \beta) \) for all the pairs of nodes \( i(\in V) \) and \( j(\in V) \) by considering the first derivatives of the expectation values \( \sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i P(z_1, z_2, \ldots, z_{|V|}) \) with respect of \( \Delta \beta_j \) for all the nodes \( i(\in V) \).

By setting \( W_{(i,j)}(a_i, a_j) \) and \( W_i(a_i) \) as

\[
\begin{align*}
W_{(i,j)}(a_i, a_j) &\equiv \exp(\alpha_{(i,j)} a_i a_j), \quad (9.61) \\
W_i(a_i) &\equiv \exp(\beta_i a_i), \quad (9.62)
\end{align*}
\]

the probability density function (9.49) can be reduced to Eq. (3.18). By substituting Eqs. (9.61) and (9.62) to Eqs. (9.41) and (9.42), the approximate marginal probability distribution in the mean-field method can be derived as follows:

\[
\begin{align*}
\hat{Q}_i(a_i) &= \exp \left( \beta_i + \sum_{j \in \partial i} \alpha_{(i,j)} \hat{m}_j(\alpha, \beta) a_i \right) \quad (i \in V), \quad (9.63) \\
\hat{m}_i(\alpha, \beta) &\equiv \sum_{z_i \in \{+1,-1\}} z_i \hat{Q}_i(z_i) \quad (i \in V). \quad (9.64)
\end{align*}
\]

By substituting Eq. (9.63) to Eq. (9.64), we can reduce the deterministic equations of \( \{a_i\} \) to the following simultaneous fixed-point equations as

\[
\hat{m}_i(\alpha, \beta) = \tanh \left( \beta_i + \sum_{j \in \partial i} \alpha_{(i,j)} \hat{m}_j(\alpha, \beta) \right) \quad (i \in V). \quad (9.65)
\]

By defining \( \Omega_i \equiv \{+1,-1\} \) and \( \overline{\Omega}_i \equiv \{0,1\} \), we introduce a set of orthonormal polynomials \( \{\Phi_i(\zeta, a_i) \} \) for each node \( i(\in V) \) as

\[
\begin{align*}
\Phi_i(0, a_i) &\equiv \frac{1}{\sqrt{2}}, \quad \Phi_i(1, a_i) \equiv \frac{1}{\sqrt{2}} a_i \quad (a_i \in \{+1,-1\}), \quad (9.66)
\end{align*}
\]

The orthonormal polynomials satisfy the following relations:

\[
\sum_{z_i \in \{+1,-1\}} \Phi_i(\zeta, z_i) \Phi_i'(\zeta', z_i) = \delta_{\zeta, \zeta'} \quad (\zeta \in \{0,1\}, \zeta' \in \{0,1\}, i \in V). \quad (9.67)
\]

By using these orthonormal polynomials, the marginal probability distribution \( Q_i(a_i) \) in Eq. (9.31) can be expressed as
the following orthonormal expansion:

\[ Q_i(a_i) = \frac{1}{\sqrt{2}} \Phi_i(0, a_i) + \frac{1}{\sqrt{2}} \bar{m}_i \Phi_i(0, a_i) = \frac{1}{2} (1 + \bar{m}_i a_i) \ (i \in V) \]  \ (9.68)

where

\[ \bar{m}_i \equiv \sqrt{2} \sum_{z_i \in \{-1, 1\}} \Phi_i(z_i) Q_i(z_i) = \sum_{z_i \in \{-1, 1\}} z_i Q_i(z_i). \]  \ (9.69)

We remark that \[ \sum_{z_i \in \{-1, 1\}} \Phi_i(z_i) Q_i(z_i) = \frac{1}{\sqrt{2}} \] because \( Q_i(a_i) \) should satisfy the normalization constant

\[ \sum_{z_i \in \{-1, 1\}} Q_i(z_i) = 1 \] for every node \( i \in \langle V \rangle \).

By substituting Eq. (9.68) to Eq. (9.34), we have

\[ \mathcal{F}_\text{MF}[\{ Q_i \mid i \in V \}] = \mathcal{F}_\text{MF}[\{ \bar{m}_i \mid i \in V \}] \]

\[ = - \sum_{i \in V} \beta_i \bar{m}_i - \sum_{\{ i,j \} \in E} \alpha_{\{i,j\}} \bar{m}_i \bar{m}_j + \sum_{i \in V, z_i \in \{-1, 1\}} \frac{1}{2} (1 + \bar{m}_i z_i) \ln \left( \frac{1}{2} (1 + \bar{m}_i z_i) \right). \]  \ (9.70)

Instead of the conditional minimization (9.35) of the mean-field free energy functional \( \mathcal{F}_\text{MF}[\{ Q_i \mid i \in V \}] \) under the normalization condition of \( Q_i(a_i) \), we consider the minimization of \( \mathcal{F}_\text{MF}[\{ \bar{m}_i \mid i \in V \}] \) as follows:

\[ \{ \bar{m}_i(\alpha, \beta) \mid i \in V \} = \arg \min_{\{ \bar{m}_i \mid i \in V \}} \mathcal{F}_\text{MF}[\{ \bar{m}_i \mid i \in V \}]. \]  \ (9.71)

From the above minimization, we derive the same simultaneous mean-field fixed point equations to determine the set \( \{ \bar{m}_i \mid i \in V \} \) as Eq. (9.65).

In the mean-field method, the approximate values \( \hat{u}_{\{i,j\}}(\alpha, \beta) \) of the correlation functions in Eq. (9.55) are given as

\[ \hat{u}_{\{i,j\}}(\alpha, \beta) = \sum_{z_i \in \{-1, 1\}} \sum_{z_j \in \{-1, 1\}} \cdots \sum_{z_V \in \{-1, 1\}} \prod_{k \in V} \hat{Q}_k(z_k) \]

\[ = \left( \prod_{k=1}^{j-1} \sum_{z_k \in \{-1, 1\}} \hat{Q}_k(z_k) \right) \left( \sum_{z_i \in \{-1, 1\}} z_i \hat{Q}_i(z_i) \right) \left( \prod_{k=j+1}^{j-1} \sum_{z_k \in \{-1, 1\}} \hat{Q}_k(z_k) \right) \]

\[ = \hat{m}_i(\alpha, \beta) \hat{m}_j(\alpha, \beta) \ (i \in V, j \in V) \]  \ (9.72)

It means that the covariances \( \hat{u}_{\{i,j\}}(\alpha, \beta) - \hat{m}_i(\alpha, \beta) \hat{m}_j(\alpha, \beta) \) between any pairs of nodes \( i \) and \( j \) are always zero and is one of difficulties of applications for statistical inferences by means of the mean-field method.

The approximate one-body marginal probabilities in Eq. (9.63) are replaced by

\[ \hat{Q}_i(a_i | \Delta \beta_j) = \frac{\exp \left( \beta_i + \delta_i \Delta \beta_j + \sum_{\{i,k\} \in \partial i} \alpha_{\{i,k\}} (\hat{m}_j(\alpha, \beta) + \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j)) a_i \right)}{\sum_{z_i \in \{-1, 1\}} \exp \left( \left( \beta_i + \delta_i \Delta \beta_j + \sum_{\{i,k\} \in \partial i} \alpha_{\{i,k\}} (\hat{m}_j(\alpha, \beta) + \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j)) z_i \right) \right)} \ (i \in V). \]  \ (9.73)

for the probability distribution \( P(a | \Delta \beta_j) \) in Eq. (9.58).

For an infinitesimal \( \Delta \beta_j \), the numerator of the left-hand side in Eq. (9.73) is expanded to

\[ \exp \left( \beta_i + \delta_i \Delta \beta_j + \sum_{\{i,k\} \in \partial i} \alpha_{\{i,k\}} (\hat{m}_j(\alpha, \beta) + \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j)) a_i \right) \]

\[ = \left( 1 + \left( \delta_i \Delta \beta_j + \sum_{\{i,k\} \in \partial i} \alpha_{\{i,k\}} \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) \right) z_i + \mathcal{O} ( (\Delta \beta_j)^2 ) \right) \]

\[ \times \exp \left( (\beta_i + \sum_{\{i,k\} \in \partial i} \alpha_{\{i,k\}} \hat{m}_j(\alpha, \beta) a_i) (\Delta \beta_j \to 0) \right). \]  \ (9.74)
By using it, the expectation value \( \sum_{z_i \in \Omega} \alpha_i \hat{Q}_i(z_i | \Delta \beta_j) \) of \( a_i \) is expanded to

\[
\hat{m}_i(\alpha, \beta) + \Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j) \equiv \sum_{z_i \in \{+1, -1\}} \alpha_i \hat{Q}_i(z_i | \Delta \beta_j) \\
= \sum_{z_i \in \{+1, -1\}} z_i \exp \left( \beta_i + \sum_{\{k \in \partial i\}} \alpha_{i,k} (\hat{m}_i(\alpha, \beta) + \Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j)) z_j \right) \\
= \left( \sum_{z_i \in \{+1, -1\}} z_i \hat{Q}_i(z_i) + \sum_{\{k \in \partial i\}} \alpha_{i,k} \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) + O((\Delta \beta_j)^2) \right) \\
\times \left( 1 + \sum_{\{k \in \partial i\}} \alpha_{i,k} \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) \right) \\
= \hat{m}_i(\alpha, \beta) + \Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j) \\
= \hat{m}_i(\alpha, \beta) + \sum_{\{k \in \partial i\}} \alpha_{i,k} \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) \\
= \hat{m}_i(\alpha, \beta) + \sum_{\{k \in \partial i\}} \alpha_{i,k} \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) \\
= \hat{m}_i(\alpha, \beta) + \Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j) \\
\]

From Eq. (9.75), we derive the simultaneous linear equations of \( \{\Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j) | i \in V\} \) as follows:

\[
\Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j) = \sum_{\{k \in \partial i\}} \alpha_{i,k} \left( 1 - \hat{m}_i(\alpha, \beta)^2 \right) \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) \\
= \delta_{i,j} \Delta \beta_j \left( 1 - \hat{m}_i(\alpha, \beta)^2 \right) (i \in V, j \in V). \tag{9.76}
\]

such that

\[
\frac{1}{1 - \hat{m}_i(\alpha, \beta)^2} \Delta \hat{m}_i(\alpha, \beta | \Delta \beta_j) - \sum_{\{k \in \partial i\}} \alpha_{i,k} \Delta \hat{m}_k(\alpha, \beta | \Delta \beta_j) = \delta_{i,j} \Delta \beta_j (i \in V, j \in V). \tag{9.77}
\]

By using Eq. (9.77), the correlation function between every pair of nodes \( i \) and \( j \) are expressed by means of the linear response formulas as follows:

\[
\pi_{i,j}^r(\alpha, \beta) - \pi_i^r(\alpha, \beta) \pi_j^r(\alpha, \beta) \simeq (i | G(\alpha, \beta)^{-1} | j) (i \in V, j \in V), \tag{9.78}
\]

\[
\langle i | G(\alpha, \beta) | j \rangle \equiv \begin{cases} 
(1 - \hat{m}_i(\alpha, \beta)^2)^{-1} & (i = j), \\
-\alpha_{i,j} & (\{i, j\} \in E), \\
0 & (\text{otherwise})
\end{cases} (i \in V, j \in V). \tag{9.79}
\]

On the other hand, Eqs. (3.59), (3.60) and (3.53) are reduced to

\[
\hat{Q}_i(a_i) = \frac{\exp(\hat{\beta}_i a_i) \prod_{k \in \partial i} \mu_{i,k}^{a_i}(a_i)}{\sum_{z_i \in \{+1, -1\}} \exp(\hat{\beta}_i z_i) \prod_{k \in \partial i} \mu_{i,k}^{z_i}(z_i)} (i \in V, |\partial i| \geq 2), \tag{9.80}
\]
\[ Q_{[i,j]}(a_i, a_j) = \left( \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} \exp(\beta_i z_i + \beta_j z_j + \alpha_{[i,j]} a_i a_j) \prod_{\langle i,k \rangle \in \partial \setminus \{i,j\}} \mu_{[i,k]}(z_i) \prod_{\langle j,l \rangle \in \partial \setminus \{i,j\}} \mu_{[j,l]}(z_j) \right) \]

\[ \mu_{[i,j] \rightarrow i}(a_i) = \left( \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} \exp(\beta_i z_i + \alpha_{[i,j]} a_i z_j) \prod_{\langle j,l \rangle \in \partial \setminus \{i,j\}} \mu_{[j,l]}(z_j) \right) \]

Eq. (9.82) can be rewritten as follows:

\[ \frac{\sqrt{\mu_{[i,j] \rightarrow (+1)}}}{\sqrt{\mu_{[i,j] \rightarrow (-1)}}} = \frac{\exp(\alpha_{[i,j]}) - \exp(-\alpha_{[i,j]})}{\exp(\alpha_{[i,j]}) + \exp(-\alpha_{[i,j]})} \]

By introducing new parameter

\[ \tilde{h}_{[i,j] \rightarrow i}(\alpha, \beta) = \frac{1}{2} \ln \left( \frac{\mu_{[i,j] \rightarrow (+1)}}{\mu_{[i,j] \rightarrow (-1)}} \right) \]

Eq. (9.82) is reduced to

\[ \tilde{h}_{[i,j] \rightarrow i}(\alpha, \beta) = \arctanh \left( \tanh(\alpha_{[i,j]} + \sum_{\langle i,k \rangle \in \partial \setminus \{i,j\}} \tilde{h}_{[i,k] \rightarrow j}(\alpha, \beta)) \right) \]

where \( \tilde{h}_{[i,j] \rightarrow i}(\alpha, \beta) \) is referred to as an effective field from the \( j \)-th node to the \( i \)-th node in the statistical-mechanical informatics. The approximate marginal probability distributions in Eqs. (9.80) and (9.81) are also expressed in terms of effective fields as follows:

\[ \tilde{Q}_{i}(a_i) = \left( \sum_{z_i \in \{+1,-1\}} \exp \left( \beta_i + \sum_{\langle i,k \rangle \in \partial \setminus i} \tilde{h}_{[i,k] \rightarrow i}(\alpha, \beta) a_i \right) z_i \right) \]

\[ \tilde{Q}_{[i,j]}(a_i, a_j) = \left( \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} \exp \left( \beta_i + \sum_{\langle i,k \rangle \in \partial \setminus \{i,j\}} \tilde{h}_{[i,k] \rightarrow i}(\alpha, \beta) a_i + \sum_{\langle j,k \rangle \in \partial \setminus \{i,j\}} \tilde{h}_{[j,k] \rightarrow j}(\alpha, \beta) a_j + \alpha_{[i,j]} a_i a_j \right) \right) \]

The loopy belief propagation is often referred to as Bethe approximation in the statistical-mechanical informatics.
By substituting Eqs. (9.61) and (9.62), to Eq. (3.35), the Bethe free energy functional is reduced to

$$\mathcal{F}_{\text{Bethe}}[\{Q_i|i \in V\}, \{Q_{(i,j)}|\{i,j\} \in E\}]
= -\sum_{i \in V} \beta_i \sum_{z_i \in \mathbb{Z}} z_i Q_i(z_i) - \sum_{\{i,j\} \in E} \alpha_{(i,j)} \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} z_i z_j Q_{(i,j)}(z_i, z_j)
+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \{+1,-1\}} Q_i(z_i) \ln Q_i(z_i)
+ \sum_{\{i,j\} \in E} \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} Q_{(i,j)}(z_i, z_j) \ln Q_{(i,j)}(z_i, z_j),$$

(9.88)

By using these orthonormal polynomials in Eq. (9.66), the marginal probability distribution $Q_i(a_i)$ in Eq. (3.31) can be expressed as Eq. (9.68) with Eq. (9.69). Moreover, the marginal probability distribution $Q_{(i,j)}(a_i, a_j)$ in Eq. (3.32) can be also expressed as the following orthonormal expansion:

$$Q_{(i,j)}(a_i, a_j) = \frac{1}{2} \Phi_i(0, a_i) \Phi_j(0, a_j) + \frac{1}{2} \overline{a}_i \Phi_i(1, a_i) \Phi_j(0, a_j) + \frac{1}{2} \overline{a}_j \Phi_i(0, a_i) \Phi_j(1, a_j) + \frac{1}{2} \overline{a}_i \overline{a}_j \Phi_i(1, a_i) \Phi_j(1, a_j)
= \frac{1}{4} (1 + a_i a_j + a_j a_i) + \Pi_{(i,j)} a_i a_j \{(i,j) \in E\}$$

(9.89)

where

$$\Pi_{(i,j)} = 2 \sum_{z_i \in \{+1,-1\}} \Phi_i(1, z_i) \Phi_j(1, z_j) Q_{(i,j)}(z_i, z_j) - \sum_{z_i \in \{+1,-1\}} z_i z_j Q_{(i,j)}(z_i, z_j).$$

(9.90)

We remark that $\sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} Q_{(i,j)}(z_i, z_j) = 1$ for every edge $\{i,j\} \in E$. Moreover, in deriving Eq. (9.89) we use the following consistencies for every edge $\{i,j\} \in E$:

$$\begin{align*}
\overline{m}_i &= \sqrt{2} \sum_{z_i \in \{+1,-1\}} \Phi_i(1, z_i) Q_i(z_i) = 2 \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} \Phi_i(1, z_i) \Phi_j(0, z_j) Q_{(i,j)}(z_i, z_j), \\
\overline{m}_j &= \sqrt{2} \sum_{z_j \in \{+1,-1\}} \Phi_j(1, z_j) Q_j(z_j) = 2 \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} \Phi_i(0, z_i) \Phi_j(1, z_j) Q_{(i,j)}(z_i, z_j),
\end{align*}$$

(9.91)

By substituting Eqs. (9.68) and (9.89) to Eq. (9.88), we have

$$\mathcal{F}_{\text{Bethe}}[\{\overline{m}_i|i \in V\}, \{\Pi_{(i,j)}|\{i,j\} \in E\}]
= -\sum_{i \in V} \beta_i \overline{m}_i - \sum_{\{i,j\} \in E} \alpha_{(i,j)} \overline{m}_{(i,j)}
+ \sum_{i \in V} (1 - |\partial i|) \sum_{z_i \in \{+1,-1\}} \left( \frac{1}{2} + \frac{1}{2} \overline{m}_i z_i \right) \ln \left( \frac{1}{2} + \frac{1}{2} \overline{m}_i z_i \right)
+ \sum_{\{i,j\} \in E} \sum_{z_i \in \{+1,-1\}} \sum_{z_j \in \{+1,-1\}} \left( \frac{1}{4} + \frac{1}{4} \overline{m}_i z_i + \frac{1}{4} \overline{m}_j z_j + \frac{1}{4} \Pi_{(i,j)} z_i z_j \right)
\times \ln \left( \frac{1}{4} + \frac{1}{4} \overline{m}_i z_i + \frac{1}{4} \overline{m}_j z_j + \frac{1}{4} \Pi_{(i,j)} z_i z_j \right).$$

(9.92)

Instead of the conditional minimization (9.35) of the Bethe free energy free functional

$$\mathcal{F}_{\text{Bethe}}[\{Q_i|i \in V\}, \{Q_{(i,j)}|\{i,j\} \in E\}]$$
under the normalization conditions and the reducibility conditions in Eqs. (3.36)-(3.39), we consider the minimization of $\mathcal{F}_{\text{Bethe}}[\{\overline{m}_i|i \in V\}, \{\Pi_{(i,j)}|\{i,j\} \in E\}]$ as follows:

$$\arg \min_{\{\overline{m}_i|i \in V\}, \{\Pi_{(i,j)}|\{i,j\} \in E\}} \mathcal{F}_{\text{Bethe}}[\{\overline{m}_i|i \in V\}, \{\Pi_{(i,j)}|\{i,j\} \in E\}].$$

(9.93)

From the above minimization, we derive the same simultaneous fixed point equations to determine the set $\{\overline{m}_i|i \in V\}$ as
Eq. (9.65). The extremum conditions \( \mathcal{F}_{\text{Bethe}}[\{\mathbf{m}_i|i\in V\}, \{\mathbf{r}_{i,j}\}|\{i,j\}\in E] \) with respect to \( \{\mathbf{m}_i|i\in V\} \) and \( \{\mathbf{r}_{i,j}\}|\{i,j\}\in E \) are reduced to

\[
\beta_i = - (1 - |\partial i|) \sum_{z_i \in \{\pm 1\}} z_i \ln \hat{Q}_i(z_i) - \sum_{\{i,k\}\in \partial \partial i, \in \{\pm 1\}} \sum_{z_j \in \{\pm 1\}} z_j \ln \hat{Q}_{i,k}(z_i, z_k) \quad (i\in V),
\]

(9.94)

\[
\alpha_{(i,j)} = - \sum_{z_i \in \{\pm 1\}} \sum_{z_j \in \{\pm 1\}} z_i z_j \ln \hat{Q}_{i,j}(z_i, z_j) \quad (\{i,j\}\in E),
\]

(9.95)

\[
\hat{Q}_i(a_i) = \frac{1}{4} + \frac{1}{2} \tilde{m}_i(\alpha, \beta, a_i),
\]

(9.96)

\[
\hat{Q}_{i,j}(a_i, a_j) = \frac{1}{4} + \frac{1}{4} \tilde{m}_i(\alpha, \beta, a_i) + \frac{1}{4} \tilde{m}_j(\alpha, \beta, a_j) + \frac{1}{4} \tilde{u}_{i,j}(\alpha, \beta, a_i) a_j.
\]

(9.97)

By using the solutions \( \hat{a}(\alpha, \beta) \) and \( \hat{c}(\alpha, \beta) \) of Eqs. (9.94)-(9.97) the minimum of \( \mathcal{F}_{\text{Bethe}}[\{\tilde{m}_i(\alpha, \beta)\}i\in V\}, \{\tilde{u}_{i,j}(\alpha, \beta)\}\{i,j\}\in E] \) is given as

\[
\mathcal{F}_{\text{Bethe}}(\alpha, \beta) = \min_{\{\mathbf{m}_i|i\in V\}, \{\mathbf{r}_{i,j}\}|\{i,j\}\in E} \mathcal{F}_{\text{Bethe}}[\{\mathbf{m}_i|i\in V\}, \{\mathbf{r}_{i,j}\}|\{i,j\}\in E] \\
= - \sum_{i\in V} \beta_i \tilde{m}_i(\alpha, \beta) - \sum_{\gamma \in E} \alpha_{(i,j)} \tilde{u}_{i,j}(\alpha, \beta) \\
+ \sum_{i\in V} (1 - |\partial i|) \sum_{z_i \in \{\pm 1\}} \left( \frac{1}{2} + \frac{1}{2} \tilde{m}_i(\alpha, \beta, z_i) \right) \ln \left( \frac{1}{2} + \frac{1}{2} \tilde{m}_i(\alpha, \beta, z_i) \right) \\
+ \sum_{\{i,j\}\in E} \sum_{z_i \in \{\pm 1\}} \sum_{z_j \in \{\pm 1\}} \left( \frac{1}{4} + \frac{1}{4} \tilde{m}_i(\alpha, \beta, z_i) + \frac{1}{4} \tilde{m}_j(\alpha, \beta, z_j) + \frac{1}{4} \tilde{u}_{i,j}(\alpha, \beta, z_i) z_j \right) \\
\times \ln \left( \frac{1}{4} + \frac{1}{4} \tilde{m}_i(\alpha, \beta, z_i) + \frac{1}{4} \tilde{m}_j(\alpha, \beta, z_j) + \frac{1}{4} \tilde{u}_{i,j}(\alpha, \beta, z_i) z_j \right).
\]

(9.98)

By substituting Eq. (9.97) to Eq. (9.95), \( \tilde{u}_{i,j}(\alpha, \beta) \) can be expressed in terms of \( \alpha_{(i,j)}, \tilde{m}_i(\alpha, \beta) \) and \( \tilde{m}_j(\alpha, \beta) \) as follows:

\[
\tilde{u}_{i,j}(\alpha, \beta) = \frac{1}{\tanh(2\alpha_{(i,j)})} \left( 1 - \left( 1 - (1 - \tilde{m}_i(\alpha, \beta)^2 - \tilde{m}_j(\alpha, \beta)^2) \tanh^2(2\alpha_{(i,j)}) - 2\tilde{m}_i(\alpha, \beta) \tilde{m}_j(\alpha, \beta) \tanh(2\alpha_{(i,j)}) \right)^{1/2} \right)
\]

\[
\quad (\{i,j\}\in E).
\]

(9.99)

By using Eqs. (9.100), (9.101) and (9.102), the correlation function between every pair of nodes \( i \) and \( j \) is expressed by means of the linear response formulas as follows:

\[
\mathcal{W}_{i,j}(\alpha, \beta) - \mathcal{M}_i(\alpha, \beta) \mathcal{M}_j(\alpha, \beta) \simeq \langle {ij} | G(\alpha, \beta)^{-1} | j \rangle \quad (i\in V, j\in V, |\partial i| \geq 2, |\partial j| \geq 2),
\]

(9.100)

\[
\mathcal{W}_{i,j}(\alpha, \beta) - \mathcal{M}_i(\alpha, \beta) \mathcal{M}_j(\alpha, \beta) \simeq \sum_{\{k|k\in \partial i, |k\in \partial i| \geq 2\}} \langle {ij} | G^{-1} | k \rangle \langle k | R | j \rangle \quad (i\in V, j\in V, |\partial i| \geq 2, |\partial j| \geq 2),
\]

(9.101)

\[
\mathcal{W}_{i,j}(\alpha, \beta) - \mathcal{M}_i(\alpha, \beta) \mathcal{M}_j(\alpha, \beta) \simeq \sum_{\{k|k\in \partial j, |k\in \partial j| \geq 2\}} \sum_{\{l|l\in V, |l\in \partial j| \geq 2\}} \langle {ij} | R | k \rangle \langle k | G^{-1} | l \rangle \langle l | R | j \rangle \quad (i\in V, j\in V, |\partial i| = 1, |\partial j| = 1),
\]

(9.102)
where

\[
\langle i|G(\alpha, \beta)|j \rangle = \begin{cases} 
(1 - |\partial i|)(1 - \hat{m}_i(\alpha, \beta)^2)^{-1} + \sum_{\{k,l\} \in \partial i} \langle i|X_{i,k}(\alpha, \beta)^{-1}|i \rangle & (i = j) \\
0 & (\{i, j\} \in E) \\
\text{(otherwise)}
\end{cases}
\]  

(\{i, j\} \in E, |\partial i| \geq 2, |\partial j| \geq 2), \quad (9.103)

\[
X_{i,j}(\alpha, \beta) = \left( \begin{array}{c} \langle i|X_{i,j}(\alpha, \beta)|i \rangle \\ \langle j|X_{i,j}(\alpha, \beta)|j \rangle \\ \langle j|X_{i,j}(\alpha, \beta)|i \rangle \\ \langle i|X_{i,j}(\alpha, \beta)|j \rangle \end{array} \right)
\]

\[
= \begin{pmatrix} 1 - \hat{m}_i(\alpha, \beta)^2 & \hat{u}_{i,j}(\alpha, \beta) - \hat{m}_i(\alpha, \beta)\hat{m}_j(\alpha, \beta) \\
\hat{u}_{i,j}(\alpha, \beta) - \hat{m}_i(\alpha, \beta)\hat{m}_j(\alpha, \beta) & 1 - \hat{m}_j(\alpha, \beta)^2 \\
\hat{m}_j(\alpha, \beta) & \hat{m}_j(\alpha, \beta) \\
\hat{m}_j(\alpha, \beta) & \hat{m}_j(\alpha, \beta) \end{pmatrix}
\]  

(\{i, j\} \in E, |\partial i| \geq 2, |\partial j| \geq 2), \quad (9.104)

\[
\langle i|R|j \rangle = \langle j|R|i \rangle = \langle i|X_{i,j}^{-1}|i \rangle \left( \hat{u}_{i,j}(\alpha, \beta) - \hat{m}_i(\alpha, \beta)\hat{m}_j(\alpha, \beta) \right)
\]

(\{i, j\} \in E, |\partial i| \geq 2, |\partial j| = 1). \quad (9.105)

### 9.4 Perturbative Interpretation of Mean Field Method for Ising Model

In the mean-field method, the approximate free energy \( F_{\text{MF}}(\alpha, \beta) \) and the approximate probability distribution \( P_{\text{MF}}(a) = P_{\text{MF}}(a_1, a_2, \cdots, a_V) \) of Ising model in Eq. (9.49) on the square grid graph \((V, E)\) is expressed as

\[
F_{\text{MF}}(\alpha, \beta) = -\ln \left\{ \sum_{z_1 \in \{\pm 1\}} \sum_{z_2 \in \{\pm 1\}} \cdots \sum_{z_V \in \{\pm 1\}} \exp \left( H_{\text{MF}}(z_1, z_2, \cdots, z_V) \right) \right\}
\]

\[
= -\sum_{i \in V} \beta_1 \hat{m}_i(\alpha, \beta) - \sum_{\{i,j\} \in E} \hat{m}_i(\alpha, \beta)\hat{m}_j(\alpha, \beta)
\]

\[
+ \sum_{i \in V} \ln \left( \cosh \left( \sum_{\{i,j\} \in \partial i} \alpha_{i,j} \hat{m}_i(\alpha, \beta) \right) \right)
\]

\[
= -\sum_{i \in V} \beta_1 \hat{m}_i(\alpha, \beta) - \sum_{\{i,j\} \in E} \alpha_{i,j} \hat{m}_i(\alpha, \beta)\hat{m}_j(\alpha, \beta)
\]

\[
+ \sum_{i \in V} \sum_{\xi \in \{\pm 1\}} \frac{1}{2} \left( 1 + \hat{m}_i(\alpha, \beta)\xi \right) \ln \left( \frac{1}{2} \left( 1 + \hat{m}_i(\alpha, \beta)\xi \right) \right), \quad (9.106)
\]

\[
P_{\text{MF}}(a_1, a_2, \cdots, a_V) \equiv \frac{\exp \left( -H_{\text{MF}}(a_1, a_2, \cdots, a_V) \right)}{\sum_{z_1 \in \{\pm 1\}} \sum_{z_2 \in \{\pm 1\}} \cdots \sum_{z_V \in \{\pm 1\}} \exp \left( -H_{\text{MF}}(z_1, z_2, \cdots, z_V) \right)}, \quad (9.107)
\]

where

\[
H_{\text{MF}}(a) = H_{\text{MF}}(a_1, a_2, \cdots, a_V)
\]

\[
\equiv -\sum_{\{i,j\} \in E} \alpha_{i,j} \left( \hat{m}_i(\alpha, \beta) a_i + \hat{m}_j(\alpha, \beta) a_j - \hat{m}_i(\alpha, \beta)\hat{m}_j(\alpha, \beta) \right) - \sum_{i \in V} \beta_i a_i, \quad (9.108)
\]

By substituting the solution of the simultaneous fixed point equations

\[
\hat{m}_i(\alpha, \beta) = \tanh (\beta_1 + \sum_{\{i,j\} \in \partial i} \alpha_{i,j} \hat{m}_j(\alpha, \beta)) \ (i \in V), \quad (9.109)
\]

to Eqs. (9.106)-(9.107), we obtain some approximate values of statistical quantities in the mean-field method for the probability distribution in Eq. (9.49). Eqs. (9.106)-(9.109) are derived also by substituting \( W_{\{i,j\}}(a_i, a_j) \) and \( W_i(a_i) \) in Eqs. (9.61) and (9.62) to Eqs. (9.41) and (9.42).

We consider a perturbation expansion of the free energy

\[
F(\alpha, \beta) \equiv -\ln \left( \sum_{z_1 \in \{\pm 1\}} \sum_{z_2 \in \{\pm 1\}} \cdots \sum_{z_V \in \{\pm 1\}} \exp \left( -H(z_1, z_2, \cdots, z_V) \right) \right), \quad (9.110)
\]
from the mean-field free energy $F_{\text{MF}}(\alpha, \beta)$ in Eq.\((9.106)\) as follows:

\[
F(\alpha, \beta) = F_{\text{MF}}(\alpha, \beta) \\
-\ln \left\{ \sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{-1,1\}} \cdots \sum_{z_{|V|} \in \{-1,1\}} P_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) \times \exp \left( H_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) - H(z_1, z_2, \ldots, z_{|V|}) \right) \right\}
\]  

(9.111)

We expand the right-hand side of Eq.\((9.111)\) to

\[
F(\alpha, \beta) = F_{\text{MF}}(\alpha, \beta) \\
-\ln \left( \sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{-1,1\}} \cdots \sum_{z_{|V|} \in \{-1,1\}} P_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) \times \left( 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left( H_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) - H(z_1, z_2, \ldots, z_{|V|}) \right)^n \right) \right).
\]  

(9.112)

By using the following equalities:

\[
H_{\text{MF}}(a_1, a_2, \ldots, a_{|V|}) - H(a_1, a_2, \ldots, a_{|V|}) = \sum_{\{i,j\} \in E} \alpha_{\{i,j\}} (a_i - \hat{m}_i(\alpha, \beta))(a_j - \hat{m}_j(\alpha, \beta)),
\]  

(9.113)

\[
\sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{-1,1\}} \cdots \sum_{z_{|V|} \in \{-1,1\}} (z_i - \hat{m}_i(\alpha, \beta)) P_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) = 0,
\]  

(9.114)

\[
\sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{-1,1\}} \cdots \sum_{z_{|V|} \in \{-1,1\}} (z_i - \hat{m}_i(\alpha, \beta))^2 P_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) = 1 - \hat{m}_i(\alpha, \beta)^2,
\]  

(9.115)

\[
\sum_{z_1 \in \{+1,-1\}} \sum_{z_2 \in \{-1,1\}} \cdots \sum_{z_{|V|} \in \{-1,1\}} (z_i - \hat{m}_i(\alpha, \beta))^3 P_{\text{MF}}(z_1, z_2, \ldots, z_{|V|}) = -2 \hat{m}_i(\alpha, \beta)(1 - \hat{m}_i(\alpha, \beta)^2),
\]  

(9.116)

the expansion of the free energy can be rewritten as

\[
F(\alpha, \beta) = F_{\text{MF}}(\alpha, \beta) \\
-\frac{1}{2} \sum_{\{i,j\} \in E} \left( \alpha_{\{i,j\}}^2 (1 - \hat{m}_i(\alpha, \beta)^2)(1 - \hat{m}_j(\alpha, \beta)^2) + \mathcal{O}(\alpha_{\{i,j\}}^3) \right).
\]  

(9.117)

By using Eq.\((9.117)\), we introduce $\mathcal{F}_{\text{TAP}}[\{a_i \in V\}]$ defined by

\[
\mathcal{F}_{\text{TAP}}[\{a_i \in V\}] \equiv F_{\text{MF}}[\{a_i \in V\}] - \frac{1}{2} \sum_{\{i,j\} \in E} \left( \alpha_{\{i,j\}}^2 (1 - a_i^2)(1 - a_j^2) + \mathcal{O}(\alpha_{\{i,j\}}^3) \right)
\]

\[= -\sum_{i \in V} \beta_i a_i - \sum_{\{i,j\} \in E} \alpha_{\{i,j\}} a_i a_j + \sum_{i \in V} \sum_{z_i \in \{-1,1\}} a_i \frac{1}{2} \ln \left( \frac{1}{2} (1 + a_i z_i) \right) \]

\[\quad -\frac{1}{2} \sum_{\{i,j\} \in E} \left( \alpha_{\{i,j\}}^2 (1 - a_i^2)(1 - a_j^2) + \mathcal{O}(\alpha_{\{i,j\}}^3) \right).
\]  

(9.118)

Instead of Eq.\((9.71)\), we consider the following minimization problem

\[
\{\hat{m}_i(\alpha, \beta) | i \in V\} = \arg \min_{\{\overline{m}_i | i \in V\}} \mathcal{F}_{\text{TAP}}[\{\overline{m}_i | i \in V\}],
\]  

(9.119)

to determine the set $\hat{m}_i(\alpha, \beta) | i \in V \}$. The extremum conditions of Eq.\((9.118)\) is reduced to the following simultaneous fixed point equations:

\[
\hat{m}_i(\alpha, \beta) = \tanh \left( \beta_i + \sum_{\{i,j\} \in \partial_i} \alpha_{\{i,j\}} \hat{m}_j(\alpha, \beta) \right)
\]
By substituting Eq.(9.120) to Eq.(9.118), we obtain

\[ F_{\text{TAP}}(\alpha, \beta) = \min_{\{\bar{m}_i \in \mathbb{V}\}} F_{\text{TAP}}[\{\bar{m}_i \mid i \in \mathbb{V}\}], \]

\[ = -\sum_{i \in \mathbb{V}} \beta_i \bar{m}_i(\alpha, \beta) - \sum_{\{i,j\} \in \mathcal{E}} \alpha_{\{i,j\}} \bar{m}_i(\alpha, \beta) \bar{m}_j(\alpha, \beta) \]

\[ + \sum_{i \in \mathbb{V}} \sum_{z_i \in \{+1, -1\}} \frac{1}{2} (1 + \bar{m}_i(\alpha, \beta) z_i) \ln \left( \frac{1}{2} (1 + \bar{m}_i(\alpha, \beta) z_i) \right), \]

\[ - \frac{1}{2} \sum_{\{i,j\} \in \mathcal{E}} \left( \alpha_{\{i,j\}} \right)^2 \left( 1 - \bar{m}_i(\alpha, \beta)^2 \right) \left( 1 - \bar{m}_j(\alpha, \beta)^2 \right) + \mathcal{O}(\alpha_{\{i,j\}}^3). \]  

(9.121)

where \( \bar{m}(\alpha, \beta) \equiv \{ \bar{m}_i(\alpha, \beta) \mid i \in \mathbb{V} \} \) are determined so as to satisfy the simultaneous fixed point equations in Eq.(9.120). Eq.(9.120) and Eq.(9.121) are referred to as a \textit{TAP equation} and a \textit{TAP free energy}, respectively. The term \(- \sum_{\{i,j\} \in \mathcal{E}} \alpha_{\{i,j\}} \bar{m}_i(\alpha, \beta) \bar{m}_j(\alpha, \beta) \) in Eq.(9.120) is often referred to as an \textit{Onsager reaction term}.

In order to derive the correlation \( \bar{m}_{\{i,j\}}(\alpha, \beta) \) in Eq.(9.55) from the TAP free energy \( F_{\text{TAP}}(\alpha, \beta) \), we introduce the Legendre transformation \( G_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) \) as

\[ G_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) = F_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) + \sum_{i \in \mathbb{V}} \bar{\beta}_i(\alpha, m) m_i, \]

(9.122)

where \( \bar{\beta}(\alpha, m) \equiv \{ \bar{\beta}_i(\alpha, m) \mid i \in \mathbb{V} \} \) are determined so as to satisfy the following equations:

\[ \bar{m}_i = -\left[ \frac{\partial}{\partial \bar{\beta}_i} F_{\text{TAP}}(\alpha, \bar{\beta}) \right]_{\bar{\beta} = \bar{\beta}(\alpha, m)} (i \in \mathbb{V}). \]

(9.123)

We have the following relationship between \( G_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) \) and \( F_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) \):

\[ \frac{\partial}{\partial \alpha_{\{i,j\}}} G_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) = \frac{\partial}{\partial \alpha_{\{i,j\}}} F_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) \]

\[ + \sum_{k \in \mathbb{V}} \left[ \frac{\partial}{\partial \bar{\beta}_k} F_{\text{TAP}}(\alpha, \bar{\beta}) \right]_{\bar{\beta} = \bar{\beta}(\alpha, a)} \left( \frac{\partial}{\partial \alpha_{\{i,j\}}} \bar{\beta}_k(\alpha, a) \right) \]

\[ + \sum_{\{i,j\} \in \mathcal{E}} \left( \frac{\partial}{\partial \alpha_{\{i,j\}}} \bar{\beta}_k(\alpha, a) \right) m_k \]

\[ = \frac{\partial}{\partial \alpha_{\{i,j\}}} F_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, a)). \]

(9.124)

By using Eq.(9.55), the corresponding approximate correlation \( \hat{u}_{\{i,j\}}(\alpha, \beta) \) in the TAP free energy \( F_{\text{TAP}}(\alpha, \beta) \) is defined by

\[ \hat{u}_{\{i,j\}}(\alpha, \beta) = -\frac{\partial}{\partial \alpha_{\{i,j\}}} F_{\text{TAP}}(\alpha, \beta). \]

(9.125)

By substituting Eq.(9.124) to Eq.(9.125), the approximate value \( \hat{u}_{\{i,j\}}(\alpha, \beta) \) of the correlation is derived as

\[ \hat{u}_{\{i,j\}}(\alpha, \beta) = -\left[ \frac{\partial}{\partial \alpha_{\{i,j\}}} G_{\text{TAP}}(\alpha, \bar{\beta}(\alpha, m)) \right]_{\bar{m} = \bar{m}(\alpha, \beta)} \]

\[ = \hat{m}_i(\alpha, \beta) \hat{m}_j(\alpha, \beta) + \alpha_{\{i,j\}} (1 - \hat{m}_i(\alpha, \beta)^2)(1 - \hat{m}_j(\alpha, \beta)^2) \]

\[ + 2 \alpha_{\{i,j\}} (1 - \hat{m}_i(\alpha, \beta)^2)(1 - \hat{m}_j(\alpha, \beta)^2) \hat{m}_i(\alpha, \beta) \hat{m}_j(\alpha, \beta) \]

\[ + \mathcal{O}(\alpha_{\{i,j\}}^3). \]

(9.126)

On the other hand, we can derive the same expansion as Eq.(9.121) from Eq.(9.92) and Eq.(9.99). Eq.(9.99) can be expanded as

\[ \hat{u}_{\{i,j\}}(\alpha, \beta) = \hat{m}_i(\alpha, \beta) \hat{m}_j(\alpha, \beta) + \alpha_{\{i,j\}} (1 - \hat{m}_i(\alpha, \beta)^2)(1 - \hat{m}_j(\alpha, \beta)^2) \]
\[ +2\alpha_{(i,j)}^2 (1 - \hat{m}_i(\alpha, \beta)^2)(1 - \hat{m}_j(\alpha, \beta)^2) \hat{m}_i(\alpha, \beta) \hat{m}_j(\alpha, \beta) + \mathcal{O}(\alpha_{(i,j)}^3). \] (9.127)

By substituting Eq. (9.127) to Eq. (9.98) and expanding it with respect to \( \hat{a}(\alpha, \beta) \), we can derive the same expression as Eq. (9.121).

### 9.5 Thermodynamic Limit and Phase Transition in Ising Model

We consider a \( M \times N \) square grid graph \((V, E)\) with periodic boundary conditions along both the abscissa and the ordinate in the two-dimensional space. In the square grid graph, the position vector \( \mathbf{r}_i \) of each node \( i \) is assigned by Eq. (2.42). When the abscissa and the ordinate of \( \mathbf{r}_i \) are denoted by \( m \) and \( n \), such that \( \mathbf{r}_i = (m, n) \), the periodic boundary conditions for the abscissa and the ordinate means that \( m = M \) and \( n = N \) are interpreted as \( m = 0 \) and \( n = 0 \), respectively. In Eq. (9.49) with Eq. (9.50), we set \( \alpha_{(i,j)} = \alpha \) and \( \beta_i = \beta \) as follows:

\[ P(\alpha) = P(a_1, a_2, \ldots, a_{|V|}) \equiv \frac{\exp\left(-H(a_1, a_2, \ldots, a_{|V|})\right)}{\sum_{z_1 \in \{+1, -1\}} \sum_{z_2 \in \{+1, -1\}} \cdots \sum_{z_{|V|} \in \{+1, -1\}} \exp\left(-H(z_1, z_2, \ldots, z_{|V|})\right)}, \] (9.128)

where

\[ H(a_1, a_2, \ldots, a_{|V|}) \equiv -\alpha \sum_{i \in V} a_i a_j - \beta \sum_{i \in V} a_i. \] (9.129)

The probabilistic model in Eq. (9.49) is referred to as an \textit{Ising model} in the statistical-mechanical informatics. For the probabilistic model in Eq. (9.49), \( f(\alpha, \beta) \) is introduced as follows:

\[ f(\alpha, \beta) \equiv -\lim_{M \to +\infty, N \to +\infty} \frac{1}{MN} \ln \left( \sum_{z_1 \in \{+1, -1\}} \sum_{z_2 \in \{+1, -1\}} \cdots \sum_{z_{|V|} \in \{+1, -1\}} \exp\left(-H(z_1, z_2, \ldots, z_{|V|})\right) \right). \] (9.130)

We remark that the limit of the right-hand side is referred to as a thermodynamic limit in the statistical-mechanical informatics and \( f(\alpha, \beta) \) corresponds to the free energy per node in the thermodynamic limit. From derivatives of the free energy, some statistical quantities are also expressed as follows:

\[ m(\alpha, \beta) \equiv \lim_{M \to +\infty, N \to +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_{z_1 \in \{+1, -1\}} \sum_{z_2 \in \{+1, -1\}} \cdots \sum_{z_{|V|} \in \{+1, -1\}} z_i P(z_1, z_2, \ldots, z_{|V|}) = \frac{\partial}{\partial \beta} f(\alpha, \beta), \] (9.131)

\[ u(\alpha, \beta) \equiv \lim_{M \to +\infty, N \to +\infty} \frac{1}{|E|} \sum_{i \in E} \sum_{z_1 \in \{+1, -1\}} \sum_{z_2 \in \{+1, -1\}} \cdots \sum_{z_{|V|} \in \{+1, -1\}} z_i z_j P(z_1, z_2, \ldots, z_{|V|}) = \frac{1}{2} \frac{\partial}{\partial \alpha} f(\alpha, \beta). \] (9.132)

\( m(\alpha, \beta) \) and \( u(\alpha, \beta) \) are referred to as a \textit{magnetization} per node and a \textit{internal energy} per edge, respectively.

For \( \beta = 0 \), the closed expressions of free energy \( f(\alpha, 0) \) and the internal energy \( u(\alpha, 0) \) have been derived exactly as follows:

\[ f(\alpha, 0) = -\frac{1}{2\pi^2} \int_0^{\pi} \int_0^{\pi} \ln \left( \cosh^2(2\alpha) + \sinh(2\alpha) \left( \cos(\phi) + \cos(\psi) \right) \right) d\phi d\psi; \] (9.133)

\[ u(\alpha, 0) = \frac{1}{2} \frac{\partial}{\partial \alpha} f(\alpha, 0). \]

\[ = \alpha \coth(2\alpha) \left( 1 + \frac{2}{\pi} (2\tanh^2(2\alpha) - 1) \right)^{\pi/2} \sqrt{1 - \left( \frac{2\tanh(2\alpha)}{\cosh(2\alpha)} \right)^2} \sin(\phi) d\phi. \] (9.134)

These expressions are referred to as Onsager solutions for the two-dimensional Ising model. It is known that the derivative of \( u(\alpha, 0) \) with respect to \( \alpha \) diverges at

\[ \alpha = \alpha_c \equiv \frac{1}{2} \arcsinh(1) = 0.440686... \] (9.135)
\( \alpha_c \) is referred to as a critical point. Moreover, \( \lim_{\beta \to 0} m(\alpha, \beta) \) has been derived exactly as follows:

\[
\lim_{\beta \to 0} m(\alpha, \beta) = \begin{cases} 
(1 - \sinh^4(2\alpha))^{1/8} & (\alpha > \alpha_c), \\
0 & (\alpha < \alpha_c). 
\end{cases}
\]  

(9.136)

\( m(\alpha, +0) = \lim_{\beta \to 0} m(\alpha, \beta) \) is referred to as a Spontaneous Magnetization of the Ising model on the square grid graph.

The correlation \( \Pi_{(i,j)}(\alpha, \beta) \) for \( r_i - r_j = (0, n) \ (n = 1, 2, \cdots) \) is given in terms of a Toeplitz determinant as follows:

\[
u_{(i,j)}(\alpha, \beta) = \lim_{M \to \infty, N \to \infty} \frac{1}{2\pi i} \int_{|z|=1} z^{-n} g(z) dz,
\]

(9.138)

\[
g(z) = \sqrt{\frac{(1 - \kappa_1/z)(1 - \kappa_2/z)}{(1 - \kappa_1/z)(1 - \kappa_2/z)}} g(-1) = 1,\]

(9.139)

\[
\kappa_1 \equiv \frac{1}{\tanh(\alpha)} \left( \frac{1 - \tanh(\alpha)}{1 + \tanh(\alpha)} \right),
\]

(9.140)

\[
\kappa_2 \equiv \tanh(\alpha) \left( \frac{1 - \tanh(\alpha)}{1 + \tanh(\alpha)} \right).
\]

(9.141)

The complex function \( g(z) \) of the complex variable \( z \) has four branch points \( z = \kappa_1, z = \kappa_1^{-1}, z = \kappa_2 \) and \( z = \kappa_2^{-1} \) on the real axis of the \( z \)-plane, and they have the following relationships:

\[
\begin{cases} 
0 < \kappa_2^{-1} < \kappa_1^{-1} < 1 < \kappa_1 < \kappa_2 & (0 < \alpha < \alpha_c), \\
0 < \kappa_2^{-1} < \kappa_1 < 1 < \kappa_1^{-1} < \kappa_2 & (\alpha_c < \alpha). 
\end{cases}
\]

(9.142)

In the case of \( \alpha > \alpha_c \), \( \text{ln}(g(z)) \) is analytic in the region \( \kappa_1 < |z| < \kappa_1^{-1} \) on the \( z \)-plane. If we choose \( \text{arg}(1 - \kappa_1/z), \text{arg}(1 - \kappa_2/z), \text{arg}(1 - \kappa_1/z) \) and \( \text{arg}(1 - \kappa_2/z) \) to be their principal values and the branch of \( \text{ln}g(z) \) to be \( \text{ln}(-1) = 0 \), \( \text{ln}(g(z)) \) can be expanded as the following Laurent series around \( z = 0 \):

\[
\ln(g(z)) = \sum_{l=1}^{\infty} \frac{1}{2l}(\kappa_1^l - \kappa_2^l) z^l + \sum_{l=1}^{\infty} \frac{1}{2l}(-\kappa_1^l + \kappa_2^l) z^{-l} (\kappa_1^{-1} < |z| < \kappa_1). 
\]

(9.143)

In the case of \( 0 < \alpha < \alpha_c \), \( \text{ln}(g(z)) \) is not analytic in the region \( \kappa_1^{-1} < |z| < \kappa_1 \). In this case, if we introduce

\[
w(z) \equiv -zg(z) = \sqrt{\frac{(1 - \kappa_1^{-1}/z)(1 - \kappa_2/z)}{(1 - \kappa_1^{-1}/z)(1 - \kappa_2/z)}} w(-1) = 1,
\]

(9.144)

\( \text{ln}(w(z)) \) is analytic in the region \( \kappa_1 < |z| < \kappa_1^{-1} \) on the \( z \)-plane. If we choose \( \text{arg}(1 - \kappa_1^{-1}/z), \text{arg}(1 - \kappa_2/z), \text{arg}(1 - \kappa_1^{-1}/z) \) and \( \text{arg}(1 - \kappa_2/z) \) to be their principal values and the branch of \( \text{ln}w(z) \) to be \( \text{ln}(-1) = 0 \), \( \text{ln}(w(z)) \) can be expanded as the following Laurent series around \( z = 0 \):

\[
\ln(w(z)) = -\sum_{l=1}^{\infty} \frac{1}{2l}(\kappa_1^{-l} + \kappa_2^l) z^l + \sum_{l=1}^{\infty} \frac{1}{2l}(\kappa_1^{-l} + \kappa_2^l) z^{-l} (\kappa_1 < |z| < \kappa_1^{-1}).
\]

(9.145)
the long range limit $n \to +\infty$ of the correlation $\mathbf{\tau}_{(i,j)}(\alpha, \beta)$ for $r_i - r_j = (0, n)$ in Eq. (9.55) is derived as

$$
\lim_{n \to +\infty} \mathbf{\tau}_{(i,j)}(\alpha, \beta) = \begin{cases} 
(m(\alpha, \beta))^2 & (\alpha > \alpha_C), \\
0 & (0 < \alpha < \alpha_C).
\end{cases} 
$$  

(9.146)

For a finite system in Eq. (9.128),

$$
\sum_{z_i \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} \exp(-H(z_1, z_2, \cdots, z_{|V|}))
$$

always takes a positive value and is analytic with respect to $\alpha$ and $\beta$, because it is expressed as a summation of a finite number of exponential functions. In such cases, $\mathbf{\tau}_{i}(\alpha, \beta)$ ($i \in V$) and $\mathbf{\tau}_{(i,j)}(\alpha, \beta)$ ($\{i, j\} \in E$) are always analytic functions with respect to $\alpha$ and $\beta$. Moreover, $\mathbf{\tau}_{i}(\alpha, \beta)$ ($i \in V$) is also an analytic function with respect to $\alpha$ and $\beta$. Because $P(a) = P(-a)$ for every possible configuration $a = (a_1, a_2, \cdots, a_{|V|})$ in the case of $\beta = 0$, $\lim_{\beta \to +0} \mathbf{\tau}_{i}(\alpha, \beta)$ is always zero as follows:

$$
\lim_{\beta \to +0} \lim_{M \to +\infty} \lim_{N \to +\infty} \sum_{z_1 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i P(z_1, z_2, \cdots, z_{|V|}) = 0 \quad (i \in V).
$$

(9.147)

However, in the thermodynamic limit, the commutativity between $\lim_{M \to +\infty} \lim_{N \to +\infty}$ and $\lim_{\beta \to +0}$ does not be guaranteed as follows:

$$
\lim_{M \to +\infty} \lim_{N \to +\infty} \lim_{\beta \to +0} \sum_{z_1 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i P(z_1, z_2, \cdots, z_{|V|}) \neq \lim_{\beta \to +0} \lim_{M \to +\infty} \lim_{N \to +\infty} \sum_{z_1 \in \{+1,-1\}} \cdots \sum_{z_{|V|} \in \{+1,-1\}} z_i P(z_1, z_2, \cdots, z_{|V|}) \quad (i \in V).
$$

(9.148)

This is the reason why $\lim_{\beta \to +0} m(\alpha, \beta)$ is not always zero for any value of $\alpha$ as shown in Eq. (9.136). Eventually, the Ising model in Eq. (9.128) has the phase transition only in the thermodynamic limit.

In the square grid graph $G = (V, E)$ with the periodic boundary conditions in the thermodynamic limit $M \to +\infty$ and $N \to +\infty$, $\mathbf{\tau}_{i}(\alpha, \beta)$ ($i \in V$) and $\mathbf{\tau}_{(i,j)}(\alpha, \beta)$ ($\{i, j\} \in E$) are not depend on $i$ and $\{i, j\}$, respectively. Then we replace $\mathbf{\tau}_{i}(\alpha, \beta)$ ($i \in V$) and $\mathbf{\tau}_{(i,j)}(\alpha, \beta)$ ($\{i, j\} \in E$) by

$$
m(\alpha, \beta) = \lim_{M \to +\infty} \lim_{N \to +\infty} \mathbf{\tau}_{i}(\alpha, \beta),
$$

(9.149)

and

$$
u(\alpha, \beta) = \lim_{M \to +\infty} \lim_{N \to +\infty} \mathbf{\tau}_{(i,j)}(\alpha, \beta),
$$

(9.150)

respectively.

Now we consider the mean-field method of the Ising model in Eq. (9.128) in the thermodynamic limit. In the mean-field method and the loopy belief propagation, the approximate values $\hat{m}(\alpha, \beta)$ and $\hat{u}(\alpha, \beta)$ of $m(\alpha, \beta)$ and $u(\alpha, \beta)$ are given as the solutions of the following fixed point equations:

$$
\begin{cases}
\hat{m}(\alpha, \beta) = \tanh\left(\beta + 4\alpha\hat{m}(\alpha, \beta)\right), \\
\hat{u}(\alpha, \beta) = \hat{m}(\alpha, \beta)^2,
\end{cases}
$$

(9.151)

and

$$
\begin{cases}
\tilde{m}(\alpha, \beta) = \tanh\left(\beta + 4\tilde{m}(\alpha, \beta)\right), \\
\tilde{u}(\alpha, \beta) = \frac{e^{2\alpha}\cosh(2(\beta + 3\tilde{m}(\alpha, \beta))) - 1}{e^{2\alpha}\cosh(2(\beta + 3\tilde{m}(\alpha, \beta))) + 1}, \\
\tilde{h}(\alpha, \beta) = \arctanh\left(\tanh(\alpha)\tanh(\beta + 3\tilde{m}(\alpha, \beta))\right),
\end{cases}
$$

(9.152)

respectively.

The covariance $\mathbf{\tau}_{(i,j)}(\alpha, \beta) - a(i, \alpha, \beta)a(j, \alpha, \beta)$ by means of the linear response formulas for the mean-field method and for the loopy belief propagation are given as

$$
\mathbf{\tau}_{(i,j)}(\alpha, \beta) - m(\alpha, \beta)\mathbf{m}_{j}(\alpha, \beta) \sim \langle i|G(\alpha, \beta)^{-1}|j \rangle \quad (i \in V, j \in V),
$$

(9.153)

where

$$
\langle i|G(\alpha, \beta)|j \rangle \equiv \begin{cases}
(1 - \hat{m}(\alpha, \beta))^{-1} & (i = j), \\
\alpha & (\{i, j\} \in E), \\
0 & (\text{otherwise}),
\end{cases}
$$
We remark that the position vector \( r \) for the mean-field method and \( u(\alpha, \beta) \) can be rewritten as follows:

\[
\langle i|G(\alpha, \beta)j \rangle \equiv \begin{cases} 
-3(1 - \hat{m}(\alpha, \beta)^2)^{-1} + 4|\langle i|X_{(i,j)}(\alpha, \beta)^{-1}|i \rangle| & (i = j) \\
0 & (\{i, j\}\in E) \\
0 & (\{i, j\}\in \text{otherwise}) 
\end{cases}
\]

(\[i\in V, j\in V, \text{Mean Field Method}\),

(9.154)

for the mean-field method and

\[
X_{(i,j)}(\alpha, \beta) = \begin{cases} 
\langle i|X_{(i,j)}(\alpha, \beta)^{-1}|i \rangle & (\{i, j\}\in E) \\
\langle j|X_{(i,j)}(\alpha, \beta)^{-1}|j \rangle & (\{i, j\}\in \text{otherwise}) 
\end{cases}
\]

(9.155)

\[
\langle i|G(\alpha, \beta)j \rangle = \begin{cases} 
\frac{-3(\hat{u}(\alpha, \beta)^2 + (1 - \hat{m}(\alpha, \beta)^2)^2)}{(\hat{u}(\alpha, \beta)^2 + (1 - \hat{m}(\alpha, \beta)^2)^2)} & (i = j) \\
0 & (\{i, j\}\in \text{otherwise}) 
\end{cases}
\]

(9.156)

for the loopy belief propagation. By substituting Eq.(9.156) to Eq.(9.155), \( G(\alpha, \beta) \) is expressed in terms of \( \hat{m}(\alpha, \beta) \) and \( \hat{u}(\alpha, \beta) \) as follows:

Because the present system in Eq.(9.128) with the periodic boundary conditions along the abcissa and the ordinate directions has the translational and reflectional symmetries, the representation of each component in \( G(\alpha, \beta) \) can be rewritten as follows:

\[
\langle (i-1)\text{mod}(M), \frac{j-1}{M} | G(\alpha, \beta)^{-1} | (j-1)\text{mod}(M), \frac{j-1}{M} \rangle = \langle (i-1)\text{mod}(M), \frac{j-1}{M} | G(\alpha, \beta)^{-1} | (i-1)\text{mod}(M), \frac{j-1}{M} \rangle
\]

(9.158)

where the position vector \( r_i \) of the node \( i \) is represented by Eq.(2.42). We introduce \( MN \times MN \) unitary matrix \( U \) and its conjugate matrix \( U^\dagger \) as follows:

\[
\langle m, n|U|k, l \rangle \equiv \frac{1}{\sqrt{MN}} \exp\left(-i\frac{2\pi km}{M} - i\frac{2\pi ln}{N}\right),
\]

(9.159)

\[
\langle k, l|U^\dagger|m, n \rangle \equiv \frac{1}{\sqrt{MN}} \exp\left(i\frac{2\pi km}{M} + i\frac{2\pi ln}{N}\right).
\]

(9.160)

We remark that \( \langle m, n|G(\alpha, \beta)^{-1}|m, n \rangle \) depends only on \( |m - m'| \) and \( |n - n'| \), because it has both spatially translational and reflectional symmetries. By multiply \( G(\alpha, \beta) \) by \( U \) from the right and by \( U^\dagger \) from the left, the matrix \( G(\alpha, \beta) \) can be diagonalized as follows:

\[
\langle k, l|U^\dagger G(\alpha, \beta)U|k', l' \rangle = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \exp\left(i\frac{2\pi km'}{M} + i\frac{2\pi ln'}{N}\right)
\]

\[
\times \langle m, n|G(\alpha, \beta)|m', n'\rangle \exp\left(-i\frac{2\pi km}{M} - i\frac{2\pi ln}{N}\right)
\]

\[
= \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \exp\left(i\frac{2\pi (k' - k)m}{M} + i\frac{2\pi (l' - l)n}{N}\right)
\]

\[
\times \langle m, n|G(\alpha, \beta)|m', n'\rangle \exp\left(-i\frac{2\pi k(m - m')}{M} - i\frac{2\pi l(n - n')}{N}\right)
\]

\[
= \left( \frac{1}{\sqrt{MN}} \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \langle m, n|G(\alpha, \beta)|m', n'\rangle \right)
\]

\[
\left( \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle m, n|G(\alpha, \beta)|m', n'\rangle \right).
\]

(9.157)
\begin{align*}
\times \exp \left( -\frac{2\pi k(m-m')}{M} - i\frac{2\pi l(n-n')}{N} \right) \\
= \left( \frac{1}{\sqrt{MN}} \sum_{m'=0}^{N-1} \sum_{m=0}^{N-1} \exp \left( i\frac{2\pi (k'-k)m'}{M} + i\frac{2\pi (l'-l)n'}{N} \right) \right) \\
\times \left( \frac{1}{\sqrt{MN}} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \exp \left( -\frac{2\pi km}{M} - i\frac{2\pi ln}{N} \right) \right) \\
= \delta_{k,k'} \delta_{l,l'} \frac{1}{\sqrt{MN}} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \langle m,n \rangle \exp \left( -\frac{2\pi km}{M} - i\frac{2\pi ln}{N} \right) \\
= \delta_{k,k'} \delta_{l,l'} G_{k,l}(\alpha, \beta), \tag{9.161}
\end{align*}

where

\[ G_{k,l}(\alpha, \beta) \equiv \frac{1}{\sqrt{MN}} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \langle m,n \rangle \exp \left( -\frac{2\pi km}{M} - i\frac{2\pi ln}{N} \right) \]

\[ = \frac{1}{\sqrt{MN}} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \langle m,n \rangle \exp \left( -\frac{2\pi km}{M} - i\frac{2\pi ln}{N} \right). \tag{9.162} \]

The above calculations can be reduced to

\[ G(\alpha, \beta) = U \Lambda(\alpha, \beta) U^\dagger, \tag{9.163} \]

\[ G(\alpha, \beta)^{-1} = U \Lambda(\alpha, \beta)^{-1} U^\dagger, \tag{9.164} \]

where \( \Lambda(\alpha, \beta) \) is a diagonal matrix defined by

\[ \langle k, l | \Lambda(\alpha, \beta) | k', l' \rangle \equiv \delta_{k,k'} \delta_{l,l'} G_{k,l}(\alpha, \beta). \tag{9.165} \]

Eq. \( 9.166 \) is reduced to

\[ \pi_{(i,j)}(\alpha, \beta) - a_i(\alpha, \beta) a_j(\alpha, \beta) \simeq \langle ij | U \Lambda(\alpha, \beta)^{-1} U^\dagger | j \rangle \]

\[ = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{L-1} \sum_{k'=0}^{M-1} \sum_{l'=0}^{L-1} \exp \left( -\frac{2\pi km'}{M} - \frac{2\pi l'n'}{N} \right) \]

\[ \times \delta_{k,k'} \delta_{l,l'} G_{k,l}(\alpha, \beta)^{-1} \exp \left( \frac{2\pi km}{M} + i\frac{2\pi ln}{N} \right) \]

\[ = \frac{1}{\sqrt{MN}} \sum_{k=0}^{M-1} \sum_{l=0}^{L-1} \exp \left( \frac{2\pi km}{M} + i\frac{2\pi ln}{N} \right) \]

\[ (i \in V, j \in V). \tag{9.166} \]

Here \( G_{k,l}(\alpha, \beta) \) in Eq. \( 9.153 \) is expressed as

\[ G_{k,l}(\alpha, \beta) = \frac{1}{1 - \tilde{m}(\alpha, \beta)^2} - 2\alpha \left( \cos \left( \frac{2\pi km}{M} \right) + \cos \left( \frac{2\pi ln}{N} \right) \right), \tag{9.167} \]

for the mean-field method and is

\[ G_{k,l}(\alpha, \beta) = \frac{3(\tilde{u}(\alpha, \beta) - \tilde{m}(\alpha, \beta)^2)^2 + (1 - \tilde{m}(\alpha, \beta)^2)^2}{(1 - \tilde{m}(\alpha, \beta)^2)(1 - \tilde{u}(\alpha, \beta))} \]

\[ - \frac{2(\tilde{u}(\alpha, \beta) - \tilde{m}(\alpha, \beta)^2)}{(1 - \tilde{u}(\alpha, \beta))} \left( \cos \left( \frac{2\pi km}{M} \right) + \cos \left( \frac{2\pi ln}{N} \right) \right), \tag{9.168} \]

for the loopy belief propagation, such that Bethe approximation.

References


[18] Y. Weiss and W. T. Freeman, On the optimality of solutions of the max-product belief propagation algorithm in


