

Chapter 10

Review of Sublinear Modeling in Probabilistic Graphical Models by Statistical Mechanical Informatics and Statistical Machine Learning Theory



Kazuyuki Tanaka

Abstract We review sublinear modeling in probabilistic graphical models by statistical mechanical informatics and statistical machine learning theory. Our statistical mechanical informatics schemes are based on advanced mean-field methods including loopy belief propagations. This chapter explores how phase transitions appear in loopy belief propagations for prior probabilistic graphical models. The frameworks are mainly explained for loopy belief propagations in the Ising model which is one of the elementary versions of probabilistic graphical models. We also expand the schemes to quantum statistical machine learning theory. Our framework can provide us with sublinear modeling based on the momentum space renormalization group methods.

10.1 Introduction

Statistical machine learning frameworks using **probabilistic graphical models** are useful for many applications, including information communication technologies [1–3], compressed sensing [4, 5] and neural information processing systems [6–10] in data-driven sciences.

Most probabilistic graphical models belong to the exponential family [11] and can be regarded as classical spin systems in **statistical mechanical informatics** [12–17]. However, it is well known that many applicable formulations in data sciences as well as computational sciences can be reduced to combinatorial problems with some constraint conditions which can be regarded as an **Ising Model** in statistical mechanical informatics [18, 19]. Moreover, much interest has focused on applying **quantum annealing** as a novel high-speed optimization technology to massive optimization problems [20–24].

K. Tanaka (✉)

Graduate School of Information Sciences, Tohoku University, Sendai, Japan
e-mail: kazu@tohoku.ac.jp

© The Author(s) 2022
N. Katoh et al. (eds.), *Sublinear Computation Paradigm*,
https://doi.org/10.1007/978-981-16-4095-7_10

10.2 Statistical Machine Learning

In statistical machine learning, most of the mathematical frameworks for machine learning are based on maximum likelihood frameworks [25, 26] from statistical mathematical sciences. The important points are how to assume the prior distribution and the data generative probability distribution and how to express the joint probability between the parameters and the data vector. In this section, we explore maximum likelihood frameworks in terms of model selection and parameter selection from a given data vector.

10.2.1 Bayesian Statistics and Maximization of Marginal Likelihood

Let us consider a graph specified by nodes and edges, (V, E) , where V is the set of all nodes i and E is the set of all edges $\{i, j\}$. State variables s_i and d_i are associated with

each node i . The vectors $\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_{|V|} \end{pmatrix}$ and $\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{|V|} \end{pmatrix}$ correspond to the parameters

and the data vector, respectively. The state spaces of s_i and d_i are given by Ω and $(-\infty, +\infty)$, respectively. Now $\rho(\mathbf{d}|\mathbf{s}, \beta)$ and $P(\mathbf{s}|\alpha)$ which correspond to the data generative and prior models, respectively, are assumed to be as follows:

$$\rho(\mathbf{d}|\mathbf{s}, \beta) = \prod_{i \in V} \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{1}{2}\beta(d_i - s_i)^2\right), \quad (10.1)$$

$$P(\mathbf{s}|\alpha) = \frac{\prod_{\{i,j\} \in E} \exp\left(-\frac{1}{2}\alpha(s_i - s_j)^2\right)}{\sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \prod_{\{i,j\} \in E} \exp\left(-\frac{1}{2}\alpha(s_i - s_j)^2\right)}. \quad (10.2)$$

The expressions for the posterior probability $P(\mathbf{s}|\mathbf{d}, \alpha, \beta)$, joint probability $\rho(\mathbf{s}, \mathbf{d}|\alpha, \beta)$, and marginal likelihood $\rho(\mathbf{d}|\alpha, \beta)$ are given by Bayes formulas as follows:

$$P(\mathbf{s}|\mathbf{d}, \alpha, \beta) = \frac{\rho(\mathbf{s}, \mathbf{d}|\alpha, \beta)}{\rho(\mathbf{d}|\alpha, \beta)} = \frac{\rho(\mathbf{d}|\mathbf{s}, \beta)P(\mathbf{s}|\alpha)}{\rho(\mathbf{d}|\alpha, \beta)}, \quad (10.3)$$

$$\rho(\mathbf{s}, \mathbf{d}|\alpha, \beta) = \rho(\mathbf{d}|\mathbf{s}, \beta)P(\mathbf{s}|\alpha), \quad (10.4)$$

$$\begin{aligned} \rho(\mathbf{d}|\alpha, \beta) &= \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \rho(\mathbf{s}, \mathbf{d}|\alpha, \beta) \\ &= \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \rho(\mathbf{d}|\mathbf{s}, \beta)P(\mathbf{s}|\alpha). \end{aligned} \quad (10.5)$$

Estimates of the hyperparameters and the parameter vector $\widehat{\alpha}, \widehat{\beta}, \widehat{\mathbf{s}} = (\widehat{s}_1, \widehat{s}_2, \dots, \widehat{s}_{|V|})$ are determined by

$$(\widehat{\alpha}(\mathbf{d}), \widehat{\beta}(\mathbf{d})) = \operatorname{argmax}_{(\alpha, \beta)} \rho(\mathbf{d}|\alpha, \beta), \quad (10.6)$$

$$\widehat{s}_i(\mathbf{d}) = \operatorname{argmax}_{s_i \in \Omega} P_i(s_i|\mathbf{d}, \widehat{\alpha}(\mathbf{d}), \widehat{\beta}(\mathbf{d})) \quad (i \in V). \quad (10.7)$$

Equations (10.6) and (10.7) are referred to as the **maximization of marginal likelihood (MML)** [25, 26] and the **maximization of posterior marginal (MPM)** [27], respectively.

10.2.2 Expectation-Maximization Algorithm

The expectation-maximization (EM) algorithm is often used to maximize the marginal likelihood in Eq. (10.6) [25, 26]. The \mathcal{Q} -function for the EM algorithm in the present framework is defined by

$$\mathcal{Q}(\alpha, \beta|\alpha', \beta', \mathbf{d}) \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} P(\mathbf{s}|\mathbf{d}, \alpha', \beta') \ln(\rho(\mathbf{s}, \mathbf{d}|\alpha, \beta)). \quad (10.8)$$

The EM algorithm is a procedure that performs the following procedures of **E-** and **M-step** repeatedly for $t = 0, 1, 2, \dots$ until $\widehat{\alpha}(\mathbf{d})$ and $\widehat{\beta}(\mathbf{d})$ converge:

E-step: Compute $\mathcal{Q}(\alpha, \beta|\alpha(\mathbf{d}, t), \beta(\mathbf{d}, t), \mathbf{d})$ for various values of α and β .

M-step: Determine $(\alpha(\mathbf{d}, t+1), \beta(\mathbf{d}, t+1))$ so as to satisfy the extremum conditions of $\mathcal{Q}(\alpha, \beta|\alpha(\mathbf{d}, t), \beta(\mathbf{d}, t), \mathbf{d})$ with respect to α and β . Update $\widehat{\alpha}(\mathbf{d}) \leftarrow \alpha(\mathbf{d}, t+1)$ and $\widehat{\beta}(\mathbf{d}) \leftarrow \beta(\mathbf{d}, t+1)$.

The update rule from $(\alpha(\mathbf{d}, t), \beta(\mathbf{d}, t))$ to $(\alpha(\mathbf{d}, t+1), \beta(\mathbf{d}, t+1))$ for the extremum conditions can be written as

$$\begin{aligned}
& \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} (s_i - s_j)^2 P(s_1, s_2, \dots, s_{|V|} | \alpha(\mathbf{d}, t+1)) \\
&= \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} (s_i - s_j)^2 P(s_1, s_2, \dots, s_{|V|} | \mathbf{d}, \alpha(\mathbf{d}, t), \beta(\mathbf{d}, t)),
\end{aligned} \tag{10.9}$$

$$\frac{1}{\beta(\mathbf{d}, t+1)} = \frac{1}{|V|} \sum_{i \in V} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} (s_i - d_i)^2 P(s_1, s_2, \dots, s_{|V|} | \mathbf{d}, \alpha(\mathbf{d}, t), \beta(\mathbf{d}, t)). \tag{10.10}$$

The marginal probability distributions of $P(\mathbf{s} | \mathbf{d}, \alpha, \beta)$ and $P(\mathbf{s} | \alpha)$ are introduced as

$$P_i(s_i | \mathbf{d}, \alpha, \beta) \equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} P(\tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{d}, \alpha, \beta) \quad (i \in V), \tag{10.11}$$

$$\begin{aligned}
P_{ij}(s_i, s_j | \mathbf{d}, \alpha, \beta) &= P_{ji}(s_j, s_i | \mathbf{d}, \alpha, \beta) \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} P(\tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{d}, \alpha, \beta) \quad (\{i, j\} \in E),
\end{aligned} \tag{10.12}$$

$$\begin{aligned}
P_{ij}(s_i, s_j | \alpha) &= P_{ji}(s_j, s_i | \alpha) \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} P(\tau_1, \tau_2, \dots, \tau_{|V|} | \alpha) \quad (\{i, j\} \in E).
\end{aligned} \tag{10.13}$$

In this way, the extremum conditions can be reduced to

$$\begin{aligned}
& \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{s_i \in \Omega} \sum_{s_j \in \Omega} (s_i - s_j)^2 P_{ij}(s_i, s_j | \alpha(\mathbf{d}, t+1)) \\
&= \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{s_i \in \Omega} \sum_{s_j \in \Omega} (s_i - s_j)^2 P_{ij}(s_i, s_j | \mathbf{d}, \alpha(\mathbf{d}, t), \beta(\mathbf{d}, t)),
\end{aligned} \tag{10.14}$$

$$\frac{1}{\beta(\mathbf{d}, t+1)} = \frac{1}{|V|} \sum_{i \in V} \sum_{s_i \in \Omega} (s_i - d_i)^2 P_i(s_i | \mathbf{d}, \alpha(\mathbf{d}, t), \beta(\mathbf{d}, t)). \tag{10.15}$$

To realize the EM procedure as a practical algorithm, **Markov chain Monte Carlo (MCMC) Methods** are often used, which are powerful probabilistic methods [28, 29]. In some recent developments, advanced mean-field methods from statistical mechanical informatics are also used as powerful deterministic algorithms, as

shown in Sect. 10.3. Consider the expectation values for both sides of Eqs. (10.9) and (10.10) with respect to the state vector \mathbf{d} of a data point according to the following probability density function where the hyperparameters α and β are set to their true values α^* and β^* , respectively:

$$\rho(\mathbf{d}|\alpha^*, \beta^*) = \sum_{\boldsymbol{\tau} \in \Omega^{|V|}} \rho(\mathbf{d}|\boldsymbol{\tau}, \alpha^*, \beta^*) P(\boldsymbol{\tau}|\alpha^*), \quad (10.16)$$

such that

$$\begin{aligned} \rho(d_1, d_2, \dots, d_{|V|}|\alpha^*, \beta^*) &= \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \rho(d_1, d_2, \dots, d_{|V|}|\tau_1, \tau_2, \dots, \tau_{|V|}, \beta^*) \\ &\quad \times P(\tau_1, \tau_2, \dots, \tau_{|V|}|\alpha^*) \\ &= (d_1 \in (-\infty, +\infty), d_2 \in (-\infty, +\infty), \dots, d_{|V|} \in (-\infty, +\infty)). \end{aligned} \quad (10.17)$$

We can then derive simultaneous equations for the statistical trajectory $\{\bar{\alpha}(\alpha^*, \beta^*, t), \bar{\beta}(\alpha^*, \beta^*, t)|t = 1, 2, 3, \dots\}$ in the convergence process $\{(\alpha(\mathbf{d}, t), \beta(\mathbf{d}, t))|t = 1, 2, 3, \dots\}$ of the above EM algorithm.

Equations (10.9) and (10.10) can be rewritten as follows:

$$\frac{1}{|E|} \frac{\partial}{\partial \alpha(\mathbf{d}, t+1)} (\ln(Z(\alpha(\mathbf{d}, t+1)))) = \frac{1}{|E|} \frac{\partial}{\partial \alpha(\mathbf{d}, t)} (\ln(Z(\mathbf{d}, \alpha(\mathbf{d}, t), \beta(\mathbf{d}, t)))), \quad (10.18)$$

$$\frac{1}{\beta(\mathbf{d}, t+1)} = \frac{1}{|V|} \frac{\partial}{\partial \beta(\mathbf{d}, t)} (\ln(Z(\mathbf{d}, \alpha(\mathbf{d}, t), \beta(\mathbf{d}, t)))), \quad (10.19)$$

where

$$Z(\alpha) \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \prod_{\{i,j\} \in E} \exp\left(-\frac{1}{2} \alpha (s_i - s_j)^2\right), \quad (10.20)$$

$$Z(\mathbf{d}, \alpha, \beta) \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} w(s_1, s_2, \dots, s_{|V|}|\mathbf{d}, \alpha, \beta), \quad (10.21)$$

$$\begin{aligned} w(\mathbf{s}|\mathbf{d}, \alpha, \beta) &= w(s_1, s_2, \dots, s_{|V|}|d_1, d_2, \dots, d_{|V|}, \alpha, \beta) \\ &\equiv \left(\prod_{i \in V} \exp\left(-\frac{1}{2} \beta (s_i - d_i)^2\right) \right) \left(\prod_{\{i,j\} \in E} \exp\left(-\frac{1}{2} \alpha (s_i - s_j)^2\right) \right). \end{aligned} \quad (10.22)$$

By taking the expectation values of both sides of Eqs. (10.18) and (10.19) with respect to the state vector of the data point \mathbf{d} in the probability density function $\rho(\mathbf{d}|\alpha^*, \beta^*)$, the simultaneous deterministic equation for the statistical trajectory

$\{\bar{\alpha}(\alpha^*, \beta^*, t), \bar{\beta}(\alpha^*, \beta^*, t) | t = 1, 2, 3, \dots\}$ of the EM procedure can be derived as follows:

$$\begin{aligned} & \frac{1}{|E|} \frac{\partial}{\partial \bar{\alpha}(\alpha^*, \beta^*, t+1)} (\ln(Z(\bar{\alpha}(\alpha^*, \beta^*, t+1)))) \\ &= \frac{1}{|E|} \frac{\partial}{\partial \bar{\alpha}(\alpha^*, \beta^*, t)} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) (\ln(Z(\mathbf{d}, \bar{\alpha}(\alpha^*, \beta^*, t), \bar{\beta}(\alpha^*, \beta^*, t)))) dd_1 dd_2 \dots dd_{|V|} \right), \end{aligned} \quad (10.23)$$

$$\begin{aligned} \frac{1}{\bar{\beta}(\alpha^*, \beta^*, t+1)} &= \frac{1}{|V|} \frac{\partial}{\partial \bar{\beta}(\alpha^*, \beta^*, t)} \\ & \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) (\ln(Z(\mathbf{d}, \bar{\alpha}(\alpha^*, \beta^*, t), \bar{\beta}(\alpha^*, \beta^*, t)))) dd_1 dd_2 \dots dd_{|V|} \right). \end{aligned} \quad (10.24)$$

In the case of a continuous state space $\Omega = (-\infty, +\infty)$, the posterior and prior probabilistic models correspond to Gaussian graphical models, and the statistical trajectory in Eqs. (10.23) and (10.24) can be exactly computed by means of the multi-dimensional Gaussian integral formula [30].

For a discrete state space Ω , it is generally hard to treat Eqs. (10.23) and (10.24) analytically. To estimate Eqs. (10.23) and (10.24), the following quantity is often introduced in statistical mechanical informatics [13, 17]:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) (\ln(Z(\mathbf{d}, \alpha, \beta))) dd_1 dd_2 \dots dd_{|V|}. \quad (10.25)$$

The quantity in Eq. (10.25) can be rewritten as follows:

$$\begin{aligned} & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) \left(\lim_{n \rightarrow +0} \frac{1}{n} (Z(\mathbf{d}, \alpha, \beta)^n - 1) \right) dd_1 dd_2 \dots dd_{|V|} \\ &= \lim_{n \rightarrow +0} \frac{1}{n} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) ((Z(\mathbf{d}, \alpha, \beta)^n - 1)) dd_1 dd_2 \dots dd_{|V|} \\ &= \lim_{n \rightarrow +0} \frac{1}{n} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) \left(\sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \dots \sum_{s_{|V|} \in \Omega} w(\mathbf{s} | \mathbf{d}, \alpha, \beta) \right)^n dd_1 dd_2 \dots dd_{|V|} - 1 \\ &= \lim_{n \rightarrow +0} \frac{1}{n} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho(\mathbf{d} | \alpha^*, \beta^*) \right. \\ & \quad \times \prod_{j=1}^n \left(\sum_{s_{1,j} \in \Omega} \sum_{s_{2,j} \in \Omega} \dots \sum_{s_{|V|,j} \in \Omega} w(s_{1,j}, s_{2,j}, \dots, s_{|V|,j} | \mathbf{d}, \alpha, \beta) \right) dd_1 dd_2 \dots dd_{|V|} \left. - 1 \right) \\ &= \frac{1}{Z(\alpha^*)} \left(\sqrt{\frac{\beta^*}{2\pi}} \right)^{|V|} \\ & \times \left\{ \lim_{n \rightarrow +0} \frac{1}{n} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left(\sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \dots \sum_{\tau_{|V|} \in \Omega} w(\tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{d}, \alpha^*, \beta^*) \right) \right. \right. \\ & \quad \times \prod_{j=1}^n \left(\sum_{s_{1,j} \in \Omega} \sum_{s_{2,j} \in \Omega} \dots \sum_{s_{|V|,j} \in \Omega} w(s_{1,j}, s_{2,j}, \dots, s_{|V|,j} | \mathbf{d}, \alpha, \beta) \right) \left. \left. \right) dd_1 dd_2 \dots dd_{|V|} \right\} - 1. \end{aligned} \quad (10.26)$$

Equation (10.26) means that computation of the statistical quantity in Eq. (10.25) can be reduced, up to some normalization constant, to computation of the statistical quantity in the probabilistic model given by the weight factor

$$w(\tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{d}, \alpha^*, \beta^*) \prod_{j=1}^n w(s_{1,j}, s_{2,j}, \dots, s_{|V|,j} | \mathbf{d}, \alpha, \beta). \quad (10.27)$$

We remark that the weight factor (10.27) is expressed by considering some replicas of the posterior probabilistic model $P(s | \mathbf{d}, \alpha, \beta)$ and the analysis starting from the weight factor (10.27) is referred to as a **replica method** [13, 17]. One possible case for analytical treatment is the EM algorithm with the prior and posterior probabilistic models in Eqs. (10.2) and (10.3) for the complete graph (V, E) . The dynamics of the EM algorithm with the MCMC method can be analyzed by using the replica method and the **master equations** for **Glauber dynamics** [31].¹

10.2.3 Expectation-Maximization Algorithm for Probabilistic Image Segmentations

This section extends the previous section to the statistical machine learning framework for probabilistic image segmentation. In probabilistic image segmentations, we consider a square grid graph (V, E) in which a light intensity vector $\mathbf{d}_i = (d_{iR}, d_{iG}, d_{iB})$ for the three components red d_{iR} , green d_{iG} and blue d_{iB} is assigned to each node i . The state vector s for the labeled configuration and the data matrix \mathbf{D} for the color image configuration are expressed as

$$s = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_{|V|} \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \\ \vdots \\ \mathbf{d}_{|V|} \end{pmatrix} = \begin{pmatrix} d_{1R} & d_{1G} & d_{1B} \\ d_{2R} & d_{2G} & d_{2B} \\ d_{3R} & d_{3G} & d_{3B} \\ \vdots & \vdots & \vdots \\ d_{|V|R} & d_{|V|G} & d_{|V|B} \end{pmatrix}. \quad (10.28)$$

Here $\rho(\mathbf{D} | s, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ and $P(s | \alpha)$ are assumed to be as follows:

$$\rho(\mathbf{D} | s, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) = \prod_{i \in V} g(\mathbf{d}_i | s_i, \mathbf{a}(s_i), \mathbf{C}(s_i)), \quad (10.29)$$

¹ **Glauber dynamics** was proposed in Ref. [32].

$$P(\mathbf{s}|\alpha) = \frac{\prod_{\{i,j\} \in E} \exp(-2\alpha(1 - \delta_{s_i, s_j}))}{\sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \prod_{\{i,j\} \in E} \exp(-2\alpha(1 - \delta_{s_i, s_j}))}, \quad (10.30)$$

where

$$g(\mathbf{d}_i | s_i, \mathbf{a}(s_i), \mathbf{C}(s_i)) \equiv \sqrt{\frac{1}{\det(2\pi \mathbf{C}(s_i))}} \exp\left(-\frac{1}{2}(\mathbf{d}_i - \mathbf{a}(s_i))\mathbf{C}^{-1}(s_i)(\mathbf{d}_i - \mathbf{a}(s_i))^T\right), \quad (10.31)$$

$$\mathbf{a}(+1) = \begin{pmatrix} a_R(+1) \\ a_G(+1) \\ a_B(+1) \end{pmatrix}, \quad \mathbf{a}(-1) = \begin{pmatrix} a_R(-1) \\ a_G(-1) \\ a_B(-1) \end{pmatrix}, \quad (10.32)$$

$$\mathbf{C}(+1) = \begin{pmatrix} C_{RR}(+1) & C_{RG}(+1) & C_{RB}(+1) \\ C_{GR}(+1) & C_{GG}(+1) & C_{GB}(+1) \\ C_{BR}(+1) & C_{BG}(+1) & C_{BB}(+1) \end{pmatrix}, \quad \mathbf{C}(-1) = \begin{pmatrix} C_{RR}(-1) & C_{RG}(-1) & C_{RB}(-1) \\ C_{GR}(-1) & C_{GG}(-1) & C_{GB}(-1) \\ C_{BR}(-1) & C_{BG}(-1) & C_{BB}(-1) \end{pmatrix}. \quad (10.33)$$

Note that the probabilistic graphical model in Eq. (10.30) is referred to as a **Potts mode** [33].

In probabilistic segmentation and clustering, $\rho(\mathbf{D}|\mathbf{s}, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ in Eq. (10.29) and $P(\mathbf{s}|\alpha)$ in Eq. (10.30) correspond to the data generative and prior models, respectively. The joint probability of \mathbf{s} and \mathbf{D} is expressed in terms of the data generative and prior distributions, $\rho(\mathbf{D}|\mathbf{s}, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ and $P(\mathbf{s}|\alpha)$, as follows:

$$\rho(\mathbf{s}, \mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \equiv \rho(\mathbf{D}|\mathbf{s}, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))P(\mathbf{s}|\alpha). \quad (10.34)$$

By using the joint probability distribution, the posterior probability $P(\mathbf{s}|\mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ and the marginal likelihood $\rho(\mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ are defined by using Bayes formulas as follows:

$$P(\mathbf{s}|\mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \equiv \frac{\rho(\mathbf{s}, \mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))}{\rho(\mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))}, \quad (10.35)$$

$$\begin{aligned} & \rho(\mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \\ & \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \rho(\mathbf{s}, \mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)). \end{aligned} \quad (10.36)$$

Estimates of the hyperparameters and parameter vector, namely, $\widehat{\alpha}(\mathbf{D})$, $\widehat{\mathbf{a}}(+1|\mathbf{D})$, $\widehat{\mathbf{a}}(-1|\mathbf{D})$, $\widehat{\mathbf{C}}(+1|\mathbf{D})$, $\widehat{\mathbf{C}}(-1|\mathbf{D})$, $\widehat{\mathbf{s}}(\mathbf{D}) = (\widehat{s}_1(\mathbf{D}), \widehat{s}_2(\mathbf{D}), \dots, \widehat{s}_{|V|}(\mathbf{D}))$ are determined by

$$\begin{aligned} & \left(\widehat{\alpha}(\mathbf{D}), \widehat{\mathbf{a}}(+1|\mathbf{D}), \widehat{\mathbf{a}}(-1|\mathbf{D}), \widehat{\mathbf{C}}(+1|\mathbf{D}), \widehat{\mathbf{C}}(-1|\mathbf{D}) \right) \\ &= \arg \max_{(\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))} \rho(\mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)), \end{aligned} \quad (10.37)$$

$$\widehat{s}_i(\mathbf{D}) = \arg \max_{s_i \in \Omega} P_i(s_i | \mathbf{D}, \widehat{\alpha}(\mathbf{D}), \widehat{\mathbf{a}}(+1|\mathbf{D}), \widehat{\mathbf{a}}(-1|\mathbf{D}), \widehat{\mathbf{C}}(+1|\mathbf{D}), \widehat{\mathbf{C}}(-1|\mathbf{D})) \quad (i \in V). \quad (10.38)$$

The \mathcal{Q} -function for the EM algorithm in the present framework is defined by

$$\begin{aligned} & \mathcal{Q}(\alpha, \mathbf{s}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1) | \mathbf{s}'(+1), \mathbf{s}'(-1), \mathbf{C}'(+1), \mathbf{C}'(-1), \mathbf{D}) \\ & \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} P(\mathbf{s} | \mathbf{D}, \alpha, \mathbf{a}'(+1), \mathbf{a}'(-1), \mathbf{C}'(+1), \mathbf{C}'(-1)) \\ & \quad \times \ln(\rho(\mathbf{s}, \mathbf{D} | \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))). \end{aligned} \quad (10.39)$$

The EM algorithm is a procedure that performs the following **E-step** and **M-step** repeatedly for $t = 0, 1, 2, \dots$ until $\widehat{\alpha}(\mathbf{D})$, $\widehat{\mathbf{a}}(+1, \mathbf{D})$, $\widehat{\mathbf{a}}(-1, \mathbf{D})$, $\widehat{\mathbf{C}}(+1, \mathbf{D})$, $\widehat{\mathbf{C}}(-1, \mathbf{D})$ converge:

E-step: Compute $\mathcal{Q}(\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1) | \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t))$ for various values of $\mathbf{a}(+1)$, $\mathbf{a}(-1)$, $\mathbf{C}(+1)$, and $\mathbf{C}(-1)$.

M-step: Determine $\alpha(t+1)$, $\mathbf{a}(+1, t+1)$, $\mathbf{a}(-1, t+1)$, $\mathbf{C}(+1, t+1)$, and $\mathbf{C}(-1, t+1)$ that satisfy the extremum conditions of \mathcal{Q} -function with respect to $\mathbf{a}(+1)$, $\mathbf{a}(-1)$, $\mathbf{C}(+1)$ and $\mathbf{C}(-1)$ as follows:

$$\begin{aligned} & (\alpha(t+1), \mathbf{a}(+1, t+1), \mathbf{a}(-1, t+1), \mathbf{C}(+1, t+1), \mathbf{C}(-1, t+1)) \\ & \leftarrow \underset{(\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))}{\text{extremum}} \\ & \mathcal{Q}(\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1) | \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t), \mathbf{D}). \end{aligned} \quad (10.40)$$

Update $\widehat{\alpha}(\mathbf{D}) \leftarrow \alpha(t+1)$, $\widehat{\mathbf{a}}(+1, \mathbf{D}) \leftarrow \mathbf{a}(+1, t+1)$, $\widehat{\mathbf{a}}(-1, \mathbf{D}) \leftarrow \mathbf{a}(-1, t+1)$, $\widehat{\mathbf{C}}(+1, \mathbf{D}) \leftarrow \mathbf{C}(+1, t+1)$ and $\widehat{\mathbf{C}}(-1, \mathbf{D}) \leftarrow \mathbf{C}(-1, t+1)$.

By using the equalities in Eqs. (10.29), (10.30), (10.34), and (10.35), the EM algorithm by the \mathcal{Q} -function can be reduced to the following simultaneous update rules:

$$\begin{aligned} & \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{s_i \in \Omega} \sum_{s_j \in \Omega} (1 - \delta_{s_i, s_j}) P_{ij}(s_i, s_j | \alpha(t+1)) \\ &= \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{s_i \in \Omega} \sum_{s_j \in \Omega} (1 - \delta_{s_i, s_j}) P_{ij}(s_i, s_j | \mathbf{D}, \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)), \end{aligned} \quad (10.41)$$

$$\mathbf{a}(s_i, t+1) = \frac{\sum_{i \in V} \mathbf{d}_i P_i(s_i | \mathbf{D}, \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t))}{\sum_{i \in V} P_i(s_i | \mathbf{D}, \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t))} \quad (s_i \in \Omega), \quad (10.42)$$

$$\begin{aligned} & \mathbf{C}(s_i, t+1) \\ &= \frac{\sum_{i \in V} (\mathbf{d}_i - \mathbf{a}(s_i, t))^T (\mathbf{d}_i - \mathbf{a}(s_i, t)) P_i(s_i | \mathbf{D}, \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t))}{\sum_{i \in V} P_i(s_i | \mathbf{D}, \alpha(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t))} \quad (s_i \in \Omega), \end{aligned} \quad (10.43)$$

where

$$\begin{aligned} & P_i(s_i | \mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \\ & \equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} P(\tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \quad (i \in V), \end{aligned} \quad (10.44)$$

$$\begin{aligned} & P_{ij}(s_i, s_j | \mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \\ &= P_{ji}(s_j, s_i | \mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \\ & \equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} P(\tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{D}, \alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \quad (\{i, j\} \in E), \end{aligned} \quad (10.45)$$

$$\begin{aligned} P_{ij}(s_i, s_j | \alpha) &= P_{ji}(s_j, s_i | \alpha) \\ & \equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} P(\tau_1, \tau_2, \dots, \tau_{|V|} | \alpha) \quad (\{i, j\} \in E). \end{aligned} \quad (10.46)$$

10.3 Statistical Mechanical Informatics

In statistical mechanical informatics [13–17], Ising models are very familiar probabilistic models for which computations are done by statistical mechanical techniques, including advanced mean-field methods, renormalization group methods, Monte Carlo simulations, and replica methods [36, 37]. This section reviews the framework of the Ising model and associated advanced mean-field methods.²

² A review of both exact results and approximate results as well as perturbative computations for Ising models is given in Refs. [34, 35].

10.3.1 Ising Model

Let us consider an Ising model defined by the following probability distribution for the state space $\Omega = \{+1, -1\}$ for the state variable s_i at each node $i \in V$:

$$\begin{aligned}
 P\left(s \mid d, \frac{J}{k_B T}, \frac{h}{k_B T}\right) &= P\left(s_1, s_2, \dots, s_{|V|} \mid d_1, d_2, \dots, d_{|V|}, \frac{J}{k_B T}, \frac{h}{k_B T}\right) \\
 &\equiv \frac{\exp\left(-\frac{1}{k_B T} \left(\frac{1}{2} J \sum_{\{i,j\} \in E} (s_i - s_j)^2 + \frac{1}{2} h \sum_{i \in V} (s_i - d_i)^2\right)\right)}{\sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \dots \sum_{s_{|V|} \in \Omega} \exp\left(-\frac{1}{k_B T} \left(\frac{1}{2} J \sum_{\{i,j\} \in E} (s_i - s_j)^2 + \frac{1}{2} h \sum_{i \in V} (s_i - d_i)^2\right)\right)} \\
 &\quad (J > 0, T > 0, d_i \in (-\infty, +\infty) (\forall i \in V)). \tag{10.47}
 \end{aligned}$$

Because $s_i^2 = 1$ ($i \in V$), the probability distribution $P(s)$ can be reduced to

$$P\left(s \mid d, \frac{J}{k_B T}, \frac{h}{k_B T}\right) = \frac{1}{Z} \exp\left(-\frac{1}{k_B T} H(s)\right) \quad (T > 0), \tag{10.48}$$

$$\begin{aligned}
 H(s) &= H(s_1, s_2, \dots, s_{|V|}) \\
 &\equiv -J \sum_{\{i,j\} \in E} s_i s_j - h \sum_{i \in V} d_i s_i \quad (J > 0, h \geq 0, d_i \in (-\infty, +\infty) (\forall i \in V)), \tag{10.49}
 \end{aligned}$$

$$Z \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \dots \sum_{s_{|V|} \in \Omega} \exp\left(-\frac{1}{k_B T} H(s)\right), \tag{10.50}$$

where $H(s)$ and Z are referred to in statistical mechanical informatics as the **energy function** (or **Hamiltonian**) and the **partition function**, respectively, the probability distribution in Eq. (10.47) is called the **Gibbs distribution**, k_B is the **Boltzmann constant**, T is the (**absolute**) **temperature**, J is the (**ferromagnetic**) **interaction**, and h is the **external field**.

Let us suppose the **Kullback-Leibler Divergence**

$$\text{KL}[P \parallel R] \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \dots \sum_{s_{|V|} \in \Omega} R(s) \ln \left(\frac{R(s)}{P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T}\right)} \right), \tag{10.51}$$

which is always non-negative for two probability distributions $P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T}\right)$ and $R(s)$ and is regarded as a pseudo-distance between them. By substituting the explicit expression for $P(s)$ in Eqs. (10.48), (10.49) and (10.50) into Eq. (10.51), the expression for the Kullback-Leibler divergence (10.51) in terms of the partition function Z and the free energy functional $\mathcal{F}[R]$ can be derived as follows:

$$\text{KL}[P||R] = \frac{1}{k_B T} (k_B T \ln(Z) + \mathcal{F}[R]), \quad (10.52)$$

where

$$\mathcal{F}[R] \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} H(s) R(s) + k_B T \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} R(s) \ln(R(s)). \quad (10.53)$$

For the free energy functional $\mathcal{F}[R]$, it is valid that

$$\operatorname{argmin}_R \left\{ \mathcal{F}[R] \left| \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} R(\tau_1, \tau_2, \dots, \tau_{|V|}) = 1 \right. \right\} = P \left(s \left| \frac{J}{k_B T}, \frac{h}{k_B T} \right. \right), \quad (10.54)$$

$$\min_R \left\{ \mathcal{F}[R] \left| \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} R(\tau_1, \tau_2, \dots, \tau_{|V|}) = 1 \right. \right\} = -k_B T \ln(Z). \quad (10.55)$$

Note that $-k_B T \ln(Z)$ is referred to as the **free energy** for the Gibbs distribution in Eq. (10.47).

10.3.2 Advanced Mean-Field Method

This section reviews the fundamental framework of advanced mean-field methods [12], including the **mean-field approximation** [35–37] and the **Bethe approximation** [35, 39–41]. Our framework is given for the Ising model in Eqs. (10.48), (10.49), and (10.50). It is known that a generalization of the present framework can be realized by using the **cluster variation method** in Refs. [42–45].

We introduce a trial probability distribution $R(s) = R(s_1, s_2, \dots, s_{|V|})$ which is restricted to the following functional form:

$$R(s) = R(s_1, s_2, \dots, s_{|V|}) = \prod_{i \in V} R_i(s_i), \quad (10.56)$$

$$R_i(s_i) = \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} R(\tau_1, \tau_2, \dots, \tau_{|V|}) \quad (i \in V). \quad (10.57)$$

By using the definition of $R_i(s_i)$ and the normalization condition

$$\sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} R(\tau_1, \tau_2, \dots, \tau_{|V|}) = 1, \quad (10.58)$$

we confirm that

$$\sum_{\tau_i \in \Omega} R_i(\tau_i) = R_i(+1) + R_i(-1) = 1 \quad (i \in V). \quad (10.59)$$

By substituting the expression $R(s)$ in terms of the marginal probability distribution $R_i(s_i)$ ($i \in V$), such that $\mathbf{R}_i = \begin{pmatrix} R_i(+1) & 0 \\ 0 & R_i(-1) \end{pmatrix}$ into Eqs. (10.56)–(10.53), the free energy functional $\mathcal{F}[R]$ can be reduced to the following mean-field free energy functional:

$$\mathcal{F}[R] = \mathcal{F}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}], \quad (10.60)$$

where

$$\begin{aligned} \mathcal{F}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] &= \mathcal{F}_{\text{MF}}\left[\left\{\begin{pmatrix} R_i(+1) & 0 \\ 0 & R_i(-1) \end{pmatrix} \middle| i \in V\right\}\right] \\ &\equiv -J \sum_{\{i,j\} \in E} \left(\sum_{\tau_i \in \Omega} \tau_i R_i(\tau_i) \right) \left(\sum_{\tau_j \in \Omega} \tau_j R_j(\tau_j) \right) - h \sum_{i \in V} d_i \left(\sum_{\tau_i \in \Omega} \tau_i R_i(\tau_i) \right) \\ &\quad + k_B T \sum_{i \in V} \sum_{\tau_i \in \Omega} R_i(\tau_i) \ln(R_i(\tau_i)). \end{aligned} \quad (10.61)$$

Let us suppose the following conditional minimization of the free energy functional:

$$\begin{aligned} \widehat{\mathbf{R}}_i &= \left\{ \begin{pmatrix} \widehat{R}_i(+1) & 0 \\ 0 & \widehat{R}_i(-1) \end{pmatrix} \middle| i \in V, s_i \in \Omega \right\} \\ &= \arg \min_{\{\mathbf{R}_i | i \in V\}} \left\{ \mathcal{F}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] \middle| \sum_{\tau_i \in \Omega} R_i(\tau_i) = 1, i \in V \right\}. \end{aligned} \quad (10.62)$$

First we introduce the Lagrange multiplier λ_i ($i \in V$) to ensure the normalization conditions $\sum_{\tau \in \Omega} R_i(\tau) = 1$ ($i \in V$) as follows:

$$\mathcal{L}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] \equiv \mathcal{F}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] - \sum_{i \in V} \lambda_i \left(\sum_{\tau_i \in \Omega} R_i(\tau_i) - 1 \right). \quad (10.63)$$

$\widehat{\mathbf{R}}_i$ ($i \in V$) are determined so as to satisfy the following extremum condition:

$$\left[\frac{\partial}{\partial \mathbf{R}_i(-1)} \mathcal{L}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] \right]_{\{\mathbf{R}_i = \widehat{\mathbf{R}}_i | i \in V\}} = 0 \quad (i \in V), \quad (10.64)$$

$$\left[\frac{\partial}{\partial \mathbf{R}_i(+1)} \mathcal{L}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] \right]_{\{\mathbf{R}_i = \widehat{\mathbf{R}}_i | i \in V\}} = 0 \quad (i \in V), \quad (10.65)$$

such that

$$\left[\frac{\partial}{\partial \mathbf{R}_i} \mathcal{L}_{\text{MF}}[\{\mathbf{R}_i | i \in V\}] \right]_{\{\mathbf{R}_i = \widehat{\mathbf{R}}_i | i \in V\}} = 0 \quad (i \in V). \quad (10.66)$$

It needs to be shown that $\widehat{\mathbf{R}}_i$ ($i \in V$) are derived as follows:

$$\begin{aligned} & \widehat{\mathbf{R}}_i(s_i) \\ &= \exp\left(-1 + \frac{\lambda_i}{k_B T}\right) \exp\left(\frac{1}{k_B T} \left(h d_i + J \sum_{j \in \partial i} \left(\sum_{\tau_j \in \Omega} \tau_j \widehat{\mathbf{R}}_j(\tau_j) \right) \right) s_i\right) \quad (i \in V, s_i \in \Omega). \end{aligned} \quad (10.67)$$

Finally, λ_i needs to be determined such that it satisfies the normalization condition of the marginal probability $\widehat{\mathbf{R}}_i(s_i)$. The marginal probabilities $\{\widehat{\mathbf{R}}_i | i \in V\}$ are derived as

$$\widehat{\mathbf{R}}_i(s_i) = \frac{\exp\left(\frac{1}{k_B T} \left(h d_i + J \sum_{j \in \partial i} \left(\sum_{\tau_j \in \Omega} \tau_j \widehat{\mathbf{R}}_j(\tau_j) \right) \right) s_i\right)}{\sum_{\tau_i \in \Omega} \exp\left(\frac{1}{k_B T} \left(h d_i + J \sum_{j \in \partial i} \left(\sum_{\tau_j \in \Omega} \tau_j \widehat{\mathbf{R}}_j(\tau_j) \right) \right) \tau_i\right)} \quad (i \in V, s_i \in \Omega). \quad (10.68)$$

We introduce the local magnetization

$$m_i \equiv \sum_{\tau_i \in \Omega} \tau_i \widehat{\mathbf{R}}_i(\tau_i). \quad (10.69)$$

By solving the simultaneous equations

$$\sum_{\tau_i \in \Omega} \widehat{\mathbf{R}}_i(\tau_i) = \widehat{\mathbf{R}}_i(+1) + \widehat{\mathbf{R}}_i(-1) = 1 \quad (i \in V), \quad (10.70)$$

$$\sum_{\tau_i \in \Omega} \tau_i \widehat{\mathbf{R}}_i(\tau_i) = \widehat{\mathbf{R}}_i(+1) - \widehat{\mathbf{R}}_i(-1) = m_i \quad (i \in V), \quad (10.71)$$

with respect to $\widehat{\mathbf{R}}_i(+1)$ and $\widehat{\mathbf{R}}_i(-1)$, we derive the following expression for the marginal probability:

$$\widehat{R}_i(s_i) = \frac{1}{2}(1 + m_i s_i) \quad (i \in V). \quad (10.72)$$

The extremum conditions in Eq. (10.68) can be reduced to the following simultaneous deterministic equation of $\{m_i | i \in V\}$:

$$m_i = \tanh \left(\frac{1}{k_B T} \left(h d_i + J \sum_{j \in \partial i} m_j \right) \right) \quad (i \in V), \quad (10.73)$$

which is referred to as the **mean-field equation**.³

By substituting Eq. (10.72) into Eq. (10.61), the mean-field free energy functional can be reduced to

$$\mathcal{F}_{\text{MF}}[\{\widehat{R}_i(-1), \widehat{R}_i(+1) | i \in V\}] = F_{\text{MF}}(m_1, m_2, \dots, m_{|V|}), \quad (10.74)$$

$$\begin{aligned} F_{\text{MF}}(m_1, m_2, \dots, m_{|V|}) &\equiv -J \sum_{\{i,j\} \in E} m_i m_j - h \sum_{i \in V} d_i m_i \\ &\quad + k_B T \sum_{i \in V} \frac{1}{2} (1 + m_i) \ln \left(\frac{1}{2} (1 + m_i) \right) \\ &\quad + k_B T \sum_{i \in V} \frac{1}{2} (1 - m_i) \ln \left(\frac{1}{2} (1 - m_i) \right). \end{aligned} \quad (10.75)$$

The extremum conditions

$$\frac{\partial}{\partial m_i} F_{\text{MF}}(m_1, m_2, \dots, m_{|V|}) = 0 \quad (i \in V) \quad (10.76)$$

can be reduced to the mean-field equations in Eq. (10.73).

We now explore the framework of the Bethe approximation for the Ising model in Eqs. (10.48), (10.49), and (10.50). Our framework is based on the cluster variation method [39, 42–45].

We introduce a trial probability distribution $R(\mathbf{s}) = R(s_1, s_2, \dots, s_{|V|})$ that is restricted to the following functional form:

³ Equation (10.73) is often referred to as the **naive mean-field equation** in statistical machine learning theory [2, 3, 12].

$$\begin{aligned}
R(\mathbf{s}) = R(s_1, s_2, \dots, s_{|V|}) &= \left(\prod_{i \in V} R_i(s_i) \right) \left(\prod_{\{i, j\} \in E} \frac{R_{ij}(s_i, s_j)}{R_i(s_i) R_j(s_j)} \right) \\
&= \left(\prod_{i \in V} R_i(s_i)^{1-|\partial i|} \right) \left(\prod_{\{i, j\} \in E} R_{ij}(s_i, s_j) \right), \quad (10.77)
\end{aligned}$$

where

$$\begin{aligned}
R_{ij}(s_i, s_j) &= R_{ji}(s_j, s_i) \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} R(\tau_1, \tau_2, \dots, \tau_{|V|}) \quad (\{i, j\} \in E). \quad (10.78)
\end{aligned}$$

By using Eqs. (10.57) and (10.78), we can derive the normalization and reducibility conditions in the marginal probabilities as follows:

$$\sum_{\tau_i \in \Omega} R_i(\tau_i) = 1 \quad (i \in V), \quad \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} R_{ij}(\tau_i, \tau_j) = 1 \quad (\{i, j\} \in E), \quad (10.79)$$

$$R_i(s_i) = \sum_{\tau_j \in \Omega} R_{ij}(s_i, \tau_j), \quad R_j(s_j) = \sum_{\tau_i \in \Omega} R_{ij}(\tau_i, s_j) \quad (\{i, j\} \in E). \quad (10.80)$$

By substituting the explicit expression for $P(\mathbf{s})$ and the expression $\ln(R(\mathbf{s}))$ in terms of the marginal probability distributions $\mathbf{R}_i = \begin{pmatrix} R_i(+1) & 0 \\ 0 & R_i(-1) \end{pmatrix}$ ($i \in V$) and $\mathbf{R}_{ij} = \begin{pmatrix} R_{ij}(+1, +1) & 0 & 0 & 0 \\ 0 & R_{ij}(+1, -1) & 0 & 0 \\ 0 & 0 & R_{ij}(-1, +1) & 0 \\ 0 & 0 & 0 & R_{ij}(-1, -1) \end{pmatrix}$ ($\{i, j\} \in E$) in Eq. (10.77) into Eq. (10.51), the Kullback-Leibler divergence can be reduced to the following expression in terms of the partition function Z and the Bethe free energy functional $\mathcal{F}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}]$:

$$\text{KL}[P||R] = \frac{1}{k_B T} (k_B T \ln Z + \mathcal{F}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}]), \quad (10.81)$$

where

$$\begin{aligned}
& \mathcal{F}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}] \\
& \equiv -J \sum_{\{i, j\} \in E} \left(\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \tau_i \tau_j R_{\{i, j\}}(\tau_i, \tau_j) \right) - h \sum_{i \in V} \left(\sum_{\tau_i \in \Omega} \tau_i R_i(\tau_i) \right) \\
& \quad + k_B T \sum_{i \in V} (1 - |\partial i|) \left(\sum_{\tau_i \in \Omega} R_i(\tau_i) \ln(R_i(\tau_i)) \right) \\
& \quad + k_B T \sum_{\{i, j\} \in E} \left(\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} R_{ij}(\tau_i, \tau_j) \ln(R_{ij}(\tau_i, \tau_j)) \right). \tag{10.82}
\end{aligned}$$

Let us suppose the following conditional minimization of the Bethe free energy functional:

$$\begin{aligned}
& (\{\widehat{\mathbf{R}}_i | i \in V\}, \{\widehat{\mathbf{R}}_{\{i, j\}} | \{i, j\} \in E\}) \\
& = \arg \min_{\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{\{i, j\}} | \{i, j\} \in E\}} \left\{ \mathcal{F}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{\{i, j\}} | \{i, j\} \in E\}] \right. \\
& \quad \left. \begin{aligned}
& \sum_{\tau_i \in \Omega} R_i(\tau_i) = 1 \quad (i \in V), \quad \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} R_{ij}(\tau_i, \tau_j) = 1 \quad (\{i, j\} \in E), \\
& R_i(-1) = \sum_{\tau_j \in \Omega} R_{ij}(-1, \tau_j) \quad (j \in \partial i, i \in V), \\
& R_i(+1) = \sum_{\tau_j \in \Omega} R_{ij}(+1, \tau_j) \quad (j \in \partial i, i \in V) \Big\}. \tag{10.83}
\end{aligned}
\right.
\end{aligned}$$

We introduce the Lagrange multiplier λ_i ($i \in V$), $\lambda_{\{i, j\}}$, $\lambda_{i, ij}(-1) = \lambda_{i, ji}(-1)$, $\lambda_{i, ij}(+1) = \lambda_{i, ji}(+1)$ ($\{i, j\} \in E$) to ensure the normalization and reducibility conditions as follows:

$$\begin{aligned}
\mathcal{L}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}] & \equiv \mathcal{F}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}] \\
& \quad - \sum_{i \in V} \lambda_i \left(\sum_{\tau_i \in \Omega} R_i(\tau_i) - 1 \right) \\
& \quad - \sum_{\{i, j\} \in E} \lambda_{\{i, j\}} \left(\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} R_{ij}(\tau_i, \tau_j) - 1 \right) \\
& \quad - \sum_{i \in V} \sum_{j \in \partial i} \lambda_{i, ij}(-1) (R_i(-1) - \sum_{\tau_j \in \Omega} R_{ij}(-1, \tau_j)) \\
& \quad - \sum_{i \in V} \sum_{j \in \partial i} \lambda_{i, ij}(+1) (R_i(+1) - \sum_{\tau_j \in \Omega} R_{ij}(+1, \tau_j)). \tag{10.84}
\end{aligned}$$

The marginal probabilities $\widehat{\mathbf{R}}_i$ ($i \in V$) and $\widehat{\mathbf{R}}_{ij}$ ($\{i, j\} \in E$) are determined so as to satisfy the following extremum condition:

$$\left[\frac{\partial}{\partial \mathbf{R}_i} \mathcal{L}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}] \right]_{\{\mathbf{R}_i = \widehat{\mathbf{R}}_i | i \in V\}, \{\mathbf{R}_{ij} = \widehat{\mathbf{R}}_{ij} | \{i, j\} \in E\}} = 0 \quad (i \in V), \tag{10.85}$$

$$\left[\frac{\partial}{\partial \mathbf{R}_{ij}} \mathcal{L}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i, j\} \in E\}] \right]_{\{\mathbf{R}_i = \widehat{\mathbf{R}}_i | i \in V\}, \{\mathbf{R}_{ij} = \widehat{\mathbf{R}}_{ij} | \{i, j\} \in E\}} = 0 \quad (\{i, j\} \in E). \quad (10.86)$$

It needs to be shown that $\widehat{\mathbf{R}}_i(s_i)$ ($i \in V$) and $\widehat{\mathbf{R}}_{ij}(s_i, s_j) = \widehat{\mathbf{R}}_{ji}(s_j, s_i)$ ($\{i, j\} \in E$) are derived as follows:

$$\begin{aligned} \widehat{\mathbf{R}}_i(s_i) &= \exp\left(-1 - \frac{\lambda_i}{k_B T (|\partial i| - 1)}\right) \\ &\quad \times \exp\left(\frac{1}{k_B T} \left(-\frac{1}{|\partial i| - 1} h d_i s_i - \frac{1}{|\partial i| - 1} \sum_{j \in \partial i} \lambda_{i,ij}(s_i)\right)\right) \quad (i \in V), \end{aligned} \quad (10.87)$$

$$\begin{aligned} \widehat{\mathbf{R}}_{ij}(s_i, s_j) &= \widehat{\mathbf{R}}_{ji}(s_j, s_i) = \exp\left(-1 + \frac{\lambda_{\{i,j\}}}{k_B T}\right) \\ &\quad \times \exp\left(\frac{1}{k_B T} (J s_i s_j - \lambda_{i,ij}(s_i) - \lambda_{j,ij}(s_j))\right) \quad (\{i, j\} \in E). \end{aligned} \quad (10.88)$$

Finally, λ_i and $\lambda_{\{i,j\}}$ need to be determined so as to satisfy the normalization condition of the marginal probabilities $\widehat{\mathbf{R}}_i(s_i)$ and $\widehat{\mathbf{R}}_{ij}(s_i, s_j)$.

By introducing the messages $\mu_{k \rightarrow i}(s_i)$ and $\mu_{l \rightarrow j}(s_j)$ in the transformations

$$\exp\left(-\frac{\lambda_{i,ij}(s_i)}{k_B T}\right) = \left(\prod_{k \in \partial i \setminus \{j\}} \mu_{k \rightarrow i}(s_i)\right) \exp\left(\frac{h}{k_B T} d_i s_i\right), \quad (10.89)$$

$$\exp\left(-\frac{\lambda_{j,ij}(s_j)}{k_B T}\right) = \left(\prod_{l \in \partial j \setminus \{i\}} \mu_{l \rightarrow j}(s_j)\right) \exp\left(\frac{h}{k_B T} d_j s_j\right). \quad (10.90)$$

The expressions of the marginal probabilities $\widehat{\mathbf{R}}_i$ and $\widehat{\mathbf{R}}_{ij}$ in Eqs. (10.87) and (10.88) can be reduced to the following expressions:

$$\widehat{\mathbf{R}}_i(s_i) = \frac{1}{Z_i} \left(\prod_{k \in \partial i} \mu_{k \rightarrow i}(s_i)\right) \exp\left(\frac{h}{k_B T} d_i s_i\right) \quad (i \in V), \quad (10.91)$$

$$\begin{aligned} \widehat{\mathbf{R}}_{ij}(s_i, s_j) = \widehat{\mathbf{R}}_{ji}(s_j, s_i) &= \frac{1}{Z_{\{i,j\}}} \left(\prod_{k \in \partial i \setminus \{j\}} \mu_{k \rightarrow i}(s_i) \right) \left(\prod_{l \in \partial j \setminus \{i\}} \mu_{l \rightarrow j}(s_j) \right) \\ &\times \exp\left(\frac{1}{k_B T} (J s_i s_j + h d_i s_i + h d_j s_j)\right) \quad (\{i, j\} \in E), \end{aligned} \quad (10.92)$$

$$Z_i \equiv \sum_{\tau_i \in \Omega} \left(\prod_{k \in \partial i} \mu_{k \rightarrow i}(\tau_i) \right) \exp\left(\frac{h}{k_B T} d_i \tau_i\right) \quad (i \in V), \quad (10.93)$$

$$\begin{aligned} Z_{\{i,j\}} &\equiv \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \left(\prod_{k \in \partial i \setminus \{j\}} \mu_{k \rightarrow i}(\tau_i) \right) \left(\prod_{l \in \partial j \setminus \{i\}} \mu_{l \rightarrow j}(\tau_j) \right) \\ &\times \exp\left(\frac{1}{k_B T} (J \tau_i \tau_j + h d_i \tau_i + h d_j \tau_j)\right) \quad (\{i, j\} \in E). \end{aligned} \quad (10.94)$$

By substituting Eqs. (10.91) and (10.92) into the reducibility conditions in Eq. (10.80), the simultaneous deterministic equations for the messages can be derived as follows:

$$\mu_{j \rightarrow i}(s_i) = \frac{Z_i}{Z_{\{i,j\}}} \sum_{\tau_j \in \Omega} \left(\prod_{l \in \partial j \setminus \{i\}} \mu_{l \rightarrow j}(\tau_j) \right) \exp\left(\frac{1}{k_B T} (J s_i \tau_j + h d_j \tau_j)\right) \quad (\{i, j\} \in E), \quad (10.95)$$

$$\mu_{i \rightarrow j}(s_j) = \frac{Z_j}{Z_{\{i,j\}}} \sum_{\tau_i \in \Omega} \left(\prod_{k \in \partial i \setminus \{j\}} \mu_{k \rightarrow i}(\tau_i) \right) \exp\left(\frac{1}{k_B T} (J \tau_i s_j + h d_i \tau_i)\right) \quad (\{i, j\} \in E). \quad (10.96)$$

The Bethe free energy functional is given by

$$\mathcal{F}_{\text{Bethe}}[\{\widehat{\mathbf{R}}_i | i \in V\}, \{\widehat{\mathbf{R}}_{ij} | \{i, j\} \in E\}] = -k_B T \sum_{i \in V} (1 - |\partial i|) \ln Z_i - k_B T \sum_{\{i,j\} \in E} \ln Z_{\{i,j\}}. \quad (10.97)$$

The framework in the Bethe approximation using Eqs. (10.91), (10.92), (10.93), and (10.94) with Eqs. (10.95) and (10.96) is referred to as a **loopy belief propagation** in statistical machine learning theory [12, 46–48]. The present derivation is based on the cluster variation method in Refs. [39, 42–44], and [45]. Recently, some novel approaches for loopy belief propagation methods have been proposed, including the **approximate message passing algorithm** [49], and **replica cluster variation method** [50, 51]. A review summarizing recent developments in loopy belief propagation methods is given in Ref. [52].

By solving

$$\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \widehat{R}_{ij}(\tau_i, \tau_j) = \widehat{R}_{ij}(+1, +1) + \widehat{R}_{ij}(-1, +1) + \widehat{R}_{ij}(+1, -1) + \widehat{R}_{ij}(-1, -1) = 1, \quad (10.98)$$

$$\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \tau_i \widehat{R}_{ij}(\tau_i, \tau_j) = \widehat{R}_{ij}(+1, +1) - \widehat{R}_{ij}(-1, +1) + \widehat{R}_{ij}(+1, -1) - \widehat{R}_{ij}(-1, -1) = m_i, \quad (10.99)$$

$$\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \tau_j \widehat{R}_{ij}(\tau_i, \tau_j) = \widehat{R}_{ij}(+1, +1) + \widehat{R}_{ij}(-1, +1) - \widehat{R}_{ij}(+1, -1) - \widehat{R}_{ij}(-1, -1) = m_j, \quad (10.100)$$

$$\begin{aligned} \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \tau_i \tau_j \widehat{R}_{ij}(\tau_i, \tau_j) &= \widehat{R}_{ij}(+1, +1) - \widehat{R}_{ij}(-1, +1) \\ &\quad - \widehat{R}_{ij}(+1, -1) + \widehat{R}_{ij}(-1, -1) = c_{\{i,j\}} = c_{\{j,i\}}, \end{aligned} \quad (10.101)$$

as simultaneous linear equations for $\widehat{R}_{ij}(+1, +1)$, $\widehat{R}_{ij}(-1, +1)$, $\widehat{R}_{ij}(+1, -1)$, and $\widehat{R}_{ij}(-1, -1)$, we can confirm the following equality:

$$\widehat{R}_{ij}(s_i, s_j) = \frac{1}{4} (1 + m_i s_i + m_j s_j + c_{\{i,j\}} s_i s_j). \quad (10.102)$$

By substituting Eqs. (10.72) and (10.102) into Eq. (10.82), the Bethe free energy functional can be reduced to

$$\mathcal{F}_{\text{Bethe}} \left[\left\{ \widehat{R}_i \mid i \in V \right\}, \left\{ \widehat{R}_{ij} \mid \{i, j\} \in E \right\} \right] = F_{\text{Bethe}} \left(\{m_i \mid i \in V\}, \{c_{\{i,j\}} \mid \{i, j\} \in E\} \right), \quad (10.103)$$

$$\begin{aligned}
F_{\text{Bethe}}(\{m_i | i \in V\}, \{c_{\{i,j\}} | \{i,j\} \in E\}) &\equiv -J \sum_{\{i,j\} \in E} c_{\{i,j\}} - h \sum_{i \in V} d_i m_i \\
&+ k_B T \sum_{i \in V} (1 - |\partial i|) \sum_{s_i = \pm 1} \widehat{R}_i(s_i) \ln(\widehat{R}_i(s_i)) + k_B T \sum_{\{i,j\} \in E} \sum_{s_j = \pm 1} \widehat{R}_{ij}(s_i, s_j) \ln(\widehat{R}_{ij}(s_i, s_j)).
\end{aligned} \tag{10.104}$$

The extremum conditions

$$\frac{\partial}{\partial m_k} F_{\text{Bethe}}(\{m_i | i \in V\}, \{c_{\{i,j\}} | \{i,j\} \in E\}) = 0 \quad (k \in V), \tag{10.105}$$

$$\frac{\partial}{\partial c_{\{k,l\}}} F_{\text{Bethe}}(\{m_i | i \in V\}, \{c_{\{i,j\}} | \{i,j\} \in E\}) = 0 \quad (\{k,l\} \in E) \tag{10.106}$$

can be reduced to the following simultaneous equations:

$$\frac{h}{k_B T} d_i = \frac{1}{2} (1 - |\partial i|) \sum_{\tau_i \in \Omega} \tau_i \ln(\widehat{R}_i(\tau_i)) + \frac{1}{4} \sum_{j \in \partial i} \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \tau_i \ln(\widehat{R}_{ij}(\tau_i, \tau_j)) \quad (i \in V), \tag{10.107}$$

$$\frac{J}{k_B T} = \frac{1}{4} \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \tau_i \tau_j \ln(\widehat{R}_{ij}(\tau_i, \tau_j)) \quad (\{i,j\} \in E). \tag{10.108}$$

The schemes for the derivations of Eqs. (10.107) and (10.108) from the Bethe free energy (Eqs. (10.103)–(10.104)) are given in Refs. [41, 53–55].

In the advanced mean-field method, some researchers are interested in perturbative computation of the correction terms with respect to $\frac{J}{k_B T}$ from the mean-field free energy [56, 57], which is referred to as a **Thouless-Anderson-Palmar (TAP) free energy**. The scheme used in the derivations has been extended to a classical Heisenberg model [58]. One familiar perturbative method in statistical mechanical informatics is the Plefka expansion, in which we obtain higher-order correction terms with respect to $\frac{J}{k_B T}$ from the mean-field free energy [12]. By substituting Eq. (10.102) into Eq. (10.108), $c_{\{i,j\}}$ can be expressed in terms of m_i , m_j , and $\frac{J}{k_B T}$. It is known that the TAP equation can be derived by expanding the expression for $c_{\{i,j\}}$ up to the second-order term $\left(\frac{J}{k_B T}\right)^2$ with respect to an infinitesimal $\frac{J}{k_B T}$ and by substituting it into Eq. (10.82) with Eqs. (10.72) and (10.102) [41]. The fundamental framework of the TAP free energy and its expansion using the advanced mean-field method has been clarified [59]. The Bethe free energy functional and the TAP free energy as well as loopy belief propagation have been applied to Boltzmann machine learning [53, 60–64]. Some recent developments appear in Chap. 7 of Part 3 in this book.

The EM schemes with advanced mean-field methods in the previous sections have been applied to noise reduction in probabilistic image processing [30, 51, 65–69]. The basic frameworks are based on Eqs. (10.14) and (10.15) with the two-body and

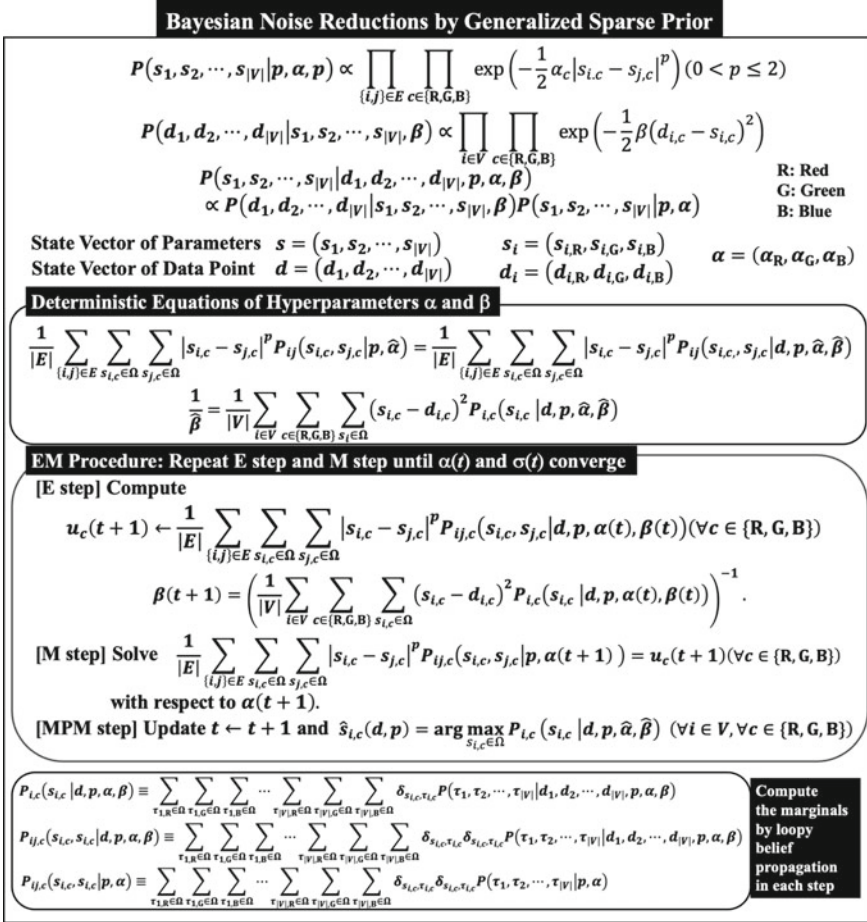


Fig. 10.1 Fundamental framework of Bayesian noise reduction by generalized sparse prior and additive white Gaussian noise

one-body posterior marginal probability distributions in Eqs. (10.11) and (10.12) as well as the two-body prior marginal probability distribution in Eq. (10.13). They can be computed by means of the message passing algorithms in Eqs. (10.91) and (10.92) with Eqs. (10.93), (10.94), (10.95), and (10.96) for the Ising model in Eqs. (10.47), (10.48), and (10.50) with the prior and posterior probability distributions in Eqs. (10.2) and (10.3), respectively. The framework and some numerical experimental results are shown in Figs. 10.1 and 10.2, respectively. Moreover, the loopy belief propagation is applicable to Bayesian image segmentation in the framework of Sect. 10.2.3 [70]. They are also useful for community detection by means of the stochastic block model for modular networks [71–73].

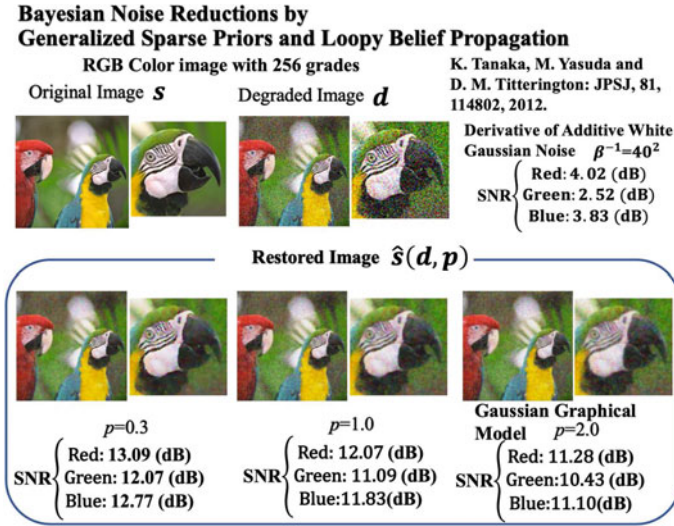


Fig. 10.2 Numerical experiments in Bayesian noise reduction by the generalized sparse prior and additive white Gaussian noise

10.3.3 Free Energy Landscapes and Phase Transitions in the Thermodynamic Limit

In this section, we consider the Ising model defined by

$$P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T}\right) \equiv \frac{1}{Z\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right)} \exp\left(\frac{1}{k_B T} H(s)\right), \quad (10.109)$$

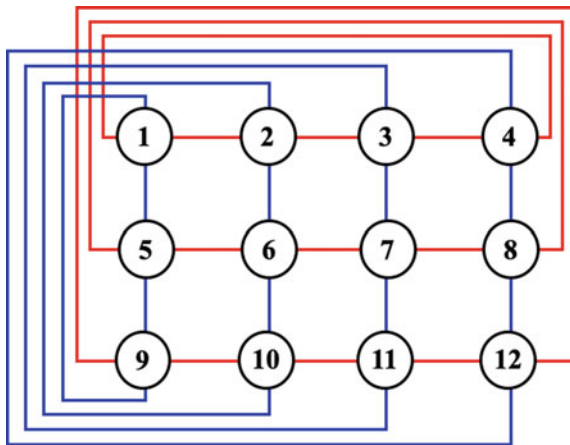
where

$$\begin{aligned} H(s) &= H(s_1, s_2, \dots, s_{|V|}) \\ &\equiv -J \sum_{\{i, j\} \in E} s_i s_j - h \sum_{i \in V} s_i \quad (J > 0), \end{aligned} \quad (10.110)$$

$$Z\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \dots \sum_{s_{|V|} \in \Omega} \exp\left(-\frac{1}{k_B T} H(s)\right). \quad (10.111)$$

The energy function in Eq. (10.110) corresponds to the one in Eq. (10.49) for the case of $d_i = 1$ for every node $i \in V$. In the present section, we consider a regular of degree 4 that includes a square grid graph with periodic boundary conditions along the x - and y -direction as shown in Fig. 10.3.

Fig. 10.3 Square grid graph (V, E) with periodic boundary conditions along the x - and y -direction in the case $V = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$



For the Ising model in Eq. (10.110) and its partition function in Eq. (10.50), we have the free energy per node

$$f(J, h, T) = -k_B T \times \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \ln(Z), \quad (10.112)$$

the internal energy for zero external field

$$\begin{aligned} Ju &\equiv \frac{\partial}{\partial \left(\frac{1}{k_B T}\right)} \left(\frac{1}{k_B T} f(J, h = 0, T) \right) \\ &= \lim_{|V| \rightarrow +\infty} \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_s (-s_i s_j) P \left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T} = 0 \right), \end{aligned} \quad (10.113)$$

and the spontaneous magnetization

$$\begin{aligned} m_{\pm} &\equiv \lim_{h \rightarrow \pm 0} \frac{\partial}{\partial \left(\frac{h}{k_B T}\right)} \left(\frac{1}{k_B T} f(J, h, T) \right) \\ &= \lim_{h \rightarrow \pm 0} \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_s s_i P \left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T} \right), \end{aligned} \quad (10.114)$$

as important statistical quantities in the thermodynamic limit $|V| \rightarrow +\infty$. The existence of the thermodynamic limit $|V| \rightarrow +\infty$ means that the limit of the right-hand side in Eq. (10.110) converges. Sufficient conditions for the existence of the thermodynamic limit of the Ising model of Eqs. (10.109), (10.110), and (10.111) have been given by Ruelle in Ref. [38].

In the thermodynamic limit $|V| \rightarrow +\infty$ for the Ising model in Eq. (10.110) on a square grid graph with periodic boundary conditions along the x - and y -direction as

shown in Fig. 10.3,

$$\frac{u}{J} = -\coth\left(\frac{2J}{k_B T}\right) \left(1 + \left(2 \tanh^2\left(\frac{2J}{k_B T}\right) - 1 \right) \left(\frac{2}{\pi}\right) \int_0^{\frac{\pi}{2}} \left(1 - \left(\frac{2 \sinh\left(\frac{2J}{k_B T}\right)}{\cosh^2\left(\frac{2J}{k_B T}\right)} \right)^2 \sin^2(\theta) \right)^{-\frac{1}{2}} d\theta \right), \quad (10.115)$$

$$\begin{aligned} m_{\pm}^2 &= \lim_{|r_i - r_j| \rightarrow +\infty} \lim_{|V| \rightarrow +\infty} \sum_s s_i s_j P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T} = 0\right) \\ &= \begin{cases} 0 & \left(\frac{J}{k_B T} < \frac{1}{2} \operatorname{arcsinh}(1)\right) \\ \left(1 - \sinh^{-4}\left(\frac{2J}{k_B T}\right)\right)^{\frac{1}{4}} & \left(\frac{J}{k_B T} > \frac{1}{2} \operatorname{arcsinh}(1)\right) \end{cases}, \quad (10.116) \end{aligned}$$

where r_i is the position vector of each node $i \in V$ [34, 74, 75]. In Eq. (10.116), the spontaneous magnetizations m_+ and m_- correspond to each branch of $m_+ \geq 0$ and $m_- \leq 0$, respectively. They are as shown in Fig. 10.4. Note that for the Ising model in Eq. (10.110) on such regular graphs,

$$m_{i,V}\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) \equiv \sum_s s_i P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T}\right), \quad (10.117)$$

for every $i \in V$, does not depend on i but can be expressed as $m_V\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right)$.

In the mean-field approximation of the previous subsection, the spontaneous magnetizations

$$m_{\pm} = \lim_{h \rightarrow \pm 0} \lim_{|V| \rightarrow +\infty} m_V\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right), \quad (10.118)$$

are given as solutions of the following mean-field equation:

$$m_{\pm} = \tanh\left(\frac{1}{k_B T}(h + 4Jm_{\pm})\right) \quad (h \rightarrow \pm 0), \quad (10.119)$$

and the internal energy Ju in Eq. (10.113) in the mean-field approximation is given as

$$u = -m_{\pm}^2. \quad (10.120)$$

The solutions of Eq. (10.119) correspond to the extremum values of the following mean-field free energy:

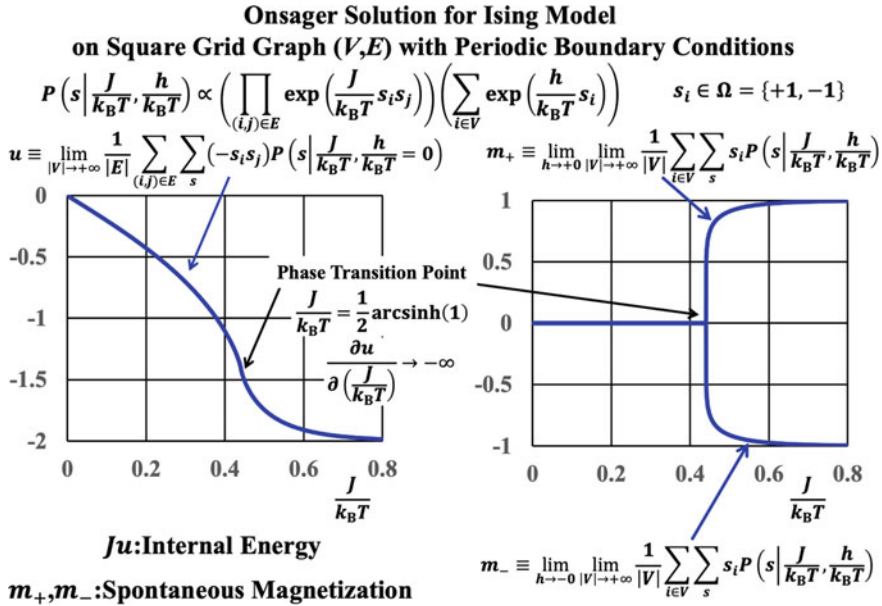


Fig. 10.4 Internal energy Ju in Eq. (10.113) and magnetization m_{\pm} in Eq. (10.114) in the Onsager solution in the Ising model of Eqs. (10.48) and (10.50) with Eq. (10.110) on the square grid graph (V, E) with the periodic boundary conditions along the x - and y -direction

$$\begin{aligned} f_{\text{MF}}(m) &\equiv \frac{1}{|V|} F_{\text{MF}}(m_1, m_2, \dots, m_{|V|}) \\ &= -2Jm^2 + k_B T \left(\frac{1}{2}(1+m) \right) \ln \left(\frac{1}{2}(1+m) \right) \\ &\quad + k_B T \left(\frac{1}{2}(1-m) \right) \ln \left(\frac{1}{2}(1-m) \right), \quad (10.121) \end{aligned}$$

which corresponds to $\frac{1}{|V|} F_{\text{MF}}(m_1, m_2, \dots, m_{|V|})$ in Eq. (10.75) for $h = 0$. The spontaneous magnetization m_{\pm} and the internal energy u for $h = 0$ are computed by setting $0 < \left| \frac{h}{k_B T} \right| < 10^{-5}$ and using the iteration method for Eq. (10.119) numerically. The graphs of $\left(\frac{J}{k_B T}, u \right)$ and $\left(\frac{J}{k_B T}, m_{\pm} \right)$ are shown in Fig. 10.5. Moreover, the graphs of $\left(m, \frac{1}{k_B T} f_{\text{MF}}(m) \right)$ for $\frac{J}{k_B T} = 0.20, 0.25,$ and 0.40 are shown in Fig. 10.6. It is known that the mean-field equation always has the trivial solution $m_{\pm} = 0$, and begins to have some non-trivial solutions for $m_+ > 0$ and $m_- < 0$. The mean-field equation (10.119) begins to have some non-trivial solutions in the region of $\frac{J}{k_B T} > \frac{1}{4}$ by expanding the right-hand side of Eq. (10.119) around $m = 0$ and keeping the first-order term of m .

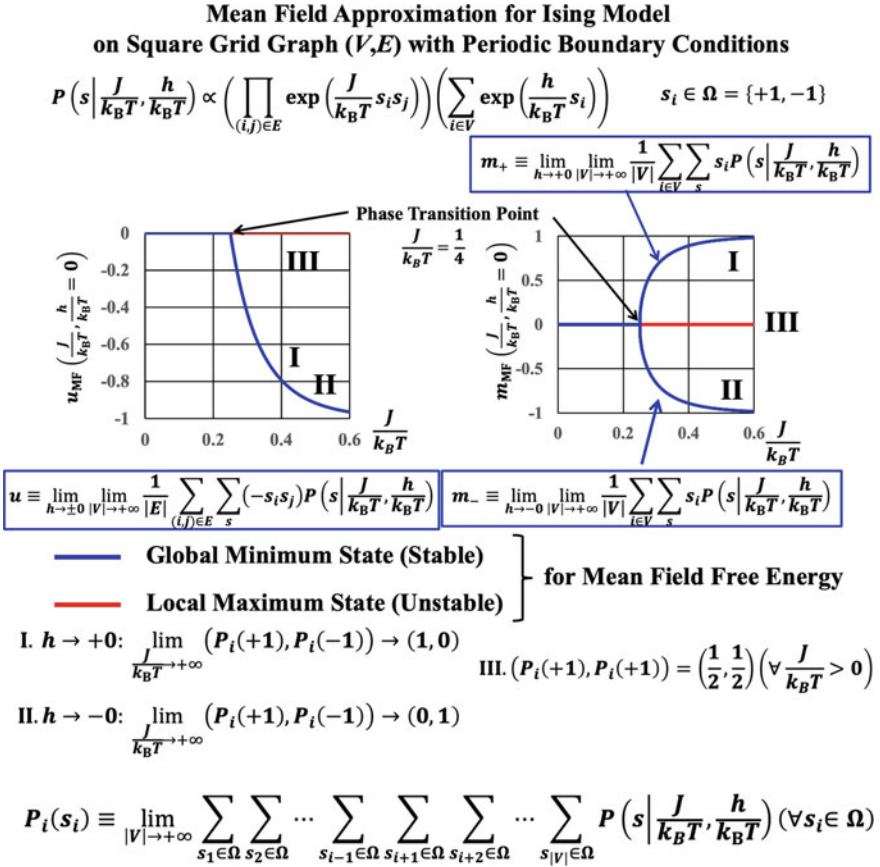


Fig. 10.5 Internal energy u from Eq. (10.113) and magnetization m_{\pm} from Eq. (10.118) in mean-field approximation for the Ising model in Eqs. (10.48), (10.50), and (10.110) on the regular graph (V, E) of degree 4

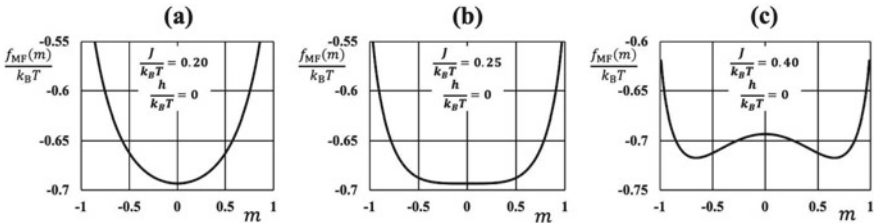


Fig. 10.6 Free energy from Eq. (10.121) in mean-field approximation for the Ising model in Eqs. (10.48), (10.50) and (10.110) on the regular graph (V, E) of degree 4. **a** $\frac{J}{k_B T} = 0.2$, $\frac{h}{k_B T} = 0$. **b** $\frac{J}{k_B T} = 0.25$, $\frac{h}{k_B T} = 0$. **c** $\frac{J}{k_B T} = 0.4$, $\frac{h}{k_B T} = 0$

Next, we consider the Bethe approximation for the Ising model in Eqs. (10.48), (10.50), and (10.110) on the regular graph (V, E) of degree 4. In this case, the average m_i , the correlation c_{ij} and the messages $\mu_{i \rightarrow j}(+1)$ and $\mu_{i \rightarrow j}(-1)$ do not depend on i and j , and can be expressed as $m, c, \mu(+1)$ and $\mu(-1)$. We now introduce

$$\Lambda \equiv \frac{1}{2} k_B T \ln \left(\frac{\mu(+1)}{\mu(-1)} \right). \quad (10.122)$$

The message passing equations in Eqs. (10.95) and (10.96) and the magnetization are reduced to

$$\frac{\Lambda}{k_B T} = \operatorname{arctanh} \left(\tanh \left(\frac{J}{k_B T} \right) \tanh \left(\frac{h + 3\Lambda}{k_B T} \right) \right). \quad (10.123)$$

Moreover, since the marginal probabilities $\widehat{R}_i(+1)$ and $\widehat{R}_i(-1)$ are also independent of i , we can derive the expression for the magnetization in terms of Λ as follows:

$$m_{\text{Bethe}} \left(\frac{J}{k_B T}, \frac{h}{k_B T} \right) \equiv \frac{1}{|V|} \sum_{i \in V} \left(\sum_{s_i \in \Omega} s_i \widehat{R}(s_i) \right) = \tanh \left(\frac{h + 4\Lambda}{k_B T} \right). \quad (10.124)$$

For the infinitesimal small limits of h , such that $h \rightarrow +0$ and $h \rightarrow -0$, the magnetization $m_{\text{Bethe}} \left(\frac{J}{k_B T}, \frac{h}{k_B T} \right)$ in Eq. (10.124) can be computed numerically by using the iteration method. Moreover, Eqs. (10.104) and (10.108) can be reduced to

$$\begin{aligned} f_{\text{Bethe}}(m, c) &= \frac{1}{|V|} F_{\text{Bethe}}(\{m_i | i \in V\}, \{c_{\{i,j\}} | \{i, j\} \in E\}) \\ &= -2Jc - hm \\ &\quad - 3k_B T \left(\frac{1}{2}(1+m) \right) \ln \left(\frac{1}{2}(1+m) \right) - 3k_B T \left(\frac{1}{2}(1-m) \right) \ln \left(\frac{1}{2}(1-m) \right) \\ &\quad + 2k_B T \left(\frac{1}{4}(1+2m+c) \right) \ln \left(\frac{1}{4}(1+2m+c) \right) \\ &\quad + 4k_B T \left(\frac{1}{4}(1-c) \right) \ln \left(\frac{1}{4}(1-c) \right) \\ &\quad + 2k_B T \left(\frac{1}{4}(1-2m+c) \right) \ln \left(\frac{1}{4}(1-2m+c) \right), \end{aligned} \quad (10.125)$$

and

$$\frac{J}{k_B T} = \ln \frac{(1+c)^2 - 4m^2}{(1-c)^2}, \quad (10.126)$$

such that

$$c = \frac{1}{\tanh\left(\frac{2J}{k_B T}\right)} \left(1 - \sqrt{1 - (1 - 2m^2)\tanh^2\left(\frac{2J}{k_B T}\right) - 2m^2\tanh\left(\frac{2J}{k_B T}\right)} \right). \quad (10.127)$$

The graphs of

$$\left(m, \frac{1}{k_B T} f_{\text{Bethe}} \left(m, \frac{1}{\tanh\left(\frac{2J}{k_B T}\right)} \left(1 - \sqrt{1 - (1 - 2m^2)\tanh^2\left(\frac{2J}{k_B T}\right) - 2m^2\tanh\left(\frac{2J}{k_B T}\right)} \right) \right) \right), \quad (10.128)$$

for $\frac{J}{k_B T} = 0.25$, $\frac{J}{k_B T} = \text{arctanh}\left(\frac{1}{3}\right)$, and $\frac{k_B T}{J} = 0.40$ in the case of $h = 0$ are shown in Figs. 10.7, 10.8, and 10.9, respectively. Figure 10.7 shows the internal energy u from Eq. (10.113) and the spontaneous magnetization m_{\pm} in Eq. (10.118) in loopy belief propagation (Bethe approximation) for the Ising model in Eqs. (10.48), (10.50), and (10.110) on the regular graph (V, E) of degree 4. These quantities u and m_{\pm} are obtained by

$$u_{\text{Bethe}} \left(\frac{J}{k_B T}, \frac{h}{k_B T} = 0 \right) = \frac{\cosh\left(\frac{6\Lambda}{k_B T}\right) - \exp\left(-\frac{2J}{k_B T}\right)}{\cosh\left(\frac{6\Lambda}{k_B T}\right) + \exp\left(-\frac{2J}{k_B T}\right)}, \quad (10.129)$$

$$m_{\text{Bethe}} \left(\frac{J}{k_B T}, \frac{h}{k_B T} = 0 \right) = \tanh\left(\frac{4\Lambda}{k_B T}\right), \quad (10.130)$$

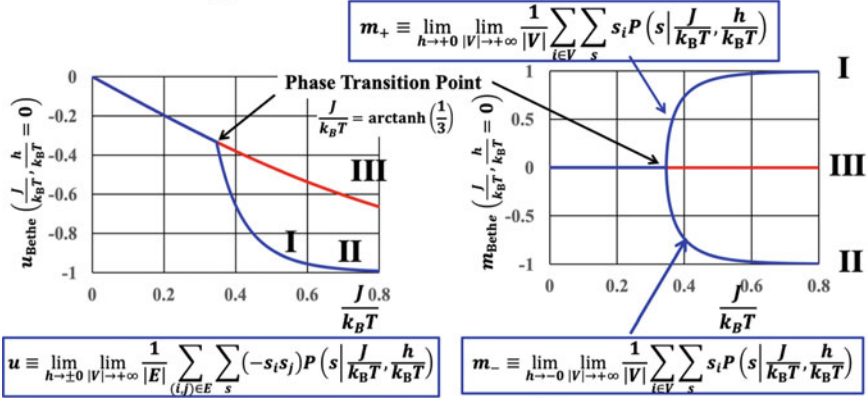
where

$$\frac{\Lambda}{k_B T} = \text{arctanh} \left(\tanh\left(\frac{J}{k_B T}\right) \tanh\left(\frac{3\Lambda}{k_B T}\right) \right). \quad (10.131)$$

These always give the same results as in Eq. (10.110) on the regular graph (V, E) of degree 4. In particular, it is known that the results for Eqs. (10.48), (10.50), and (10.110) on the regular tree graph (V, E) of degree 4 are exact. It is known that Eq. (10.123) always has the trivial solution $\Lambda = 0$, but begins to have some non-trivial solutions in the region of $\frac{J}{k_B T} > \text{arctanh}\left(\frac{1}{3}\right)$ by expanding the right-hand side of Eq. (10.123) around $\Lambda = 0$ and keeping the first-order term of Λ . In Fig. 10.7, the blue curves correspond to global minimum states that are stable states and the red lines correspond to the local maximum state that are unstable states for each value of $\frac{J}{k_B T}$ in the Bethe free energy $f_{\text{Bethe}}(m, c)$ of Eq. (10.125) for the case of $h = 0$. The Bethe free energy landscapes $f_{\text{Bethe}}(m, c)$ of Eq. (10.125) in the case of $h = 0$ for several values of $\frac{J}{k_B T}$ are shown in Figs. 10.8 and 10.9. It is known that Eq. (10.123) always has the trivial solution $\Lambda = 0$, but begins to have some non-trivial solutions in the region of $\frac{J}{k_B T} > \text{arctanh}\left(\frac{1}{3}\right)$ by expanding the right-hand side of Eq. (10.123) around $\Lambda = 0$ and keeping the first-order term of Λ . In Fig. 10.7, the blue curves correspond to global minimum states that are stable states and the red lines

**Loopy Belief Propagation for Ising Model
on Regular Graph (V, E) with Degree 4**

$$P(s | \frac{J}{k_B T}, \frac{h}{k_B T}) \propto \left(\prod_{(i,j) \in E} \exp\left(\frac{J}{k_B T} s_i s_j\right) \right) \left(\sum_{i \in V} \exp\left(\frac{h}{k_B T} s_i\right) \right) \quad s_i \in \Omega = \{+1, -1\}$$



— **Global Minimum State (Stable)**
— **Local Maximum State (Unstable)**

} for Bethe Free Energy

I. $h \rightarrow +0$: $\lim_{\frac{J}{k_B T} \rightarrow +\infty} (P_i(+1), P_i(-1)) \rightarrow (1, 0)$ III. $(P_i(+1), P_i(+1)) = \left(\frac{1}{2}, \frac{1}{2}\right) (\forall \frac{J}{k_B T} > 0)$

II. $h \rightarrow -0$: $\lim_{\frac{J}{k_B T} \rightarrow +\infty} (P_i(+1), P_i(-1)) \rightarrow (0, 1)$

$$P_i(s_i) \equiv \lim_{|V| \rightarrow +\infty} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{i-1} \in \Omega} \sum_{s_{i+1} \in \Omega} \sum_{s_{i+2} \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T}\right) (\forall s_i \in \Omega)$$

Fig. 10.7 Internal energy u from Eq. (10.113) and magnetization m from Eq. (10.118) in loopy belief propagation (Bethe approximation) for the Ising model in Eqs. (10.48), (10.50), and (10.110) on the regular graph (V, E) of degree 4

correspond to the local maximum state that are unstable states for each value of $\frac{J}{k_B T}$ in the Bethe free energy $f_{\text{Bethe}}(m, c)$ in Eq. (10.125) for the case of $h = 0$. The Bethe free energy landscapes $f_{\text{Bethe}}(m, c)$ in Eq. (10.125) in the case of $h = 0$ for several values of $\frac{J}{k_B T}$ are shown in Figs. 10.8 and 10.9.

Now we consider the $|\Omega|$ -state Potts model [33] given by

$$\begin{aligned}
 & P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \dots, \frac{h_{|\Omega|-1}}{k_B T}\right) \\
 &= \frac{1}{Z} \left(\prod_{\{i,j\} \in E} \exp\left(\frac{J}{k_B T} \delta_{s_i s_j}\right) \right) \left(\prod_{i \in V} \prod_{n \in \Omega} \exp\left(\frac{h_n}{k_B T} \delta_{s_i, n}\right) \right), \quad (10.132)
 \end{aligned}$$

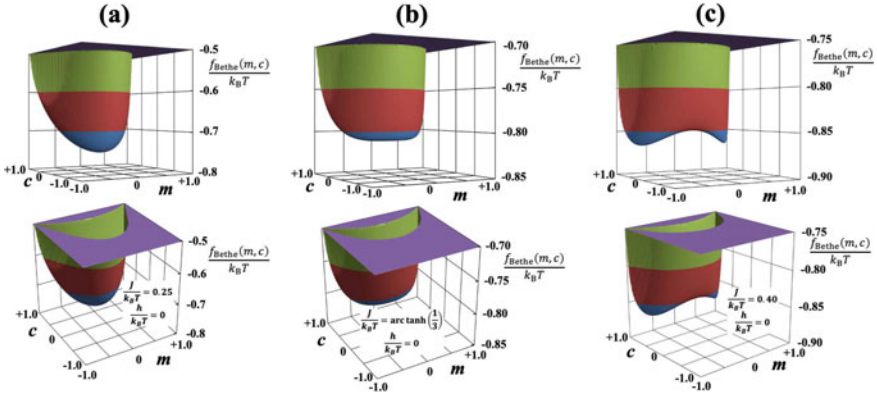


Fig. 10.8 Bethe free energy for $f_{\text{Bethe}}(m, c)$ for the Ising model in Eqs. (10.48), (10.50), and (10.110) on the regular graph (V, E) of degree 4. **a** $\frac{J}{k_B T} = 0.25$, $\frac{h}{k_B T} = 0$. **b** $\frac{J}{k_B T} = \text{arctanh}(\frac{1}{3})$, $\frac{h}{k_B T} = 0$. **c** $\frac{J}{k_B T} = 0.4$, $\frac{h}{k_B T} = 0$

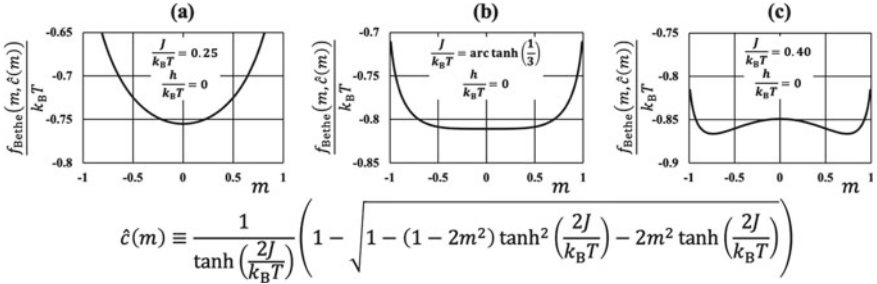


Fig. 10.9 Bethe free energy Extremum $f_{\text{Bethe}}(m, \hat{c})$ for the Ising model of Eqs. (10.48), and (10.50) with Eqs. (10.110) on the regular graph (V, E) with degree 4. **a** $\frac{J}{k_B T} = 0.25$, $\frac{h}{k_B T} = 0$. **b** $\frac{J}{k_B T} = \text{arctanh}(\frac{1}{3})$, $\frac{h}{k_B T} = 0$. **c** $\frac{J}{k_B T} = 0.4$, $\frac{h}{k_B T} = 0$

$$Z \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \left(\prod_{\{i, j\} \in E} \exp\left(\frac{J}{k_B T} \delta_{s_i s_j}\right) \right) \left(\prod_{i \in V} \prod_{n \in \Omega} \exp\left(\frac{h_n}{k_B T} \delta_{s_i, n}\right) \right), \quad (10.133)$$

where $\Omega = \{0, 1, 2, \dots, |\Omega| - 1\}$. By similar arguments to those for Eqs. (10.72) and (10.102), the marginal probabilities $\hat{R}_i(s_i)$ and $\hat{R}_{ij}(s_i, s_j)$ can be expressed as orthonormal expansions as follows:

$$\hat{R}_i(s_i) = \left(\frac{1}{|\Omega|} \right) + \sum_{k \in \Omega \setminus \{0\}} m_i^{(k)} \Phi_k(s_i), \quad (10.134)$$

$$\begin{aligned} \widehat{R}_{ij}(s_i, s_j) &= \left(\frac{1}{|\Omega|}\right)^2 + \left(\frac{1}{|\Omega|}\right) \sum_{k \in \Omega \setminus \{0\}} m_i^{(k)} \Phi_k(s_i) \\ &+ \left(\frac{1}{|\Omega|}\right) \sum_{l \in \Omega \setminus \{0\}} m_j^{(l)} \Phi_l(s_j) + \sum_{k \in \Omega \setminus \{0\}} \sum_{l \in \Omega \setminus \{0\}} c_{ij}^{(k,l)} \Phi_k(s_i) \Phi_l(s_j), \end{aligned} \quad (10.135)$$

where $\{\Phi_k(s_i) | s_i \in \Omega, k \in \Omega\}$ is the set of orthonormal polynomials satisfying the following relationships:

$$\Phi_0(s_i) \equiv \left(\frac{1}{|\Omega|}\right), \quad (10.136)$$

$$\sum_{s_i \in \Omega} \Phi_k(s_i) \Phi_l(s_i) = \delta_{k,l} \quad (k \in \Omega, l \in \Omega). \quad (10.137)$$

Because it is valid that

$$\sum_{s_j \in \Omega} \sum_{s_i \in \Omega} \Phi_k(s_i) \Phi_l(s_j) \delta_{s_i, s_j} = \sum_{s_i \in \Omega} \Phi_k(s_i) \Phi_l(s_i) = \delta_{k,l} \quad (k \in \Omega, l \in \Omega), \quad (10.138)$$

we have the following orthonormal expansion of δ_{s_i, s_j} :

$$\begin{aligned} \delta_{s_i, s_j} &= \sum_{k \in \Omega} \sum_{l \in \Omega} \left(\sum_{s'_i \in \Omega} \sum_{s'_j \in \Omega} \Phi_k(s'_i) \Phi_l(s'_j) \delta_{s'_i, s'_j} \right) \Phi_k(s_i) \Phi_l(s_j) \\ &= \sum_{k \in \Omega} \Phi_k(s_i) \Phi_k(s_j) \quad (s_i \in \Omega, s_j \in \Omega). \end{aligned} \quad (10.139)$$

By using Eqs. (10.134) and (10.135) and the orthonormal expansion of the two-body interaction part of the Potts model, the Bethe free energy functional for the Potts model in Eqs. (10.132) and (10.133) can be reduced to

$$\begin{aligned} &\mathcal{F}_{\text{Bethe}} \left[\left\{ \widehat{R}_i \mid i \in V \right\}, \left\{ \widehat{R}_{ij} \mid \{i, j\} \in E \right\} \right] \\ &= F_{\text{Bethe}} \left(\left\{ m_i^{(k)} \mid i \in V, k \in \Omega \setminus \{0\} \right\}, \left\{ c_{\{i,j\}}^{(k,l)} \mid \{i, j\} \in E, k, l \in \Omega \setminus \{0\} \right\} \right), \end{aligned} \quad (10.140)$$

where

$$\begin{aligned}
& F_{\text{Bethe}}\left(\{m_i^{(k)} \mid i \in V, k \in \Omega \setminus \{0\}\}, \{c_{\{i,j\}}^{(k,l)} \mid \{i,j\} \in E, k, l \in \Omega \setminus \{0\}\}\right) \\
& \equiv -J \sum_{i \in V} |\partial i| \left(\frac{1}{|\Omega|}\right)^2 - J \sum_{i \in V} |\partial i| \frac{1}{|\Omega|} \sum_{k \in \Omega \setminus \{0\}} m_i^{(k)} - J \sum_{\{i,j\} \in E} \sum_{k \in \Omega \setminus \{0\}} c_{i,j}^{(k,k)} \\
& + k_B T \sum_{i \in V} (1 - |\partial i|) \widehat{R}_i(s_i) \ln(\widehat{R}_i(s_i)) \\
& + k_B T \sum_{\{i,j\} \in E} \widehat{R}_{ij}(s_i, s_j) \ln(\widehat{R}_{ij}(s_i, s_j)). \tag{10.141}
\end{aligned}$$

For the case of spatial uniformity, $m_i^{(k)}$ and $c_{ij}^{(k,l)}$ are independent of i and $\{i, j\}$ and can be represented by $m^{(k)}$ and $c^{(k,l)}$, respectively, in the Bethe free energy in Eq. (10.141). For the three-state and four-state Potts model, the Bethe free energy in Eq. (10.141) can be represented by

$$F_{\text{Bethe}}\left(\left(\begin{matrix} m^{(1)} \\ m^{(2)} \end{matrix}\right), \left(\begin{matrix} c^{(1,1)} & c^{(1,2)} \\ c^{(2,1)} & c^{(2,2)} \end{matrix}\right)\right), \tag{10.142}$$

and

$$F_{\text{Bethe}}\left(\left(\begin{matrix} m^{(1)} \\ m^{(2)} \\ m^{(3)} \end{matrix}\right), \left(\begin{matrix} c^{(1,1)} & c^{(1,2)} & c^{(1,3)} \\ c^{(2,1)} & c^{(2,2)} & c^{(2,3)} \\ c^{(3,1)} & c^{(3,2)} & c^{(3,3)} \end{matrix}\right)\right), \tag{10.143}$$

respectively. Figures 10.10 and 10.11 show the internal energy with no external fields

$$u = \lim_{|V| \rightarrow +\infty} \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_s (-\delta_{s_i, s_j}) P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T} = 0, \frac{h_1}{k_B T} = 0, \dots, \frac{h_{|\Omega|-1}}{k_B T} = 0\right), \tag{10.144}$$

in loopy belief propagation (Bethe approximation) on the regular graph (V, E) of degree 4. We now consider also the moments $m^{(2)}$ and $m^{(1)}$ as **order parameters** for the three-state and four-state Potts model, respectively, for the following cases:

$$\left\{ \begin{array}{l} \text{(I)} \quad \lim_{h_0 \rightarrow +0} m^{(2)}, h_1 = h_2 = 0, \\ \text{(II)} \quad \lim_{h_1 \rightarrow +0} m^{(2)}, h_0 = h_2 = 0, \\ \text{(III)} \quad \lim_{h_2 \rightarrow +0} m^{(2)}, h_0 = h_1 = 0, \\ \text{(IV)} \quad m^{(2)} \text{ under } h_0 = h_1 = h_2 = 0 \text{ and } \mu(0) = \mu(1) = \mu(2) = \frac{1}{3}, \end{array} \right. \tag{10.145}$$

for the three-state Potts model, and

$$\left\{ \begin{array}{l} \text{(I)} \quad \lim_{h_0 \rightarrow +0} m^{(2)}, h_1 = h_2 = h_3 = 0, \\ \text{(II)} \quad \lim_{h_1 \rightarrow +0} m^{(2)}, h_0 = h_2 = h_3 = 0, \\ \text{(III)} \quad \lim_{h_2 \rightarrow +0} m^{(2)}, h_0 = h_1 = h_3 = 0, \\ \text{(IV)} \quad \lim_{h_3 \rightarrow +0} m^{(2)}, h_0 = h_1 = h_2 = 0, \\ \text{(V)} \quad m^{(2)} \text{ under } h_0 = h_1 = h_2 = 0 \text{ and } \mu(0) = \mu(1) = \mu(2) = \mu(3) = \frac{1}{4}, \end{array} \right. \quad (10.146)$$

for the four-state Potts model. These are also shown in Figs. 10.10 and 10.11. In Figs. 10.10 and 10.11, blue, green, and red lines show the global minimum states, local minimum states, and local maximum states, respectively, of the Bethe free energies which are given by Eq. (10.142) for the three-state Potts model and by Eq. (10.143) for the four-state Potts model. In the global minimum states, there exist discontinuous points in $m^{(2)}$ and $m^{(1)}$ as well as u . Although the first derivative Ju of the free energy with respect to $\frac{1}{k_B T}$ is always continuous, the second derivative diverges or has discontinuity in the Ising model as shown in Figs. 10.4, 10.5, and 10.7. This kind of singularity is referred to as a **second-order phase transition** in statistical mechanics. However, the first derivative Ju of the free energy with respect to $\frac{1}{k_B T}$ has a discontinuity as shown in Figs. 10.10 and 10.11. This singularity is referred to as a **first-order phase transition** in statistical mechanics. Figures 10.12 and 10.13 show the Bethe free energy landscapes

$$f_{\text{Bethe}}(m^{(1)}, m^{(2)}) \equiv \frac{1}{|V|} \underset{\begin{pmatrix} c^{(1,1)} & c^{(1,2)} \\ c^{(2,1)} & c^{(2,2)} \end{pmatrix}}{\text{extremum}} F_{\text{Bethe}} \left(\begin{pmatrix} m^{(1)} \\ m^{(2)} \end{pmatrix}, \begin{pmatrix} c^{(1,1)} & c^{(1,2)} \\ c^{(2,1)} & c^{(2,2)} \end{pmatrix} \right), \quad (10.147)$$

for the three-state Potts model and

$$f_{\text{Bethe}}(m^{(1)}, m^{(3)}) \equiv \frac{1}{|V|} \underset{\begin{pmatrix} c^{(1,1)} & c^{(1,2)} & c^{(1,3)} \\ c^{(2,1)} & c^{(2,2)} & c^{(2,3)} \\ c^{(3,1)} & c^{(3,2)} & c^{(3,3)} \end{pmatrix}}{\text{extremum}} F_{\text{Bethe}} \left(\begin{pmatrix} m^{(1)} \\ m^{(2)} \\ m^{(3)} \end{pmatrix}, \begin{pmatrix} c^{(1,1)} & c^{(1,2)} & c^{(1,3)} \\ c^{(2,1)} & c^{(2,2)} & c^{(2,3)} \\ c^{(3,1)} & c^{(3,2)} & c^{(3,3)} \end{pmatrix} \right), \quad (10.148)$$

for the four-state Potts model, respectively.

10.3.4 Ising Model on a Complete Graph

This section considers a complete graph (V, E) for which the energy function $H(s)$ is defined by

Loopy Belief Propagation for Three State Potts Model on Regular Graph (V,E) with Degree 4

$$s_i \in \Omega = \{0, 1, 2\}$$

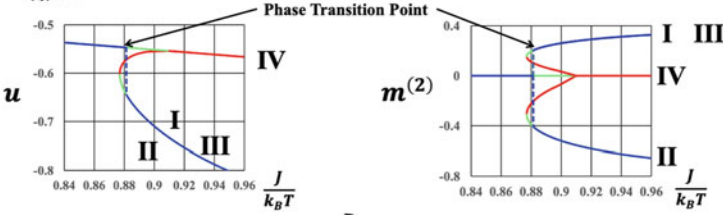
$$P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}\right) \propto \left(\prod_{(i,j) \in E} \exp\left(\frac{J}{k_B T} \delta_{s_i, s_j}\right) \right) \left(\prod_{i \in V} \exp\left(\sum_{n \in \Omega} \frac{h_n}{k_B T} \delta_{s_i, n}\right) \right)$$

$$m^{(k)} \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_s \Phi_k(s_i) P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}\right)$$

$$c^{(k,l)} \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|E|} \sum_{(l,j) \in E} \sum_s \Phi_k(s_l) \Phi_l(s_j) P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}\right)$$

$$(\Phi_0(0), \Phi_0(1), \Phi_0(2)) \equiv \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) \quad (\Phi_1(0), \Phi_1(1), \Phi_1(2)) \equiv \left(\frac{1}{2}, 0, -\frac{1}{2}\right) \quad (\Phi_2(0), \Phi_2(1), \Phi_2(2)) \equiv \left(\frac{1}{6}, -\frac{2}{3}, \frac{1}{6}\right)$$

$$u \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|E|} \sum_{(l,j) \in E} \sum_s (-\delta_{s_l, s_j}) P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}\right) \quad m^{(2)} \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_s \Phi_2(s_i) P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}\right)$$



— Global Minimum State (Stable)
— Local Minimum State (Metastable)
— Local Maximum State (Unstable)

$$F_{\text{Bethe}} \left(\begin{pmatrix} m^{(1)} \\ m^{(2)} \end{pmatrix}, \begin{pmatrix} c^{(1,1)} & c^{(1,2)} \\ c^{(2,1)} & c^{(2,2)} \end{pmatrix} \right)$$

I. $h_0 \rightarrow +0, h_1 = h_2 = 0$: $\lim_{\frac{J}{k_B T} \rightarrow +\infty} (P_i(0), P_i(1), P_i(2)) = (1, 0, 0)$

II. $h_1 \rightarrow +0, h_0 = h_2 = 0$: $\lim_{\frac{J}{k_B T} \rightarrow +\infty} (P_i(0), P_i(1), P_i(2)) = (0, 1, 0)$

III. $h_2 \rightarrow +0, h_0 = h_1 = 0$: $\lim_{\frac{J}{k_B T} \rightarrow +\infty} (P_i(0), P_i(1), P_i(2)) = (0, 0, 1)$

IV. $(P_i(0), P_i(1), P_i(2)) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \left(\forall \frac{J}{k_B T} > 0\right)$

$$P_i(s_i) \equiv \lim_{|V| \rightarrow +\infty} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{i-1} \in \Omega} \sum_{s_{i+1} \in \Omega} \sum_{s_{i+2} \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} P\left(s \mid \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}\right) \quad (\forall s_i \in \Omega)$$

Fig. 10.10 Internal energy u in Eq. (10.144) and (I) $\lim_{h_0 \rightarrow +0} m^{(2)}$ for $h_1 = h_2 = 0$, (II) $\lim_{h_1 \rightarrow +0} m^{(2)}$ for $h_0 = h_2 = 0$, (III) $\lim_{h_2 \rightarrow +0} m^{(2)}$ for $h_0 = h_1 = 0$, (IV) $m^{(2)}$ under $h_0 = h_1 = h_2 = 0$ and $\mu(0) = \mu(1) = \mu(2) = \frac{1}{3}$ such that $P_i(0) = P_i(1) = P_i(2)$ in loopy belief propagation (Bethe approximation) for the three-state Potts model in Eqs. (10.132) and (10.133) on the regular graph (V, E) of degree 4

$$H(s) = H(s_1, s_2, \dots, s_{|V|}) \equiv -\frac{J}{|V|} \sum_{\{i,j\} \in E} s_i s_j - h \sum_{i \in V} s_i \quad (J > 0), \quad (10.149)$$

instead of Eq. (10.110) in Eqs. (10.109) and (10.111). Note that the interaction between every pair of connected nodes is set to $\frac{J}{|V|}$ to guarantee the existence of the thermodynamic limit in $|V| \rightarrow +\infty$ for the complete graph in the sense of Ruelle in Ref. [38].

Loopy Belief Propagation for Four State Potts Model

on Regular Graph (V, E) with Degree 4 $s_i \in \Omega = \{0, 1, 2, 3\}$

$$P\left(s \left| \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}, \frac{h_3}{k_B T} \right. \right) \propto \left(\prod_{(i,j) \in E} \exp\left(\frac{J}{k_B T} \delta_{s_i, s_j}\right) \right) \left(\prod_{i \in V} \exp\left(\sum_{n \in \Omega} \frac{h_n}{k_B T} \delta_{s_i, n}\right) \right)$$

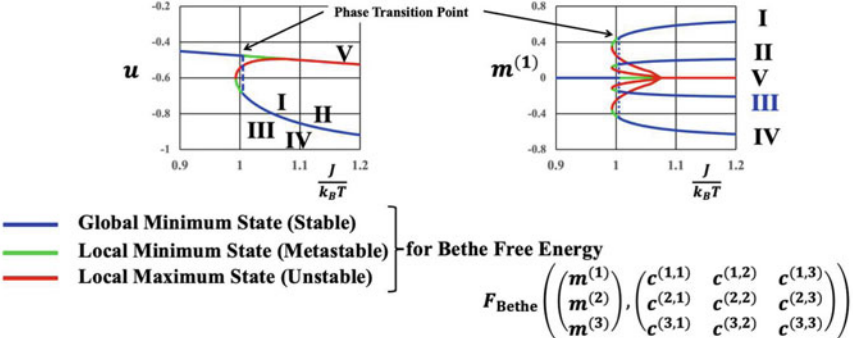
$$m^{(k)} \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_s \Phi_k(s_i) P\left(s \left| \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}, \frac{h_3}{k_B T} \right. \right) \quad c^{(k,l)} \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|E|} \sum_{(j,l) \in E} \sum_s \Phi_k(s_i) \Phi_l(s_j) P\left(s \left| \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}, \frac{h_3}{k_B T} \right. \right)$$

$$(\Phi_0(0), \Phi_0(1), \Phi_0(2), \Phi_0(3)) \equiv \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$$

$$(\Phi_1(0), \Phi_1(1), \Phi_1(2), \Phi_1(3)) \equiv \left(\frac{3}{2\sqrt{5}}, \frac{1}{2\sqrt{5}}, -\frac{1}{2\sqrt{5}}, -\frac{3}{2\sqrt{5}}\right) \quad (\Phi_2(0), \Phi_2(1), \Phi_2(2), \Phi_2(3)) \equiv \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)$$

$$(\Phi_3(0), \Phi_3(1), \Phi_3(2), \Phi_3(3)) \equiv \left(\frac{1}{2\sqrt{5}}, -\frac{3}{2\sqrt{5}}, \frac{3}{2\sqrt{5}}, -\frac{1}{2\sqrt{5}}\right)$$

$$u \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|E|} \sum_{(j,l) \in E} \sum_s (-\delta_{s_i, s_j}) P\left(s \left| \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}, \frac{h_3}{k_B T} \right. \right) \quad m^{(1)} \equiv \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_s \Phi_1(s_i) P\left(s \left| \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}, \frac{h_3}{k_B T} \right. \right)$$



- I. $h_0 \rightarrow +0, h_1 = h_2 = h_3 = 0$: $\lim_{J/k_B T \rightarrow +\infty} (P_i(0), P_i(1), P_i(2), P_i(3)) = (1, 0, 0, 0)$
- II. $h_1 \rightarrow +0, h_0 = h_2 = h_3 = 0$: $\lim_{J/k_B T \rightarrow +\infty} (P_i(0), P_i(1), P_i(2), P_i(3)) = (0, 1, 0, 0)$, $V. (P_i(0), P_i(1), P_i(2), P_i(3)) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \left(\forall \frac{J}{k_B T} > 0\right)$
- III. $h_2 \rightarrow +0, h_0 = h_1 = h_3 = 0$: $\lim_{J/k_B T \rightarrow +\infty} (P_i(0), P_i(1), P_i(2), P_i(3)) = (0, 0, 1, 0)$
- IV. $h_3 \rightarrow +0, h_0 = h_1 = h_2 = 0$: $\lim_{J/k_B T \rightarrow +\infty} (P_i(0), P_i(1), P_i(2), P_i(3)) = (0, 0, 0, 1)$

$$P_i(s_i) \equiv \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \dots \sum_{s_{i-1} \in \Omega} \sum_{s_{i+1} \in \Omega} \sum_{s_{i+2} \in \Omega} \dots \sum_{s_{|V|} \in \Omega} P\left(s \left| \frac{J}{k_B T}, \frac{h_0}{k_B T}, \frac{h_1}{k_B T}, \frac{h_2}{k_B T}, \frac{h_3}{k_B T} \right. \right) (\forall i \in V, \forall s_i \in \Omega)$$

Fig. 10.11 Internal energy u in Eq. (10.144) and order parameter $m^{(1)}$ in loopy belief propagation (Bethe approximation) for the four-state Potts model in Eqs. (10.132) and (10.133) on the regular graph (V, E) of degree 4

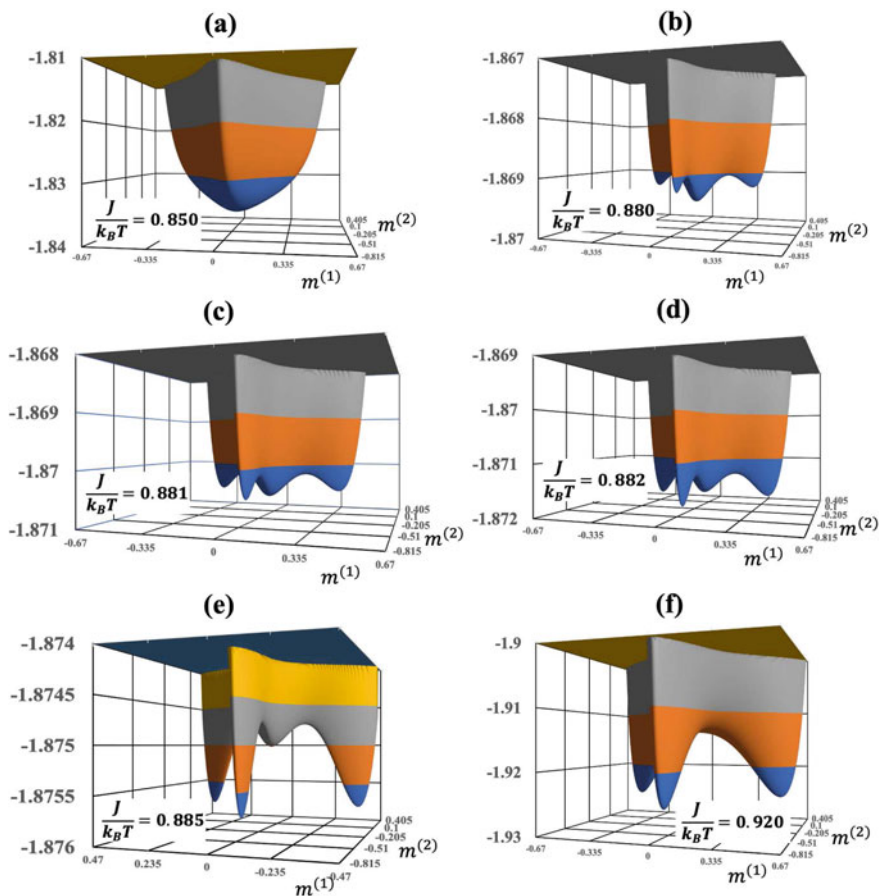


Fig. 10.12 Bethe free energy for $f_{\text{Bethe}}(m^{(1)}, m^{(2)})$ for the three-state Potts model in Eqs. (10.132) and (10.133) on the regular graph (V, E) of degree 4. **a** $J/k_B T = 0.850$. **b** $J/k_B T = 0.880$. **c** $J/k_B T = 0.881$. **d** $J/k_B T = 0.882$. **e** $J/k_B T = 0.885$. **f** $J/k_B T = 0.920$

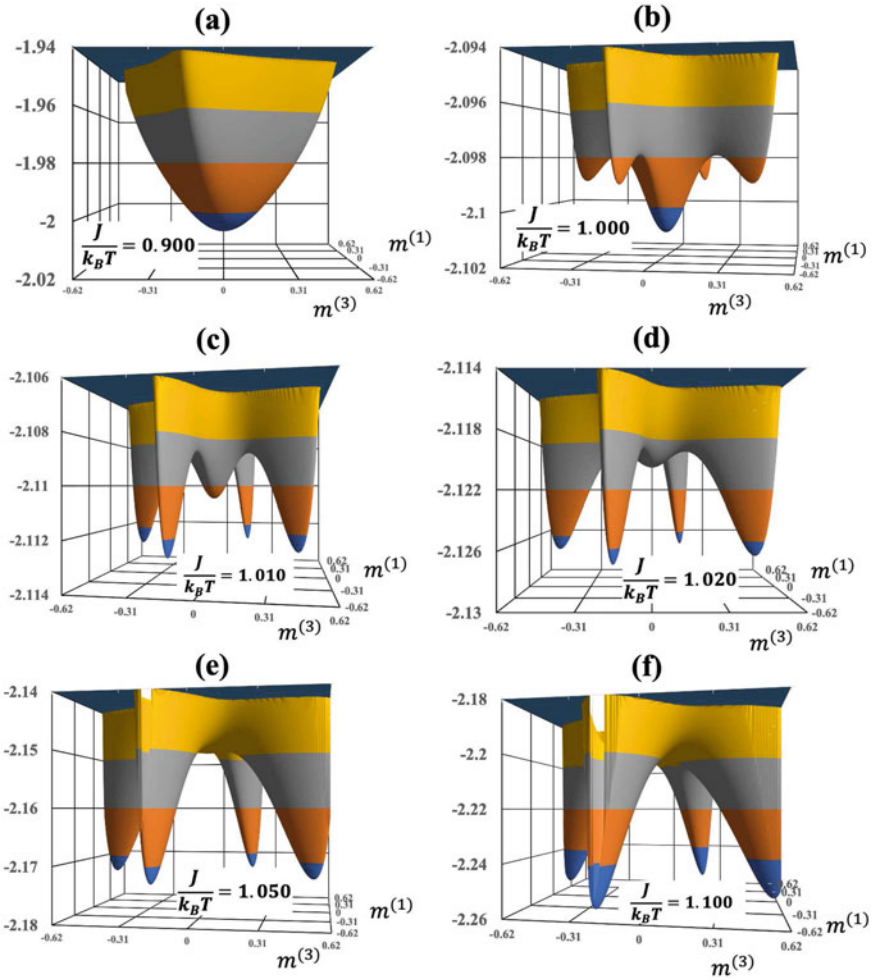


Fig. 10.13 Bethe free energy for $f_{\text{Bethe}}(m^{(1)}, m^{(2)})$ for the four-state Potts model in Eqs. (10.132) and (10.133) on the regular graph (V, E) of degree 4. **a** $\frac{J}{k_B T} = 0.900$. **b** $\frac{J}{k_B T} = 1.000$. **c** $\frac{J}{k_B T} = 1.010$. **d** $\frac{J}{k_B T} = 1.020$. **e** $\frac{J}{k_B T} = 1.050$. **f** $\frac{J}{k_B T} = 1.100$

The free energy in Eq. (10.55) is expressed as follows:

$$\begin{aligned}
& -k_B T \ln \left(Z \left(\frac{J}{k_B T}, \frac{h}{k_B T} \right) \right) \\
&= -k_B T \ln \left(\sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \exp \left(\frac{h}{k_B T} \sum_{i \in V} \tau_i + \frac{J}{|V| k_B T} \sum_{\{i, j\} \in E} \tau_i \tau_j \right) \right) \\
&= \frac{J}{2} - k_B T \ln \left(\sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \exp \left(\frac{h}{k_B T} \sum_{i \in V} \tau_i \right) \exp \left(\frac{1}{2} \left(\sqrt{\frac{J}{|V| k_B T}} \sum_{i \in V} \tau_i \right)^2 \right) \right). \tag{10.150}
\end{aligned}$$

By using the Gauss integral formula

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \left(-\frac{1}{2} x^2 + ax \right) dx = \exp \left(\frac{1}{2} a^2 \right), \tag{10.151}$$

the expression for the free energy is rewritten as

$$\begin{aligned}
& -k_B T \ln \left(Z \left(\frac{J}{k_B T}, \frac{h}{k_B T} \right) \right) \\
&= \frac{1}{2} J - \ln \left(\sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} x^2} \right. \\
&\quad \left. \times \exp \left(\sum_{i \in V} \left(\sqrt{\frac{J}{|V| k_B T}} x + \frac{h}{k_B T} \right) \tau_i \right) dx \right) \\
&= \frac{1}{2} J - \ln \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} x^2} \prod_{i \in V} \left(\sum_{\tau_i \in \Omega} \exp \left(\left(\sqrt{\frac{J}{|V| k_B T}} x + \frac{h}{k_B T} \right) \tau_i \right) \right) dx \right) \\
&= \frac{1}{2} J - \ln \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} x^2} \prod_{i \in V} \left(2 \cosh \left(\sqrt{\frac{J}{|V| k_B T}} x + \frac{h}{k_B T} \right) \right) dx \right). \tag{10.152}
\end{aligned}$$

Note that the procedure in which a new continuous variable x is introduced in Eq. (10.152) is referred to as a **Hubbard-Stratonovich transformation** [13]. Moreover,

by replacing the variable x by $y = x \sqrt{|V|} \left(\frac{k_B T}{J} \right)$, the free energy can be written as

$$-k_B T \ln \left(Z \left(\frac{J}{k_B T}, \frac{h}{k_B T} \right) \right) = \frac{1}{2} J - \ln \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(|V| \psi(y)) dy \right), \tag{10.153}$$

where

$$\psi(y) \equiv -\frac{1}{2} \left(\frac{J}{k_B T} \right) y^2 + \frac{1}{|V|} \sum_{i \in V} \ln \left(2 \cosh \left(\frac{1}{k_B T} (Jy + h) \right) \right). \tag{10.154}$$

We now consider the magnetization

$$m\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) = \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \sum_{i \in V} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} s_i P\left(s \mid \frac{J}{k_B T}, \frac{h}{k_B T}\right) \quad \text{for}$$

Eqs. (10.109) and (10.111) with Eq. (10.149) as follows:

$$\begin{aligned} m\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) &= \lim_{|V| \rightarrow +\infty} \frac{1}{|V|} \frac{\partial}{\partial \left(\frac{h}{k_B T}\right)} \left(-k_B T \ln\left(Z\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right)\right)\right) \\ &= \lim_{|V| \rightarrow +\infty} \frac{\int_{-\infty}^{+\infty} \exp\left(|V| \left(\psi(y) + \frac{1}{|V|} \ln\left(\tanh\left(\frac{1}{k_B T}(Jy + h)\right)\right)\right)\right) dy}{\int_{-\infty}^{+\infty} \exp(|V| \psi(y)) dy}. \end{aligned} \quad (10.155)$$

Because it is valid that

$$\begin{aligned} \lim_{|V| \rightarrow +\infty} \frac{\partial}{\partial y} \psi(y) &= \lim_{|V| \rightarrow +\infty} \frac{\partial}{\partial y} \left(\psi(y) + \frac{1}{|V|} \ln\left(\tanh\left(\frac{1}{k_B T}(Jy + h)\right)\right)\right) \\ &= -\frac{J}{k_B T} \left(y - \tanh\left(\frac{1}{k_B T}(Jy + h)\right)\right), \end{aligned} \quad (10.156)$$

we obtain the magnetization as

$$\begin{aligned} m\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) &= \lim_{|V| \rightarrow +\infty} \frac{\exp\left(|V| \left(\psi(y_{\max}) + \frac{1}{|V|} \ln\left(\tanh\left(\frac{1}{k_B T}(Jy_{\max} + h)\right)\right)\right)\right)}{\exp(|V| \psi(y_{\max}))} \\ &= \tanh\left(\frac{1}{k_B T}(Jy_{\max} + h)\right), \end{aligned} \quad (10.157)$$

where

$$y_{\max} = \tanh\left(\frac{1}{k_B T}(Jy_{\max} + h)\right) \quad (10.158)$$

by using a **saddle point method** [37]. Equations (10.157) and (10.158) reduce to the following mean-field equation for $m\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right)$:

$$m\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) = \tanh\left(\frac{1}{k_B T} \left(Jm\left(\frac{J}{k_B T}, \frac{h}{k_B T}\right) + h\right)\right). \quad (10.159)$$

This means that it is possible to treat the Ising model on the complete graph in the thermodynamic limit analytically using the mean-field method.

By combining the replica method with the Hubbard-Stratonovich transformation and the saddle point method, it is possible to treat the random average in Eq. (10.25) for the Ising model with non-uniform external fields on the complete graph analytically [13, 76]. In statistical mechanics, this kind of approach has been developed as the **spin glass theory** [77–80]. Such computational techniques that use the replica

method for Ising models with spatially non-uniform interactions and external fields on the complete graph have been used to estimate statistical performance analysis for many probabilistic information processing systems [13, 15–17].

Next, we consider the belief propagation method for the Ising model on the complete graph in Eq. (10.149) with Eqs. (10.109) and (10.111). For an infinitesimal small $|V|^{-1}$, the message passing rule in Eq. (10.95) can be expanded to

$$\begin{aligned}
 \mu_{j \rightarrow i}(s_i) &= \frac{Z_i}{Z_{(i,j)}} \sum_{\tau_j \in \Omega} \left(\prod_{l \in \partial j \setminus \{i\}} \mu_{l \rightarrow j}(\tau_j) \right) \exp\left(\frac{1}{k_B T} \left(\frac{J}{|V|} s_i \tau_j + h \tau_j \right)\right) \\
 &= \frac{Z_i}{Z_{(i,j)}} \sum_{\tau_j \in \Omega} \left(\prod_{l \in \partial j \setminus \{i\}} \mu_{l \rightarrow j}(\tau_j) \right) \exp\left(\frac{h}{k_B T} \tau_j\right) \left(1 + \frac{1}{|V|} \frac{J}{k_B T} s_i \tau_j + \mathcal{O}(|V|^{-2})\right) \\
 &= \frac{Z_i Z_j}{Z_{(i,j)}} \sum_{\tau_j \in \Omega} \widehat{R}_j(\tau_j) \left(1 + \frac{1}{|V|} \frac{J}{k_B T} s_i \tau_j + \mathcal{O}(|V|^{-2})\right) \\
 &= \frac{Z_i Z_j}{Z_{(i,j)}} \left(1 + \frac{1}{k_B T} \left(\frac{J}{|V|} \sum_{\tau_j \in \Omega} \tau_j \widehat{R}_j(\tau_j)\right) s_i + \mathcal{O}(|V|^{-2})\right) \\
 &= \frac{Z_i Z_j}{Z_{(i,j)}} \left(\exp\left(\frac{1}{k_B T} \left(\frac{J}{|V|} \sum_{\tau_j \in \Omega} \tau_j \widehat{R}_j(\tau_j)\right) s_i\right) + \mathcal{O}(|V|^{-2})\right). \tag{10.160}
 \end{aligned}$$

By substituting Eq. (10.160) into Eq. (10.91), the marginal probabilities can be expressed as follows:

$$\begin{aligned}
 \widehat{R}_i(s_i) &= \frac{\exp\left(\frac{1}{k_B T} \left(\frac{J}{|V|} \sum_{j \in V \setminus \{i\}} \sum_{\tau_j \in \Omega} \tau_j \widehat{R}_j(\tau_j) + h\right) s_i\right)}{\sum_{\tau_i \in \Omega} \exp\left(\frac{1}{k_B T} \left(\frac{J}{|V|} \sum_{j \in V \setminus \{i\}} \sum_{\tau_j \in \Omega} \tau_j \widehat{R}_j(\tau_j) + h\right) \tau_i\right)} \\
 &\quad + \mathcal{O}(|V|^{-1}) \quad (|V| \rightarrow +\infty, s_i \in \Omega, i \in V), \tag{10.161}
 \end{aligned}$$

$$\widehat{R}_{ij}(s_i, s_j) = \widehat{R}_i(s_i) \widehat{R}_j(s_j) + \mathcal{O}(|V|^{-1}) \quad (|V| \rightarrow +\infty, s_i \in \Omega, s_j \in \Omega, \{i, j\} \in E). \tag{10.162}$$

Equation (10.161) can be regarded as a system of simultaneous deterministic equations for $\{\widehat{R}_i(s_i) | s_i \in \Omega, i \in V\}$ and is equivalent to the mean-field equation in Eq. (10.68) for Eq. (10.149) with Eqs. (10.109) and (10.111).

10.3.5 Probabilistic Segmentation by Potts Prior and Loopy Belief Propagation

In Sect. 10.2.3, we gave the fundamental framework of probabilistic segmentation based on the Potts prior, and reduced the framework of the EM procedure for estimating hyperparameters to the extremum conditions of the \mathcal{Q} -function as shown in Eqs. (10.41), (10.42), and (10.43) with Eqs. (10.44), (10.45), and (10.46). These frameworks can be realized by combining them with the loopy belief propagation in Sect. 10.3.2 to give the following practical procedures [70]:

Probabilistic segmentation algorithm (Input : D , Output : $\hat{\alpha}(D)$, $\hat{u}(D)$, $\hat{a}(D)$, $\hat{C}(D)$, $\hat{s}(D)$)

Step 1: Input the data vector \mathbf{d} and set the initial values of hyperparameters $\hat{\alpha}(D)$, $\hat{a}(D)$, $\hat{C}(D)$ and messages in the loopy belief propagation $\{\hat{\mu}_{j \rightarrow i}(s_i, D) | i \in V, j \in \partial i, s_i \in \Omega\}$ for the posterior probability distribution. We set $t \leftarrow 0$ as the number of iterations of the EM procedure.

Step 2 (E-step) Set $t \leftarrow t + 1$ and update $\hat{u}(D)$, $\hat{a}(D)$, $\hat{C}(D)$, $\{\hat{\mu}_{j \rightarrow i}(s_i, D) | s_i \in \Omega, i \in V, j \in \partial i\}$ using the following procedures:

$$\mu_{j \rightarrow i}(s_i) \leftarrow \frac{\sum_{\tau_j \in \Omega} \exp(2\hat{\alpha}(D)\delta_{s_i, \tau_j})g(\mathbf{d}_j | \tau_j, \hat{a}(\tau_j, D), \hat{C}(\tau_j, D)) \prod_{k \in \partial j \setminus \{i\}} \hat{\mu}_{k \rightarrow j}(\tau_j, D)}{\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \exp(2\hat{\alpha}(D)\delta_{\tau_i, \tau_j})g(\mathbf{d}_j | \tau_j, \hat{a}(\tau_j, D), \hat{C}(\tau_j, D)) \prod_{k \in \partial j \setminus \{i\}} \hat{\mu}_{k \rightarrow j}(\tau_j, D)} \quad (s_i \in \Omega, i \in V, j \in \partial i), \quad (10.163)$$

$$\hat{\mu}_{j \rightarrow i}(s_i, D) \leftarrow \mu_{j \rightarrow i}(s_i) \quad (s_i \in \Omega, i \in V, j \in \partial i), \quad (10.164)$$

$$B_i \leftarrow \sum_{\tau_i \in \Omega} g(\mathbf{d}_i | \tau_i, \hat{a}(\tau_i, D), \hat{\sigma}(\tau_i, D)) \prod_{k \in \partial i} \hat{\mu}_{k \rightarrow i}(\tau_i, D) \quad (i \in V), \quad (10.165)$$

$$\begin{aligned} B_{\{i, j\}} \leftarrow & \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \left(\prod_{k \in \partial i \setminus \{j\}} \hat{\mu}_{k \rightarrow i}(\tau_i, D) \right) g(\mathbf{d}_i | \tau_i, \hat{a}(\tau_i, D), \hat{C}(\tau_i, D)) \\ & \times \exp(2\hat{\alpha}(D)\delta_{\tau_i, \tau_j}) \\ & \times g(\mathbf{d}_j | \tau_j, \hat{a}(\tau_j, D), \hat{C}(\tau_j, D)) \left(\prod_{k \in \partial j \setminus \{i\}} \hat{\mu}_{k \rightarrow j}(\tau_j, D) \right) \quad (\{i, j\} \in E), \end{aligned} \quad (10.166)$$

$$\mathbf{a}(s_i) \leftarrow \frac{\sum_{i \in V} \frac{1}{B_i} \mathbf{d}_i g(\mathbf{d}_i | s_i, \widehat{\mathbf{a}}(s_i, \mathbf{D}), \widehat{\mathbf{C}}(s_i, \mathbf{D})) \left(\prod_{k \in \partial i} \widehat{\mu}_{k \rightarrow i}(s_i, \mathbf{D}) \right)}{\sum_{i \in V} \frac{1}{B_i} g(\mathbf{d}_i | s_i, \mathbf{a}(s_i, \mathbf{D}), \mathbf{C}(s_i, \mathbf{D})) \left(\prod_{k \in \partial i} \widehat{\mu}_{k \rightarrow i}(s_i, \mathbf{D}) \right)} (s_i \in \Omega), \quad (10.167)$$

$$\mathbf{C}(s_i) \leftarrow \frac{\sum_{i \in V} \frac{1}{B_i} (\mathbf{d}_i - \widehat{\mathbf{a}}(s_i, \mathbf{D})) (\mathbf{d}_i - \widehat{\mathbf{a}}(s_i, \mathbf{D}))^T g(\mathbf{d}_i | s_i, \widehat{\mathbf{a}}(s_i, \mathbf{D}), \widehat{\mathbf{C}}(s_i, \mathbf{D})) \left(\prod_{k \in \partial i} \widehat{\mu}_{k \rightarrow i}(s_i, \mathbf{D}) \right)}{\sum_{i \in V} \frac{1}{B_i} g(\mathbf{d}_i | s_i, \widehat{\mathbf{a}}(s_i, \mathbf{D}), \widehat{\mathbf{C}}(s_i, \mathbf{D})) \left(\prod_{k \in \partial i} \widehat{\mu}_{k \rightarrow i}(s_i, \mathbf{D}) \right)} (s_i \in \Omega), \quad (10.168)$$

$$\begin{aligned} \widehat{\mathbf{u}}(\mathbf{D}) \leftarrow \frac{1}{|E|} \sum_{\{i,j\} \in E} \left(\frac{1}{B_{\{i,j\}}} \sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} (-\delta_{\tau_i, \tau_j}) \exp(2\widehat{\alpha}(\mathbf{D}) \delta_{\tau_i, \tau_j}) \right. \\ \times \left(\prod_{k \in \partial i \setminus \{j\}} \widehat{\mu}_{k \rightarrow i}(\tau_i, \mathbf{D}) \right) g(\mathbf{d}_i | \tau_i, \widehat{\mathbf{a}}(\tau_i, \mathbf{D}), \widehat{\mathbf{C}}(\tau_i, \mathbf{D})) \\ \times \left. \left(\prod_{k \in \partial j \setminus \{i\}} \widehat{\mu}_{k \rightarrow j}(\tau_j, \mathbf{D}) \right) g(\mathbf{d}_j | \tau_j, \widehat{\mathbf{a}}(\tau_j, \mathbf{D}), \widehat{\mathbf{C}}(\tau_j, \mathbf{D})) \right), \end{aligned} \quad (10.169)$$

$$\widehat{\mathbf{a}}(s_i, \mathbf{D}) \leftarrow \mathbf{a}(s_i) \quad (s_i \in \Omega), \quad (10.170)$$

$$\widehat{\mathbf{C}}(s_i, \mathbf{D}) \leftarrow \mathbf{C}(s_i) \quad (s_i \in \Omega). \quad (10.171)$$

Here, $g(\mathbf{d}_i | \xi, \widehat{\mathbf{a}}(\xi, \mathbf{D}), \widehat{\mathbf{C}}(\xi, \mathbf{D}))$ is defined by Eq. (10.31) for each state $\xi (\in \Omega)$.

Step 3 (M-step): Set the initial values of the messages $\{\widehat{\lambda}(\xi) | \xi \in \Omega\}$ in the loopy belief propagation for the Potts prior and repeat the following procedure until $\widehat{\alpha}(\mathbf{d})$ and $\{\widehat{\lambda}(\xi) | \xi \in \Omega\}$ converge:

$$\lambda(s_i) \leftarrow \frac{\sum_{\tau_j \in \Omega} \exp(2\widehat{\alpha}(\mathbf{D}) \delta_{s_i, \tau_j}) \widehat{\lambda}(\tau_j, \mathbf{D})^3}{\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \exp(2\widehat{\alpha}(\mathbf{D}) \delta_{\tau_i, \tau_j}) \widehat{\lambda}(\tau_j, \mathbf{D})^3} (s_i \in \Omega), \quad (10.172)$$

$$\widehat{\lambda}(s_i, \mathbf{d}) \leftarrow \lambda(s_i) \quad (s_i \in \Omega), \quad (10.173)$$

$$\widehat{\alpha}(\mathbf{D}) \leftarrow \widehat{\alpha}(\mathbf{D}) \times \left(\frac{1}{1 + \widehat{\alpha}(\mathbf{D})} \frac{\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} (1 - \delta_{\tau_i, \tau_j}) \widehat{\lambda}(\tau_i, \mathbf{D})^3 \exp(2\widehat{\alpha}(\mathbf{D})\delta_{\tau_i, \tau_j}) \widehat{\lambda}(\tau_j, \mathbf{D})^3}{\sum_{\tau_i \in \Omega} \sum_{\tau_j \in \Omega} \widehat{\lambda}(\tau_i, \mathbf{D})^3 \exp(2\widehat{\alpha}(\mathbf{D})\delta_{\tau_i, \tau_j}) \widehat{\lambda}(\tau_j, \mathbf{D})^3} \right)^{1/4}. \quad (10.174)$$

Step 4 Compute the output $\widehat{\mathbf{s}}(\mathbf{D}) = (\widehat{s}_1(\mathbf{D}), \widehat{s}_2(\mathbf{D}), \dots, \widehat{s}_{|V|}(\mathbf{D}))$ as follows:

$$\widehat{s}_i(\mathbf{D}) \leftarrow \operatorname{argmax}_{s_i \in \Omega} g(\mathbf{d}_i | s_i, \widehat{\alpha}(s_i, \mathbf{D}), \widehat{\mathbf{C}}(s_i, \mathbf{D})) \prod_{k \in \partial i} \widehat{\mu}_{k \rightarrow i}(s_i, \mathbf{D}) \quad (i \in V). \quad (10.175)$$

Stop if the hyperparameters $\widehat{\alpha}(\mathbf{D})$, $\widehat{\alpha}(s_i, \mathbf{D})$ ($s_i \in \Omega$), and $\widehat{\mathbf{C}}(s_i, \mathbf{D})$ ($s_i \in \Omega$) converge and return to **Step 2** otherwise.

Some of the numerical experimental results are shown in Fig. 10.14. The Potts prior has the first-order phase transition as shown in Sect. 10.3.6. Figure 10.14 shows how the hyperparameter $2\alpha = \frac{J}{k_B T}$ converges in the EM procedure with loopy belief propagation under the first-order phase transition.

10.3.6 Real-Space Renormalization Group Method and Sublinear Modeling of Statistical Machine Learning

First, we explore the most fundamental real-space renormalization procedure for the Ising model in Eq. (10.49) on the ring graph (V, E) , where

$$E \equiv \left\{ \{1, 2\}, \{2, 3\}, \{3, 4\}, \dots, \{|V| - 1, |V|\}, \{|V|, 1\} \right\}, \quad (10.176)$$

in the case of $|V| = 2^L$. We have the following equality:

$$\begin{aligned} & \sum_{s_2 \in \Omega} \sum_{s_4 \in \Omega} \sum_{s_6 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \prod_{s_{|V|} \in \Omega} \exp\left(\frac{1}{k_B T} J s_i s_{i+1}\right) \\ &= \left(\sum_{s_2 \in \Omega} \exp\left(\frac{1}{k_B T} J (s_1 + s_3) s_2\right) \right) \left(\sum_{s_4 \in \Omega} \exp\left(\frac{1}{k_B T} J (s_3 + s_5) s_4\right) \right) \\ & \quad \times \cdots \times \left(\sum_{s_{|V|} \in \Omega} \exp\left(\frac{1}{k_B T} J (s_{|V|-3} + s_{|V|-1}) s_{|V|-2}\right) \right) \end{aligned}$$

Bayesian Image Segmentations by Potts Model and Loopy Belief Propagation

K. Tanaka et al:
JPSJ, 34, 124002, 2014.

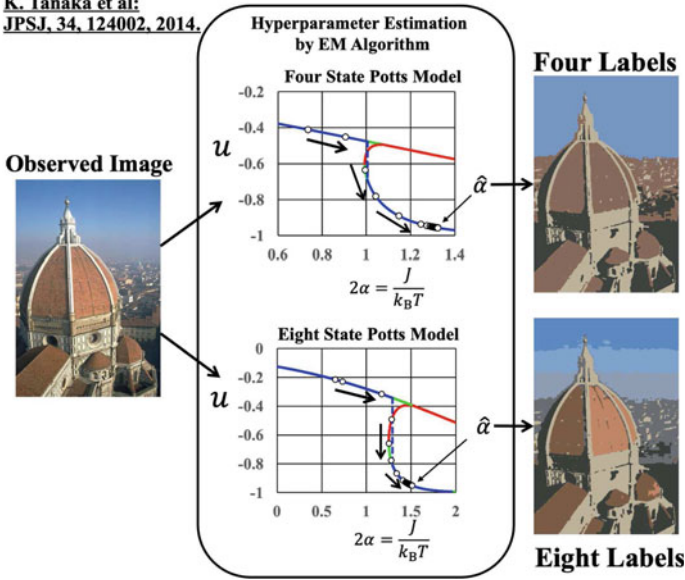


Fig. 10.14 Numerical experimental results of probabilistic segmentations by Potts prior and loopy belief propagation. The graph (V, E) is a square grid graph with periodic boundary conditions along the x - and y -directions

$$\begin{aligned}
 & \left(\sum_{s_{|V|} \in \Omega} \exp\left(\frac{1}{k_B T} J (s_{|V|-1} + s_1) s_{|V|}\right) \right) \\
 &= 2^{\frac{|V|}{2}} \left(\cosh\left(\frac{1}{k_B T} J\right) \right)^{\frac{1}{2}(1+s_1 s_3)} \left(\cosh\left(\frac{1}{k_B T} J\right) \right)^{\frac{1}{2}(1+s_3 s_5)} \\
 & \quad \times \cdots \times \left(\cosh\left(\frac{1}{k_B T} J\right) \right)^{\frac{1}{2}(1+s_{|V|-3} s_{|V|-1})} \left(\cosh\left(\frac{1}{k_B T} J\right) \right)^{\frac{1}{2}(1+s_{|V|-1} s_1)} \\
 &= 2^{\frac{|V|}{2}} \left(\prod_{i=0}^{\frac{|V|}{2}-2} \exp\left(\left(1 + s_{2i+1} s_{2i+3}\right) \times \frac{1}{2} \ln\left(\cosh\left(\frac{2}{k_B T} J\right)\right)\right) \right) \\
 & \exp\left(\left(1 + s_{|V|-1} s_1\right) \times \frac{1}{2} \ln\left(\cosh\left(\frac{2}{k_B T} J\right)\right)\right).
 \end{aligned} \tag{10.177}$$

For the $\frac{|V|}{2}$ -dimensional state vector $(a_1, a_3, a_5, \dots, a_{|V|-3}, a_{|V|-1})$, the marginal probability distribution $P_{\{1,3,5,\dots,|V|-3,|V|-1\}}(a_1, a_3, a_5, \dots, a_{|V|-3}, a_{|V|-1} | \alpha)$ is

expressed as

$$\begin{aligned}
& P_{\{1,3,5,\dots,|V|-3,|V|-1\}}(s_1, s_3, s_5, \dots, s_{|V|-3}, s_{|V|-1}) \\
& \equiv \sum_{s_2 \in \Omega} \sum_{s_4 \in \Omega} \sum_{s_6 \in \Omega} \cdots \sum_{s_{|V|-2} \in \Omega} \sum_{s_{|V|-1} \in \Omega} P(s_1, s_2, s_3, s_4, s_5, s_6, \dots, s_{|V|-3}, s_{|V|-2}, s_{|V|-1}, s_{|V|}) \\
& = \frac{\left(\prod_{i=0}^{\frac{|V|}{2}-2} \exp(\alpha^{(1)} s_{2i+1}, s_{2i+3}) \right) \exp(\alpha^{(1)} s_{|V|-1}, s_1)}{\sum_{a_1 \in \Omega} \sum_{a_3 \in \Omega} \cdots \sum_{a_{|V|-1} \in \Omega} \left(\prod_{i=0}^{\frac{|V|}{2}-2} \exp(\alpha^{(1)} s_{2i+1}, s_{2i+3}) \right) \exp(\alpha^{(1)} s_{|V|-1}, s_1)}, \quad (10.178)
\end{aligned}$$

where

$$\alpha^{(1)} \equiv \frac{1}{2} \ln \left(\cosh \left(\frac{2}{k_B T} J \right) \right). \quad (10.179)$$

The remaining nodes, which are denoted by odd numbers, are now renumbered by replacing i with $\frac{i-1}{2}$ for $i = 1, 3, 5, \dots, |V| - 3, |V| - 1$ and new sets $V^{(1)}$ and $E^{(1)}$ of nodes and edges and a new state vector $\mathbf{s}^{(1)} = (s_1^{(1)}, s_2^{(1)}, s_3^{(1)}, \dots, s_{\frac{|V|}{2}-1}^{(1)}, s_{\frac{|V|}{2}}^{(1)})^T$ are introduced as follows:

$$V^{(1)} \equiv \left\{ 1, 2, 3, 4, \dots, \frac{|V|}{2} - 1, \frac{|V|}{2} \right\}, \quad (10.180)$$

$$E^{(1)} \equiv \left\{ \{1, 2\}, \{2, 3\}, \{3, 4\}, \dots, \left\{ \frac{|V|}{2} - 1, \frac{|V|}{2} \right\}, \left\{ \frac{|V|}{2}, 1 \right\} \right\}, \quad (10.181)$$

$$s_i^{(1)} = s_{2i-1} \quad (i = 1, 2, \dots, |V|/2). \quad (10.182)$$

For the $\frac{|V|}{2}$ -dimensional state vector $\mathbf{s}^{(1)} = (s_1^{(1)}, s_2^{(1)}, s_3^{(1)}, \dots, s_{\frac{|V|}{2}-1}^{(1)}, s_{\frac{|V|}{2}}^{(1)})^T$, we define a new renormalized probability distribution by

$$P^{(1)}(\mathbf{s}^{(1)}) \equiv \frac{\prod_{\{i,j\} \in E^{(1)}} \exp(\alpha^{(1)} s_i^{(1)} s_j^{(1)})}{\sum_{a_1^{(1)} \in \Omega} \sum_{a_2^{(1)} \in \Omega} \cdots \sum_{a_{\frac{|V|}{2}}^{(1)} \in \Omega} \prod_{\{i,j\} \in E^{(1)}} \exp(\alpha^{(1)} s_i^{(1)} s_j^{(1)})}. \quad (10.183)$$

By repeating the above renormalizing procedures,

$$\begin{aligned}
& \sum_{s_2^{(r-1)} \in \Omega} \sum_{s_4^{(r-1)} \in \Omega} \sum_{s_6^{(r-1)} \in \Omega} \cdots \sum_{s_{|V|^{(r-1)}} \in \Omega} \prod_{\{i,j\} \in E^{(r-1)}} \exp\left(\frac{1}{k_B T} J s_i^{(r-1)} s_{i+1}^{(r-1)}\right) \\
&= 2^{\frac{|V|}{2^r}} \left(\prod_{i=0}^{\frac{|V|}{2^r}-2} \exp\left((1 + s_{2i+1} s_{2i+3}) \times \frac{1}{2} \ln(\cosh(2\alpha^{(r-1)}))\right) \right) \\
& \quad \exp\left((1 + s_{|V|-1} s_1) \times \frac{1}{2} \ln\left(\cosh\left(\frac{2}{k_B T} J\right)\right)\right),
\end{aligned} \tag{10.184}$$

$$\alpha^{(r)} \equiv \frac{1}{2} \ln(\cosh(2\alpha^{(r-1)})), \tag{10.185}$$

$$s_i^{(r)} = s_{2i-1}^{(r-1)} \left(i = 1, 2, \dots, \frac{|V|}{2^r} \right), \tag{10.186}$$

the renormalized probability of the r -th step is generated as follows:

$$P^{(r)}(\mathbf{s}^{(r)}) \equiv \frac{\prod_{\{i,j\} \in E^{(r)}} \exp(\alpha^{(r)} s_i^{(r)} s_{i+1}^{(r)})}{\sum_{s_1^{(r)} \in \Omega} \sum_{s_2^{(r)} \in \Omega} \cdots \sum_{s_{|V|/2^r}^{(r)} \in \Omega} \prod_{\{i,j\} \in E^{(r)}} \exp(\alpha^{(r)} s_i^{(r)} s_{i+1}^{(r)})}, \tag{10.187}$$

where

$$V^{(r)} \equiv \left\{ 1, 2, \dots, \frac{|V|}{2^r} \right\}, \tag{10.188}$$

$$E^{(r)} \equiv \left\{ \{1, 2\}, \{2, 3\}, \{3, 4\}, \dots, \left\{ \frac{|V|}{2^r} - 1, \frac{|V|}{2^r} \right\}, \left\{ \frac{|V|}{2^r}, 1 \right\} \right\}. \tag{10.189}$$

Note that $V^{(0)} = V$, $E^{(0)} = E$, $\alpha^{(0)} = \frac{J}{k_B T}$, $\mathbf{s}^{(0)} = \mathbf{s}$, and $P^{(0)}(\mathbf{s}^{(0)}) = P(\mathbf{s})$.

Equation (10.185) corresponds to the update rule from $\alpha^{(r-1)}$ to $\alpha^{(r)}$. By solving Eq. (10.185) with respect to $\alpha^{(r-1)}$, we can derive the inverse transformation rule of the real-space renormalization group procedure as follows:

$$\alpha^{(r-1)} = \frac{1}{2} \operatorname{arccosh}(\exp(2\alpha^{(r)})). \quad (10.190)$$

If the hyperparameter $\alpha^{(r)}$ in the r -th renormalized probability distribution $P^{(r)}(\mathbf{s}^{(r)})$ has been estimated from given data vectors by means of the EM algorithm for renormalized probabilistic graphical models on ring graphs $(V^{(r)}, E^{(r)})$, we can estimate the hyperparameter $\alpha^{(0)} = \frac{J}{k_B T}$ of the probabilistic graphical models (10.49) on ring graphs (V, E) by using the inverse transformation rule of the real-space renormalization group procedure (10.190).

Now, we extend the real-space renormalization group scheme for the probabilistic graphical model on the ring graph to the square grid graph as a pair approximation in the real-space renormalization group framework as follows:

$$\exp(\alpha^{(r)} s_1 s_3) \propto \sum_{s_2 \in \Omega} \sum_{s_4 \in \Omega} \exp(\alpha^{(r-1)} (s_1 s_2 + s_2 s_3 + s_1 s_4 + s_4 s_3)). \quad (10.191)$$

Equation (10.191) can be reduced to

$$\alpha^{(r)} = \ln(\cosh(2\alpha^{(r-1)})). \quad (10.192)$$

The r -th renormalized probability distribution for Eq. (10.49) is expressed as

$$P^{(r)}(\mathbf{s}^{(r)}) \propto \prod_{\{i,j\} \in E^{(r)}} \exp(\alpha^{(r)} s_i^{(r)} s_j^{(r)}). \quad (10.193)$$

The inversion formula in Eq. (10.192) can be derived as

$$\alpha^{(r-1)} = \frac{1}{2} \operatorname{arccosh}(\exp(\alpha^{(r)})). \quad (10.194)$$

The above framework can be extended to the $|\Omega|$ -state Potts model, as shown in Fig. 10.15. The inverse renormalization group transformation can also be applied to the probabilistic segmentations in Eqs. (10.41), (10.42), and (10.43) with Eqs. (10.44), (10.45), and (10.46) in Sect. 10.2.3 [81]. One of the numerical experimental results in the inverse renormalization group transformation in probabilistic segmentations is shown in Fig. 10.16.

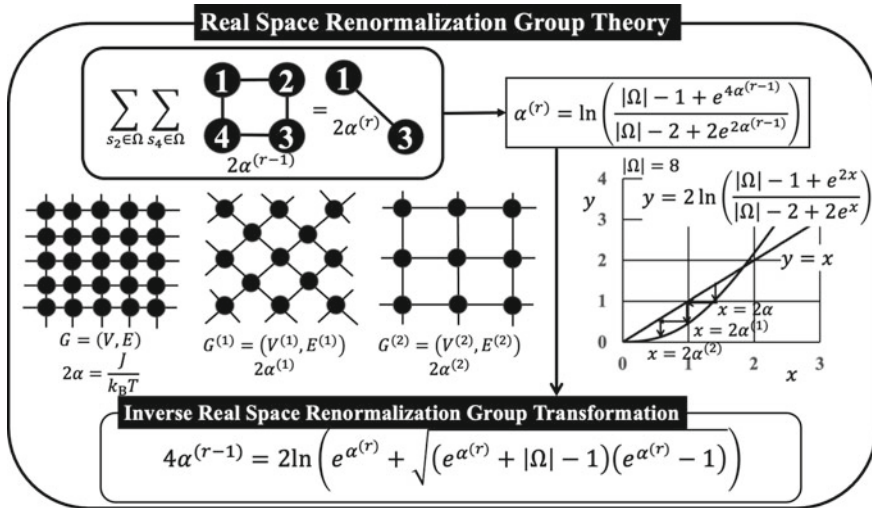


Fig. 10.15 Fundamental framework of sublinear computational modeling by the inverse renormalization group transformation in probabilistic segmentations

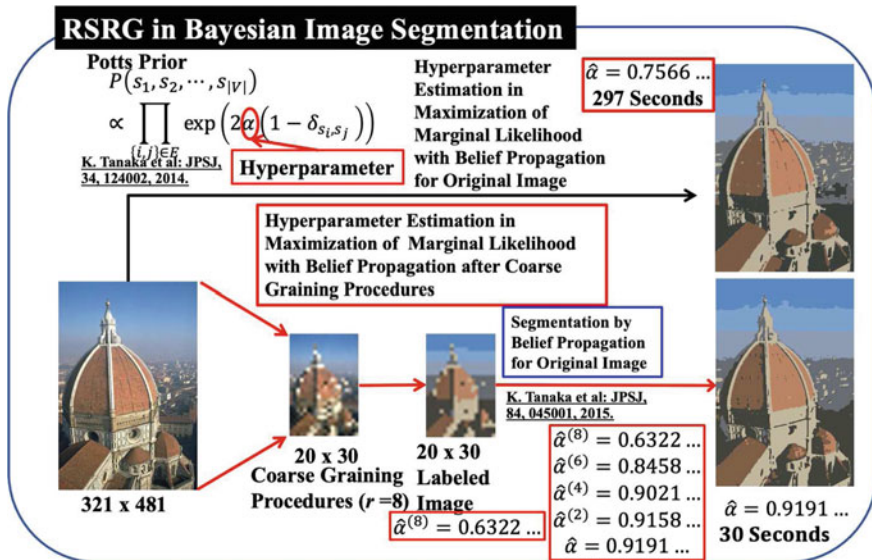


Fig. 10.16 Numerical experimental results of sublinear computational modeling in the inverse renormalization group transformation in probabilistic segmentations

10.4 Quantum Statistical Machine Learning

This section explores the fundamental frameworks of quantum probabilistic graphical models based on energy matrices and density matrices. Note that every energy matrix needs to be Hermitian and have a density matrix that is defined by all the eigenvalues and all the eigenvectors of each energy matrix. If all the off-diagonal elements of the density matrix are zero, the diagonal elements correspond to the probability distribution in the probabilistic graphical model. First, we explain general frameworks of density matrices and their differentiations and define the minimization of free energies of density matrices. Second, we give the definitions of tensor products of matrices as well as vectors. By using Pauli spin matrices as well as tensor products, we introduce quantum probabilistic graphical models. Finally, we extend the conventional EM algorithm to a quantum expectation-maximization (QEM) algorithm.

10.4.1 Elementary Function and Differentiations of Hermitian Matrices

Before proceeding with the quantum statistical mechanical extension of statistical machine learning, we need to explore some essential formulas for Hermitian matrices and their derivatives. Some fundamental properties of matrices for statistical inference have appeared in Ref. [82]. In the present section, we give some useful formulas for treating the entropy in quantum probabilistic graphical models.

We consider the $M \times M$ Hermitian matrix A

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1M} \\ A_{21} & A_{22} & \cdots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MM} \end{pmatrix}, \quad (10.195)$$

which satisfies $A = \overline{A}^T$. Here we remark that A^T and \overline{A} are the transpose and conjugate matrix of A , respectively. We introduce vertical and horizontal basis vectors in the M -dimensional space as follows:

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \dots, |M-1\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, |M\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad (10.196)$$

and

$$\left\{ \begin{array}{l} \langle 1| = (1, 0, 0, 0, \dots, 0, 0, 0), \\ \langle 2| = (0, 1, 0, 0, \dots, 0, 0, 0), \\ \langle 3| = (0, 0, 1, 0, \dots, 0, 0, 0), \\ \vdots \\ \langle M-1| = (0, 0, 0, 0, \dots, 0, 1, 0), \\ \langle M| = (0, 0, 0, 0, \dots, 0, 0, 1). \end{array} \right. \quad (10.197)$$

We can confirm that

$$\langle i|\mathbf{A}|j\rangle = A_{ij} \quad (i \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, M\}). \quad (10.198)$$

The Hermitian matrix \mathbf{A} is diagonalized as

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}, \quad (10.199)$$

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_M \end{pmatrix}, \quad (10.200)$$

where all the eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_M$, are always real numbers. For the eigen-

vector $\mathbf{u}_i = \begin{pmatrix} U_{1i} \\ U_{2i} \\ \vdots \\ U_{Mi} \end{pmatrix}$ corresponding to the eigenvalue λ_i , such that $\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i$, for every $i \in \{1, 2, 3, \dots, M\}$ the matrix \mathbf{U} is defined by

$$\mathbf{U} \equiv (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_M) = \begin{pmatrix} U_{11} & U_{12} & U_{13} & \dots & U_{1M} \\ U_{21} & U_{22} & U_{23} & \dots & U_{2M} \\ U_{31} & U_{32} & U_{33} & \dots & U_{3M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{M1} & U_{M2} & U_{M3} & \dots & U_{MM} \end{pmatrix}. \quad (10.201)$$

It is known that \mathbf{U} is a unitary matrix that satisfies $\mathbf{U}^{-1} = \overline{\mathbf{U}}^T$ for any Hermitian matrix s . If λ_1 is the maximum eigenvalue, its corresponding eigenvector \mathbf{u}_1 is expressed using the following notation:

$$\mathbf{u}_1 = \operatorname{argmax} \mathbf{A}. \quad (10.202)$$

Note that $\text{argmax} \mathbf{A}$ is the eigenvector that corresponds to the maximum eigenvalue of \mathbf{A} .

For any Hermitian matrix \mathbf{A} , the exponential function is defined by

$$\begin{aligned} \exp(\mathbf{A}) &\equiv \sum_{n=0}^{+\infty} \frac{1}{n!} \mathbf{A}^n \\ &= \mathbf{U} \begin{pmatrix} \exp(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & \exp(\lambda_2) & 0 & \cdots & 0 \\ 0 & 0 & \exp(\lambda_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \exp(\lambda_M) \end{pmatrix} \mathbf{U}^{-1}, \end{aligned} \quad (10.203)$$

and $\ln(\mathbf{A})$ is defined by the inverse function of $\exp(\mathbf{A})$ such that

$$\exp(\ln(\mathbf{A})) = \mathbf{A}. \quad (10.204)$$

In the present definition, we have

$$\exp(\mathbf{A} \otimes \mathbf{I}) = (\exp(\mathbf{A})) \otimes \mathbf{I}, \exp(\mathbf{I} \otimes \mathbf{A}) = \mathbf{I} \otimes (\exp(\mathbf{A})), \quad (10.205)$$

where \mathbf{I} is an identity matrix.

For $|1 - \lambda_1| < 1, |1 - \lambda_2| < 1, \dots, |1 - \lambda_N| < 1$, $\ln(\mathbf{A})$ is defined by

$$\begin{aligned} \ln(\mathbf{A}) = \ln(\mathbf{I} - (\mathbf{I} - \mathbf{A})) &\equiv - \sum_{n=1}^{+\infty} \frac{1}{n} (\mathbf{I} - \mathbf{A})^n \\ &= \mathbf{U} \begin{pmatrix} \ln(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & \ln(\lambda_2) & 0 & \cdots & 0 \\ 0 & 0 & \ln(\lambda_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \ln(\lambda_M) \end{pmatrix} \mathbf{U}^{-1}. \end{aligned} \quad (10.206)$$

By using Eqs. (10.203) and (10.206), we can confirm that

$$\begin{aligned} \exp(\ln(\mathbf{A})) &= \sum_{n=0}^{+\infty} \frac{1}{n!} (\ln(\mathbf{A}))^n \\ &= \sum_{n=0}^{+\infty} \frac{1}{n!} \mathbf{U} \begin{pmatrix} (\ln(\lambda_1))^n & 0 & 0 & \cdots & 0 \\ 0 & (\ln(\lambda_2))^n & 0 & \cdots & 0 \\ 0 & 0 & (\ln(\lambda_3))^n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (\ln(\lambda_M))^n \end{pmatrix} \mathbf{U}^{-1} \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{U} \begin{pmatrix} \exp(\ln(\lambda_1)) & 0 & 0 & \cdots & 0 \\ 0 & \exp(\ln(\lambda_2)) & 0 & \cdots & 0 \\ 0 & 0 & \exp(\ln(\lambda_3)) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \exp(\ln(\lambda_M)) \end{pmatrix} \mathbf{U}^{-1} \\
 &= \mathbf{U} \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_M \end{pmatrix} \mathbf{U}^{-1} = \mathbf{A}.
 \end{aligned} \tag{10.207}$$

Moreover, we have

$$\begin{aligned}
 &\sum_{n=0}^N (\mathbf{I} - \mathbf{A})^n - (\mathbf{I} - s) \sum_{n=0}^N (\mathbf{I} - \mathbf{A})^n \\
 &= \mathbf{I} + (\mathbf{I} - s) + (\mathbf{I} - \mathbf{A})^2 + \cdots + (\mathbf{I} - \mathbf{A})^N \\
 &\quad - (\mathbf{I} - s) - (\mathbf{I} - \mathbf{A})^2 - \cdots - (\mathbf{I} - s)^N - (\mathbf{I} - \mathbf{A})^{N+1} \\
 &= \mathbf{I} - (\mathbf{I} - \mathbf{A})^{N+1},
 \end{aligned} \tag{10.208}$$

such that

$$\sum_{n=0}^N (\mathbf{I} - \mathbf{A})^n = (\mathbf{I} - (\mathbf{I} - \mathbf{A}))^{-1} (\mathbf{I} - (\mathbf{I} - \mathbf{A})^{N+1}). \tag{10.209}$$

We have

$$\mathbf{A}^{N+1} = \mathbf{U} \begin{pmatrix} (1 - \lambda_1)^{N+1} & 0 & 0 & \cdots & 0 \\ 0 & (1 - \lambda_2)^{N+1} & 0 & \cdots & 0 \\ 0 & 0 & (1 - \lambda_3)^{N+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (1 - \lambda_M)^{N+1} \end{pmatrix} \mathbf{U}^{-1} \rightarrow 0 \quad (N \rightarrow +\infty), \tag{10.210}$$

so it is valid that

$$\mathbf{A}^{-1} = (\mathbf{I} - (\mathbf{I} - \mathbf{A}))^{-1} = \sum_{n=0}^{+\infty} (\mathbf{I} - \mathbf{A})^n. \tag{10.211}$$

Note that $\exp(\mathbf{A})$ and $\ln(\mathbf{A})$ as well as \mathbf{A}^{-1} are also Hermitian matrices in the present case. (This can be shown by using $\overline{\mathbf{A}^n}^T = (\overline{\mathbf{A}}^T)^n$.)

We now introduce a Hermitian matrix function $\mathbf{G}(x)$ for any real number x as follows:

$$\mathbf{G}(x) \equiv \begin{pmatrix} G_{11}(x) & G_{12}(x) & G_{13}(x) & \cdots & G_{1M}(x) \\ G_{21}(x) & G_{22}(x) & G_{23}(x) & \cdots & G_{2M}(x) \\ G_{31}(x) & G_{32}(x) & G_{33}(x) & \cdots & G_{3M}(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{M1}(x) & G_{M2}(x) & G_{M3}(x) & \cdots & G_{MM}(x) \end{pmatrix}. \quad (10.212)$$

We have

$$G_{ij}(x) = \overline{G_{ji}(x)} \quad (i \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, M\}), \quad (10.213)$$

such that

$$\langle i | \mathbf{G}(x) | j \rangle = \langle j | \overline{\mathbf{G}(x)} | i \rangle \quad (i \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, M\}). \quad (10.214)$$

It is obvious that the derivative of the matrix $\mathbf{G}(x)$ with respect to x , namely,

$$\frac{d}{dx} \mathbf{G}(x) \equiv \begin{pmatrix} \frac{d}{dx} G_{11}(x) & \frac{d}{dx} G_{12}(x) & \frac{d}{dx} G_{13}(x) & \cdots & \frac{d}{dx} G_{1M}(x) \\ \frac{d}{dx} G_{21}(x) & \frac{d}{dx} G_{22}(x) & \frac{d}{dx} G_{23}(x) & \cdots & \frac{d}{dx} G_{2M}(x) \\ \frac{d}{dx} G_{31}(x) & \frac{d}{dx} G_{32}(x) & \frac{d}{dx} G_{33}(x) & \cdots & \frac{d}{dx} G_{3M}(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{d}{dx} G_{M1}(x) & \frac{d}{dx} G_{M2}(x) & \frac{d}{dx} G_{M3}(x) & \cdots & \frac{d}{dx} G_{MM}(x) \end{pmatrix} \quad (10.215)$$

is also a Hermitian matrix such that

$$\langle i | \frac{d}{dx} \mathbf{G}(x) | j \rangle = \langle j | \frac{d}{dx} \overline{\mathbf{G}(x)} | i \rangle \quad (i \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, M\}). \quad (10.216)$$

We have the following equalities:

$$\frac{d}{dx} (\text{Tr}[\mathbf{G}(x)]) = \text{Tr} \left[\frac{d}{dx} \mathbf{G}(x) \right], \quad (10.217)$$

and

$$\text{Tr} \left(\left(\frac{d}{dx} \mathbf{G}(x) \right) \mathbf{G}(x) \right) = \text{Tr} \left(\mathbf{G}(x) \left(\frac{d}{dx} \mathbf{G}(x) \right) \right). \quad (10.218)$$

Equation (10.218) can be confirmed as follows:

$$\begin{aligned}
\text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)\mathbf{G}(x)\right) &= \sum_{i=1}^M \langle i | \left(\frac{d}{dx}\mathbf{G}(x)\right) \mathbf{G}(x) | i \rangle = \sum_{i=1}^M \sum_{j=1}^M \langle i | \left(\frac{d}{dx}\mathbf{G}(x)\right) | j \rangle \langle j | \mathbf{G}(x) | i \rangle \\
&= \sum_{j=1}^M \sum_{i=1}^M \langle j | \mathbf{G}(x) | i \rangle \langle i | \left(\frac{d}{dx}\mathbf{G}(x)\right) | j \rangle = \text{Tr}\left(\mathbf{G}(x)\left(\frac{d}{dx}\mathbf{G}(x)\right)\right).
\end{aligned} \tag{10.219}$$

By using Eqs. (10.217) and (10.219), we derive the following fundamental formula

$$\begin{aligned}
\frac{d}{dx} \text{tr}(\mathbf{G}(x)^n) &= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)(\mathbf{G}(x)^{n-1})\right) + \text{Tr}\left(\mathbf{G}(x)\left(\frac{d}{dx}\mathbf{G}(x)^{n-1}\right)\right) \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)(\mathbf{G}(x)^{n-1})\right) \\
&\quad + \text{Tr}\left(\mathbf{G}(x)\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)(\mathbf{G}(x)^{n-2}) + \mathbf{G}(x)\left(\frac{d}{dx}\mathbf{G}(x)^{n-2}\right)\right)\right) \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)(\mathbf{G}(x)^{n-1})\right) \\
&\quad + \text{Tr}\left(\mathbf{G}(x)\left(\frac{d}{dx}\mathbf{G}(x)\right)(\mathbf{G}(x)^{n-2})\right) + \text{Tr}\left(\mathbf{G}(x)^2\left(\frac{d}{dx}\mathbf{G}(x)^{n-2}\right)\right) \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)\mathbf{G}(x)^{n-1}\right) \\
&\quad + \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)\mathbf{G}(x)\mathbf{G}(x)^{n-2}\right) + \text{Tr}\left(\mathbf{G}(x)^2\left(\frac{d}{dx}\mathbf{G}(x)^{n-2}\right)\right) \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)2\mathbf{G}(x)^{n-1}\right) + \text{Tr}\left(\mathbf{G}(x)^2\left(\frac{d}{dx}\mathbf{G}(x)^{n-2}\right)\right) \\
&= \dots \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)(n-1)\mathbf{G}(x)^{n-1}\right) + \text{Tr}\left(\mathbf{G}(x)^{n-1}\left(\frac{d}{dx}\mathbf{G}(x)\right)\right) \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)(n-1)\mathbf{G}(x)^{n-1}\right) + \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)\mathbf{G}(x)^{n-1}\right) \\
&= \text{Tr}\left(\left(\frac{d}{dx}\mathbf{G}(x)\right)n\mathbf{G}(x)^{n-1}\right).
\end{aligned} \tag{10.220}$$

From Eqs. (10.217) and (10.220), we can confirm the following equality:

$$\begin{aligned}
\frac{d}{dx} \text{Tr}(\ln(\mathbf{G}(x))) &= \frac{d}{dx} \text{Tr}(\ln(\mathbf{I} - (\mathbf{I} - \mathbf{G}(x)))) \\
&= \frac{d}{dx} \text{Tr} \left(- \sum_{n=1}^{+\infty} \frac{1}{n} (\mathbf{I} - \mathbf{G}(x))^n \right) \\
&= - \sum_{n=1}^{+\infty} \frac{1}{n} \frac{d}{dx} \text{Tr}((\mathbf{I} - \mathbf{G}(x))^n) \\
&= - \sum_{n=1}^{+\infty} \frac{1}{n} \text{Tr} \left(\left(\frac{d}{dx} (\mathbf{I} - \mathbf{G}(x)) \right) n (\mathbf{I} - \mathbf{G}(x))^{n-1} \right) \\
&= \text{Tr} \left(\left(- \frac{d}{dx} (\mathbf{I} - \mathbf{G}(x)) \right) \left(\sum_{n=1}^{+\infty} (\mathbf{I} - \mathbf{G}(x))^{n-1} \right) \right) \\
&= \text{Tr} \left(\left(\frac{d}{dx} \mathbf{G}(x) \right) \left(\sum_{n=1}^{+\infty} (\mathbf{I} - \mathbf{G}(x))^{n-1} \right) \right) \\
&= \text{Tr} \left(\left(\frac{d}{dx} \mathbf{G}(x) \right) (\mathbf{I} - (\mathbf{I} - \mathbf{G}(x)))^{-1} \right). \\
&= \text{Tr} \left(\left(\frac{d}{dx} \mathbf{G}(x) \right) (\mathbf{G}(x))^{-1} \right). \tag{10.221}
\end{aligned}$$

By using Eqs. (10.221), we can confirm the following equality:

$$\begin{aligned}
\frac{d}{d\mathbf{A}} \text{Tr}[\mathbf{A}(\ln(\mathbf{A}))] &\equiv \begin{pmatrix} \frac{d}{dA_{11}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] & \frac{d}{dA_{12}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] & \cdots & \frac{d}{dA_{1M}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] \\ \frac{d}{dA_{21}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] & \frac{d}{dA_{22}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] & \cdots & \frac{d}{dA_{2M}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d}{dA_{M1}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] & \frac{d}{dA_{M2}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] & \cdots & \frac{d}{dA_{MM}} \text{Tr}[\mathbf{A} \ln(\mathbf{A})] \end{pmatrix} \\
&= \ln(\mathbf{A}) + \mathbf{I}. \tag{10.222}
\end{aligned}$$

10.4.2 Minimization of Free Energy Functionals for Density Matrices

For any $M \times M$ Hermitian matrix \mathbf{H} that satisfies $\mathbf{H} = \overline{\mathbf{H}}^T$, the free energy functional for an $M \times M$ trial density matrix

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1M} \\ R_{21} & R_{22} & \cdots & R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ R_{M1} & R_{M2} & \cdots & R_{MM} \end{pmatrix} \tag{10.223}$$

is defined by

$$\mathcal{F}[\mathbf{R}] = \text{Tr}[\mathbf{R}(\mathbf{H} + k_{\text{B}}T \ln(\mathbf{R}))]. \quad (10.224)$$

The density matrix \mathbf{P} is determined so as to satisfy the following conditional minimization with the normalization condition as follows:

$$\mathbf{P} = \underset{\mathbf{R}}{\text{argmin}} \{ \mathcal{F}[\mathbf{R}] | \text{Tr}[\mathbf{R}] = 1 \}, \quad (10.225)$$

and this reduces to

$$\mathbf{P} = \frac{1}{Z} \exp\left(-\frac{1}{k_{\text{B}}T} \mathbf{H}\right), \quad (10.226)$$

$$Z \equiv \text{Tr}\left[\exp\left(-\frac{1}{k_{\text{B}}T} \mathbf{H}\right)\right]. \quad (10.227)$$

First, we introduce the Lagrange multiplier λ to ensure the normalization condition as follows:

$$\mathcal{L}[\mathbf{R}] \equiv \mathcal{F}[\mathbf{R}] - \lambda(\text{Tr}[\mathbf{R}] - 1). \quad (10.228)$$

$\hat{\mathbf{R}}$ are determined so as to satisfy the following extremum condition:

$$\frac{\partial}{\partial R_{mm'}} \mathcal{L}[\mathbf{R}] = 0 \quad (m = 1, 2, \dots, M, \quad m' = 1, 2, \dots, M). \quad (10.229)$$

Finally, by determining λ so as to satisfy the normalization condition $\text{Tr}[\hat{\mathbf{R}}] = 1$, Eqs. (10.226) and (10.227) can be derived.

Because the energy matrix \mathbf{H} is a Hermitian matrix, all the eigenvalues h_m are always real numbers and all the eigenvectors $\begin{pmatrix} \psi^{(m)}(1) \\ \psi^{(m)}(2) \\ \vdots \\ \psi^{(m)}(M) \end{pmatrix}$ can be chosen as real vectors and are defined by

$$\mathbf{H} \begin{pmatrix} \psi^{(m)}(1) \\ \psi^{(m)}(2) \\ \vdots \\ \psi^{(m)}(M) \end{pmatrix} = h^{(m)} \begin{pmatrix} \psi^{(m)}(1) \\ \psi^{(m)}(2) \\ \vdots \\ \psi^{(m)}(M) \end{pmatrix} \quad (m = 1, 2, \dots, M), \quad (10.230)$$

where

$$\left(\psi^{(m)}(1), \psi^{(m)}(2), \dots, \psi^{(m)}(M)\right) \begin{pmatrix} \psi^{(m)}(1) \\ \psi^{(m)}(2) \\ \vdots \\ \psi^{(m)}(M) \end{pmatrix} = 1 \quad (m = 1, 2, \dots, M). \quad (10.231)$$

By using these eigenvalues and eigenvectors of \mathbf{H} , the density matrix can be expressed as

$$\hat{\mathbf{R}} = \begin{pmatrix} \psi^{(1)}(1) & \psi^{(2)}(1) & \dots & \psi^{(M)}(1) \\ \psi^{(1)}(2) & \psi^{(2)}(2) & \dots & \psi^{(M)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi^{(1)}(M) & \psi^{(2)}(M) & \dots & \psi^{(M)}(M) \end{pmatrix} \begin{pmatrix} p^{(1)} & 0 & \dots & 0 \\ 0 & p^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p^{(M)} \end{pmatrix} \begin{pmatrix} \psi^{(1)}(1) & \psi^{(2)}(1) & \dots & \psi^{(M)}(1) \\ \psi^{(1)}(2) & \psi^{(2)}(2) & \dots & \psi^{(M)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi^{(1)}(M) & \psi^{(2)}(M) & \dots & \psi^{(M)}(M) \end{pmatrix}^T, \quad (10.232)$$

where

$$p^{(m)} = \frac{\exp\left(-\frac{1}{k_B T} h^{(m)}\right)}{\text{Tr}\left[\exp\left(-\frac{1}{k_B T} h^{(m)}\right)\right]} \quad (m = 1, 2, \dots, M). \quad (10.233)$$

This means that the probability of each state $\begin{pmatrix} \psi^{(m)}(1) \\ \psi^{(m)}(2) \\ \vdots \\ \psi^{(m)}(M) \end{pmatrix}$ is $p^{(m)}$ for $m = 1, 2, \dots, M$.

10.4.3 Tensor Products

This section explores **tensor products (Kronecker products)** [82]. Tensor products include some fundamental mathematical concepts for achieving quantum statistical mechanical extensions of probabilistic graphical models.

We introduce tensor products for matrices and vectors by the following definitions:

$$\begin{aligned} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \otimes \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} &= \begin{pmatrix} A_{11} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} & A_{12} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\ A_{21} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} & A_{22} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{pmatrix}, \end{aligned} \quad (10.234)$$

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \otimes \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} A_1 \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \\ A_2 \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} A_{11} B_{11} \\ A_{11} B_{21} \\ A_{21} B_{11} \\ A_{21} B_{21} \end{pmatrix}. \tag{10.235}$$

We remark that

$$\begin{aligned} & \left(\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \otimes \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \right) \left(\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \otimes \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \right) \\ &= \left(\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \right) \otimes \left(\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \right). \end{aligned} \tag{10.236}$$

Moreover, for the following general matrices \mathbf{A} and \mathbf{B} ,

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1M} \\ A_{21} & A_{22} & \cdots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MM} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1N} \\ B_{21} & B_{22} & \cdots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{N1} & B_{N2} & \cdots & B_{NN} \end{pmatrix}, \tag{10.237}$$

we define the tensor product $\mathbf{A} \otimes \mathbf{B}$ as

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B} &= \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1M} \\ A_{21} & A_{22} & \cdots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MM} \end{pmatrix} \otimes \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1N} \\ B_{21} & B_{22} & \cdots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{N1} & B_{N2} & \cdots & B_{NN} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \cdots & A_{1M}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \cdots & A_{2M}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1}\mathbf{B} & A_{M2}\mathbf{B} & \cdots & A_{MM}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & \cdots & A_{11}B_{1N} & A_{12}B_{11} & A_{12}B_{12} & \cdots & A_{12}B_{1N} & \cdots & A_{1M}B_{11} & A_{1M}B_{12} & \cdots & A_{1M}B_{1N} \\ A_{11}B_{21} & A_{11}B_{22} & \cdots & A_{11}B_{2N} & A_{12}B_{21} & A_{12}B_{22} & \cdots & A_{12}B_{2N} & \cdots & A_{1M}B_{21} & A_{1M}B_{22} & \cdots & A_{1M}B_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{11}B_{N1} & A_{11}B_{N2} & \cdots & A_{11}B_{NN} & A_{12}B_{N1} & A_{12}B_{N2} & \cdots & A_{12}B_{NN} & \cdots & A_{1M}B_{N1} & A_{1M}B_{N2} & \cdots & A_{1M}B_{NN} \\ A_{21}B_{11} & A_{21}B_{12} & \cdots & A_{21}B_{1N} & A_{22}B_{11} & A_{22}B_{12} & \cdots & A_{22}B_{1N} & \cdots & A_{2M}B_{11} & A_{2M}B_{12} & \cdots & A_{2M}B_{1N} \\ A_{21}B_{21} & A_{21}B_{22} & \cdots & A_{21}B_{2N} & A_{22}B_{21} & A_{22}B_{22} & \cdots & A_{22}B_{2N} & \cdots & A_{2M}B_{21} & A_{2M}B_{22} & \cdots & A_{2M}B_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{21}B_{N1} & A_{21}B_{N2} & \cdots & A_{21}B_{NN} & A_{22}B_{N1} & A_{22}B_{N2} & \cdots & A_{22}B_{NN} & \cdots & A_{2M}B_{N1} & A_{2M}B_{N2} & \cdots & A_{2M}B_{NN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{M1}B_{11} & A_{M1}B_{12} & \cdots & A_{M1}B_{1N} & A_{M2}B_{11} & A_{M2}B_{12} & \cdots & A_{M2}B_{1N} & \cdots & A_{MM}B_{11} & A_{MM}B_{12} & \cdots & A_{MM}B_{1N} \\ A_{M1}B_{21} & A_{M1}B_{22} & \cdots & A_{M1}B_{2N} & A_{M2}B_{21} & A_{M2}B_{22} & \cdots & A_{M2}B_{2N} & \cdots & A_{MM}B_{21} & A_{MM}B_{22} & \cdots & A_{MM}B_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{M1}B_{N1} & A_{M1}B_{N2} & \cdots & A_{M1}B_{NN} & A_{M2}B_{N1} & A_{M2}B_{N2} & \cdots & A_{M2}B_{NN} & \cdots & A_{MM}B_{N1} & A_{MM}B_{N2} & \cdots & A_{MM}B_{NN} \end{pmatrix}, \end{aligned} \tag{10.238}$$

Similarly, for vectors

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}, \quad (10.239)$$

the tensor product $\mathbf{a} \otimes \mathbf{b}$ is defined as

$$\mathbf{a} \otimes \mathbf{b} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix} \otimes \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} = \begin{pmatrix} a_1 \mathbf{b} \\ a_2 \mathbf{b} \\ \vdots \\ a_M \mathbf{b} \end{pmatrix} = \begin{pmatrix} a_1 \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} \\ a_2 \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} \\ \vdots \\ a_M \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \\ \vdots \\ a_1 b_N \\ a_2 b_1 \\ a_2 b_2 \\ \vdots \\ a_2 b_N \\ \vdots \\ a_M b_1 \\ a_M b_2 \\ \vdots \\ a_M b_N \end{pmatrix}, \quad (10.240)$$

$$\begin{aligned} \mathbf{a}^T \otimes \mathbf{b}^T &= (a_1, a_2, \dots, a_M) \otimes (b_1, b_2, \dots, b_N) = (a_1 \mathbf{b}^T, a_2 \mathbf{b}^T, \dots, a_M \mathbf{b}^T) \\ &= (a_1 b_1, a_1 b_2, \dots, a_1 b_N, a_2 b_1, a_2 b_2, \dots, a_2 b_N, \dots, a_M b_1, a_M b_2, \dots, a_M b_N). \end{aligned} \quad (10.241)$$

We introduce the following two-dimensional fundamental vectors:

$$|1\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |2\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (10.242)$$

$$|1\rangle \equiv (1, 0), \quad |2\rangle \equiv (0, 1). \quad (10.243)$$

By using the fundamental vectors in two-dimensional space, we define the vertical and horizontal fundamental vectors in four-dimensional space by using the tensor product as follows:

$$\left\{ \begin{array}{l} |1, 1\rangle \equiv |1\rangle \otimes |1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, |1, 2\rangle \equiv |1\rangle \otimes |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \\ |2, 1\rangle \equiv |2\rangle \otimes |1\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, |2, 2\rangle \equiv |2\rangle \otimes |2\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \end{array} \right. \quad (10.244)$$

$$\left\{ \begin{array}{l} \langle 1, 1| \equiv \langle 1| \otimes \langle 1| = (1, 0, 0, 0), \langle 1, 2| \equiv \langle 1| \otimes \langle 2| = (0, 1, 0, 0), \\ \langle 2, 1| \equiv \langle 2| \otimes \langle 1| = (0, 0, 1, 0), \langle 2, 2| \equiv \langle 2| \otimes \langle 2| = (0, 0, 0, 1), \end{array} \right. \quad (10.245)$$

It is easy to confirm the following equality:

$$\langle i, j| \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \otimes \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} |i', j'\rangle = A_{i,i'} B_{j,j'}. \quad (10.246)$$

By extending the above example to general-dimensional fundamental vectors, the $(i, j|i', j')$ -components of $\mathbf{A} \times \mathbf{B}$ for any $M \times M$ matrix \mathbf{A} and $N \times N$ matrix \mathbf{B} are expressed as

$$\langle i, j| \mathbf{A} \otimes \mathbf{B} |i', j'\rangle = \langle i| \mathbf{A} |i'\rangle \langle j| \mathbf{B} |j'\rangle = A_{i,i'} B_{j,j'}. \quad (10.247)$$

For $M \times M$ matrices \mathbf{A} and \mathbf{C} and $N \times N$ matrices \mathbf{B} and \mathbf{D} , we have

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}), \quad (10.248)$$

and

$$\text{Tr}[\mathbf{A} \otimes \mathbf{B}] = (\text{Tr}[\mathbf{A}])(\text{Tr}[\mathbf{B}]). \quad (10.249)$$

In deriving the equality in Eq. (10.248), the $(i, j|i', j')$ -components of the $MN \times MN$ matrix $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D})$ are given by

$$\begin{aligned} \langle i, j| (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) |i', j'\rangle &= \sum_{i''=1}^M \sum_{j''=1}^N \langle i, j| (\mathbf{A} \otimes \mathbf{B}) |i'', j''\rangle \langle i'', j''| (\mathbf{C} \otimes \mathbf{D}) |i', j'\rangle \\ &= \langle i \in \{1, 2, \dots, M\}, i' \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, N\}, j' \in \{1, 2, \dots, N\} \rangle. \end{aligned} \quad (10.250)$$

For the $M \times M$ and $N \times N$ identity matrices $\mathbf{I}^{(M)}$ and $\mathbf{I}^{(N)}$, it is valid that

$$(A \otimes I^{(N)})(I^{(M)} \otimes B) = (I^{(M)} \otimes B)(A \otimes I^{(N)}) = A \otimes B. \quad (10.251)$$

Moreover, by using mathematical induction, we can confirm the following binomial expansion:

$$(A \otimes I^{(N)} + I^{(M)} \otimes B)^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} (A \otimes I^{(N)})^k (I^{(M)} \otimes B)^{n-k}. \quad (10.252)$$

By using Eq. (10.252), we can derive the following equality:

$$\begin{aligned} \exp(A \otimes I^{(N)} + I^{(M)} \otimes B) &= \sum_{n=0}^{+\infty} \frac{1}{n!} (A \otimes I^{(N)} + I^{(M)} \otimes B)^n \\ &= \sum_{n=0}^{+\infty} \frac{1}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} (A \otimes I^{(N)})^k (I^{(M)} \otimes B)^{n-k} \\ &= \sum_{k=0}^{+\infty} \sum_{n=k}^{+\infty} \frac{1}{n!} \frac{n!}{k!(n-k)!} (A \otimes I^{(N)})^k (I^{(M)} \otimes B)^{n-k} \\ &= \sum_{k=0}^{+\infty} \sum_{n=k}^{+\infty} \frac{1}{k!(n-k)!} (A \otimes I^{(N)})^k (I^{(M)} \otimes B)^{n-k} \\ &= \sum_{k=0}^{+\infty} \sum_{l=0}^{+\infty} \frac{1}{k!l!} (A \otimes I^{(N)})^k (I^{(M)} \otimes B)^l \\ &= \left(\sum_{k=0}^{+\infty} \frac{1}{k!} (A \otimes I^{(N)})^k \right) \left(\sum_{l=0}^{+\infty} \frac{1}{l!} (I^{(M)} \otimes B)^l \right) \\ &= \left(\sum_{k=0}^{+\infty} \frac{1}{k!} (A^k \otimes I^{(N)}) \right) \left(\sum_{l=0}^{+\infty} \frac{1}{l!} (I^{(M)} \otimes B^l) \right) \\ &= \left(\left(\sum_{k=0}^{+\infty} \frac{1}{k!} A^k \right) \otimes I^{(N)} \right) \left(I^{(M)} \otimes \left(\sum_{l=0}^{+\infty} \frac{1}{l!} B^l \right) \right) \\ &= (\exp(A) \otimes I^{(N)}) (I^{(M)} \otimes \exp(B)) \\ &= \exp(A) \otimes \exp(B). \end{aligned} \quad (10.253)$$

By taking the logarithm of both sides of Eq. (10.253), we have

$$\ln(\exp(A) \otimes I^{(N)} + I^{(M)} \otimes \ln(\exp(B))) = \ln(\exp(A) \otimes \exp(B)). \quad (10.254)$$

10.4.4 Quantum Probabilistic Graphical Models and Quantum Expectation-Maximization Algorithm

This section explores a type of probabilistic graphical modeling based on Pauli spin matrices from the quantum statistical mechanical point of view. Our review focuses on the transverse Ising model in statistical mechanical informatics [37, 83, 84]. Note that generalization of the framework is possible.

Consider a graph specified by nodes and edges (V, E) where V is the set of all nodes i and E is the set of all edges $\{i, j\}$. We introduce Pauli spin matrices σ^z and σ^x as well as an identity matrix I defined by

$$\sigma^z = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma^x = \begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix}, \quad I = \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix}. \quad (10.255)$$

The Pauli spin matrices at each node $i \in V \equiv \{1, 2, \dots, N\}$ are defined by

$$\left\{ \begin{array}{l} \sigma_1^x \equiv \sigma^x \otimes I \otimes I \otimes \dots \otimes I \otimes I, \quad \sigma_1^y \equiv \sigma^y \otimes I \otimes I \otimes \dots \otimes I \otimes I, \quad \sigma_1^z \equiv \sigma^z \otimes I \otimes I \otimes \dots \otimes I \otimes I, \\ \sigma_2^x \equiv I \otimes \sigma^x \otimes I \otimes \dots \otimes I \otimes I, \quad \sigma_2^y \equiv I \otimes \sigma^y \otimes I \otimes \dots \otimes I \otimes I, \quad \sigma_2^z \equiv I \otimes \sigma^z \otimes I \otimes \dots \otimes I \otimes I, \\ \vdots \\ \sigma_{|V|}^x \equiv I \otimes I \otimes I \otimes \dots \otimes I \otimes \sigma^x, \quad \sigma_{|V|}^y \equiv I \otimes I \otimes I \otimes \dots \otimes I \otimes \sigma^y, \quad \sigma_{|V|}^z \equiv I \otimes I \otimes I \otimes \dots \otimes I \otimes \sigma^z. \end{array} \right. \quad (10.256)$$

The vertical and horizontal N -dimensional state vectors are defined by

$$|s_1, s_2, \dots, s_{|V|}\rangle \equiv |s_1\rangle \otimes |s_2\rangle \otimes \dots \otimes |s_{|V|}\rangle \quad (s_1 \in \Omega, s_2 \in \Omega, \dots, s_{|V|} \in \Omega), \quad (10.257)$$

$$\langle s_1, s_2, \dots, s_{|V|}| \equiv \langle s_1| \otimes \langle s_2| \otimes \dots \otimes \langle s_{|V|}| \quad (s_1 \in \Omega, s_2 \in \Omega, \dots, s_{|V|} \in \Omega), \quad (10.258)$$

where

$$\langle +1| \equiv (1, 0), \quad \langle -1| \equiv (0, 1), \quad | +1\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad | -1\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (10.259)$$

By using the state vector representations, $(s_1, s_2, \dots, s_{|V|}|s'_1, s'_2, \dots, s'_{|V|})$ -elements of σ_i^x , σ_j^z and $\sigma_i^z \sigma_j^z$ are given as

$$\begin{aligned} \langle s_1, s_2, \dots, s_{|V|}| \sigma_i^x |s'_1, s'_2, \dots, s'_{|V|}\rangle &= \left(\prod_{k \in V \setminus \{i\}} \delta_{s_k, s'_k} \right) \langle s_i | \sigma^x |s'_i\rangle \\ &(s_1 \in \Omega, s_2 \in \Omega, \dots, s_{|V|} \in \Omega; s'_1 \in \Omega, s'_2 \in \Omega, \dots, s'_{|V|} \in \Omega), \end{aligned} \quad (10.260)$$

$$\begin{aligned} \langle s_1, s_2, \dots, s_{|V|} | \sigma_i^z | s'_1, s'_2, \dots, s'_{|V|} \rangle &= \left(\prod_{k \in V \setminus \{i\}} \delta_{s_k, s'_k} \right) \langle s_i | \sigma_i^z | s'_i \rangle \\ (s_1 \in \Omega, s_2 \in \Omega, \dots, s_{|V|} \in \Omega; s'_1 \in \Omega, s'_2 \in \Omega, \dots, s'_{|V|} \in \Omega), \end{aligned} \quad (10.261)$$

$$\begin{aligned} \langle s_1, s_2, \dots, s_{|V|} | \sigma_i^z \sigma_j^z | s'_1, s'_2, \dots, s'_{|V|} \rangle &= \left(\prod_{k \in V \setminus \{i, j\}} \delta_{s_k, s'_k} \right) \langle s_i, s_j | (\sigma_i^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma_j^z) | s'_i, s'_j \rangle \\ (s_1 \in \Omega, s_2 \in \Omega, \dots, s_{|V|} \in \Omega; s'_1 \in \Omega, s'_2 \in \Omega, \dots, s'_{|V|} \in \Omega). \end{aligned} \quad (10.262)$$

The prior density matrix $\mathbf{P}(\alpha, \gamma)$ and the data generative density matrix $\mathbf{P}(\mathbf{d}|\beta)$ for a given data vector \mathbf{d} are assumed to be

$$\mathbf{P}(\alpha, \gamma) = \frac{\exp\left(-\frac{1}{2}\alpha \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 + \gamma \sum_{i \in V} \sigma_i^x\right)}{\text{Tr}\left[\exp\left(-\alpha \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 + \gamma \sum_{i \in V} \sigma_i^x\right)\right]}, \quad (10.263)$$

$$\mathbf{P}(\mathbf{d}|\beta) = \left(\sqrt{\frac{\beta}{2\pi}}\right)^{|V|} \exp\left(-\frac{1}{2}\beta \sum_{i \in V} (d_i \mathbf{I}^{(2^{|V|})} - \sigma_i^z)^2\right), \quad (10.264)$$

where α , β , and γ are hyperparameters. The data generative density matrix $\mathbf{P}(\mathbf{d}|\beta)$ is expressed as a $|\Omega|^{|V|} \times |\Omega|^{|V|}$ diagonal matrix in which all the off-diagonal elements are zero. Each diagonal element $\langle s_1, s_2, \dots, s_{|V|} | \mathbf{P}(\mathbf{d}|\beta) | s_1, s_2, \dots, s_{|V|} \rangle$ ($(s_1, s_2, \dots, s_{|V|})^T \in \Omega^{|V|}$) corresponds to the probability of the data vector \mathbf{d} according to additive white Gaussian noise when the state vector $(s_1, s_2, \dots, s_{|V|})$ is given, and β corresponds to the inverse of variance in the additive white Gaussian noise. By considering a quantum statistical mechanical extension of the Bayes formula, a posterior density matrix $\mathbf{P}(\mathbf{d}, \alpha, \beta, \gamma)$ and a joint density matrix $\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)$ can be expressed as follows:

$$\begin{aligned} \mathbf{P}(\mathbf{d}, \alpha, \beta, \gamma) &\equiv \frac{\exp\left(\ln(\mathbf{P}(\mathbf{d}|\beta)) + \ln(\mathbf{P}(\alpha, \gamma))\right)}{\text{Tr}\left[\exp\left(\ln(\mathbf{P}(\mathbf{d}|\beta)) + \ln(\mathbf{P}(\alpha, \gamma))\right)\right]} \\ &= \frac{\exp\left(-\frac{1}{2}\alpha \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 - \frac{1}{2}\beta \sum_{i \in V} (d_i \mathbf{I}^{(2^{|V|})} - \sigma_i^z)^2 + \gamma \sum_{i \in V} \sigma_i^x\right)}{\text{Tr}\left[\exp\left(-\frac{1}{2}\alpha \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 - \frac{1}{2}\beta \sum_{i \in V} (d_i \mathbf{I}^{(2^{|V|})} - \sigma_i^z)^2 + \gamma \sum_{i \in V} \sigma_i^x\right)\right]}, \end{aligned} \quad (10.265)$$

$$\begin{aligned}
\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma) &\equiv \exp\left(\ln(\mathbf{P}(\mathbf{d}|\beta)) + \ln(\mathbf{P}(\alpha, \gamma))\right) \\
&= \frac{\exp\left(-\frac{1}{2}\alpha \sum_{(i,j) \in E} (\sigma_i^z - \sigma_j^z)^2 + \frac{1}{2}\beta \sum_{i \in V} (d_i \mathbf{I}^{(2|V|)} - \sigma_i^z)^2 + \gamma \sum_{i \in V} \sigma_i^x\right)}{\left(\sqrt{\frac{2\pi}{\beta}}\right)^{|V|} \text{Tr}\left[\exp\left(-\frac{1}{2}\alpha \sum_{(i,j) \in E} (\sigma_i^z - \sigma_j^z)^2 + \gamma \sum_{i \in V} \sigma_i^x\right)\right]}.
\end{aligned} \tag{10.266}$$

The estimates of the states of the hyperparameters $(\widehat{\alpha}(\mathbf{d}), \widehat{\beta}(\mathbf{d}), \widehat{\gamma}(\mathbf{d}))$ are found that maximize the marginal likelihood $\text{Tr}[\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)]$ as follows:

$$(\widehat{\alpha}(\mathbf{d}), \widehat{\beta}(\mathbf{d}), \widehat{\gamma}(\mathbf{d})) \equiv \arg \max_{(\alpha, \beta, \gamma)} \text{Tr}[\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)]. \tag{10.267}$$

To achieve the estimation criteria for hyperparameters α, β , and γ in Eqs. (10.267), we extend the \mathcal{Q} -function in Eq. (10.8) to the following expression from a quantum statistical mechanical point of view:

$$\mathcal{Q}(\alpha, \beta, \gamma | \alpha', \beta', \gamma', \mathbf{d}) \equiv \text{Tr}[\mathbf{P}(\mathbf{d}, \alpha', \beta', \gamma') \ln(\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma))] \tag{10.268}$$

The quantum EM algorithm can be summarized as a procedure consisting of the following **E-** and **M-step** which are repeated for $t = 0, 1, 2, \dots$ until $\widehat{\alpha}$ and $\widehat{\beta}$ converge:

- E-step:** Compute $\mathcal{Q}(\alpha, \beta | \alpha(t), \beta(t), \mathbf{d})$ for various values of α and β .
M-step: Determine $(\alpha(t+1), \beta(t+1))$ so as to satisfy the extremum of conditions of $\mathcal{Q}(\alpha, \beta | \alpha(t), \beta(t), \mathbf{d})$ with respect to α and β . Update $\widehat{\alpha} \leftarrow \alpha(t+1)$ and $\widehat{\beta} \leftarrow \beta(t+1)$.

The quantum EM algorithm can obtain the solution of the extremum condition in the marginal likelihood $\text{Tr}[\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)]$, because we have the following equalities:

$$\begin{cases} \left[\frac{\partial}{\partial \alpha} \mathcal{Q}(\alpha, \beta, \gamma | \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \mathbf{d}) \right]_{(\alpha, \beta, \gamma) = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})} = \left[\frac{\partial}{\partial \alpha} \ln(\text{Tr}[\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)]) \right]_{(\alpha, \beta, \gamma) = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})}, \\ \left[\frac{\partial}{\partial \beta} \mathcal{Q}(\alpha, \beta, \gamma | \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \mathbf{d}) \right]_{(\alpha, \beta, \gamma) = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})} = \left[\frac{\partial}{\partial \beta} \ln(\text{Tr}[\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)]) \right]_{(\alpha, \beta, \gamma) = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})}, \\ \left[\frac{\partial}{\partial \gamma} \mathcal{Q}(\alpha, \beta, \gamma | \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \mathbf{d}) \right]_{(\alpha, \beta, \gamma) = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})} = \left[\frac{\partial}{\partial \alpha} \ln(\text{Tr}[\mathbf{P}(\mathbf{d}|\alpha, \beta, \gamma)]) \right]_{(\alpha, \beta, \gamma) = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})}. \end{cases} \tag{10.269}$$

By substituting Eq. (10.265) into Eq. (10.268), the \mathcal{Q} -function can be rewritten as follows:

$$\mathcal{Q}(\alpha, \beta, \gamma | \alpha', \beta', \gamma', \mathbf{d}) = -\frac{1}{2}\alpha \sum_{(i,j) \in E} \text{Tr}(\sigma^z \otimes \mathbf{I} - \mathbf{I} \otimes \sigma^z)^2 \mathbf{P}_{ij}(\mathbf{d}, \alpha', \beta', \gamma')$$

$$\begin{aligned}
& -\frac{1}{2}\beta \sum_{i \in V} \text{Tr}[d_i \mathbf{I} - \boldsymbol{\sigma}^z]^2 \mathbf{P}_i(\mathbf{d}, \alpha', \beta', \gamma') \\
& + \gamma \sum_{i \in V} \text{Tr}[\boldsymbol{\sigma}_i^x \mathbf{P}_i(\mathbf{d}, \alpha', \beta', \gamma')] \\
& + |V| \ln \left(\sqrt{\frac{2\pi}{\beta}} \right) \\
& + \ln \left(\text{Tr} \left[\exp \left(-\frac{1}{2} \alpha \sum_{\{i,j\} \in E} (\boldsymbol{\sigma}^z \otimes \mathbf{I} - \mathbf{I} \otimes \boldsymbol{\sigma}^z)^2 + \gamma \sum_{i \in V} \boldsymbol{\sigma}_i^x \right) \right] \right).
\end{aligned} \tag{10.270}$$

The extremum conditions of $\mathcal{Q}(\alpha, \beta, \gamma | \alpha(t), \beta(t), \gamma(t), \mathbf{d})$ with respect to α , β and γ , such that,

$$\begin{cases} \frac{\partial}{\partial \alpha} \mathcal{Q}(\alpha, \beta, \gamma | \alpha(t), \beta(t), \gamma(t), \mathbf{d}) = 0, \\ \frac{\partial}{\partial \beta} \mathcal{Q}(\alpha, \beta, \gamma | \alpha(t), \beta(t), \gamma(t), \mathbf{d}) = 0, \\ \frac{\partial}{\partial \gamma} \mathcal{Q}(\alpha, \beta, \gamma | \alpha(t), \beta(t), \gamma(t), \mathbf{d}) = 0 \end{cases} \tag{10.271}$$

can be reduced to the following simultaneous update rules in the quantum EM algorithm:

$$\begin{aligned}
& \sum_{\{i,j\} \in E} \text{Tr} \left[(\boldsymbol{\sigma}^z \otimes \mathbf{I} - \mathbf{I} \otimes \boldsymbol{\sigma}^z)^2 \mathbf{P}_{ij}(\alpha(t+1), \gamma(t+1)) \right] \\
& = \sum_{\{i,j\} \in E} \text{Tr} \left[(\boldsymbol{\sigma}^z \otimes \mathbf{I} - \mathbf{I} \otimes \boldsymbol{\sigma}^z)^2 \mathbf{P}_{ij}(\mathbf{d}, \alpha(t), \beta(t), \gamma(t)) \right],
\end{aligned} \tag{10.272}$$

$$\frac{1}{\beta(t+1)} = \sum_{i \in V} \text{Tr} \left[(d_i \mathbf{I} - \boldsymbol{\sigma}^z)^2 \mathbf{P}_i(\mathbf{d}, \alpha(t), \beta(t), \gamma(t)) \right], \tag{10.273}$$

$$\sum_{i \in V} \text{Tr}[\boldsymbol{\sigma}_i^x \mathbf{P}_i(\alpha(t+1), \gamma(t+1))] = \sum_{i \in V} \text{Tr}[\boldsymbol{\sigma}_i^x \mathbf{P}_i(\mathbf{d}, \alpha(t), \beta(t), \gamma(t))], \tag{10.274}$$

where

$$\begin{aligned}
\langle s_i | \mathbf{P}_i(\mathbf{d}, \alpha, \beta, \gamma) | s'_i \rangle & = \langle s_i | \text{Tr}_i \mathbf{P}(\mathbf{d}, \alpha, \beta, \gamma) | s'_i \rangle \\
& \equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s'_i, \tau'_i} \\
& \quad \times \left(\prod_{j \in V \setminus \{i\}} \delta_{\tau_j, \tau'_j} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\mathbf{d}, \alpha, \beta, \gamma) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\
& \quad (s_i \in \Omega, s'_i \in \Omega, i \in V),
\end{aligned} \tag{10.275}$$

$$\begin{aligned}
\langle s_i, s_j | \mathbf{P}_{ij}(\mathbf{d}, \alpha, \beta, \gamma) | s'_i, s'_j \rangle &= \langle s_i, s_j | \mathbf{P}_{ji}(\mathbf{d}, \alpha, \beta, \gamma) | s'_i, s'_j \rangle \\
&= \langle s_i, s_j | \text{Tr}_{\setminus\{i,j\}} \mathbf{P}(\mathbf{d}, \alpha, \beta, \gamma) | s'_i, s'_j \rangle \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} \delta_{s'_i, \tau'_i} \delta_{s'_j, \tau'_j} \\
&\quad \times \left(\prod_{k \in V \setminus \{i,j\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\mathbf{d}, \alpha, \beta, \gamma) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\
&\quad (s_i \in \Omega, s_j \in \Omega, s'_i \in \Omega, s'_j \in \Omega, i \in V, j \in V, i < j), \quad (10.276)
\end{aligned}$$

$$\begin{aligned}
\langle s_i, s_j | \mathbf{P}_{ij}(\alpha, \gamma) | s'_i, s'_j \rangle &= \langle s_i, s_j | \mathbf{P}_{ji}(\alpha, \gamma) | s'_i, s'_j \rangle \\
&= \langle s_i, s_j | \text{Tr}_{\setminus\{i,j\}} \mathbf{P}(\alpha, \gamma) | s'_i, s'_j \rangle \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} \delta_{s'_i, \tau'_i} \delta_{s'_j, \tau'_j} \\
&\quad \times \left(\prod_{k \in V \setminus \{i,j\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\alpha, \gamma) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\
&\quad (s_i \in \Omega, s_j \in \Omega, s'_i \in \Omega, s'_j \in \Omega, i \in V, j \in V, i < j). \quad (10.277)
\end{aligned}$$

Finally, we explain how the state at each node is estimated from the reduced posterior density matrix $\mathbf{P}_i(\mathbf{d}, \alpha, \beta, \gamma)$ in Eq. (10.276) for each node $i \in V$. The reduced posterior density matrix $\mathbf{P}_i(\mathbf{d}, \alpha, \beta, \gamma)$ is a real symmetric matrix and can be diagonalized as

$$\begin{aligned}
\mathbf{P}_i(\mathbf{d}, \alpha, \beta, \gamma) &= \begin{pmatrix} \psi_i^{(1)}(+1|\mathbf{d}, \alpha, \beta, \gamma) & \psi_i^{(2)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(1)}(-1|\mathbf{d}, \alpha, \beta, \gamma) & \psi_i^{(2)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix} \\
&\quad \times \begin{pmatrix} P_i^{(1)}(\mathbf{d}, \alpha, \beta, \gamma) & 0 \\ 0 & P_i^{(2)}(\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix} \\
&\quad \times \begin{pmatrix} \psi_i^{(1)}(+1|\mathbf{d}, \alpha, \beta, \gamma) & \psi_i^{(2)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(1)}(-1|\mathbf{d}, \alpha, \beta, \gamma) & \psi_i^{(2)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix}^T, \quad (10.278)
\end{aligned}$$

where the eigenvalues, $P_i^{(1)}(\mathbf{d}, \alpha, \beta, \gamma)$ and $P_i^{(2)}(\mathbf{d}, \alpha, \beta, \gamma)$, are always real numbers. The vectors $\begin{pmatrix} \psi_i^{(1)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(1)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix}$ and $\begin{pmatrix} \psi_i^{(2)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(2)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix}$ correspond to the eigenvectors for the eigenvalues $P_i^{(1)}(\mathbf{d}, \alpha, \beta, \gamma)$ and $P_i^{(2)}(\mathbf{d}, \alpha, \beta, \gamma)$, such that

$$P_i(\mathbf{d}, \alpha, \beta, \gamma) \begin{pmatrix} \psi_i^{(n)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(n)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix} = P_i(n|\mathbf{d}, \alpha, \beta, \gamma) \begin{pmatrix} \psi_i^{(n)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(n)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix} \quad (i \in V, n \in \{1, 2\}). \quad (10.279)$$

This means that the eigenvectors correspond to all possible states and probabilities of the states $\begin{pmatrix} \psi_i^{(1)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(1)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix}$ and $\begin{pmatrix} \psi_i^{(2)}(+1|\mathbf{d}, \alpha, \beta, \gamma) \\ \psi_i^{(2)}(-1|\mathbf{d}, \alpha, \beta, \gamma) \end{pmatrix}$ are $P_i^{(2)}(\mathbf{d}, \alpha, \beta, \gamma)$ and $P_i^{(1)}(\mathbf{d}, \alpha, \beta, \gamma)$, respectively, in the reduced density matrix $P_i(\mathbf{d}, \alpha, \beta, \gamma)$. The estimates for the state at each node $i \in V$, $\begin{pmatrix} \widehat{\psi}_i(+1|\mathbf{d}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) \\ \widehat{\psi}_i(-1|\mathbf{d}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) \end{pmatrix}$, are given by

$$\begin{pmatrix} \widehat{\psi}_i(+1|\mathbf{d}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) \\ \widehat{\psi}_i(-1|\mathbf{d}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) \end{pmatrix} \equiv \operatorname{argmax} P_i(\mathbf{d}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) \quad (i \in V). \quad (10.280)$$

These estimation criteria in Eqs. (10.267) and (10.280) correspond to quantum statistical mechanical extensions of the maximizations of marginal likelihood and posterior marginal.

10.4.5 Quantum Expectation-Maximization (EM) Algorithm for Probabilistic Image Segmentation

This section applies the framework of Sect. 10.4.5 to the EM algorithm for probabilistic image segmentations in Sect. 10.2.3. In our present framework, **Hubbard Operators** [85] are used instead of Pauli spin matrices.

First, we introduce Hubbard operators $X_i^{\tau, \tau'}$ at each node $i \in V$ as follows:

$$\begin{cases} X_1^{(\tau, \tau')} \equiv X^{(\tau, \tau')} \otimes I \otimes I \otimes \dots \otimes I \otimes I, \\ X_2^{(\tau, \tau')} \equiv I \otimes X^{(\tau, \tau')} \otimes I \otimes \dots \otimes I \otimes I, \\ \vdots \\ X_{|V|}^{(\tau, \tau')} \equiv I \otimes I \otimes I \otimes \dots \otimes I \otimes X_{|V|}^{(\tau, \tau')}, \end{cases} \quad (\tau \in \Omega, \tau' \in \Omega), \quad (10.281)$$

where

$$X^{(+1, +1)} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad X^{(+1, -1)} \equiv \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad X^{(-1, +1)} \equiv \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X^{(-1, -1)} \equiv \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (10.282)$$

In probabilistic segmentation and clustering, $\rho(\mathbf{D}|s, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ in Eq. (10.29) and $P(s|\alpha)$ in Eq. (10.30) correspond to the data generative and prior models, respectively. By using the Hubbard operators and extending Eq. (10.29) and Eq. (10.30) from the standpoint of quantum statistical mechanical informatics,

the density matrices of the data generative model and the prior model in quantum machine learning systems for probabilistic image processing can be expressed as follows:

$$\begin{aligned}
& \mathbf{R}(\mathbf{D}|\mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \\
&= \prod_{i \in V} \sum_{s_i \in \Omega} \mathbf{X}_i^{(s_i, s_i)} \sqrt{\frac{1}{\det(2\pi \mathbf{C}(s_i))}} \exp\left(-\frac{1}{2}(\mathbf{d}_i - \mathbf{a}(s_i))\mathbf{C}^{-1}(s_i)(\mathbf{d}_i - \mathbf{a}(s_i))^T\right) \\
&= \exp\left(-\frac{1}{2} \sum_{i \in V} \sum_{s_i \in \Omega} \left((\mathbf{d}_i - \mathbf{a}(s_i))\mathbf{C}^{-1}(s_i)(\mathbf{d}_i - \mathbf{a}(s_i))^T + \ln(\det(2\pi \mathbf{C}(s_i)))\right) \mathbf{X}_i^{(s_i, s_i)}\right),
\end{aligned} \tag{10.283}$$

$$\mathbf{R}(\alpha, \gamma) = \frac{\exp\left(-2\alpha \sum_{(i,j) \in E} (I^{(2|V)} - X_i^{(+1,+1)} X_j^{(+1,+1)} - X_i^{(-1,-1)} X_j^{(-1,-1)}) + \gamma \sum_{i \in V} (X_i^{(+1,-1)} + X_i^{(-1,+1)})\right)}{\text{Tr}\left[\exp\left(-2\alpha \sum_{(i,j) \in E} (I^{(2|V)} - X_i^{(+1,+1)} X_j^{(+1,+1)} - X_i^{(-1,-1)} X_j^{(-1,-1)}) + \gamma \sum_{i \in V} (X_i^{(+1,-1)} + X_i^{(-1,+1)})\right)\right]}. \tag{10.284}$$

where

$$\mathbf{a}(+1) = \begin{pmatrix} a_R(+1) \\ a_G(+1) \\ a_B(+1) \end{pmatrix}, \quad \mathbf{a}(-1) = \begin{pmatrix} a_R(-1) \\ a_G(-1) \\ a_B(-1) \end{pmatrix}, \tag{10.285}$$

$$\mathbf{C}(+1) = \begin{pmatrix} C_{RR}(+1) & C_{RG}(+1) & C_{RB}(+1) \\ C_{GR}(+1) & C_{GG}(+1) & C_{GB}(+1) \\ C_{BR}(+1) & C_{BG}(+1) & C_{BB}(+1) \end{pmatrix}, \quad \mathbf{C}(-1) = \begin{pmatrix} C_{RR}(-1) & C_{RG}(-1) & C_{RB}(-1) \\ C_{GR}(-1) & C_{GG}(-1) & C_{GB}(-1) \\ C_{BR}(-1) & C_{BG}(-1) & C_{BB}(-1) \end{pmatrix}. \tag{10.286}$$

The joint density matrix of \mathbf{s} and \mathbf{D} is expressed in terms of the data generative and prior density matrix as follows:

$$\begin{aligned}
& \mathbf{P}(\mathbf{D}|\alpha, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \\
& \equiv \exp(\ln(\mathbf{P}(\mathbf{D}|\mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))) + \ln(\mathbf{P}(\alpha, \gamma))).
\end{aligned} \tag{10.287}$$

By using the joint density matrix $\mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$, the posterior density matrix $\mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))$ is defined by using Bayes formulas as follows:

$$\mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) \equiv \frac{\mathbf{P}(\mathbf{D}|\alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))}{\mathbf{P}(\mathbf{D}|\alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1))}, \tag{10.288}$$

Estimates of the hyperparameters and parameter vector, $\widehat{\alpha}(\mathbf{D})$, $\widehat{\gamma}(\mathbf{D})$, $\widehat{\mathbf{a}}(+1|\mathbf{D})$, $\widehat{\mathbf{a}}(-1|\mathbf{D})$, $\widehat{\mathbf{C}}(+1|\mathbf{D})$, $\widehat{\mathbf{C}}(-1|\mathbf{D})$, are given by

$$\begin{aligned} & \left(\widehat{\alpha}(\mathbf{D}), \widehat{\gamma}(\mathbf{D}), \widehat{\mathbf{a}}(+1|\mathbf{D}), \widehat{\mathbf{a}}(-1|\mathbf{D}), \widehat{\mathbf{C}}(+1|\mathbf{D}), \widehat{\mathbf{C}}(-1|\mathbf{D}) \right) \\ & = \arg \max_{(\alpha, \gamma, \boldsymbol{\mu}^{(+1)}, \boldsymbol{\mu}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)})} \text{Tr} \left[\mathbf{P}(\mathbf{D}|\alpha, \gamma, \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)}) \right], \end{aligned} \quad (10.289)$$

The parameter vector $\widehat{\mathbf{s}}(\mathbf{D}) = (\widehat{s}_1(\mathbf{D}), \widehat{s}_2(\mathbf{D}), \dots, \widehat{s}_{|V|}(\mathbf{D}))$ can be estimated from the reduced posterior marginal density matrix at each node i of $\mathbf{P}(\mathbf{D}, \alpha, \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)})$ by similar arguments to those for Eqs. (10.278), (10.279), and (10.280).

The \mathcal{Q} -function for the EM algorithm in the present framework is defined by

$$\begin{aligned} & \mathcal{Q}(\alpha', \gamma', \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)} | \alpha', \gamma', \mathbf{a}'^{(+1)}, \mathbf{a}'^{(-1)}, \mathbf{C}'^{(+1)}, \mathbf{C}'^{(-1)}, \mathbf{D}) \\ & \equiv \text{Tr} \left[\mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}'^{(+1)}, \mathbf{a}'^{(-1)}, \mathbf{C}'^{(+1)}, \mathbf{C}'^{(-1)}) \right. \\ & \quad \left. \times \ln(\mathbf{P}(\mathbf{D} | \alpha, \gamma, \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)})) \right]. \end{aligned} \quad (10.290)$$

The EM algorithm is a procedure that performs the following **E-** and **M-step** repeatedly for $t = 0, 1, 2, \dots$ until $\widehat{\alpha}(\mathbf{D})$, $\widehat{\mathbf{a}}(+1, \mathbf{D})$, $\widehat{\mathbf{a}}(-1, \mathbf{D})$, $\widehat{\mathbf{C}}(+1, \mathbf{D})$, $\widehat{\mathbf{C}}(-1, \mathbf{D})$ converge:

E-step: Compute $\mathcal{Q}_{(\alpha, \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)})} | \alpha(t), \mathbf{a}^{(+1, t)}, \mathbf{a}^{(-1, t)}, \mathbf{C}^{(+1, t)}, \mathbf{C}^{(-1, t)}$ for various values of $\mathbf{a}^{(+1)}$, $\mathbf{a}^{(-1)}$, $\mathbf{C}^{(+1)}$ and $\mathbf{C}^{(-1)}$.

M-step: Determine $\alpha(t+1)$, $\mathbf{a}^{(+1, t+1)}$, $\mathbf{a}^{(-1, t+1)}$, $\mathbf{C}^{(+1, t+1)}$ and $\mathbf{C}^{(-1, t+1)}$ so as to satisfy the extremum conditions of \mathcal{Q} -function with respect to $\mathbf{a}^{(+1)}$, $\mathbf{a}^{(-1)}$, $\mathbf{C}^{(+1)}$ and $\mathbf{C}^{(-1)}$ as follows:

$$\begin{aligned} & (\alpha(t+1), \mathbf{a}^{(+1, t+1)}, \mathbf{a}^{(-1, t+1)}, \mathbf{C}^{(+1, t+1)}, \mathbf{C}^{(-1, t+1)}) \\ & \leftarrow \underset{\alpha, \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)}}{\text{extremum}} \\ & \mathcal{Q}(\alpha, \mathbf{a}^{(+1)}, \mathbf{a}^{(-1)}, \mathbf{C}^{(+1)}, \mathbf{C}^{(-1)} | \alpha(t), \mathbf{a}^{(+1, t)}, \mathbf{a}^{(-1, t)}, \mathbf{C}^{(+1, t)}, \mathbf{C}^{(-1, t)}, \mathbf{D}). \end{aligned} \quad (10.291)$$

Update $\widehat{\alpha}(\mathbf{D}) \leftarrow \alpha(t+1)$, $\widehat{\mathbf{a}}(+1, \mathbf{D}) \leftarrow \mathbf{a}^{(+1, t+1)}$, $\widehat{\mathbf{a}}(-1, \mathbf{D}) \leftarrow \mathbf{a}^{(-1, t+1)}$, $\widehat{\mathbf{C}}(+1, \mathbf{D}) \leftarrow \mathbf{C}^{(+1, t+1)}$ and $\widehat{\mathbf{C}}(-1, \mathbf{D}) \leftarrow \mathbf{C}^{(-1, t+1)}$.

By using some equalities in Eqs. (10.283), (10.284), (10.287), and (10.288), the EM algorithm using the \mathcal{Q} -function can be reduced to the following expression:

$$\begin{aligned} & \frac{1}{|E|} \sum_{\{i,j\} \in E} \text{Tr} \left[\left(\mathbf{I} \otimes \mathbf{I} - \mathbf{X}^{(+1,+1)} \otimes \mathbf{X}^{(+1,+1)} - \mathbf{X}^{(+1,+1)} \otimes \mathbf{X}^{(+1,+1)} \right) \mathbf{P}_{ij}(\alpha(t+1), \gamma(t+1)) \right] \\ &= \frac{1}{|E|} \sum_{\{i,j\} \in E} \text{Tr} \left[\left(\mathbf{I} \otimes \mathbf{I} - \mathbf{X}^{(+1,+1)} \otimes \mathbf{X}^{(+1,+1)} - \mathbf{X}^{(+1,+1)} \otimes \mathbf{X}^{(+1,+1)} \right) \right. \\ & \quad \left. \times \mathbf{P}_{ij}(\mathbf{D}, \alpha(t), \gamma(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)) \right], \quad (10.292) \end{aligned}$$

$$\begin{aligned} & \frac{1}{|V|} \sum_{i \in V} \text{Tr} \left[\left(\mathbf{X}^{(+1,-1)} + \mathbf{X}^{(-1,+1)} \right) \mathbf{P}_i(\alpha(t+1), \gamma(t+1)) \right] \\ &= \frac{1}{|V|} \sum_{i \in V} \text{Tr} \left[\left(\mathbf{X}^{(+1,-1)} + \mathbf{X}^{(-1,+1)} \right) \right. \\ & \quad \left. \times \mathbf{P}_i(\mathbf{D}, \alpha(t), \gamma(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)) \right], \quad (10.293) \end{aligned}$$

$$\begin{aligned} \mu(\xi, t+1) &= \frac{\sum_{i \in V} d_i \text{Tr} \left[\mathbf{X}^{(\xi, \xi)} \mathbf{P}_i(\mathbf{D}, \alpha(t), \gamma(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)) \right]}{\sum_{i \in V} \text{Tr} \left[\mathbf{X}^{(\xi, \xi)} \mathbf{P}_i(\mathbf{D}, \alpha(t), \gamma(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)) \right]} \quad (\xi \in \Omega), \quad (10.294) \end{aligned}$$

$$\begin{aligned} & \mathbf{C}(\xi; t+1) \\ &= \frac{\sum_{i \in V} (d_i - \mathbf{a}(\xi; t))^T (d_i - \mathbf{a}(\xi; t)) \text{Tr} \left[\mathbf{X}^{(\xi, \xi')} \mathbf{P}_i(\mathbf{D}, \alpha(t), \gamma(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)) \right]}{\sum_{i \in V} \text{Tr} \left[\mathbf{X}^{(\xi, \xi')} \mathbf{P}_i(\mathbf{D}, \alpha(t), \gamma(t), \mathbf{a}(+1, t), \mathbf{a}(-1, t), \mathbf{C}(+1, t), \mathbf{C}(-1, t)) \right]} \quad (\xi \in \Omega), \quad (10.295) \end{aligned}$$

where

$$\begin{aligned} & \langle s_i | \mathbf{P}_i(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | s'_i \rangle \\ &= \langle s_i | \text{Tr}_i \mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | s'_i \rangle \\ &\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s'_i, \tau'_i} \\ & \times \left(\prod_{j \in V \setminus \{i\}} \delta_{\tau_j, \tau'_j} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\ & (s_i \in \Omega, s'_i \in \Omega, i \in V), \quad (10.296) \end{aligned}$$

$$\begin{aligned} & \langle s_i, s_j | \mathbf{P}_{ij}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | s'_i, s'_j \rangle \\ &= \langle s_i, s_j | \mathbf{P}_{ji}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | s'_i, s'_j \rangle \end{aligned}$$

$$\begin{aligned}
&= \langle s_i, s_j | \text{Tr}_{\setminus\{i,j\}} \mathbf{P}(\mathbf{D}, \alpha, \beta, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | s'_i, s'_j \rangle \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} \delta_{s'_i, \tau'_i} \delta_{s'_j, \tau'_j} \\
&\times \left(\prod_{k \in V \setminus \{i,j\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\mathbf{D}, \alpha, \gamma, \mathbf{a}(+1), \mathbf{a}(-1), \mathbf{C}(+1), \mathbf{C}(-1)) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\
&(s_i \in \Omega, s_j \in \Omega, s'_i \in \Omega, s'_j \in \Omega, i \in V, j \in V, i < j), \tag{10.297}
\end{aligned}$$

$$\begin{aligned}
\langle s_i | \mathbf{P}_i(\alpha, \gamma) | s'_i \rangle &= \langle s_i | \text{Tr}_{\setminus\{i\}} \mathbf{P}(\alpha, \gamma) | s'_i \rangle \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s'_i, \tau'_i} \\
&\times \left(\prod_{k \in V \setminus \{i\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\alpha, \gamma) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\
&(s_i \in \Omega, s_j \in \Omega, s'_i \in \Omega, s'_j \in \Omega, i \in V), \tag{10.298}
\end{aligned}$$

$$\begin{aligned}
\langle s_i, s_j | \mathbf{P}_{ij}(\alpha, \gamma) | s'_i, s'_j \rangle &= \langle s_i, s_j | \mathbf{P}_{ji}(\alpha, \gamma) | s'_i, s'_j \rangle \\
&= \langle s_i, s_j | \text{Tr}_{\setminus\{i,j\}} \mathbf{P}(\alpha, \gamma) | s'_i, s'_j \rangle \\
&\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} \delta_{s'_i, \tau'_i} \delta_{s'_j, \tau'_j} \\
&\times \left(\prod_{k \in V \setminus \{i,j\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{P}(\alpha, \gamma) | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\
&(s_i \in \Omega, s_j \in \Omega, s'_i \in \Omega, s'_j \in \Omega, i \in V, j \in V, i < j). \tag{10.299}
\end{aligned}$$

10.5 Quantum Statistical Mechanical Informatics

This section explains some quantum graphical modeling using some quantum mechanical extensions of statistical mechanical informatics, such as quantum statistical mechanical informatics, and particularly, advanced quantum mean-field methods. Fundamental frameworks and recent developments have been explored in some textbooks in statistical mechanics [37, 86]. In some applications of quantum annealing to massive optimization problems, a transverse Ising model is an important quantum probabilistic graphical model [83, 84] and it is known that the density matrices, for example, in Eqs. (10.263), (10.265), and (10.266), in some familiar quantum statistical machine learning systems can be reduced to transverse Ising models.

In quantum statistical mechanical informatics, one of most important schemes is Suzuki-Trotter decompositions [87, 88]. This was used to realize the quantum Monte Carlo methods by mapping d -dimensional density matrices to corresponding $(d + 1)$ -dimensional probability distributions [89]. Recently, some quantum annealing schemes have been realized as actual quantum computers, for example, the d -wave machine.

In the first part of this section, we explain some basic frameworks in advanced quantum mean-field methods for realizing familiar quantum statistical machine learning systems for the transverse Ising models, including conventional frameworks of quantum belief propagations. In the second part, we propose a quantum adaptive Thouless-Anderson-Palmar (TAP) method and a new approach using the momentum space renormalization group method to realize coarse graining for the transverse Ising model not only for regular graphs but also for random graphs. In the third part, we introduce Suzuki-Trotter decompositions [87, 88], and show the basic scheme for mapping a d -dimensional transverse Ising model to a $(d + 1)$ -dimensional Ising model and apply the scheme to the message passing rules of the conventional quantum belief propagation.

10.5.1 Advanced Mean-Field Methods for the Transverse Ising Model

This section explores the detailed derivation of the deterministic equations in both the quantum mean-field method and the quantum loopy belief propagation method for the transverse Ising model [83, 84]. Note that the present framework of the quantum mean-field method and the quantum loopy belief propagation method are constructed in real space, while other familiar frameworks in quantum statistical mechanics such as spin wave theory are constructed in momentum space.

For a graph (V, E) with a set of nodes V and set of edges E , we consider a density matrix \mathbf{P} as

$$\mathbf{P} = \frac{\exp\left(-\frac{1}{k_B T} \left(\frac{1}{2} J \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 + \frac{1}{2} h \sum_{i \in V} (\sigma_i^z - d_i \mathbf{I}^{(2^{|V|})})^2 - \gamma \sum_{i \in V} \sigma_i^x \right)\right)}{\text{Tr} \left[\exp\left(-\frac{1}{k_B T} \left(\frac{1}{2} J \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 + \frac{1}{2} h \sum_{i \in V} (\sigma_i^z - d_i \mathbf{I}^{(2^{|V|})})^2 - \Gamma \sum_{i \in V} \sigma_i^x \right)\right)\right]}. \quad (10.300)$$

Because $\sigma_i^z \sigma_i^z = \mathbf{I}^{(2^{|V|})}$, the density matrix in Eq. (10.300) can be reduced to Eqs. (10.226) and (10.227) with

$$\mathbf{H} = -J \sum_{\{i,j\} \in E} \sigma_i^z \sigma_j^z - h \sum_{i \in V} d_i \sigma_i^z - \Gamma \sum_{i \in V} \sigma_i^x. \quad (10.301)$$

Here, all the nodes j connected with the node i by an edge $\{i, j\}$ are referred to as neighboring nodes of the node i , and the set of all neighboring nodes of the node i is denoted by the notation ∂i . The quantum probabilistic graphical model in Eqs. (10.300) and (10.301) is referred to as the **Transverse Ising Model** [83, 84].

First, we explain the conventional quantum mean-field method for the transverse Ising model. We introduce a $2^N \times 2^N$ trial density matrix \mathbf{R} and its 2×2 trial reduced density matrix \mathbf{R}_i for each node $i (i \in V)$ defined by

$$\mathbf{R}_i = \text{Tr}_{V \setminus i} \mathbf{R} = \left(\begin{array}{cc} \langle +1 | \mathbf{R}_i | +1 \rangle & \langle +1 | \mathbf{R}_i | -1 \rangle \\ \langle -1 | \mathbf{R}_i | +1 \rangle & \langle -1 | \mathbf{R}_i | -1 \rangle \end{array} \right), \quad (10.302)$$

where

$$\begin{aligned} \langle s_i | \mathbf{R}_i | s'_i \rangle &= \langle s_i | \text{Tr}_{V \setminus i} \mathbf{R} | s'_i \rangle \\ &\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s'_i, \tau'_i} \left(\prod_{j \in V \setminus \{i\}} \delta_{\tau_j, \tau'_j} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{R} | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\ &\quad (s_i \in \Omega, s'_i \in \Omega, i \in V). \end{aligned} \quad (10.303)$$

By using Eq. (10.303), the average $\text{Tr}(\sigma_i^x \mathbf{R})$ can be expressed in terms of the reduced density matrix \mathbf{R}_i as follows:

$$\begin{aligned} \text{Tr}(\sigma_i^x \mathbf{R}) &= \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \sum_{s'_1 \in \Omega} \sum_{s'_2 \in \Omega} \cdots \sum_{s'_{|V|} \in \Omega} \langle s_1, s_2, \dots, s_{|V|} | \sigma_i^x | s'_1, s'_2, \dots, s'_{|V|} \rangle \\ &\quad \times \langle s'_1, s'_2, \dots, s'_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle \\ &= \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \sum_{s'_1 \in \Omega} \sum_{s'_2 \in \Omega} \cdots \sum_{s'_{|V|} \in \Omega} \left(\prod_{k \in V \setminus \{i\}} \delta_{s_k, s'_k} \right) \\ &\quad \times \langle s_i | \sigma^x | s'_i \rangle \langle s'_1, s'_2, \dots, s'_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle \\ &= \sum_{s_i \in \Omega} \sum_{s'_i \in \Omega} \langle s_i | \sigma^x | s'_i \rangle \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \delta_{s_i, \tau_i} \delta_{s'_i, \tau'_i} \\ &\quad \times \left(\prod_{k \in V \setminus \{i\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau'_1, \tau'_2, \dots, \tau'_{|V|} | \mathbf{R} | \tau_1, \tau_2, \dots, \tau_{|V|} \rangle \\ &= \sum_{s_i \in \Omega} \sum_{s'_i \in \Omega} \langle s_i | \sigma^x | s'_i \rangle \langle s'_i | \mathbf{R}_i | s_i \rangle \\ &= \text{Tr}(\sigma^x \mathbf{R}_i). \end{aligned} \quad (10.304)$$

By similar arguments to those for Eq. (10.304), we derive

$$\text{Tr}(\sigma_i^z \mathbf{R}) = \text{Tr}(\sigma^z \mathbf{R}_i). \quad (10.305)$$

Now, we assume that the trial density matrix \mathbf{R} is expressed as

$$\mathbf{R} = \mathbf{R}_1 \otimes \mathbf{R}_2 \otimes \cdots \otimes \mathbf{R}_{|V|}. \quad (10.306)$$

In this case, the average $\text{Tr}(\sigma_i^z \sigma_j^z \mathbf{R})$ and the entropy $-k_B \text{Tr} \mathbf{R} \ln \mathbf{R}$ can be expressed as

$$\begin{aligned} \text{Tr}(\sigma_i^z \sigma_j^z \mathbf{R}) &= \text{Tr}(\sigma_i^z \sigma_j^z (\mathbf{R}_1 \otimes \mathbf{R}_2 \otimes \cdots \otimes \mathbf{R}_{|V|})) \\ &= \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \sum_{s'_1 \in \Omega} \sum_{s'_2 \in \Omega} \cdots \sum_{s'_{|V|} \in \Omega} \sum_{s''_1 \in \Omega} \sum_{s''_2 \in \Omega} \cdots \sum_{s''_{|V|} \in \Omega} \left(\prod_{k \in V \setminus \{i\}} \delta_{s_k, s'_k} \right) \langle s_i | \sigma^z | s'_i \rangle \\ &\quad \times \left(\prod_{l \in V \setminus \{j\}} \delta_{s'_l, s''_l} \right) \langle s'_j | \sigma^z | s''_j \rangle \langle s''_1 | \mathbf{R}_1 | s_1 \rangle \langle s''_2 | \mathbf{R}_2 | s_2 \rangle \times \cdots \times \langle s''_{|V|} | \mathbf{R}_{|V|} | s_{|V|} \rangle \\ &= \left(\sum_{s_i \in \Omega} \sum_{s'_i \in \Omega} \langle s_i | \sigma^z | s'_i \rangle \langle s'_i | \mathbf{R}_i | s_i \rangle \right) \\ &\quad \times \left(\sum_{s_j \in \Omega} \sum_{s'_j \in \Omega} \langle s_j | \sigma^z | s'_j \rangle \langle s'_j | \mathbf{R}_j | s_j \rangle \right) \left(\prod_{k \in V \setminus \{i, j\}} \left(\sum_{s'_k \in \Omega} \sum_{s''_k \in \Omega} \delta_{s_k, s'_k} \delta_{s'_k, s''_k} \right) \langle s''_k | \mathbf{R}_k | s_k \rangle \right) \\ &= (\text{Tr}(\sigma^z \mathbf{R}_i)) (\text{Tr}(\sigma^z \mathbf{R}_j)) \left(\prod_{k \in V \setminus \{i, j\}} (\text{Tr}(\mathbf{R}_k)) \right) \\ &= (\text{Tr}(\sigma^z \mathbf{R}_i)) (\text{Tr}(\sigma^z \mathbf{R}_j)), \end{aligned} \quad (10.307)$$

$$\begin{aligned} -k_B \text{Tr}(\mathbf{R} \ln(\mathbf{R})) &= -\text{Tr}((\mathbf{R}_1 \otimes \mathbf{R}_2 \otimes \cdots \otimes \mathbf{R}_{|V|}) \ln(\mathbf{R}_1 \otimes \mathbf{R}_2 \otimes \cdots \otimes \mathbf{R}_{|V|})) \\ &= -k_B \sum_{i=1}^N \text{Tr}((\mathbf{R}_1 \otimes \mathbf{R}_2 \otimes \cdots \otimes \mathbf{R}_{|V|}) (\mathbf{I}^{(i-1)} \otimes \ln(\mathbf{R}_i) \otimes \mathbf{I}^{(N-i)})) \\ &= -k_B \sum_{i=1}^N \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \sum_{s'_1 \in \Omega} \sum_{s'_2 \in \Omega} \cdots \sum_{s'_{|V|} \in \Omega} \langle s_1 | \mathbf{R}_1 | s'_1 \rangle \langle s_2 | \mathbf{R}_2 | s'_2 \rangle \times \cdots \times \langle s_{|V|} | \mathbf{R}_{|V|} | s'_{|V|} \rangle \\ &\quad \times \left(\prod_{k \in V \setminus \{i\}} \delta_{s_k, s'_k} \right) \langle s'_i | \ln(\mathbf{R}_i) | s_i \rangle \\ &= -k_B \sum_{i=1}^N \left(\sum_{s_i \in \Omega} \sum_{s'_i \in \Omega} \langle s_i | \mathbf{R}_i | s'_i \rangle \langle s'_i | \ln(\mathbf{R}_i) | s_i \rangle \right) \left(\prod_{k \in V \setminus \{i\}} \left(\sum_{s_k \in \Omega} \sum_{s'_k \in \Omega} \delta_{s_k, s'_k} \langle s_k | \mathbf{R}_k | s'_k \rangle \right) \right) \\ &= -k_B \sum_{i=1}^N (\text{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i))) \left(\prod_{k \in V \setminus \{i\}} \text{Tr}(\mathbf{R}_k) \right) \\ &= -k_B \sum_{i=1}^N \text{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)). \end{aligned} \quad (10.308)$$

The free energy functional can be reduced to

$$\begin{aligned} \mathcal{F}[\mathbf{R}] = \mathcal{F}_{\text{MF}}[\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_{|V|}] \equiv & -J \sum_{(i,j) \in E} (\text{Tr}(\sigma^z \mathbf{R}_i)) (\text{Tr}(\sigma^z \mathbf{R}_j)) - h \sum_{i \in V} \text{Tr}(\sigma^z \mathbf{R}_i) \\ & - \Gamma \sum_{i \in V} \text{Tr}(\sigma^x \mathbf{R}_i) + k_B T \sum_{i \in V} \text{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)). \end{aligned} \quad (10.309)$$

We define the optimal reduced density matrix $\widehat{\mathbf{R}}_i$ for each node $i \in V$ by

$$\widehat{\mathbf{R}}_i = \arg \underset{\mathbf{R}_i}{\text{extremum}} \left\{ \mathcal{F}_{\text{MF}} \left[\widehat{\mathbf{R}}_1, \widehat{\mathbf{R}}_1, \dots, \widehat{\mathbf{R}}_{i-1}, \mathbf{R}_i, \widehat{\mathbf{R}}_{i+1}, \widehat{\mathbf{R}}_{i+2}, \dots, \widehat{\mathbf{R}}_{|V|} \right] \mid \text{Tr} \mathbf{R}_i = 1 \right\} \quad (i \in V). \quad (10.310)$$

The simultaneous self-consistent equations for reduced density matrices are expressed as

$$\widehat{\mathbf{R}}_i = \frac{1}{Z_i} \exp \left(\frac{1}{k_B T} \left(\left(J \sum_{j \in \partial i} (\text{Tr}(\sigma^z \widehat{\mathbf{R}}_j)) + h d_i \right) \sigma^z + \Gamma \sigma^x \right) \right), \quad (10.311)$$

$$Z_i \equiv \text{Tr} \left[\exp \left(\frac{1}{k_B T} \left(\left(J \sum_{j \in \partial i} (\text{Tr}(\sigma^z \widehat{\mathbf{R}}_j)) + h d_i \right) \sigma^z + \Gamma \sigma^x \right) \right) \right]. \quad (10.312)$$

From Eq. (10.311), we can derive the following simultaneous self-consistent equations for the magnetizations $\widehat{m}_i^z \equiv \text{Tr}(\sigma^z \widehat{\mathbf{R}}_i)$ ($i \in V$) and $\widehat{m}_i^x \equiv \text{Tr}(\sigma^x \widehat{\mathbf{R}}_i)$ ($i \in V$):

$$\begin{aligned} \widehat{m}_i^z &= \frac{\frac{J}{k_B T} \sum_{j \in \partial i} \widehat{m}_j^z + \frac{h}{k_B T} d_i}{\sqrt{\left(\frac{J}{k_B T} \sum_{j \in \partial i} \widehat{m}_j^z + \frac{h}{k_B T} d_i \right)^2 + \left(\frac{\Gamma}{k_B T} \right)^2}} \\ &\quad \times \tanh \left(\sqrt{\left(\frac{J}{k_B T} \sum_{j \in \partial i} \widehat{m}_j^z + \frac{h}{k_B T} d_i \right)^2 + \left(\frac{\Gamma}{k_B T} \right)^2} \right), \end{aligned} \quad (10.313)$$

$$\widehat{m}_i^x = \frac{\Gamma}{k_B T} \sqrt{\left(\frac{J}{k_B T} \sum_{j \in \partial i} \widehat{m}_j^z + \frac{h}{k_B T} d_i \right)^2 + \left(\frac{\Gamma}{k_B T} \right)^2} \times \tanh \left(\sqrt{\left(\frac{J}{k_B T} \sum_{j \in \partial i} \widehat{m}_j^z + \frac{h}{k_B T} d_i \right)^2 + \left(\frac{\Gamma}{k_B T} \right)^2} \right). \quad (10.314)$$

The mean-field free energy $\mathcal{F}_{\text{MF}}[\widehat{\mathbf{R}}_1, \widehat{\mathbf{R}}_2, \dots, \widehat{\mathbf{R}}_{|V|}]$ of the present system is expressed as

$$\begin{aligned} \mathcal{F}_{\text{MF}}[\widehat{\mathbf{R}}_1, \widehat{\mathbf{R}}_2, \dots, \widehat{\mathbf{R}}_{|V|}] &= \sum_{i \in V} (-k_B T \ln(Z_i)) \\ &= -k_B T \sum_{i \in V} \ln \left(\text{Tr} \left[\exp \left(\frac{1}{k_B T} \left(\left(\sum_{j \in \partial i} (\text{Tr}(\sigma^z \widehat{\mathbf{R}}_j)) \right) \sigma^z + \Gamma \sigma^x \right) \right) \right] \right) \\ &= -k_B T \sum_{i \in V} \ln \left(2 \cosh \left(\sqrt{\left(\frac{J}{k_B T} \sum_{j \in \partial i} \widehat{m}_j^z \right)^2 + \Gamma^2} \right) \right). \end{aligned} \quad (10.315)$$

Next, we extend the above framework for the mean-field method for the transverse Ising model to the quantum loopy belief propagation method based on the quantum cluster variation method in Ref. [90]. We introduce a $2^N \times 2^N$ trial density matrix \mathbf{R} and its 2×2 trial reduced density matrix \mathbf{R}_i for each node $i (\in V)$ defined by

$$\begin{aligned} \mathbf{R}_{ij} &= \mathbf{R}_{ji} = \text{Tr}_{\setminus\{i,j\}} \mathbf{R} \\ &= \begin{pmatrix} \langle +1, +1 | \mathbf{R}_{ij} | +1, +1 \rangle & \langle +1, +1 | \mathbf{R}_{ij} | -1, +1 \rangle & \langle +1, +1 | \mathbf{R}_{ij} | +1, -1 \rangle & \langle +1, +1 | \mathbf{R}_{ij} | -1, -1 \rangle \\ \langle +1, -1 | \mathbf{R}_{ij} | +1, +1 \rangle & \langle +1, -1 | \mathbf{R}_{ij} | -1, +1 \rangle & \langle +1, -1 | \mathbf{R}_{ij} | +1, -1 \rangle & \langle +1, -1 | \mathbf{R}_{ij} | -1, -1 \rangle \\ \langle -1, +1 | \mathbf{R}_{ij} | +1, +1 \rangle & \langle -1, +1 | \mathbf{R}_{ij} | -1, +1 \rangle & \langle -1, +1 | \mathbf{R}_{ij} | +1, -1 \rangle & \langle -1, +1 | \mathbf{R}_{ij} | -1, -1 \rangle \\ \langle -1, -1 | \mathbf{R}_{ij} | +1, +1 \rangle & \langle -1, -1 | \mathbf{R}_{ij} | -1, +1 \rangle & \langle -1, -1 | \mathbf{R}_{ij} | +1, -1 \rangle & \langle -1, -1 | \mathbf{R}_{ij} | -1, -1 \rangle \end{pmatrix} \\ &(i \in V, j \in V, i < j), \end{aligned} \quad (10.316)$$

where

$$\begin{aligned} \langle s_i, s_j | \mathbf{R}_{ij} | s'_i, s'_j \rangle &= \langle s_i, s_j | \mathbf{R}_{ji} | s'_i, s'_j \rangle \\ &= \langle s_i, s_j | \text{Tr}_{\setminus\{i,j\}} \mathbf{R} | s'_i, s'_j \rangle \\ &\equiv \sum_{\tau_1 \in \Omega} \sum_{\tau_2 \in \Omega} \cdots \sum_{\tau_{|V|} \in \Omega} \sum_{\tau'_1 \in \Omega} \sum_{\tau'_2 \in \Omega} \cdots \sum_{\tau'_{|V|} \in \Omega} \\ &\quad \times \delta_{s_i, \tau_i} \delta_{s_j, \tau_j} \delta_{s'_i, \tau'_i} \delta_{s'_j, \tau'_j} \left(\prod_{k \in V \setminus \{i,j\}} \delta_{\tau_k, \tau'_k} \right) \langle \tau_1, \tau_2, \dots, \tau_{|V|} | \mathbf{R} | \tau'_1, \tau'_2, \dots, \tau'_{|V|} \rangle \\ &(s_i \in \Omega, s_j \in \Omega, s'_i \in \Omega, s'_j \in \Omega, i \in V, j \in V, i < j). \end{aligned} \quad (10.317)$$

By similar arguments to those for Eq. (10.304), we derive

$$\mathrm{Tr}(\boldsymbol{\sigma}_i^z \boldsymbol{\sigma}_j^z \mathbf{R}) = \mathrm{Tr}((\boldsymbol{\sigma}^z \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^z) \mathbf{R}_{ij}). \quad (10.318)$$

We now assume that the free energy functional can be expressed as

$$\begin{aligned} \mathcal{F}[\mathbf{R}] &= \mathcal{F}_{\text{Bethe}}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{\{i,j\}} | \{i,j\} \in E\}] \\ &\equiv -J \sum_{\{i,j\} \in E} \mathrm{Tr}((\boldsymbol{\sigma}^z \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^z) \mathbf{R}_{\{i,j\}}) \\ &\quad - h \sum_{i \in V} d_i \mathrm{Tr}(\boldsymbol{\sigma}^z \mathbf{R}_i) - \Gamma \sum_{i \in V} \mathrm{Tr}(\boldsymbol{\sigma}^x \mathbf{R}_i) \\ &\quad + k_B T \sum_{i \in V} \mathrm{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)) \\ &\quad + k_B T \sum_{\{i,j\} \in E} (\mathrm{Tr}(\mathbf{R}_{\{i,j\}} \ln(\mathbf{R}_{\{i,j\}})) - \mathrm{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)) - \mathrm{Tr}(\mathbf{R}_j \ln(\mathbf{R}_j))) \\ &= -J \sum_{\{i,j\} \in E} \mathrm{Tr}((\boldsymbol{\sigma}^z \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^z) \mathbf{R}_{\{i,j\}}) \\ &\quad - h \sum_{i \in V} d_i \mathrm{Tr}(\boldsymbol{\sigma}^z \mathbf{R}_i) - \Gamma \sum_{i \in V} \mathrm{Tr}(\boldsymbol{\sigma}^x \mathbf{R}_i) \\ &\quad + k_B T \sum_{\{i,j\} \in E} \mathrm{Tr}(\mathbf{R}_{\{i,j\}} \ln(\mathbf{R}_{\{i,j\}})) + k_B T \sum_{i \in V} (1 - |\partial i|) \mathrm{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)). \end{aligned} \quad (10.319)$$

We define the reduced density matrix $\widehat{\mathbf{R}}_i$ for each node $i \in V$ by

$$\widehat{\mathbf{R}}_k = \arg \operatorname{extremum}_{\mathbf{R}_k} \left\{ \mathcal{F}_{\text{Bethe}} \left[\mathbf{R}_k, \{\widehat{\mathbf{R}}_i | i \in V \setminus \{k\}\}, \{\widehat{\mathbf{R}}_{\{i,j\}} | \{i,j\} \in E\} \right] \middle| \mathrm{Tr} \mathbf{R}_k = 1, \mathbf{R}_k = \mathrm{Tr}_{\setminus k} \widehat{\mathbf{R}}_{\{k,j\}} (j \in \partial k) \right\} \quad (k \in V). \quad (10.320)$$

$$\begin{aligned} \widehat{\mathbf{R}}_{\{k,l\}} &= \arg \operatorname{extremum}_{\mathbf{R}_{\{k,l\}}} \left\{ \mathcal{F}_{\text{Bethe}} \left[\mathbf{R}_{\{k,l\}}, \{\widehat{\mathbf{R}}_i | i \in V\}, \{\widehat{\mathbf{R}}_{\{i,j\}} | \{i,j\} \in E \setminus \{k,l\}\} \right] \right\} \\ &\quad \left. \mathrm{Tr} \mathbf{R}_{\{k,l\}} = 1, \widehat{\mathbf{R}}_k = \mathrm{Tr}_{\setminus \{k,l\}} \mathbf{R}_{\{k,l\}}, \widehat{\mathbf{R}}_l = \mathrm{Tr}_{\setminus \{k,l\}} \mathbf{R}_{\{k,l\}} \right\} \quad (\{k,l\} \in E). \end{aligned} \quad (10.321)$$

To ensure the constraint conditions, we introduce the Lagrange multipliers as follows:

$$\mathcal{L}[\{\mathbf{R}_i | i \in V\}, \{\mathbf{R}_{ij} | \{i,j\} \in E\}] = -J \sum_{\{i,j\} \in E} \mathrm{Tr}((\boldsymbol{\sigma}^z \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^z) \mathbf{R}_{ij})$$

$$\begin{aligned}
& -h \sum_{i \in V} d_i \text{Tr}(\sigma^z \mathbf{R}_i) - \Gamma \sum_{i \in V} \text{Tr}(\sigma^x \mathbf{R}_i) \\
& + k_B T \sum_{\{i,j\} \in E} \text{Tr}(\mathbf{R}_{ij} \ln(\mathbf{R}_{ij})) + k_B T \sum_{i \in V} (1 - |\partial i|) \text{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)). \\
& - \sum_{i \in V} \lambda_i (\text{Tr} \mathbf{R}_i - 1) - \sum_{\{i,j\} \in E} \lambda_{\{i,j\}} (\text{Tr} \mathbf{R}_{ij} - 1) \\
& - \sum_{\{i,j\} \in E} \text{Tr} \lambda_{i,ij} (\mathbf{R}_i - \text{Tr}_{\setminus i} \mathbf{R}_{ij}) - \sum_{\{i,j\} \in E} \text{Tr} \lambda_{j,ij} (\mathbf{R}_j - \text{Tr}_{\setminus j} \mathbf{R}_{ij}) \\
= & \sum_{\{i,j\} \in E} \text{Tr} \left(\mathbf{R}_{ij} \left(-J(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) + k_B \ln(\mathbf{R}_{ij}) + \lambda_{i,ij} \otimes \mathbf{I} + \mathbf{I} \otimes \lambda_{j,ij} - \lambda_{ij}(\mathbf{I} \otimes \mathbf{I}) \right) \right) \\
& + \sum_{i \in V} \text{Tr} \left(\mathbf{R}_i \left(-hd_i \sigma^z - \Gamma \sigma^x + k_B T(1 - |\partial i|) \ln(\mathbf{R}_i) - \sum_{j \in \partial i} \lambda_{i,ij} - \lambda_i \mathbf{I} \right) \right) \\
& + \sum_{i \in V} \lambda_i + \sum_{\{i,j\} \in E} \lambda_{\{i,j\}}. \tag{10.322}
\end{aligned}$$

Here we remark that $\lambda_{i,ij} = \lambda_{i,ji}$ and $\lambda_{j,ij} = \lambda_{j,ji}$ ($\{i, j\} \in E, i < j$).

We define the reduced density matrix $\widehat{\mathbf{R}}_i$ for each node $i \in V$ and $\mathbf{R}_{ij} = \mathbf{R}_{ji}$ for each edge $\{i, j\} \in E$ by

$$\widehat{\mathbf{R}}_i = \arg \text{extremum}_{\mathbf{R}_i} \left\{ \mathbf{R}_i \left(-hd_i \sigma^z - \Gamma \sigma^x - \sum_{j \in \partial i} \lambda_{i,ij} + k_B T(1 - |\partial i|) \ln(\mathbf{R}_i) - \lambda_i \mathbf{I} \right) \right\} \tag{10.323}$$

$(i \in V),$

$$\widehat{\mathbf{R}}_{ij} = \arg \text{extremum}_{\mathbf{R}_{ij}} \left\{ \mathbf{R}_{ij} \left(-J(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) + \lambda_{i,ij} \otimes \mathbf{I} + \mathbf{I} \otimes \lambda_{j,ij} + k_B T \ln(\mathbf{R}_{ij}) - \lambda_{\{i,j\}}(\mathbf{I} \otimes \mathbf{I}) \right) \right\} \tag{10.324}$$

$(\{i, j\} \in E).$

The simultaneous self-consistent equations for reduced density matrices are expressed as

$$\widehat{\mathbf{R}}_i = \exp \left(-1 + \frac{\lambda_i}{k_B T} \right) \exp \left(\frac{1}{k_B T} \left(\frac{1}{|\partial i| - 1} \right) \left(-hd_i \sigma^z - \Gamma \sigma^x - \sum_{j \in \partial i} \lambda_{i,ij} \right) \right), \tag{10.325}$$

$$\widehat{\mathbf{R}}_{ij} = \exp \left(-1 + \frac{\lambda_{\{i,j\}}}{k_B T} \right) \exp \left(\frac{1}{k_B T} (J(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) - \lambda_{i,ij} \otimes \mathbf{I} - \mathbf{I} \otimes \lambda_{j,ij}) \right), \tag{10.326}$$

$$\exp\left(1 - \frac{\lambda_i}{k_B T}\right) = \text{Tr} \left[\exp \left(\frac{1}{k_B T} \left(\frac{1}{|\partial i| - 1} \right) \left(-hd_i \sigma^z - \Gamma \sigma^x - \sum_{j \in \partial i} \lambda_{k,j} \right) \right) \right], \quad (10.327)$$

$$\exp\left(1 - \frac{\lambda_{[i,j]}}{k_B T}\right) = \text{Tr} \left[\exp \left(\frac{1}{k_B T} (J(\sigma^z \otimes I)(I \otimes \sigma^z) - \lambda_{i,j} \otimes I - I \otimes \lambda_{j,i,j}) \right) \right]. \quad (10.328)$$

By introducing the linear transformations

$$\lambda_{i,j} = \lambda_{i,ji} = -hd_i \sigma^z - \Gamma \sigma^x - \sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i}, \quad (10.329)$$

Equations (10.325) and (10.326) can be rewritten as

$$\widehat{R}_i = \frac{1}{Z_i} \exp \left(\frac{1}{k_B T} \left(hd_i \sigma^z + \Gamma \sigma^x + \sum_{k \in \partial i} \lambda_{k \rightarrow i} \right) \right), \quad (10.330)$$

$$\begin{aligned} \widehat{R}_{ij} = \frac{1}{Z_{[i,j]}} \exp \left(\frac{1}{k_B T} \left(J(\sigma^z \otimes I)(I \otimes \sigma^z) + h(d_i(\sigma^z \otimes I) + d_j(I \otimes \sigma^z)) + \Gamma(\sigma^x \otimes I + I \otimes \sigma^x) \right. \right. \\ \left. \left. + \sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i} \otimes I + \sum_{l \in \partial j \setminus \{i\}} I \otimes \lambda_{l \rightarrow j} \right) \right), \end{aligned} \quad (10.331)$$

$$Z_i = \text{Tr} \left[\exp \left(\frac{1}{k_B T} \left(hd_i \sigma^z + \Gamma \sigma^x + \sum_{k \in \partial i} \lambda_{k \rightarrow i} \right) \right) \right], \quad (10.332)$$

$$\begin{aligned} Z_{[i,j]} = \text{Tr} \left[\exp \left(\frac{1}{k_B T} \left(J(\sigma^z \otimes I)(I \otimes \sigma^z) + h(d_i(\sigma^z \otimes I) + d_j(I \otimes \sigma^z)) + \Gamma(\sigma^x \otimes I + I \otimes \sigma^x) \right. \right. \right. \\ \left. \left. \left. + \sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i} \otimes I + \sum_{l \in \partial j \setminus \{i\}} I \otimes \lambda_{l \rightarrow j} \right) \right) \right]. \end{aligned} \quad (10.333)$$

Then, by substituting Eq. (10.330) and Eq. (10.331) into

$$\mathbf{R}_i = \text{Tr}_{\setminus i} \mathbf{R}_{ij}, \quad (10.334)$$

we derive the following simultaneous self-consistent equations for the effective fields:

$$\begin{aligned} & \exp\left(\frac{1}{k_{\text{B}}T}\lambda_{j\rightarrow i} + \frac{1}{k_{\text{B}}T}\left(hd_i\sigma^z + \Gamma\sigma^x + \sum_{k\in\partial i\setminus\{j\}}\lambda_{k\rightarrow i}\right)\right) \\ &= \frac{Z_i}{Z_{(i,j)}}\text{Tr}_{\nu_i}\left[\exp\left(\frac{1}{k_{\text{B}}T}\left(J(\sigma^z\otimes\mathbf{I})(\mathbf{I}\otimes\sigma^z) + \mathbf{I}\otimes(hd_i\sigma^z + \Gamma\sigma^x + \sum_{l\in\partial j\setminus\{i\}}\lambda_{l\rightarrow j})\right)\right.\right. \\ & \quad \left.\left.+ \frac{1}{k_{\text{B}}T}\left(hd_i\sigma^z + \Gamma\sigma^x + \sum_{k\in\partial i\setminus\{j\}}\lambda_{k\rightarrow i}\right)\otimes\mathbf{I}\right)\right], \end{aligned} \quad (10.335)$$

such that

$$\begin{aligned} \frac{1}{k_{\text{B}}T}\lambda_{j\rightarrow i} &= -\frac{1}{k_{\text{B}}T}\left(hd_i\sigma^z + \Gamma\sigma^x + \sum_{k\in\partial i\setminus\{j\}}\lambda_{k\rightarrow i}\right) \\ &+ \ln\left(\frac{Z_i}{Z_{(i,j)}}\text{Tr}_{\nu_i}\left[\exp\left(\frac{1}{k_{\text{B}}T}\left(J(\sigma^z\otimes\mathbf{I})(\mathbf{I}\otimes\sigma^z) + \mathbf{I}\otimes(hd_i\sigma^z + \Gamma\sigma^x + \sum_{l\in\partial j\setminus\{i\}}\lambda_{l\rightarrow j})\right)\right.\right.\right. \\ & \quad \left.\left.\left.+ \frac{1}{k_{\text{B}}T}\left(hd_i\sigma^z + \Gamma\sigma^x + \sum_{k\in\partial i\setminus\{j\}}\lambda_{k\rightarrow i}\right)\otimes\mathbf{I}\right)\right]\right). \end{aligned} \quad (10.336)$$

Note that Eqs. (10.335) and (10.336) can be regarded as conventional message passing rule quantum loopy belief propagation. The Bethe free energy $\mathcal{F}_{\text{Bethe}}\left[\left\{\widehat{\mathbf{R}}_i|i\in V\right\},\left\{\widehat{\mathbf{R}}_{ij}|i,j\in E\right\}\right]$ of the present system is given by

$$\begin{aligned} \mathcal{F}_{\text{Bethe}}\left[\left\{\widehat{\mathbf{R}}_i|i\in V\right\},\left\{\widehat{\mathbf{R}}_{ij}|i,j\in E\right\}\right] &= \sum_{i\in V}(-k_{\text{B}}T\ln(Z_i)) \\ &+ \sum_{\{i,j\}\in E}(-k_{\text{B}}T\ln(Z_{i,j}) + k_{\text{B}}T\ln(Z_i) + k_{\text{B}}T\ln(Z_j)). \end{aligned} \quad (10.337)$$

The conventional quantum message passing rules in Eqs. (10.335) and (10.336) reduce to Eqs. (10.95) and (10.96) for the case of $\Gamma = 0$.

Because we have the orthonormal relationships

$$\begin{cases} \text{Tr}[\sigma^z\sigma^z] = \text{Tr}[\sigma^x\sigma^x] = 2, \\ \text{Tr}[\sigma^z\mathbf{I}] = \text{Tr}[\mathbf{I}\sigma^z] = \text{Tr}[\sigma^x\mathbf{I}] = \text{Tr}[\mathbf{I}\sigma^x] = 0, \\ \text{Tr}[\sigma^z\sigma^x] = \text{Tr}[\sigma^x\sigma^z] = 0, \end{cases} \quad (10.338)$$

$$\begin{cases} \text{Tr}[(\sigma^z\otimes\mathbf{I})(\sigma^z\otimes\mathbf{I})] = \text{Tr}[(\sigma^x\otimes\mathbf{I})(\sigma^x\otimes\mathbf{I})] = \text{Tr}[(\mathbf{I}\otimes\sigma^z)(\mathbf{I}\otimes\sigma^z)] = \text{Tr}[(\mathbf{I}\otimes\sigma^x)(\mathbf{I}\otimes\sigma^x)] = 4, \\ \text{Tr}[(\sigma^z\otimes\mathbf{I})(\sigma^x\otimes\mathbf{I})] = \text{Tr}[(\sigma^x\otimes\mathbf{I})(\sigma^z\otimes\mathbf{I})] = \text{Tr}[(\mathbf{I}\otimes\sigma^x)(\mathbf{I}\otimes\sigma^z)] = \text{Tr}[(\mathbf{I}\otimes\sigma^z)(\mathbf{I}\otimes\sigma^x)] = 0, \\ \text{Tr}[(\sigma^z\otimes\mathbf{I})(\mathbf{I}\otimes\sigma^z)] = \text{Tr}[(\sigma^x\otimes\mathbf{I})(\mathbf{I}\otimes\sigma^x)] = \text{Tr}[(\sigma^z\otimes\mathbf{I})(\mathbf{I}\otimes\sigma^x)] = \text{Tr}[(\sigma^x\otimes\mathbf{I})(\mathbf{I}\otimes\sigma^z)] = 0, \end{cases} \quad (10.339)$$

the reduced density matrices \mathbf{R}_i and $\mathbf{R}_{ij} = \mathbf{R}_{ji}$ expand to the following orthonormal expansions:

$$\mathbf{R}_i = \frac{1}{2}(\mathbf{I} + m_i^x \boldsymbol{\sigma}^x + m_i^z \boldsymbol{\sigma}^z), \quad (10.340)$$

$$\begin{aligned} \mathbf{R}_{ij} &= \mathbf{R}_{ji} \\ &= \frac{1}{4} \left((\mathbf{I} \otimes \mathbf{I}) + m_i^x (\boldsymbol{\sigma}^x \otimes \mathbf{I}) + m_i^z (\boldsymbol{\sigma}^z \otimes \mathbf{I}) + m_j^x (\mathbf{I} \otimes \boldsymbol{\sigma}^x) + m_j^z (\mathbf{I} \otimes \boldsymbol{\sigma}^z) \right. \\ &\quad + c_{\{i,j\}}^{zz} (\boldsymbol{\sigma}^z \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^z) + c_{\{i,j\}}^{xz} (\boldsymbol{\sigma}^x \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^z) \\ &\quad \left. + c_{\{i,j\}}^{zx} (\boldsymbol{\sigma}^z \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^x) + c_{\{i,j\}}^{xx} (\boldsymbol{\sigma}^x \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^x) \right), \end{aligned} \quad (10.341)$$

where

$$\begin{cases} m_i^v = \text{Tr}[\boldsymbol{\sigma}^v \mathbf{R}_i] = \text{Tr}[(\boldsymbol{\sigma}^v \otimes \mathbf{I}) \mathbf{R}_{ij}], \\ m_j^{v'} = \text{Tr}[\boldsymbol{\sigma}^{v'} \mathbf{R}_j] = \text{Tr}[(\mathbf{I} \otimes \boldsymbol{\sigma}^{v'}) \mathbf{R}_{ij}], \quad (\{i, j\} \in E, i < j, v \in \{x, z\}, v' \in \{x, z\}). \\ c_{\{i,j\}}^{v,v'} = \text{Tr}[(\boldsymbol{\sigma}^v \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\sigma}^{v'}) \mathbf{R}_{ij}], \end{cases} \quad (10.342)$$

By using these orthonormal expansions of the reduced density matrices, the Bethe free energy functional in Eq. (10.319) can be rewritten as

$$\begin{aligned} \mathcal{F}[\mathbf{R}] &= \mathcal{F}_{\text{Bethe}} \left[\left\{ m_i^v \mid i \in V, v \in \{x, z\} \right\}, \left\{ c_{\{i,j\}}^{v,v'} \mid \{i, j\} \in E, v \in \{x, z\}, v' \in \{x, z\} \right\} \right] \\ &\equiv -J \sum_{\{i,j\} \in E} c_{\{i,j\}}^{z,z} - h \sum_{i \in V} d_i m_i^z - \Gamma \sum_{i \in V} m_i^x \\ &\quad + k_B T \sum_{i \in V} (1 - |\partial i|) \text{Tr}(\mathbf{R}_i \ln(\mathbf{R}_i)) \\ &\quad + k_B T \sum_{\{i,j\} \in E} \text{Tr}(\mathbf{R}_{\{i,j\}} \ln(\mathbf{R}_{\{i,j\}})). \end{aligned} \quad (10.343)$$

The extremum conditions

$$\begin{aligned} &\frac{\partial}{\partial m_k^v} \mathcal{F}_{\text{Bethe}} \left[\left\{ m_i^v \mid i \in V, v \in \{x, z\} \right\}, \left\{ c_{\{i,j\}}^{v,v'} \mid \{i, j\} \in E, v \in \{x, z\}, v' \in \{x, z\} \right\} \right] \\ &= 0 \quad (k \in V, v \in \{x, z\}), \end{aligned} \quad (10.344)$$

$$\frac{\partial}{\partial c_{\{k,l\}}^{v,v'}} \mathcal{F}_{\text{Bethe}} \left[\left\{ m_i^v \mid i \in V, v \in \{x, z\} \right\}, \left\{ c_{\{i,j\}}^{v,v'} \mid \{i, j\} \in E, v \in \{x, z\}, v' \in \{x, z\} \right\} \right]$$

$$= 0 \text{ } (\{k, l\} \in E, \nu \in \{x, z\}, \nu' \in \{x, z\}), \quad (10.345)$$

can be reduced to the following simultaneous equations:

$$\begin{cases} \frac{h}{k_B T} d_i = \frac{1}{2}(1 - |\partial_i|) \text{Tr}[\sigma^z \ln(\hat{\mathbf{R}}_i)] + \frac{1}{4} \sum_{\{j \in \partial_i, j > i\}} \text{Tr}[(\sigma^z \otimes \mathbf{I}) \ln(\hat{\mathbf{R}}_{ij})] + \frac{1}{4} \sum_{\{k \in \partial_i, k < i\}} \text{Tr}[(\mathbf{I} \otimes \sigma^z) \ln(\hat{\mathbf{R}}_{ki})], \\ \frac{\Gamma}{k_B T} = \frac{1}{2}(1 - |\partial_i|) \text{Tr}[\sigma^x \ln(\hat{\mathbf{R}}_i)] + \frac{1}{4} \sum_{\{j \in \partial_i, j > i\}} \text{Tr}[(\sigma^x \otimes \mathbf{I}) \ln(\hat{\mathbf{R}}_{ij})] + \frac{1}{4} \sum_{\{k \in \partial_i, k < i\}} \text{Tr}[(\mathbf{I} \otimes \sigma^x) \ln(\hat{\mathbf{R}}_{ki})], \end{cases} \quad (i \in V), \quad (10.346)$$

$$\begin{cases} \frac{J}{k_B T} = \frac{1}{4} \text{Tr}[(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) \ln(\hat{\mathbf{R}}_{ij})], \\ 0 = \frac{1}{4} \text{Tr}[(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^x) \ln(\hat{\mathbf{R}}_{ij})] = \frac{1}{4} \text{Tr}[(\sigma^x \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) \ln(\hat{\mathbf{R}}_{ij})], \quad ((i, j) \in E), \\ 0 = \frac{1}{4} \text{Tr}[(\sigma^x \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^x) \ln(\hat{\mathbf{R}}_{ij})], \end{cases} \quad (10.347)$$

where

$$\begin{aligned} \hat{\mathbf{R}}_i &= \frac{1}{2} (\mathbf{I} + \hat{m}_i^x \sigma^x + \hat{m}_i^z \sigma^z) \quad (i \in V), \quad (10.348) \\ \hat{\mathbf{R}}_{ij} &= \hat{\mathbf{R}}_{ji} \\ &= \frac{1}{4} \left((\mathbf{I} \otimes \mathbf{I}) + \hat{m}_i^x (\sigma^x \otimes \mathbf{I}) + \hat{m}_i^z (\sigma^z \otimes \mathbf{I}) + \hat{m}_j^x (\mathbf{I} \otimes \sigma^x) + \hat{m}_j^z (\mathbf{I} \otimes \sigma^z) \right. \\ &\quad \left. + \hat{c}_{[i,j]}^{xz} (\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) + \hat{c}_{[i,j]}^{xz} (\sigma^x \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) \right. \\ &\quad \left. + \hat{c}_{[i,j]}^{xx} (\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^x) + \hat{c}_{[i,j]}^{xx} (\sigma^x \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^x) \right) \quad ((i, j) \in E, i < j). \end{aligned} \quad (10.349)$$

For $\Gamma = 0$, Eq. (10.347) with Eqs. (10.348) and (10.349) reduces to Eqs. (10.107) and (10.108) with Eqs. (10.72) and (10.102).

Before finishing the present subsection, we briefly review another framework of the quantum advanced mean-field method. As we mentioned above, advanced quantum mean-field methods have also been formulated in the momentum space. One familiar formulation is spin wave theory [91]. A general formulation of the quantum cluster variation method from the viewpoint of spin wave theory was proposed in Refs. [92, 93].

10.5.2 Real-Space Renormalization Group Method for the Transverse Ising Model

We now present sublinear modeling in statistical machine learning procedures by using the real-space renormalization group method for the transverse Ising model in Eq. (10.301) on the ring graph (V, E) of Eq. (10.176) for the case of $|V| = 2^L$ and $h = 0$. The present scheme follows the one in Refs. [37, 94]. Some extensions of the present frameworks for the ring graph in Eq. (10.176) to higher-dimensional graphs such as the torus graph may be available according to the frameworks of Ref. [94].

The important part of the transverse Ising model in Eq. (10.301), $-J(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) - \Gamma(\sigma^x \otimes \mathbf{I})$ can be diagonalized as

$$\begin{aligned}
 & -J(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) - \Gamma(\sigma^x \otimes \mathbf{I}) = \begin{pmatrix} -J & 0 & -\Gamma & 0 \\ 0 & J & 0 & -\Gamma \\ -\Gamma & 0 & J & 0 \\ 0 & -\Gamma & 0 & -J \end{pmatrix} \\
 & = \begin{pmatrix} \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 & 0 & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & -\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 \\ \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 & 0 & -\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 & \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 \end{pmatrix} \\
 & \times \begin{pmatrix} -\sqrt{J^2 + \Gamma^2} & 0 & 0 & 0 \\ 0 & -\sqrt{J^2 + \Gamma^2} & 0 & 0 \\ 0 & 0 & \sqrt{J^2 + \Gamma^2} & 0 \\ 0 & 0 & 0 & \sqrt{J^2 + \Gamma^2} \end{pmatrix} \\
 & \times \begin{pmatrix} \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 & 0 & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & -\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 \\ \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 & 0 & -\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 & \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 \end{pmatrix}^T.
 \end{aligned} \tag{10.350}$$

The eigenvalues $\varepsilon_1 = \varepsilon_2 = -\sqrt{J^2 + \Gamma^2}$, $\varepsilon_3 = \varepsilon_4 = +\sqrt{J^2 + \Gamma^2}$ have the relationship $\varepsilon_1 = \varepsilon_2 < \varepsilon_3 = \varepsilon_4$ and their corresponding eigenvectors are given by

$$\left\{ \begin{array}{l} |1\rangle = \begin{pmatrix} \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 \\ \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 \\ +\sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 \\ +\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \end{pmatrix} \\ |3\rangle = \begin{pmatrix} 0 \\ -\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 \\ +\sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \end{pmatrix}, \quad |4\rangle = \begin{pmatrix} \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 \\ -\sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \\ 0 \end{pmatrix} \end{array} \right. \quad (10.351)$$

To realize the coarse graining of the present transverse Ising model for the case of zero temperature $T = 0$ for the density matrix \mathbf{P} in Eq. (10.300), we introduce the following projection operator:

$$\mathbb{P}_i^{(2^L)} = \underbrace{\mathbb{P} \otimes \mathbb{P} \otimes \dots \otimes \mathbb{P}}_{2^L \mathbb{P}'_s} \quad (10.352)$$

where

$$\begin{aligned} \mathbb{P} &\equiv \begin{pmatrix} \langle 1| \\ \langle 2| \end{pmatrix} \begin{pmatrix} |1\rangle \langle 1| + |2\rangle \langle 2| \end{pmatrix} \begin{pmatrix} \langle 1| \\ \langle 2| \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 & \sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 \\ 0 & +\sqrt{\frac{1}{2}\left(1 - \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} & 0 & \sqrt{\frac{1}{2}\left(1 + \frac{J}{\sqrt{J^2 + \Gamma^2}}\right)} \end{pmatrix} \end{aligned} \quad (10.353)$$

Because it is valid that

$$\mathbb{P} \begin{pmatrix} -J & 0 & -\Gamma & 0 \\ 0 & J & 0 & -\Gamma \\ -\Gamma & 0 & J & 0 \\ 0 & -\Gamma & 0 & -J \end{pmatrix} \mathbb{P}^T = -\sqrt{J^2 + \Gamma^2} \begin{pmatrix} \langle 1|1\rangle & \langle 1|2\rangle \\ \langle 2|1\rangle & \langle 2|2\rangle \end{pmatrix} = -\sqrt{J^2 + \Gamma^2} \mathbf{I}, \quad (10.354)$$

we can derive the following equalities

$$\begin{aligned} &\mathbb{P}_i^{(2^L)} \left(-J \sigma_{2i-1}^z \sigma_{2i}^z - \Gamma \sigma_{2i-1}^x \right) \mathbb{P}_i^{(2^L)T} \\ &\quad \text{Tensor Products of } (i-1) \text{ Matrices } (\mathbf{I} \otimes \mathbf{I}) \\ &= \mathbb{P}_i^{(2^L)} \left(\underbrace{(\mathbf{I} \otimes \mathbf{I}) \otimes (\mathbf{I} \otimes \mathbf{I}) \otimes \dots \otimes (\mathbf{I} \otimes \mathbf{I})}_{\text{Tensor Products of } (2^{L-1}-i) \text{ Matrices } (\mathbf{I} \otimes \mathbf{I})} \otimes (-J(\sigma^z \otimes \mathbf{I})(\mathbf{I} \otimes \sigma^z) - \Gamma(\sigma^x \otimes \mathbf{I})) \right) \mathbb{P}_i^{(2^L)T} \end{aligned}$$

$$\begin{aligned}
& \text{Tensor Products of } (2^{L-1}-2) \text{ Matrices } (\mathbf{I} \otimes \mathbf{I}) \\
& \times \left((\mathbf{I} \otimes \mathbf{I}) \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I}) \otimes (\mathbf{I} \otimes \mathbf{I}) \otimes \cdots \otimes (\mathbf{I} \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ Matrices } (\mathbf{I} \otimes \mathbf{I})} \otimes (\mathbf{I} \otimes \boldsymbol{\sigma}^z) \right) \\
& - \Gamma (\mathbf{I} \otimes \mathbf{I}) \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I}) \otimes (\mathbf{I} \otimes \mathbf{I}) \otimes \cdots \otimes (\mathbf{I} \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ Matrices } (\mathbf{I} \otimes \mathbf{I})} \otimes (\mathbf{I} \otimes \boldsymbol{\sigma}^x) \Big) \mathbb{P}_i^{(2^L)^T} \\
= & - \frac{J^2}{\sqrt{J^2 + \Gamma^2}} \left(\boldsymbol{\sigma}^z \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ I's}} \otimes \mathbf{I} \right) \left(\mathbf{I} \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ I's}} \otimes \boldsymbol{\sigma}^z \right) \\
& - \frac{\Gamma^2}{\sqrt{J^2 + \Gamma^2}} \left(\mathbf{I} \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ I's}} \otimes \boldsymbol{\sigma}^x \right). \tag{10.357}
\end{aligned}$$

By using these equalities, the first step of the renormalized energy matrix $\mathbf{H}^{(2^{L-1})} \equiv \mathbb{P}_i^{(2^L)} \mathbf{H} \mathbb{P}_i^{(2^L)^T}$ can be reduced as follows:

$$\begin{aligned}
\mathbf{H}^{(2^{L-1})} & \equiv \mathbb{P}_i^{(2^L)} \mathbf{H} \mathbb{P}_i^{(2^L)^T} \\
& = -2^{L-1} \sqrt{J^2 + \Gamma^2} \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}) \text{ I's}} \\
& - \sum_{i=1}^{2^{L-1}} \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(i-1) \text{ I's}} \otimes \left(\frac{J^2}{\sqrt{J^2 + \Gamma^2}} (\boldsymbol{\sigma}^z \otimes \mathbf{I}) (\mathbf{I} \otimes \boldsymbol{\sigma}^z) + \frac{\Gamma^2}{\sqrt{J^2 + \Gamma^2}} (\boldsymbol{\sigma}^x \otimes \mathbf{I}) \right) \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-i-1) \text{ I's}} \\
& - \frac{J^2}{\sqrt{J^2 + \Gamma^2}} \left(\boldsymbol{\sigma}^z \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ I's}} \otimes \mathbf{I} \right) \left(\mathbf{I} \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ I's}} \otimes \boldsymbol{\sigma}^z \right) \\
& - \frac{\Gamma^2}{\sqrt{J^2 + \Gamma^2}} \left(\mathbf{I} \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-1}-2) \text{ I's}} \otimes \boldsymbol{\sigma}^x \right). \tag{10.358}
\end{aligned}$$

By similar arguments to those for the above procedure, the r -th step of the renormalized energy matrix $\mathbf{H}^{(2^{L-r})} \equiv \left(\mathbb{P}_i^{(2^{L-r})} \mathbb{P}_i^{(2^{L-r+1})} \cdots \mathbb{P}_i^{(2^L)} \right) \mathbf{H} \left(\mathbb{P}_i^{(2^L)^T} \cdots \mathbb{P}_i^{(2^{L-r+1})^T} \mathbb{P}_i^{(2^{L-r})^T} \right)$ can be reduced to the following recursion formulas:

$$\begin{aligned}
\mathbf{H}^{(2^{L-r})} & \equiv \mathbb{P}_i^{(2^{L-r+1})} \mathbf{H}^{(2^{L-r+1})} \mathbb{P}_i^{(2^{L-r+1})^T} \\
& = 2^{L-r} \varepsilon_1^{(r)} \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-r}) \text{ I's}} \\
& - \sum_{i=1}^{2^{L-1}} \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(i-1) \text{ I's}} \otimes \left(J^{(r)} (\boldsymbol{\sigma}^z \otimes \mathbf{I}) (\mathbf{I} \otimes \boldsymbol{\sigma}^z) + \Gamma^{(r)} (\boldsymbol{\sigma}^x \otimes \mathbf{I}) \right) \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-r}-i-1) \text{ I's}} \\
& - J^{(r)} \left(\boldsymbol{\sigma}^z \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-r}-2) \text{ I's}} \otimes \mathbf{I} \right) \left(\mathbf{I} \otimes \overbrace{(\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I})}^{(2^{L-r}-2) \text{ I's}} \otimes \boldsymbol{\sigma}^z \right)
\end{aligned}$$

$$-\Gamma^{(r)} \left(\mathbf{I} \otimes \left(\overbrace{\mathbf{I} \otimes \mathbf{I} \otimes \dots \otimes \mathbf{I}}^{(2^{L-r}-2) \text{ I's}} \right) \otimes \boldsymbol{\sigma}^x \right), \tag{10.359}$$

where

$$\begin{cases} J^{(r)} = \frac{(J^{(r-1)})^2}{\sqrt{(J^{(r-1)})^2 + (\Gamma^{(r-1)})^2}}, \\ \Gamma^{(r)} = \frac{(\Gamma^{(r-1)})^2}{\sqrt{(J^{(r-1)})^2 + (\Gamma^{(r-1)})^2}}, \end{cases} \tag{10.360}$$

$$\varepsilon_1^{(r)} = -\sqrt{(J^{(r-1)})^2 + (\Gamma^{(r-1)})^2}, \tag{10.361}$$

$$\begin{cases} \mathbf{H}^{(2^L)} \equiv \mathbf{H}, \\ J^{(0)} \equiv J, \\ \Gamma^{(0)} \equiv \Gamma. \end{cases} \tag{10.362}$$

The inverse of the real-space renormalization group is given by

$$\begin{cases} J^{(r-1)} = \sqrt{J^{(r)}(J^{(r)} + \Gamma^{(r)})}, \\ \Gamma^{(r-1)} = \sqrt{\Gamma^{(r)}(J^{(r)} + \Gamma^{(r)})}, \end{cases} \tag{10.363}$$

If the hyperparameters $J^{(r)}$ and $\Gamma^{(r)}$ in the r -th renormalized density matrix $\mathbf{H}^{(2^{L-r})}$ have been estimated from given data vectors by using the QEM algorithm for a renormalized density matrix on ring graphs $(V^{(r)}, E^{(r)})$, we can estimate the hyperparameters $J^{(0)} = J$ and $\Gamma^{(0)} = \Gamma$ of the transverse Ising model (10.301) on the ring graph E of Eq. (10.176) for the case of $|V| = 2^L$ and $h = 0$ by using the inverse transformation rule of the real-space renormalization group procedure (10.363).

10.5.3 Sublinear Modeling Using a Quantum Adaptive TAP Approach and Momentum Space Renormalization Group in the Transverse Ising Model

This section proposes a novel scheme for the momentum space renormalization group approaches in **Adaptive Thouless-Anderson-Palmar(TAP) Approaches** for the transverse Ising model on random graphs. The adaptive TAP approach is a familiar advanced mean-field method for the probabilistic graphical model and many extensions have been proposed [95–98]. Furthermore, sublinear modeling for the

EM procedure in probabilistic graphical models has been realized by introducing **Momentum Space Renormalization Group Approaches** [99, 100]. The method proposed in this section is formulated by combining the adaptive TAP approaches with the momentum space renormalization group approaches. Moreover, our method is applicable not only to regular graphs but also to random graphs.

The density matrix \mathbf{P} in Eq. (10.300) can be rewritten as

$$\mathbf{P} = \frac{\exp\left(-\frac{J}{2k_{\text{B}}T} \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 - \frac{h}{2k_{\text{B}}T} \sum_{i \in V} (\sigma_i^z - d_i \mathbf{I}^{(2|V|)})^2 - \frac{1}{2k_{\text{B}}T} \sum_{i \in V} (\sigma_i^x - \gamma \mathbf{I}^{(2|V|)})^2\right)}{\text{Tr}\left[\exp\left(-\frac{J}{2k_{\text{B}}T} \sum_{\{i,j\} \in E} (\sigma_i^z - \sigma_j^z)^2 - \frac{h}{2k_{\text{B}}T} \sum_{i \in V} (\sigma_i^z - d_i \mathbf{I}^{(2|V|)})^2 - \frac{1}{2k_{\text{B}}T} \sum_{i \in V} (\sigma_i^x - \gamma \mathbf{I}^{(2|V|)})^2\right)\right]}. \quad (10.364)$$

The density matrix \mathbf{P} satisfies the following minimization of the free energy functional:

$$\mathbf{P} = \underset{\mathbf{R}}{\text{argmin}} \left\{ \mathcal{F}[\mathbf{R}] \mid \text{Tr} \mathbf{R} = 1 \right\}, \quad (10.365)$$

$$\begin{aligned} \mathcal{F}[\mathbf{R}] \equiv & \frac{1}{2} J \sum_{\{i,j\} \in E} \text{Tr} \left[(\sigma_i^z - \sigma_j^z)^2 \mathbf{R} \right] + \frac{1}{2} h \sum_{i \in V} \text{Tr} \left[(\sigma_i^z - d_i \mathbf{I}^{(2|V|)})^2 \mathbf{R} \right] \\ & + \frac{1}{2} \sum_{i \in V} \text{Tr} \left[(\sigma_i^x - \gamma \mathbf{I}^{(2|V|)})^2 \mathbf{R} \right] + k_{\text{B}} T \text{Tr} [\mathbf{R} \ln(\mathbf{R})]. \end{aligned} \quad (10.366)$$

Because all the off-diagonal elements of $(\sigma_i^z - \sigma_j^z)^2$ are zero, we have

$$\begin{aligned} \mathcal{F}[\mathbf{R}] &= \frac{1}{2} J \sum_{\{i,j\} \in E} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \langle s_1, s_2, \dots, s_{|V|} | (\sigma_i^z - \sigma_j^z)^2 | s_1, s_2, \dots, s_{|V|} \rangle \\ & \quad \times \langle s_1, s_2, \dots, s_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle \\ & + \frac{1}{2} h \sum_{i \in V} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} \langle s_1, s_2, \dots, s_{|V|} | (\sigma_i^z - d_i \mathbf{I}^{(2|V|)})^2 | s_1, s_2, \dots, s_{|V|} \rangle \\ & \quad \times \langle s_1, s_2, \dots, s_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle \\ & + \frac{1}{2} \sum_{i \in V} \text{Tr} \left[(\sigma_i^x - \gamma \mathbf{I}^{(2|V|)})^2 \mathbf{R} \right] + k_{\text{B}} T \text{Tr} [\mathbf{R} \ln(\mathbf{R})] \\ &= \frac{1}{2} J \sum_{\{i,j\} \in E} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} (s_i - s_j)^2 \langle s_1, s_2, \dots, s_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle \\ & + \frac{1}{2} h \sum_{i \in V} \sum_{s_1 \in \Omega} \sum_{s_2 \in \Omega} \cdots \sum_{s_{|V|} \in \Omega} (s_i - d_i)^2 \langle s_1, s_2, \dots, s_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle \end{aligned}$$

$$+\frac{1}{2}\sum_{i\in V}\mathrm{Tr}\left[(\sigma_i^x - \Gamma \mathbf{I}^{(2^{|V|})})^2 \mathbf{R}\right] + k_B T \mathrm{Tr}[\mathbf{R} \ln(\mathbf{R})]. \quad (10.367)$$

By introducing the reduced density matrix \mathbf{R}_i in Eq. (10.275) and

$$\rho(s_1, s_2, \dots, s_{|V|}) \equiv \langle s_1, s_2, \dots, s_{|V|} | \mathbf{R} | s_1, s_2, \dots, s_{|V|} \rangle, \quad ((s_1, s_2, \dots, s_{|V|}) \in \Omega^{|V|}), \quad (10.368)$$

and by extending $\rho(s_1, s_2, \dots, s_{|V|})$ to

$$\rho(\boldsymbol{\phi}) = \rho(\phi_1, \phi_2, \dots, \phi_{|V|}) \quad (\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_{|V|}) \in (-\infty, +\infty)^{|V|}), \quad (10.369)$$

the free energy functional can be expressed as

$$\begin{aligned} \mathcal{F}[\mathbf{R}] &= \frac{1}{2} J \sum_{\{i,j\} \in E} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\prod_{k \in V} (\delta(\phi_k - 1) + \delta(\phi_k + 1)) \right) (\phi_i - \phi_j)^2 \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} \\ &+ \frac{1}{2} h \sum_{i \in V} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\prod_{k \in V} (\delta(\phi_k - 1) + \delta(\phi_k + 1)) \right) (\phi_i - d_i)^2 \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} \\ &+ \frac{1}{2} \sum_{i \in V} \mathrm{Tr}[(\sigma_i^x - \Gamma \mathbf{I}^{(2^{|V|})})^2 \mathbf{R}_i] + k_B T \mathrm{Tr}[\mathbf{R} \ln(\mathbf{R})]. \end{aligned} \quad (10.370)$$

Now we consider the following approximate free energy:

$$\begin{aligned} \mathcal{F}_{\text{Adaptive TAP}}[\rho, \{\mathbf{R}_i, \rho_i | i \in V\}] &\equiv \frac{1}{2} J \sum_{\{i,j\} \in E} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\phi_i - \phi_j)^2 \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} \\ &+ \frac{1}{2} h \sum_{i \in V} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\phi_i - d_i)^2 \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} \\ &+ \frac{1}{2} \sum_{i \in V} \mathrm{Tr}[(\sigma_i^x - \Gamma \mathbf{I}^{(2^{|V|})})^2 \mathbf{R}_i] \\ &+ k_B T \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \rho(\boldsymbol{\phi}) \ln(\rho(\boldsymbol{\phi})) d\phi_1 d\phi_2 \cdots d\phi_{|V|} \\ &+ k_B T \sum_{i \in V} \left(\mathrm{Tr} \mathbf{R}_i \ln \mathbf{R}_i - \int_{-\infty}^{+\infty} \rho_i(\phi_i) \ln(\rho_i(\phi_i)) d\phi_i \right), \end{aligned} \quad (10.371)$$

where

$$\rho_i(\phi_i) \equiv \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \delta(\phi_i - \phi'_i) \rho(\phi'_1, \phi'_2, \dots, \phi'_{|V|}) d\phi'_1 d\phi'_2 \cdots d\phi'_{|V|} \quad (i \in V, \phi_i \in (-\infty, +\infty)). \quad (10.372)$$

The reduced density matrix \mathbf{R}_i and the marginal probability density functions $\rho_i(\phi_i)$ and $\rho(\boldsymbol{\phi})$ need to satisfy the consistencies

$$\begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_i \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} = \int_{-\infty}^{+\infty} \phi_i \rho_i(\phi_i) d\phi_i = \text{Tr} \boldsymbol{\sigma}^z \mathbf{R}_i \quad (i \in V), \\ \sum_{i \in V} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_i^2 \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} = \sum_{i \in V} \int_{-\infty}^{+\infty} \phi_i^2 \rho_i(\phi_i) d\phi_i = 1, \end{cases} \quad (10.373)$$

and the normalizations

$$\begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} = 1, \\ \int_{-\infty}^{+\infty} \rho_i(\phi_i) d\phi_i = 1 \quad (i \in V). \\ \text{Tr} \mathbf{R}_i = 1. \end{cases} \quad (10.374)$$

\mathbf{R}_i , $\rho_i(\phi_i)$ and $\rho(\boldsymbol{\phi})$ are determined so as to minimize the above approximate free energy $\mathcal{F}[\rho, \{\mathbf{R}_i, \rho_i | i \in V\}]$ under the constraint conditions in Eqs. (10.373) and (10.374). We

introduce Lagrange multipliers $\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{|V|} \end{pmatrix}$ and $\mathbf{g} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{|V|} \end{pmatrix}$, D , L , λ , and λ_i to ensure

the constraint conditions in Eqs. (10.373) and (10.374) as follows:

$$\begin{aligned} & \mathcal{L}_{\text{Adaptive TAP}}[\rho, \{\mathbf{R}_i, \rho_i | i \in V\}] \\ & \equiv \mathcal{F}_{\text{Adaptive TAP}}[\rho, \{\mathbf{R}_i, \rho_i | i \in V\}] \\ & - \sum_{i \in V} g_i \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_i \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} - \int_{-\infty}^{+\infty} \phi_i \rho_i(\phi_i) d\phi_i \right) \\ & - \sum_{i \in V} f_i \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_i \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} - \text{Tr} \boldsymbol{\sigma}^z \mathbf{R}_i \right) \\ & - D \left(\sum_{i \in V} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_i^2 \rho(\boldsymbol{\phi}) d\phi_1 d\phi_2 \cdots d\phi_{|V|} - 1 \right) \\ & - L \left(\sum_{i \in V} \int_{-\infty}^{+\infty} \phi_i^2 \rho_i(\phi_i) d\phi_i - 1 \right) \\ & - \lambda \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \rho(\boldsymbol{\phi}) d\boldsymbol{\phi} - 1 \right) \\ & - \sum_{i \in V} \lambda_i \left(\int_{-\infty}^{+\infty} \rho_i(\phi_i) d\phi_i - 1 \right). \end{aligned} \quad (10.375)$$

By taking the first variation of the approximate free energy $\mathcal{L}_{\text{Adaptive TAP}}[\hat{\rho}, \{\hat{\mathbf{R}}_i, \hat{\rho}_i | i \in V\}]$ with respect to the marginals, we can derive the approximate expressions of $\hat{\mathbf{R}}_i$, $\hat{\rho}_i(\phi_i)$, and $\hat{\rho}(\boldsymbol{\phi})$ as follows:

$$\hat{\mathbf{R}}_i = \frac{\exp\left(\frac{1}{k_B T} (f_i \boldsymbol{\sigma}^z + \Gamma \boldsymbol{\sigma}^x)\right)}{\text{Tr} \exp\left(\frac{1}{k_B T} (h_i \boldsymbol{\sigma}^z + \Gamma \boldsymbol{\sigma}^x)\right)} \quad (i \in V), \quad (10.376)$$

$$\begin{aligned} \widehat{\rho}(\phi) &= \frac{\exp\left(\frac{1}{k_B T} \left(-\frac{1}{2} D \sum_{i \in V} \phi_i^2 + \sum_{i \in V} (f_i + g_i) \phi_i - \frac{1}{2} h \sum_{i \in V} (\phi_i - d_i)^2 - \frac{1}{2} J \sum_{(i,j) \in E} (\phi_i - \phi_j)^2\right)\right)}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \exp\left(\frac{1}{k_B T} \left(-\frac{1}{2} D \sum_{i \in V} \phi_i^2 + \sum_{i \in V} (f_i + g_i) \phi_i - \frac{1}{2} h \sum_{i \in V} (\phi_i - d_i)^2 - \frac{1}{2} J \sum_{(i,j) \in E} (\phi_i - \phi_j)^2\right)\right) d\phi_1 d\phi_2 \dots d\phi_{|V|}} \end{aligned} \tag{10.377}$$

$$\widehat{\rho}_i(\phi_i) = \frac{\exp\left(\frac{1}{k_B T} \left(-\frac{1}{2} L \phi_i^2 + g_i \phi_i - \frac{1}{2} h (\phi_i - d_i)^2\right)\right)}{\int_{-\infty}^{+\infty} \exp\left(\frac{1}{k_B T} \left(-\frac{1}{2} L \phi_i^2 + g_i \phi_i - \frac{1}{2} h (\phi_i - d_i)^2\right)\right) d\phi_i} \tag{10.378}$$

where C is the $|V| \times |V|$ matrix in which the (i, j) -elements are defined by

$$C_{ij} \equiv \begin{cases} |\partial i| & (i = j), \\ -1 & ((i, j) \in E), \\ 0 & (\text{otherwise}), \end{cases} \tag{10.379}$$

for any nodes $i \in V$ and $j \in V$. Equations (10.376), (10.377), and (10.378) can be rewritten as

$$\begin{aligned} \widehat{\mathbf{R}}_i &= \frac{1}{2 \cosh\left(\frac{1}{k_B T} \sqrt{f_i^2 + \Gamma^2}\right)} \frac{f_i + \sqrt{f_i^2 + \Gamma^2}}{2 \sqrt{f_i^2 + \Gamma^2}} \begin{pmatrix} 1 & -\frac{\Gamma}{f_i + \sqrt{f_i^2 + \Gamma^2}} \\ +\frac{\Gamma}{f_i + \sqrt{f_i^2 + \Gamma^2}} & 1 \end{pmatrix} \\ &\times \begin{pmatrix} \exp\left(+\frac{1}{k_B T} \sqrt{f_i^2 + \Gamma^2}\right) & 0 \\ 0 & \exp\left(-\frac{1}{k_B T} \sqrt{f_i^2 + \Gamma^2}\right) \end{pmatrix} \begin{pmatrix} 1 & +\frac{\Gamma}{f_i + \sqrt{f_i^2 + \Gamma^2}} \\ -\frac{\Gamma}{f_i + \sqrt{f_i^2 + \Gamma^2}} & 1 \end{pmatrix}. \end{aligned} \tag{10.380}$$

$$\begin{aligned} \widehat{\rho}(\phi) &= \sqrt{\frac{\det((h + D) \mathbf{I}^{|V|} + J \mathbf{C})}{(2\pi)^{|V|}}} \\ &\times \exp\left(-\frac{1}{2k_B T} \left(\phi - ((h + D) \mathbf{I}^{|V|} + J \mathbf{C})^{-1} (f + g + h d)\right)^T \left((h + D) \mathbf{I}^{|V|} + J \mathbf{C}\right) \right. \\ &\quad \left. \times \left(\phi - ((h + D) \mathbf{I}^{|V|} + J \mathbf{C})^{-1} (f + g + h d)\right)\right), \end{aligned} \tag{10.381}$$

$$\widehat{\rho}_i(\phi_i) = \sqrt{\frac{h + L}{2\pi}} \exp\left(-\frac{1}{2k_B T} (h + L) \left(x - \frac{g_i + h d_i}{h + L}\right)^2\right) \quad (i \in V), \tag{10.382}$$

The Lagrange multipliers f , g , L , and D are often referred to as the **effective fields** and are determined so as to satisfy the consistencies in Eq. (10.373), which reduce to the following simultaneous equations:

$$\mathbf{g} + \mathbf{h}\mathbf{d} = (\mathbf{h} + \mathbf{L})\left((\mathbf{D} - \mathbf{L})\mathbf{I}^{(2|V|)} + \mathbf{J}\mathbf{C}\right)^{-1}\mathbf{f}, \tag{10.383}$$

$$\mathbf{f} + \mathbf{g} + \mathbf{h}\mathbf{d} = \left((\mathbf{D} - \mathbf{L})\mathbf{I}^{(|V|)} + \mathbf{J}\mathbf{C}\right)^{-1} \begin{pmatrix} \frac{f_1}{\sqrt{f_1^2 + \Gamma^2}} \tanh\left(\frac{1}{k_{\text{B}}T} \sqrt{f_1^2 + \Gamma^2}\right) \\ \frac{f_2}{\sqrt{f_2^2 + \Gamma^2}} \tanh\left(\frac{1}{k_{\text{B}}T} \sqrt{f_2^2 + \Gamma^2}\right) \\ \vdots \\ \frac{f_{|V|}}{\sqrt{f_{|V|}^2 + \Gamma^2}} \tanh\left(\frac{1}{k_{\text{B}}T} \sqrt{f_{|V|}^2 + \Gamma^2}\right) \end{pmatrix}, \tag{10.384}$$

$$\mathbf{L} = -\mathbf{h} + \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{|V|}(\mathbf{f} + \mathbf{g} + \mathbf{h}\mathbf{d})^T(\mathbf{f} + \mathbf{g} + \mathbf{h}\mathbf{d})}, \tag{10.385}$$

$$\frac{1}{\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{|V|}(\mathbf{f} + \mathbf{g} + \mathbf{h}\mathbf{d})^T(\mathbf{f} + \mathbf{g} + \mathbf{h}\mathbf{d})}} = \frac{1}{|V|} \text{Tr}\left(\left((\mathbf{h} + \mathbf{D})\mathbf{I}^{(|V|)} + \mathbf{J}\mathbf{C}\right)^{-1}\right). \tag{10.386}$$

The real symmetric matrix \mathbf{C} is diagonalized as

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}, \tag{10.387}$$

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{|V|} \end{pmatrix}, \tag{10.388}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_{|V|}$. All the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{|V|}$ are always real num-

bers. For the eigenvector $\mathbf{u}_i = \begin{pmatrix} U_{1i} \\ U_{2i} \\ \vdots \\ U_{|V|i} \end{pmatrix}$ corresponding to the eigenvalue λ_i such that

$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i$, for every $i \in \{1, 2, 3, \dots, M\}$, the matrix \mathbf{U} is defined by

$$U \equiv (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_M) = \begin{pmatrix} U_{11} & U_{12} & U_{13} & \cdots & U_{1|V|} \\ U_{21} & U_{22} & U_{23} & \cdots & U_{2|V|} \\ U_{31} & U_{32} & U_{33} & \cdots & U_{3|V|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{|V|1} & U_{|V|2} & U_{|V|3} & \cdots & U_{|V||V|} \end{pmatrix}. \quad (10.389)$$

It is known that U is a unitary matrix that satisfies $U^{-1} = U^T$ for the real symmetric matrix C . By using the diagonal matrix Λ and unitary matrix U , the density matrix R in Eq. (10.390) can be represented as follows:

$$R = \frac{\exp\left(-\frac{1}{2k_B T} \xi^T \left((hI^{(2|V|)} + J\Lambda) \otimes I^{(2|V|)} \right) \xi - \frac{1}{2k_B T} \xi^T \xi \right)}{\text{Tr} \left[\exp\left(-\frac{1}{2k_B T} \xi^T \left((hI^{(2|V|)} + J\Lambda) \otimes I^{(2|V|)} \right) \xi - \frac{1}{2k_B T} \xi^T \xi \right) \right]}, \quad (10.390)$$

where

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{|V|} \end{pmatrix} \equiv (U^T \otimes I^{(2|V|)}) \begin{pmatrix} \sigma_1^z \\ \sigma_2^z \\ \vdots \\ \sigma_{|V|}^z \end{pmatrix} - \left((hI^{(|V|)} + J\Lambda)^{-1} U^T \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{|V|} \end{pmatrix} \right) \otimes I^{(2|V|)}, \quad (10.391)$$

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{|V|} \end{pmatrix} \equiv (U^T \otimes I^{(2|V|)}) \begin{pmatrix} \sigma_1^x \\ \sigma_2^x \\ \vdots \\ \sigma_{|V|}^x \end{pmatrix} - \gamma \begin{pmatrix} I^{(2|V||V|)} \\ I^{(2|V||V|)} \\ \vdots \\ I^{(2|V||V|)} \end{pmatrix}. \quad (10.392)$$

By using the Gram-Schmidt orthonormalization in the framework of Fig. 10.17, we introduce a new unitary matrix

$$\tilde{U} = \begin{pmatrix} \tilde{U}_{11} & \tilde{U}_{12} & \tilde{U}_{13} & \cdots & \tilde{U}_{1|\tilde{V}|} \\ \tilde{U}_{21} & \tilde{U}_{22} & \tilde{U}_{23} & \cdots & \tilde{U}_{2|\tilde{V}|} \\ \tilde{U}_{31} & \tilde{U}_{32} & \tilde{U}_{33} & \cdots & \tilde{U}_{3|\tilde{V}|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{U}_{|\tilde{V}|1} & \tilde{U}_{|\tilde{V}|2} & \tilde{U}_{|\tilde{V}|3} & \cdots & \tilde{U}_{|\tilde{V}||\tilde{V}|} \end{pmatrix} \equiv (\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\mathbf{u}}_3, \dots, \tilde{\mathbf{u}}_{|\tilde{V}|}), \quad (10.393)$$

where

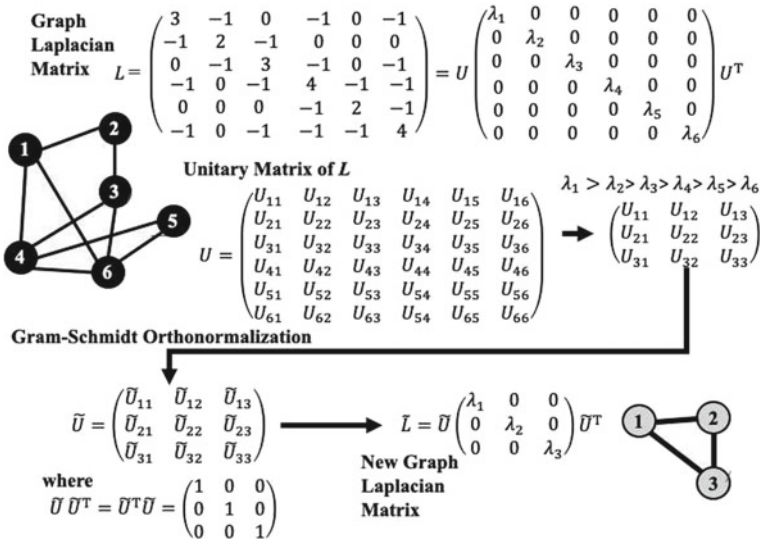


Fig. 10.17 Momentum space renormalization group for graphical models on random graphs

$$v_1 = \begin{pmatrix} U_{11} \\ U_{21} \\ \vdots \\ U_{|\tilde{V}|1} \end{pmatrix}, v_2 = \begin{pmatrix} U_{12} \\ U_{22} \\ \vdots \\ U_{|\tilde{V}|2} \end{pmatrix}, \dots, v_{|\tilde{V}|} = \begin{pmatrix} U_{1|\tilde{V}|} \\ U_{2|\tilde{V}|} \\ \vdots \\ U_{|\tilde{V}||\tilde{V}|} \end{pmatrix}, \tag{10.394}$$

$$\begin{cases} u'_1 = v_1, & \tilde{u}'_1 = \frac{u'_1}{\sqrt{u'^T_1 u'_1}}, \\ u'_2 = v_2 - \frac{u'^T_1 v_2}{u'^T_1 u'_1} u'_1, & \tilde{u}'_2 = \frac{u'_2}{\sqrt{u'^T_2 u'_2}}, \\ u'_3 = v_3 - \frac{u'^T_1 v_3}{u'^T_1 u'_1} u'_1 - \frac{u'^T_2 v_3}{u'^T_2 u'_2} u'_2, & \tilde{u}'_3 = \frac{u'_3}{\sqrt{u'^T_3 u'_3}}, \\ \vdots & \\ u'_{|\tilde{V}|} = v_{|\tilde{V}|} - \frac{u'^T_1 v_{|\tilde{V}|}}{u'^T_1 u'_1} u'_1 - \frac{u'^T_2 v_{|\tilde{V}|}}{u'^T_2 u'_2} u'_2 + \dots + \frac{u'^T_{|\tilde{V}|-1} v_{|\tilde{V}|}}{u'^T_{|\tilde{V}|-1} u'_{|\tilde{V}|-1}} u'_{|\tilde{V}|-1}, & \tilde{u}'_{|\tilde{V}|} = \frac{u'_{|\tilde{V}|}}{\sqrt{u'^T_{|\tilde{V}|} u'_{|\tilde{V}|}}}. \end{cases} \tag{10.395}$$

By using the new unitary matrix \tilde{U} and a diagonal matrix Λ

$$\tilde{\Lambda} \equiv \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{|\tilde{V}_1|} \end{pmatrix}, \quad (10.396)$$

we introduce a renormalized density matrix $\tilde{\mathcal{R}}$ from the standpoint of the momentum space renormalization group for general graphs as

$$\tilde{\mathcal{P}} \equiv \frac{\exp\left(-\frac{1}{2k_{\text{B}}T} \tilde{\zeta}^{\text{T}} \left((hI^{(2|\tilde{V}_1|)} + J\tilde{\Lambda}) \otimes I^{(2|\tilde{V}_1|)} \right) \tilde{\zeta} - \frac{1}{2k_{\text{B}}T} \tilde{\zeta}^{\text{T}} \tilde{\xi}\right)}{\text{Tr}\left[\exp\left(-\frac{1}{2k_{\text{B}}T} \tilde{\zeta}^{\text{T}} \left((hI^{(2|\tilde{V}_1|)} + J\tilde{\Lambda}) \otimes I^{(2|\tilde{V}_1|)} \right) \tilde{\zeta} - \frac{1}{2k_{\text{B}}T} \tilde{\zeta}^{\text{T}} \tilde{\xi}\right)\right]}, \quad (10.397)$$

where

$$\tilde{\zeta} = \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \\ \vdots \\ \tilde{\xi}_{|\tilde{V}_1|} \end{pmatrix} \equiv \left(\tilde{U}^{\text{T}} \otimes I^{(2|\tilde{V}_1|)} \right) \begin{pmatrix} \sigma_1^z \\ \sigma_2^z \\ \vdots \\ \sigma_{|\tilde{V}_1|}^z \end{pmatrix} - \left((hI^{(|\tilde{V}_1|)} + J\tilde{\Lambda})^{-1} \tilde{U}^{\text{T}} \begin{pmatrix} \tilde{d}_1 \\ \tilde{d}_2 \\ \vdots \\ \tilde{d}_{|\tilde{V}_1|} \end{pmatrix} \right) \otimes I^{(2|\tilde{V}_1|)}, \quad (10.398)$$

$$\tilde{\xi} = \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \\ \vdots \\ \tilde{\xi}_{|\tilde{V}_1|} \end{pmatrix} \equiv \left(U^{\text{T}} \otimes I^{(2|\tilde{V}_1|)} \right) \begin{pmatrix} \sigma_1^x \\ \sigma_2^x \\ \vdots \\ \sigma_{|\tilde{V}_1|}^x \end{pmatrix} - \gamma \begin{pmatrix} I^{(2|\tilde{V}_1||\tilde{V}_1|)} \\ I^{(2|\tilde{V}_1||\tilde{V}_1|)} \\ \vdots \\ I^{(2|\tilde{V}_1||\tilde{V}_1|)} \end{pmatrix}, \quad (10.399)$$

$$\begin{pmatrix} \tilde{d}_1 \\ \tilde{d}_2 \\ \vdots \\ \tilde{d}_{|\tilde{V}_1|} \end{pmatrix} = \begin{pmatrix} \tilde{U}_{11} & \tilde{U}_{12} & \tilde{U}_{13} & \cdots & \tilde{U}_{1|\tilde{V}_1|} \\ \tilde{U}_{21} & \tilde{U}_{22} & \tilde{U}_{23} & \cdots & \tilde{U}_{2|\tilde{V}_1|} \\ \tilde{U}_{31} & \tilde{U}_{32} & \tilde{U}_{33} & \cdots & \tilde{U}_{3|\tilde{V}_1|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{U}_{|\tilde{V}_1|1} & \tilde{U}_{|\tilde{V}_1|2} & \tilde{U}_{|\tilde{V}_1|3} & \cdots & \tilde{U}_{|\tilde{V}_1||\tilde{V}_1|} \end{pmatrix} \begin{pmatrix} U_{11} & U_{21} & U_{31} & \cdots & U_{|\tilde{V}_1|1} \\ U_{12} & U_{22} & U_{32} & \cdots & U_{|\tilde{V}_1|2} \\ U_{13} & U_{23} & U_{33} & \cdots & U_{|\tilde{V}_1|3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{1|\tilde{V}_1|} & U_{2|\tilde{V}_1|} & U_{3|\tilde{V}_1|} & \cdots & U_{|\tilde{V}_1||\tilde{V}_1|} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_{|\tilde{V}_1|} \end{pmatrix}. \quad (10.400)$$

For this density matrix $\tilde{\mathcal{P}}$ in Eq. (10.397), we can formulate the approximate reduced density matrix $\hat{\tilde{\mathcal{R}}}_i$ and the approximate Gaussian marginal probability density function $\hat{\tilde{\rho}}(\tilde{\phi}) = \hat{\tilde{\rho}}(\phi_1, \phi_2, \dots, \phi_{|\tilde{V}_1|})$ and $\hat{\tilde{\rho}}(\phi_i)$ for the corresponding quantum adaptive TAP approximation as follows:

$$\widehat{\mathbf{R}}_i = \frac{1}{2\cosh\left(\frac{1}{k_{\text{B}}T}\sqrt{\tilde{f}_i^2 + \Gamma^2}\right)} \frac{\tilde{f}_i + \sqrt{\tilde{f}_i^2 + \Gamma^2}}{2\sqrt{\tilde{f}_i^2 + \Gamma^2}} \begin{pmatrix} 1 & -\frac{\Gamma}{\tilde{f}_i + \sqrt{\tilde{f}_i^2 + \Gamma^2}} \\ +\frac{\Gamma}{\tilde{f}_i + \sqrt{\tilde{f}_i^2 + \Gamma^2}} & 1 \end{pmatrix} \\ \times \begin{pmatrix} \exp\left(+\frac{1}{k_{\text{B}}T}\sqrt{\tilde{f}_i^2 + \Gamma^2}\right) & 0 \\ 0 & \exp\left(-\frac{1}{k_{\text{B}}T}\sqrt{\tilde{f}_i^2 + \Gamma^2}\right) \end{pmatrix} \begin{pmatrix} 1 & +\frac{\Gamma}{\tilde{f}_i + \sqrt{\tilde{f}_i^2 + \Gamma^2}} \\ -\frac{\Gamma}{\tilde{f}_i + \sqrt{\tilde{f}_i^2 + \Gamma^2}} & 1 \end{pmatrix}. \quad (10.401)$$

$$\widehat{\rho}(\phi_1, \phi_2, \dots, \phi_{|\tilde{V}|}) = \sqrt{\frac{\det\left((h + \tilde{D})\mathbf{I}^{(|\tilde{V}|)} + J\tilde{\Lambda}\right)}{(2\pi k_{\text{B}}T)^{|\tilde{V}|}}} \\ \times \exp\left(-\frac{1}{2k_{\text{B}}T}\left(\tilde{\phi} - \tilde{U}\left((h + \tilde{D})\mathbf{I}^{(|\tilde{V}|)} + J\tilde{\Lambda}\right)^{-1}\tilde{U}^{\text{T}}(\tilde{f} + \tilde{g} + h\tilde{d})\right)^{\text{T}}\right. \\ \left.\times \tilde{U}\left((h + \tilde{D})\mathbf{I}^{(|\tilde{V}|)} + J\tilde{\Lambda}\right)\tilde{U}^{\text{T}}\right. \\ \left.\times \left(\tilde{\phi} - \tilde{U}\left((h + \tilde{D})\mathbf{I}^{(|\tilde{V}|)} + J\tilde{\Lambda}\right)^{-1}\tilde{U}^{\text{T}}(\tilde{f} + \tilde{g} + h\tilde{d})\right)\right), \quad (10.402)$$

$$\widehat{\rho}_i(\phi_i) = \sqrt{\frac{h + \tilde{L}}{2\pi k_{\text{B}}T}} \exp\left(-\frac{1}{2k_{\text{B}}T}(h + \tilde{L})\left(\phi_i - \frac{\tilde{g}_i + h\tilde{d}_i}{h + \tilde{L}}\right)^2\right) \quad (i \in \tilde{V}). \quad (10.403)$$

The reduced density matrix $\widehat{\mathbf{R}}_i$ and the marginal probability density functions $\widehat{\rho}_i(\phi_i)$ and $\widehat{\rho}(\phi_1, \phi_2, \dots, \phi_{|\tilde{V}|})$ need to satisfy the consistencies

$$\begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \phi_i \widehat{\rho}(\phi_1, \phi_2, \dots, \phi_{|\tilde{V}|}) d\phi_1 d\phi_2 \dots d\phi_{|\tilde{V}|} = \int_{-\infty}^{+\infty} \phi_i \widehat{\rho}_i(\phi_i) d\phi_i = \text{Tr} \sigma^z \widehat{\mathbf{R}}_i \quad (i \in \tilde{V}), \\ \sum_{i \in \tilde{V}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \phi_i^2 \widehat{\rho}(\phi_1, \phi_2, \dots, \phi_{|\tilde{V}|}) d\phi_1 d\phi_2 \dots d\phi_{|\tilde{V}|} = \sum_{i \in \tilde{V}} \int_{-\infty}^{+\infty} \phi_i^2 \widehat{\rho}_i(\phi_i) d\phi_i = 1. \end{cases} \quad (10.404)$$

The Lagrange multipliers \tilde{f} , \tilde{g} , \tilde{L} , and \tilde{D} are determined so as to satisfy the consistencies in Eq. (10.404), which reduce to the following simultaneous equations:

$$\tilde{g} + h\tilde{d} = (h + \tilde{L})\tilde{U}\left((\tilde{D} - \tilde{L})\mathbf{I}^{(|\tilde{V}|)} + J\tilde{\Lambda}\right)^{-1}\tilde{U}^{\text{T}}\tilde{f}, \quad (10.405)$$

$$\tilde{\mathbf{f}} + \tilde{\mathbf{g}} + h\mathbf{d} = \tilde{\mathbf{U}} \left((\tilde{\mathbf{D}} - \tilde{\mathbf{L}}) \mathbf{I}^{(|\tilde{\mathbf{V}}|)} + J\tilde{\mathbf{\Lambda}} \right)^{-1} \tilde{\mathbf{U}}^{\mathbf{T}} \begin{pmatrix} \frac{\tilde{f}_1}{\sqrt{\tilde{f}_1^2 + \Gamma^2}} \tanh\left(\frac{1}{k_{\text{B}}T} \sqrt{\tilde{f}_1^2 + \Gamma^2}\right) \\ \frac{\tilde{f}_2}{\sqrt{\tilde{f}_2^2 + \Gamma^2}} \tanh\left(\frac{1}{k_{\text{B}}T} \sqrt{\tilde{f}_2^2 + \Gamma^2}\right) \\ \vdots \\ \frac{\tilde{f}_{|\tilde{\mathbf{V}}|}}{\sqrt{\tilde{f}_{|\tilde{\mathbf{V}}|}^2 + \Gamma^2}} \tanh\left(\frac{1}{k_{\text{B}}T} \sqrt{\tilde{f}_{|\tilde{\mathbf{V}}|}^2 + \Gamma^2}\right) \end{pmatrix}, \quad (10.406)$$

$$\tilde{\mathbf{L}} = -h + \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{|\tilde{\mathbf{V}}|} (\tilde{\mathbf{f}} + \tilde{\mathbf{g}} + h\mathbf{d})^{\mathbf{T}} (\tilde{\mathbf{f}} + \tilde{\mathbf{g}} + h\mathbf{d})}, \quad (10.407)$$

$$\frac{1}{\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{|\tilde{\mathbf{V}}|} (\tilde{\mathbf{f}} + \tilde{\mathbf{g}} + h\mathbf{d})^{\mathbf{T}} (\tilde{\mathbf{f}} + \tilde{\mathbf{g}} + h\mathbf{d})}} = \frac{1}{|\tilde{\mathbf{V}}|} \text{Tr} \left[\left((\tilde{\mathbf{D}} - \tilde{\mathbf{L}}) \mathbf{I}^{(|\tilde{\mathbf{V}}|)} + J\tilde{\mathbf{\Lambda}} \right)^{-1} \right]. \quad (10.408)$$

10.5.4 Suzuki-Trotter Decomposition in the Transverse Ising Model

In this section, we review the Suzuki-Trotter formulas and extensions from conventional quantum loop propagation using them. In quantum probabilistic graphical models, the state space is defined by all the eigenvectors of the density matrix \mathbf{R} and the probability of each eigenvector is given by the eigenvalue as mentioned in Sect. 10.4.2. To compute some statistical quantities by using the Monte Carlo method, it is necessary to diagonalize the energy matrix \mathbf{H} , which is a massive computation. Instead of such a scheme, quantum Monte Carlo methods based on the Suzuki-Trotter formulas were proposed [89]. One important part of the quantum Monte Carlo method is the mapping from a quantum probabilistic graphical model to a conventional (classical) probabilistic graphical model by introducing the techniques of Suzuki-Trotter decompositions. It is known that some statistical quantities for conventional (classical) probabilistic graphical models can be computed by MCMC methods. This is a basic idea behind quantum Monte Carlo methods. Let us first review the Suzuki-Trotter formulas and explicitly give a detailed scheme of Suzuki-Trotter decompositions for the transverse Ising model in Eqs. (10.226) and (10.227) with Eq. (10.301).

From the definition of the exponential function for square matrices, we have

$$\exp(x(\mathbf{A} + \mathbf{B})) = \mathbf{I} + x(\mathbf{A} + \mathbf{B}) + \frac{1}{2}x^2(\mathbf{A} + \mathbf{B})^2 + \mathcal{O}(x^3) \quad (x \rightarrow 0), \tag{10.409}$$

$$\exp(x\mathbf{A}) = \mathbf{I} + x\mathbf{A} + \frac{1}{2}x^2\mathbf{A}^2 + \mathcal{O}(x^3) \quad (x \rightarrow +0), \tag{10.410}$$

$$\exp(x\mathbf{B}) = \mathbf{I} + x\mathbf{B} + \frac{1}{2}x^2\mathbf{B}^2 + \mathcal{O}(x^3) \quad (x \rightarrow +0). \tag{10.411}$$

From these equalities, the following formula can be confirmed:

$$\exp(x(\mathbf{A} + \mathbf{B})) = \exp(x\mathbf{A})\exp(x\mathbf{B}) + \mathcal{O}(x^2) \quad (x \rightarrow +0). \tag{10.412}$$

Moreover, we have

$$\exp(x\mathbf{A}) = \left[\exp\left(\frac{x}{M}\mathbf{A}\right) \right]^M + \mathcal{O}\left(\frac{x^2}{M}\right) \quad (x^2 \ll M), \tag{10.413}$$

$$\exp(x(\mathbf{A} + \mathbf{B})) = \left[\exp\left(\frac{x}{M}\mathbf{A}\right)\exp\left(\frac{x}{M}\mathbf{B}\right) \right]^M + \mathcal{O}\left(\frac{x^2}{M}\right) \quad (x^2 \ll M). \tag{10.414}$$

Generally, for a graph (V, E) with the set of nodes $V = \{1, 2, \dots, N\}$ and the set of edges $E = \{i, j\}$, we have

$$\exp\left(x \sum_{\{i,j\} \in E} \mathbf{A}_{\{i,j\}}\right) = \left[\prod_{\{i,j\} \in E} \exp\left(\frac{x}{M}\mathbf{A}_{\{i,j\}}\right) \right]^M + \mathcal{O}\left(\frac{x^2}{M}\right) \quad (x^2 \ll M), \tag{10.415}$$

$$\begin{aligned} & \exp\left(x \sum_{\{i,j\} \in E} \mathbf{A}_{\{i,j\}} + x \sum_{\{i,j\} \in E} \mathbf{B}_{\{i,j\}}\right) \\ &= \left[\left(\prod_{\{i,j\} \in E} \exp\left(\frac{x}{M}\mathbf{A}_{\{i,j\}}\right) \right) \left(\prod_{\{i,j\} \in E} \exp\left(\frac{x}{M}\mathbf{B}_{\{i,j\}}\right) \right) \right]^M + \mathcal{O}\left(\frac{x^2}{M}\right) \quad (x^2 \ll M). \end{aligned} \tag{10.416}$$

These are referred to as a **Suzuki-Trotter Decomposition** [87, 88].

For the case of $N = 2$, we consider an energy matrix \mathbf{H} defined by

$$\mathbf{H} = -J\sigma_1^z\sigma_2^z - h_1\sigma_1^z - h_2\sigma_2^z - \Gamma\sigma_1^x - \Gamma\sigma_2^x. \tag{10.417}$$

It is referred to as a quantum transverse Ising model on a chain ($V = \{1, 2\}$, $E = \{\{1, 2\}\}$) with three nodes and two edges. By using the above Suzuki-Trotter formula, we have

$$\begin{aligned}
& \langle s_{1,1}, s_{2,1} | \exp\left(-\frac{1}{k_B T} \mathbf{H}\right) | s'_{1,1}, s'_{2,2} \rangle \\
&= \lim_{M \rightarrow +\infty} \left[\exp\left(\frac{1}{k_B T M} (J \sigma_1^z \sigma_2^z + h_1 \sigma_1^z + h_2 \sigma_2^z)\right) \exp\left(\frac{1}{k_B T M} (\Gamma \sigma_1^x + \Gamma \sigma_2^x)\right) \right]^M \\
&= \lim_{M \rightarrow +\infty} \sum_{\tau_{1,1} \in \Omega} \sum_{\tau_{2,1} \in \Omega} \sum_{s_{1,2} \in \Omega} \sum_{s_{2,2} \in \Omega} \sum_{\tau_{1,2} \in \Omega} \sum_{\tau_{2,2} \in \Omega} \cdots \sum_{s_{1,M} \in \Omega} \sum_{s_{2,M} \in \Omega} \delta_{s_{1,M+1}, s'_{1,1}} \delta_{s_{2,M+1}, s'_{2,1}} \\
&\quad \times \prod_{m=1}^M \left(\langle s_{1,m}, s_{2,m} | \exp\left(\frac{1}{k_B T M} (J \sigma_1^z \sigma_2^z + h_1 \sigma_1^z + h_2 \sigma_2^z)\right) | \tau_{1,m}, \tau_{2,m} \rangle \right. \\
&\quad \left. \times \langle \tau_{1,m}, \tau_{2,m} | \exp\left(\frac{1}{k_B T M} (\Gamma \sigma_1^x + \Gamma \sigma_2^x)\right) | s_{1,m+1}, s_{2,m+1} \rangle \right). \quad (10.418)
\end{aligned}$$

Note that

$$\Gamma \sigma_1^x + \Gamma \sigma_2^x = \Gamma \sigma^x \otimes I + \Gamma I \otimes \sigma^x. \quad (10.419)$$

By using Eq. (10.253), Eq. (10.418) can be rewritten as

$$\begin{aligned}
& \langle s_{1,1}, s_{2,1} | \exp\left(-\frac{1}{k_B T} \mathbf{H}\right) | s'_{1,1}, s'_{2,1} \rangle \\
&= \lim_{M \rightarrow +\infty} \sum_{s_{1,2} \in \Omega} \sum_{s_{2,2} \in \Omega} \cdots \sum_{s_{1,M} \in \Omega} \sum_{s_{2,M} \in \Omega} \delta_{s_{1,M+1}, s'_{1,1}} \delta_{s_{2,M+1}, s'_{2,1}} \\
&\quad \times \prod_{m=1}^M \left(\exp\left(\frac{1}{k_B T M} (J s_{1,m} s_{2,m} + h_1 s_{1,m} + h_2 s_{2,m})\right) \right. \\
&\quad \left. \times \langle s_{1,m}, s_{2,m} | \left(\exp\left(\frac{1}{k_B T M} \Gamma \sigma^x\right) \otimes I \right) \left(I \otimes \exp\left(\frac{1}{k_B T M} \Gamma \sigma^x\right) \right) | s_{1,m+1}, s_{2,m+1} \rangle \right). \quad (10.420)
\end{aligned}$$

Moreover, by the definition of the tensor product for $A \otimes I$ and $I \otimes A$ for any matrix A in terms of Eq. (10.238), we have

$$\begin{aligned}
& \langle s_{1,m}, s_{2,m} | \left(\exp\left(\frac{1}{k_B T M} \Gamma \sigma^x\right) \otimes I \right) \left(I \otimes \exp\left(\frac{1}{k_B T M} \Gamma \sigma^x\right) \right) | s_{1,m+1}, s_{2,m+1} \rangle \\
&= \langle s_{1,m} | \exp\left(\frac{1}{k_B T M} \Gamma \sigma^x\right) | s_{1,m+1} \rangle \langle s_{2,m} | \exp\left(\frac{1}{k_B T M} \Gamma \sigma^x\right) | s_{2,m+1} \rangle \\
&= \langle s_{1,m} | \begin{pmatrix} \cosh\left(\frac{1}{k_B T M} \Gamma\right) & \sinh\left(\frac{1}{k_B T M} \Gamma\right) \\ \sinh\left(\frac{1}{k_B T M} \Gamma\right) & \cosh\left(\frac{1}{k_B T M} \Gamma\right) \end{pmatrix} | s_{1,m+1} \rangle \\
&\quad \times \langle s_{2,m} | \begin{pmatrix} \cosh\left(\frac{1}{k_B T M} \Gamma\right) & \sinh\left(\frac{1}{k_B T M} \Gamma\right) \\ \sinh\left(\frac{1}{k_B T M} \Gamma\right) & \cosh\left(\frac{1}{k_B T M} \Gamma\right) \end{pmatrix} | s_{2,m+1} \rangle. \quad (10.421)
\end{aligned}$$

Equation (10.420) can be rewritten in terms of the two-dimensional representation as

$$\begin{aligned}
& \langle s_{1,1}, s_{2,1} | \exp\left(-\frac{1}{k_B T} \mathbf{H}\right) | s'_{1,1}, s'_{2,1} \rangle \\
&= \lim_{M \rightarrow +\infty} \sum_{s_{1,2} \in \Omega} \sum_{s_{2,2} \in \Omega} \cdots \sum_{s_{1,M} \in \Omega} \sum_{s_{2,M} \in \Omega} \delta_{s_{1,M+1}, s'_{1,1}} \delta_{s_{2,M+1}, s'_{2,1}} \\
&\quad \times \prod_{m=1}^M \left(\exp\left(\frac{1}{k_B T M} (J s_{1,m} s_{2,m} + h_1 s_{1,m} + h_2 s_{2,m})\right) \right. \\
&\quad \times \langle s_{1,m} | \left. \begin{pmatrix} \cosh\left(\frac{1}{k_B T M} \Gamma\right) & \sinh\left(\frac{1}{k_B T M} \Gamma\right) \\ \sinh\left(\frac{1}{k_B T M} \Gamma\right) & \cosh\left(\frac{1}{k_B T M} \Gamma\right) \end{pmatrix} | s_{1,m+1} \right) \\
&\quad \times \langle s_{2,m} | \left. \begin{pmatrix} \cosh\left(\frac{1}{k_B T M} \Gamma\right) & \sinh\left(\frac{1}{k_B T M} \Gamma\right) \\ \sinh\left(\frac{1}{k_B T M} \Gamma\right) & \cosh\left(\frac{1}{k_B T M} \Gamma\right) \end{pmatrix} | s_{2,m+1} \right) \Big). \quad (10.422)
\end{aligned}$$

Eventually, the density matrix \mathbf{P} of the transverse Ising model for two nodes in Eq. (10.417), such that

$$\mathbf{P} \equiv \frac{\exp\left(-\frac{1}{k_B T} \mathbf{H}\right)}{\text{Tr}\left[\exp\left(-\frac{1}{k_B T} \mathbf{H}\right)\right]}, \quad (10.423)$$

can be reduced to the probability distribution $P^{(M)}(s_1, s_2, s_3, \dots, s_M, s_{M+1})$ for $s_m = \begin{pmatrix} s_{1,m} \\ s_{2,m} \end{pmatrix}$ ($m = 1, 2, \dots, M+1$) on the $2 \times (M+1)$ ladder graph as follows:

$$\begin{aligned}
& \langle s_{1,1}, s_{2,1} | \mathbf{P} | s'_{1,1}, s'_{2,1} \rangle \\
&= \lim_{M \rightarrow +\infty} \sum_{s_1 \in \Omega^2} \sum_{s_2 \in \Omega^2} \cdots \sum_{s_M \in \Omega^2} \sum_{s_{M+1} \in \Omega^2} \delta_{s_{1,M+1}, s'_{1,1}} \delta_{s_{2,M+1}, s'_{2,1}} P^{(M)}(s_1, s_2, s_3, \dots, s_M, s_{M+1}), \quad (10.424)
\end{aligned}$$

where

$$\begin{aligned}
& P^{(M)}(s_1, s_2, s_3, \dots, s_M, s_{M+1}) \\
&\equiv \frac{1}{Z^{(M)}} \prod_{m=1}^M \left(\exp\left(\frac{1}{k_B T M} (J s_{1,m} s_{2,m} + h_1 s_{1,m} + h_2 s_{2,m})\right) \right. \\
&\quad \times \exp\left(\frac{1}{k_B T M} (K(MT, \Gamma) s_{1,m} s_{1,m+1} + K(MT, \Gamma) s_{2,m} s_{2,m+1})\right) \Big), \quad (10.425)
\end{aligned}$$

$$\begin{aligned}
 Z^{(M)} \equiv & \sum_{s_{1,2} \in \Omega} \sum_{s_{2,2} \in \Omega} \cdots \sum_{s_{1,M} \in \Omega} \sum_{s_{2,M} \in \Omega} \delta_{s_{1,M+1}, s_{1,1}} \delta_{s_{2,M+1}, s_{2,1}} \\
 & \times \prod_{m=1}^M \left(\exp \left(\frac{1}{k_B T M} (J s_{1,m} s_{2,m} + h_1 s_{1,m} + h_2 s_{2,m}) \right) \right. \\
 & \left. \times \exp \left(\frac{1}{k_B T M} (K(MT, \Gamma) s_{1,m} s_{1,m+1} + K(MT, \Gamma) s_{2,m} s_{2,m+1}) \right) \right), \tag{10.426}
 \end{aligned}$$

$$K(T, \Gamma) \equiv k_B T \ln \left(\sqrt{\frac{\cosh\left(\frac{1}{k_B T} \Gamma\right)}{\sinh\left(\frac{1}{k_B T} \Gamma\right)}} \right). \tag{10.427}$$

The density matrix P in Eq. (10.423) of the transverse Ising model for $|V|$ nodes $V = \{1, 2, \dots, |V|\}$, which is given by Eqs. (10.226) and (10.227) with Eq. (10.301), can be reduced to the matrix representation for s probability distribution

$$P^{(M)}(s_1, s_2, s_3, \dots, s_M, s_{M+1}) \text{ for } s_m = \begin{pmatrix} s_{1,m} \\ s_{2,m} \\ \vdots \\ s_{N,m} \end{pmatrix} (m = 1, 2, \dots, M + 1) \text{ on the } |V| \times (M + 1)$$

ladder graph as follows:

$$\begin{aligned}
 & \langle s_{1,1}, s_{2,1}, \dots, s_{|V|,1} | P | s'_{1,1}, s'_{2,1}, \dots, s'_{|V|,1} \rangle \\
 = & \lim_{M \rightarrow +\infty} \sum_{s_1 \in \Omega^{|V|}} \sum_{s_2 \in \Omega^{|V|}} \cdots \sum_{s_M \in \Omega^{|V|}} \sum_{s_{M+1} \in \Omega^{|V|}} \left(\prod_{i \in V} \delta_{s_i, M+1}, s'_{i,1} \right) P^{(M)}(s_1, s_2, s_3, \dots, s_M, s_{M+1}), \tag{10.428}
 \end{aligned}$$

where

$$\begin{aligned}
 & P^{(M)}(s_1, s_2, s_3, \dots, s_M, s_{M+1}) \\
 \equiv & \frac{1}{Z^{(M)}} \prod_{m=1}^M \left(\prod_{\{i,j\} \in E} \exp \left(\frac{1}{k_B T M} J s_{i,m} s_{j,m} \right) \right) \\
 & \times \left(\prod_{i \in V} \exp \left(\frac{1}{k_B T M} K(MT, \Gamma) s_{i,m} s_{i,m+1} \right) \right) \left(\prod_{i \in V} \exp \left(\frac{1}{k_B T M} h_i s_{i,m} \right) \right), \tag{10.429}
 \end{aligned}$$

$$\begin{aligned}
Z^{(M)} \equiv & \sum_{s_1 \in \Omega^{|V|}} \sum_{s_2 \in \Omega^{|V|}} \cdots \sum_{s_M \in \Omega^{|V|}} \left(\prod_{i \in V} \delta_{s_{i,M+1}, s_{i,1}} \right) \\
& \times \prod_{m=1}^M \left(\prod_{(i,j) \in E} \exp\left(\frac{1}{k_B T M} J_{s_{i,m} s_{j,m}} \right) \right) \\
& \times \left(\prod_{i \in V} \exp\left(\frac{1}{k_B T M} K(MT, \Gamma) s_{i,m} s_{i,m+1} \right) \right) \left(\prod_{i \in V} \exp\left(\frac{1}{k_B T M} h_i s_{i,m} \right) \right).
\end{aligned} \tag{10.430}$$

The dynamics of quantum Monte Carlo methods based on Suzuki-Trotter decompositions have been analyzed by using Glauber dynamics [101, 102] and Langevin dynamics [103, 104]. Recently, these analyses are applied to some statistical machine learning systems with quantum annealing [105, 106]. Some statistical analysis of quantum Monte Carlo methods for statistical inferences based on Suzuki-Trotter decompositions [87, 88] are shown in Chaps. 12 and 13 of Part III of this book.

We now try to construct a modification of the conventional quantum message passing rule in Eq. (10.335) for the transverse Ising model in Eqs. (10.226) and (10.227) with Eq. (10.301) by imposing the assumption that all off-diagonal elements of $\lambda_{j \rightarrow i}$ and $\lambda_{i \rightarrow j}$ for any edge $(i, j) \in E$ are zero. By using the Suzuki-Trotter formulas in Eqs. (10.415)–(10.416), Eq. (10.335) can be represented as follows:

$$\begin{aligned}
& \lim_{M \rightarrow +\infty} \left[\exp\left(\frac{1}{k_B T M} \Gamma \sigma^x \right) \exp\left(\frac{1}{k_B T M} h d_i \sigma^z \right) \exp\left(\frac{1}{k_B T M} \lambda_{j \rightarrow i} \right) \exp\left(\frac{1}{k_B T M} \sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i} \right) \right]^M \\
& = \frac{Z_i}{Z_{[i,j]}} \lim_{M \rightarrow +\infty} \text{Tr}_i \left[\exp\left(\frac{1}{k_B T M} \Gamma \sigma^x \otimes I \right) \exp\left(\frac{1}{k_B T M} h d_i (\sigma^z \otimes I) \right) \right. \\
& \quad \times \exp\left(\frac{1}{k_B T M} J(\sigma^z \otimes I)(I \otimes \sigma^z) \right) \exp\left(\frac{1}{k_B T M} h d_j (I \otimes \sigma^z) \right) \exp\left(\frac{1}{k_B T M} \Gamma I \otimes \sigma^x \right) \\
& \quad \left. \times \exp\left(\frac{1}{k_B T M} I \otimes \left(\sum_{l \in \partial j \setminus \{i\}} \lambda_{l \rightarrow j} \right) \right) \exp\left(\frac{1}{k_B T M} \left(\sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i} \right) \otimes I \right) \right]^M,
\end{aligned} \tag{10.431}$$

such that,

$$\begin{aligned}
& \lim_{M \rightarrow +\infty} \sum_{s_{i,2} \in \Omega} \cdots \sum_{s_{i,M} \in \Omega} \left(\prod_{m=1}^M \langle s_{i,m} | \exp\left(\frac{1}{k_B T M} \sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i} \right) | s_{i,m} \rangle \exp\left(\frac{1}{k_B T M} h d_i s_{i,m} \right) \right) \\
& \quad \times \delta_{s'_{i,1}, s_{i,M+1}} \prod_{m=1}^M \left(\langle s_{i,m} | \exp\left(\frac{1}{k_B T M} \lambda_{j \rightarrow i} \right) | s_{i,m} \rangle \langle s_{i,m} | \exp\left(\frac{1}{k_B T M} \Gamma \sigma^x \right) | s_{i,m+1} \rangle \right) \\
& = \lim_{M \rightarrow +\infty} \sum_{s_{i,2} \in \Omega} \cdots \sum_{s_{i,M} \in \Omega} \left(\prod_{m=1}^M \langle s_{i,m} | \exp\left(\frac{1}{k_B T M} \sum_{k \in \partial i \setminus \{j\}} \lambda_{k \rightarrow i} \right) | s_{i,m} \rangle \exp\left(\frac{1}{k_B T M} h d_i s_{i,m} \right) \right) \\
& \quad \times \left(\delta_{s'_{i,1}, s_{i,M+1}} \frac{Z_i}{Z_{[i,j]}} \sum_{s_{j,1} \in \Omega} \sum_{s'_{j,1} \in \Omega} \delta_{s_{j,1}, s'_{j,1}} \right)
\end{aligned}$$

$$\begin{aligned}
& \times \sum_{s_{j,2} \in \Omega} \cdots \sum_{s_{j,M} \in \Omega} \delta_{s'_{j,1}, s_{j,M+1}} \prod_{m=1}^M \left(\exp \left(\frac{1}{k_B T M} (J s_{i,m} s_{j,m}) \right) \exp \left(\frac{1}{k_B T M} h d_j s_{j,m} \right) \right) \\
& \times \langle s_{j,m} | \exp \left(\frac{1}{k_B T M} \sum_{l \in \partial j \setminus \{i\}} \lambda_{l \rightarrow j} \right) | s_{j,m} \rangle \\
& \times \langle s_{i,m} | \exp \left(\frac{1}{k_B T M} \Gamma \sigma^x \right) | s_{i,m+1} \rangle \langle s_{j,m} | \exp \left(\frac{1}{k_B T M} \Gamma \sigma^x \right) | s_{j,m+1} \rangle \Big). \tag{10.432}
\end{aligned}$$

The sufficient conditions for Eq. (10.432) are given by

$$\begin{aligned}
& \delta_{s'_{i,1}, s_{i,M+1}} \prod_{m=1}^M \left(\langle s_{i,m} | \exp \left(\frac{1}{k_B T M} \lambda_{j \rightarrow i} \right) | s_{i,m} \rangle \langle s_{i,m} | \exp \left(\frac{1}{k_B T M} \Gamma \sigma^x \right) | s_{i,m+1} \rangle \right) \\
& = \delta_{s'_{i,1}, s_{i,M+1}} \frac{Z_i}{Z_{\{i,j\}}} \sum_{s_{j,1} \in \Omega} \sum_{s'_{j,1} \in \Omega} \delta_{s_{j,1}, s'_{j,1}} \\
& \quad \times \sum_{s_{j,2} \in \Omega} \cdots \sum_{s_{j,M} \in \Omega} \delta_{s'_{j,1}, s_{j,M+1}} \prod_{m=1}^M \left(\exp \left(\frac{1}{k_B T M} J s_{i,m} s_{j,m} \right) \exp \left(\frac{1}{k_B T M} h d_j s_{j,m} \right) \right) \\
& \quad \times \langle s_{j,m} | \exp \left(\frac{1}{k_B T M} \sum_{l \in \partial j \setminus \{i\}} \lambda_{l \rightarrow j} \right) | s_{j,m} \rangle \\
& \quad \times \langle s_{i,m} | \exp \left(\frac{1}{k_B T M} \Gamma \sigma^x \right) | s_{i,m+1} \rangle \langle s_{j,m} | \exp \left(\frac{1}{k_B T M} \Gamma \sigma^x \right) | s_{j,m+1} \rangle \Big). \tag{10.433}
\end{aligned}$$

By taking the summations $\sum_{s_{j,2} \in \Omega} \cdots \sum_{s_{j,M} \in \Omega}$ and the limit $M \rightarrow +\infty$ on both sides of Eq. (10.433), modified message passing rules can be derived as follows:

$$\begin{aligned}
& \exp \left(\frac{1}{k_B T} \lambda_{j \rightarrow i} + \frac{1}{k_B T} \Gamma \sigma^x \right) \\
& = \frac{Z_i}{Z_{\{i,j\}}} \text{Tr}_i \left[\exp \left(\frac{1}{k_B T} \left(J (\sigma^z \otimes \mathbf{I}) (\mathbf{I} \otimes \sigma^z) + \Gamma (\sigma^x \otimes \mathbf{I}) \right) \right. \right. \\
& \quad \left. \left. + \frac{1}{k_B T} \mathbf{I} \otimes \left(h d_j \sigma^z + \Gamma \sigma^x + \sum_{l \in \partial j \setminus \{i\}} \lambda_{l \rightarrow j} \right) \right) \right]. \tag{10.434}
\end{aligned}$$

We remark that the modified message passing rules of Eq. (10.434) can be derived by considering the Bethe free energy functional in the cluster variation method with a ladder-type basic cluster for the probabilistic graphical model [107] in Eqs. (10.429)–(10.430). While the conventional framework of quantum belief propagations was given as a quantum cluster variation method in Ref. [90], some extensions of loopy belief propagations have been proposed in Refs. [108–111] from a quantum statistical mechanical standpoint.

10.6 Concluding Remarks

This chapter explored sublinear modeling based on statistical mechanical informatics for statistical machine learning. In statistical machine learning, we need to compute some statistical quantities in massive probabilistic graphical models. Statistical mechanical informatics can provide us with many statistical approximate computational techniques. One is the advanced mean-field framework, which includes mean-field methods and loopy belief propagation methods such as the Bethe approximation. The advanced mean-field framework can provide good accuracy for statistical quantities, including averages and covariances. Some statistical quantities in probabilistic graphical models sometimes have phase transitions when computing the advanced mean-field method. As we have already shown in Sect. 10.3.3, we have two familiar phase transitions, namely, the first- and second-order phase transitions. Each step of the EM algorithm is often affected by the first-order phase transition because the internal energy in the prior probabilistic model has a discontinuity. This difficulty appears in the convergence procedure of the EM algorithm, in which the trajectory of a hyperparameter passes through not only the equilibrium state but also metastable and unstable states in the loopy belief propagation of probabilistic segmentations in Sect. 10.3.5. We show that some algorithms based on loopy belief propagation in probabilistic segmentations can be accelerated by the inverse real-space renormalization group techniques in Sect. 10.3.6.

The second part of this chapter explored quantum statistical machine learning and some statistical approximate algorithms in quantum statistical mechanical informatics for realizing the framework. Quantum mechanical computations for machine learning are rapidly developing in terms of both academic research and industrial implementation. In Sect. 10.4, we explained the modeling framework of density matrices and some fundamental mathematics for it and expanded the modeling framework to the quantum expectation-maximization algorithm. In Sect. 10.5, we showed the fundamental frameworks of quantum loopy belief propagation and quantum statistical mechanical extensions of the adaptive TAP method. Moreover, we reviewed the Suzuki-Trotter expansion, and the real and the momentum space renormalization group for sublinear modeling of density matrices.

Recently, we have the framework of massive fundamental mathematical modeling in the statistical machine learning theory for many practical applications, such that, mainly the sparse modeling [4, 5] and the deep learning [10]. Many academic researchers are interested in interpretations of such modelings in the stand point of probabilistic graphical models in the statistical mathematics [2, 3, 7] and the statistical mechanical informatics [8, 9, 13, 17]. Now we have novel technologies for realizing quantum computing in the stand point of quantum mechanical extensions of the statistical mechanical informatics, such that, for example, **D-wave Quantum Annealer**. Some results in which the D-Wave quantum annealers have achieved high performance computing have appeared in Refs. [112–116]. Some recent developments of the probabilistic graphical modelings and their static and dynamical analysis of the advanced mean-field methods and the Suzuki-Trotter decompositions as well as

the replica methods for realizing sublinear modeling are shown in the subsequent Chaps. 12 and 13 of the present part of this book, in the statistical mechanical point of view.

Acknowledgements This work was partly supported by the JST-CREST program (No. JPMJCR1402) of the Japan Science and Technology Agency and a JSPS KAKENHI Grant (No.18H03303) from the Ministry of Education, Culture, Sports, Science, and Technology. The author is thankful for some valuable comments from Profs. Masayuki Ohzeki and Manaka Okuyama of Tohoku University and Prof. Muneki Yasuda of Yamagata University in Japan, Prof. Federico Ricci-Tersenghi of Università di Roma, La Sapienza in Italy, and Prof. Anthony C.C. Coolen of Radboud University in the Netherlands.

References

1. J.C.D. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003)
2. D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques* (MIT Press, 2009)
3. P.K. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012)
4. I. Rish, G.Y. Grabarnik, *Sparse Modelling: Theory, Algorithms, and Applications* (Chapman & Hall/CRC, 2015)
5. T. Hastie, R. Tibshirani, M.J. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations* (Chapman & Hall/CRC, 2015)
6. J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the Theory of Neural Computation* (CRC Press, 1991)
7. W. Jim, D. Kay, *Statistics and Neural Networks: Advances at the Interface*. Royal Statistical Society Lecture Notes Series, vol. 5, ed. by M. Titterton (Oxford University Press, 2000)
8. A. Engel, C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press, 2001)
9. A.C.C. Coolen, R. Kühn, P. Sollich, *Theory of Neural Information Processing Systems* (Oxford University Press, 2005)
10. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*. Adaptive Computation and Machine Learning Series (MIT Press, 2016)
11. M.J. Wainwright, M.I. Jordan, *Graphical Models, Exponential Families, and Variational Inference*. Foundations and Trends®. Mach. Learn. **1**(1–2), 1–305 (2008). <https://doi.org/10.1561/2200000001>
12. M. Opper, D. Saad (eds.), *Advanced Mean Field Methods—Theory and Practice* (MIT Press, 2001)
13. H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: Introduction*. International Series of Monographs on Physics (Oxford Science Publications, 2001)
14. K. Tanaka, Statistical-mechanical approach to image processing (Topical Review). J. Phys. A Math. Gen. **35**(37), R81–R150 (2002). <https://doi.org/10.1088/0305-4470/35/37/201>
15. T. Tanaka, A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors. IEEE Trans. Inf. Theory **48**(11), 2888–2910 (2002). <https://doi.org/10.1109/TIT.2002.804053>
16. Y. Kabashima, D. Saad, Statistical mechanics of low-density parity-check codes (Topical Review). J. Phys. A Math. Gen. **37**(6), R1–R43 (2004). <https://doi.org/10.1088/0305-4470/37/6/R01>
17. M. Mézard, A. Montanari, *Information, Physics and Computation* (Oxford University Press, 2009)

18. A.K. Hartmann, H. Rieger, *Optimization Algorithms in Physics* (Wiley-VCH, 2001)
19. K.A. Hartmann, M. Weigt, *Phase Transitions in Combinatorial Optimization Problems* (Wiley-VCH, 2005)
20. T. Kadowaki, H. Nishimori, Quantum annealing in the transverse Ising model. *Phys. Rev. E* **58**(5), 5, 5355–5263 (1998). <https://doi.org/10.1103/PhysRevE.58.5355>
21. A. Das, B.K. Chakrabarti (eds.), *Quantum Annealing and Related Optimization Methods*. Lecture Notes in Physics, vol. 679 (Springer, 2004)
22. C.C. McGeoch, in *Adiabatic Quantum Computation and Quantum Annealing: Theory and Practice* (Morgan & Claypool, 2014)
23. T. Albash, D.A. Lidar, Adiabatic quantum computation. *Rev. Modern Phys.* **90**(1) (2018). Article ID 015002. <https://doi.org/10.1103/RevModPhys.90.015002>
24. E.K. Grant, T.S. Humble, *Adiabatic Quantum Computing and Quantum Annealing*. Oxford Research Encyclopedias (Oxford University Press and the American Institute of Physics, 2020). <https://doi.org/10.1093/acrefore/9780190871994.013.32>
25. A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodological)* **39**(1), 1–38 (1977). With discussion
26. K.L. Mengersen, C.P. Robert, D. Michael Titterton, *Mixtures: Estimation and Applications* (Wiley, 2011)
27. J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in computational vision. *J. Am. Stat. Assoc.* **82**(397), 76–89 (1987). <https://doi.org/10.1080/01621459.1987.10478393>
28. W. Qian, D. Michael Titterton, Stochastic relaxations and em algorithms for Markov random fields. *J. Stat. Comput. Simul.* **40**(1–2), 55–69 (1992) <https://doi.org/10.1080/00949659208811365>
29. C. Andrieu, N. De Freitas, A. Doucet, M.I. Jordan, An introduction to MCMC for machine learning. *Mach. Learn.* **50**(1–2), 5–43 (2003). <https://doi.org/10.1023/A:1020281327116>
30. K. Tanaka, D. Michael Titterton, Statistical trajectory of approximate EM algorithm for probabilistic image processing. *J. Phys. A Math. Theor.* **40**(37), 11285–11300 (2007). <https://doi.org/10.1088/1751-8113/40/37/007>
31. J. Inoue, K. Tanaka, Dynamics of the maximum marginal likelihood hyperparameter estimation in image restoration: gradient descent versus expectation and maximization algorithm. *Phys. Rev. E* **65**(1) (2002). Article ID.016125. <https://doi.org/10.1103/PhysRevE.65.016125>
32. J. Roy, Glauber: time-dependent statistics of the Ising model. *J. Math. Phys.* **4**(2), 294–307 (1963). <https://doi.org/10.1063/1.1703954>
33. F.Y. Wu, The Potts model. *Rev. Modern Phys.* **54**(1), 235–268 (1982). <https://doi.org/10.1103/RevModPhys.54.235>
34. C. Domb, On the theory of cooperative phenomena in crystals. *Adv. Phys.* **9**(34), 149–244 (1960). <https://doi.org/10.1080/00018736000101189>
35. C. Domb, On the theory of cooperative phenomena in crystals. *Adv. Phys.* **9**(35), 245–361 (1960). <https://doi.org/10.1080/00018736000101199>
36. G. Parisi, *Statistical Field Theory* (Addison-Wesley, 1988)
37. H. Nishimori, G. Ortiz, *Elements of Phase Transitions and Critical Phenomena* (Oxford University Press, 2011)
38. D. Ruelle, *Statistical Mechanics: Rigorous Results* (Imperial College Press, 1969)
39. T. Morita, Variational principle for the distribution function of the effective field for the random Ising model in the Bethe approximation. *Phys. A Stat. Mech. Appl.* **98**(3), 566–572 (1979). [https://doi.org/10.1016/0378-4371\(79\)90154-7](https://doi.org/10.1016/0378-4371(79)90154-7)
40. T. Morita, Variational principle for regular and random Ising models on the cactus tree or on the usual lattice in the “cactus approximation.” *Phys. A* **105**(3), 620–630 (1981). [https://doi.org/10.1016/0378-4371\(81\)90115-1](https://doi.org/10.1016/0378-4371(81)90115-1)
41. T. Horiguchi, On the Bethe approximation for the random bond Ising model. *Phys. A Stat. Mech. Appl.* **107**(2), 360–370 (1981). [https://doi.org/10.1016/0378-4371\(81\)90095-9](https://doi.org/10.1016/0378-4371(81)90095-9)
42. T. Morita, Cluster variation method of cooperative phenomena and its generalization I. *J. Phys. Soc. Jpn.* **12**(10), 753–755 (1957). <https://doi.org/10.1143/JPSJ.12.753>

43. T. Morita, General structure of the distribution functions for the Heisenberg model and the Ising model. *J. Math. Phys.* **13**(1), 115–123 (1972). <https://doi.org/10.1063/1.1665840>
44. T. Morita, Cluster variation method and Möbius inversion formula. *J. Stat. Phys.* **59**(3–4), 819–825 (1990). <https://doi.org/10.1007/BF01025852>
45. T. Morita, Cluster variation method for non-uniform Ising and Heisenberg models and spin-pair correlation function. *Progress Theor. Phys.* **85**(2), 243–255 (1991). <https://doi.org/10.1143/ptp/85.2.243>
46. Y. Kabashima, D. Saad, Belief propagation vs. TAP for decoding corrupted messages. *Europhys. Lett.* **44**(5), 668–674 (1998). <https://doi.org/10.1209/epl/i1998-00524-7>
47. J.S. Yedidia, W.T. Freeman, Y. Weiss, Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Inf. Theory* **51**(7), 2282–2312 (2005). <https://doi.org/10.1109/TIT.2005.850085>
48. A. Pelizzola, Cluster variation method in statistical physics and probabilistic graphical models (Topical Review). *J. Phys. A Math. Gen.* **38**(2005), R309–R339 (2005). <https://doi.org/10.1088/0305-4470/38/33/R01>
49. D.L. Donoho, A. Maleki, A. Montanari, Message-passing algorithms for compressed sensing, in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 45 (2009), pp. 18914–18919. <https://doi.org/10.1073/pnas.0909892106>
50. T. Rizzo, A. Lage-Castellanos, F. Ricci-Tersenghi, Replica cluster variational method. *J. Stat. Phys.* **139**(3), 367–374 (2010). <https://doi.org/10.1007/s10955-010-9938-3>
51. M. Yasuda, S. Kataoka, K. Tanaka, Statistical analysis of loopy belief propagation in random fields. *Phys. Rev. E* **92**(4) (2015). Article ID. 042120. <https://doi.org/10.1103/PhysRevE.92.042120>
52. F. Krzakala, F. Ricci-Tersenghi, L. Zdeborova, R. Zecchina, E.W. Tramel, L.F. Cugliandolo, *Statistical Physics, Optimization, Inference and Message-Passing Algorithms*. Lecture Notes of the Les Houches School of Physics. Special Issue (Oxford University Press, 2013)
53. M. Welling, Y.W. Teh, Approximate inference in Boltzmann machines. *Artif. Intell.* **143**(1), 19–50 (2003). [https://doi.org/10.1016/S0004-3702\(02\)00361-2](https://doi.org/10.1016/S0004-3702(02)00361-2)
54. M. Yasuda, S. Kataoka, K. Tanaka, Inverse problem in pairwise Markov random fields using loopy belief propagation. *J. Phys. Soc. Jpn.* **81**(4), 1–8 (2012). Article ID 044801. <https://doi.org/10.1143/JPSJ.81.044801>
55. F. Ricci-Tersenghi, The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods. *J. Stat. Mech. Theory Exp.* (2012). Article ID P08015. <https://doi.org/10.1088/1742-5468/2012/08/P08015>
56. T. Morita, T. Horiguchi, Exactly solvable model of a spin glass. *Solid State Commun.* **19**(9), 833–835 (1976). [https://doi.org/10.1016/0038-1098\(76\)90665-7](https://doi.org/10.1016/0038-1098(76)90665-7)
57. D.J. Thouless, P.W. Anderson, R.G. Palmer, Solution of ‘Solvable model of a spin glass’. *Philos. Mag. J. Theor. Exp. Appl. Phys.* **35**(3), 593–601 (1977). <https://doi.org/10.1080/14786437708235992>
58. T. Morita, T. Horiguchi, Exactly solvable model of a random classical Heisenberg magnet. *J. Phys. C Solid State Phys.* **10**(11), 1949–1961 (1977). <https://doi.org/10.1088/0022-3719/10/11/029>
59. M. Yasuda, K. Tanaka, The relationship between Plefka’s expansion and the cluster variation method. *J. Phys. Soc. Jpn.* **75**(8) (2006). Article ID 084006. <https://doi.org/10.1143/JPSJ.75.084006>
60. M. Yasuda, K. Tanaka, Approximate learning algorithm in Boltzmann machines. *Neural Comput.* **21**(11), 3130–3178 (2009). <https://doi.org/10.1162/neco.2009.08-08-844>
61. M. Yasuda, K. Tanaka, TAP equation for non-negative Boltzmann machine. *Philos. Mag.* **92**(1–3), 192–209 (2012). <https://doi.org/10.1080/14786435.2011.634856>
62. M. Yasuda, Y. Kabashima, K. Tanaka, Replica Plefka expansion of Ising systems. *J. Stat. Mech. Theory Exp.* **2012**(4) (2012). Article ID P04002. <https://doi.org/10.1088/1742-5468/2012/04/P04002>
63. E.W. Tramel, A. Drémeau, F. Krzakala, Approximate message passing with restricted Boltzmann machine priors. *J. Stat. Mech. Theory Exp.* **2016**(7) (2016). Article ID 073401 <https://doi.org/10.1088/1742-5468/2016/07/073401>

64. M. Gabrié, Mean-field inference methods for neural networks (Topical Review). *J. Phys. A Math. Theor.* **53**(23) (2020). Article ID 223002. <https://orcid.org/0000-0002-5989-1018>
65. K. Tanaka, J. Inoue, D.M. Titterton, Probabilistic image processing by means of Bethe approximation for Q-Ising model. *J. Phys. A Math. Gen.* **36**(43), 11023–11036 (2003). <https://doi.org/10.1088/0305-4470/36/43/025>
66. K. Tanaka, H. Shouno, M. Okada, D.M. Titterton, Accuracy of the Bethe approximation for hyperparameter estimation in probabilistic image processing. *J. Phys. A Math. Gen.* **37**(36), 8675–8696 (2004). <https://doi.org/10.1088/0305-4470/37/36/007>
67. S. Kataoka, M. Yasuda, K. Tanaka, Statistical performance analysis in probabilistic image processing. *J. Phys. Soc. Jpn.* **79**(2) (2010). Article ID 025001. <https://doi.org/10.1143/JPSJ.79.025001>
68. S. Kataoka, M. Yasuda, K. Tanaka, D.M. Titterton, Statistical analysis of the expectation-maximization algorithm with loopy belief propagation in Bayesian image modeling. *Philos. Mag. Study Condens Matter* **92**(1-3), 50–63 (2012). <https://doi.org/10.1080/14786435.2011.624558>
69. K. Tanaka, M. Yasuda, D. Michael Titterton, Bayesian image modelling by means of generalized sparse prior and loopy belief propagation. *J. Phys. Soc. Jpn.* **81**(11) (2012). Article ID 114802. <https://doi.org/10.1143/JPSJ.81.114802>
70. K. Tanaka, S. Kataoka, M. Yasuda, Y. Waizumi, C.-T. Hsu, Bayesian image segmentations by Potts prior and Loopy belief propagation. *J. Phys. Soc. Jpn.* **83**(12) (2014). Article ID 124002. <https://doi.org/10.7566/JPSJ.83.124002>
71. M.B. Hastings, Community detection as an inference problem. *Phys. Rev. E* **74**(3) (2006). Article ID 035102(R). <https://doi.org/10.1103/PhysRevE.74.035102>
72. A. Decelle, F. Krzakala, C. Moore, L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84**(6) (2011). Article ID 066106. <https://doi.org/10.1103/PhysRevE.84.066106>
73. S. Kataoka, T. Kobayashi, M. Yasuda, K. Tanaka, Community detection algorithm combining stochastic block model and attribute data clustering. *J. Phys. Soc. Jpn.* **85**(11) (2016). Article ID 114802. <https://doi.org/10.7566/JPSJ.85.114802>
74. B. McCoy, T.T. Wu, *The Two-Dimensional Ising Model* (Harvard University Press, 1973). ISBN: 9780674180758
75. R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, 1982). ISBN: 9780486462714
76. T. Horiguchi, Husimi-Temperley model under a random field. *J. Math. Phys.* **20**(8), 1774–1775 (1979). <https://doi.org/10.1063/1.524265>
77. M. Mezard, G. Parisi, M.A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, 1987)
78. K.H. Fisher, J.A. Hertz, *Spin Glasses* (Cambridge University Press, 1993)
79. Michel Talagrand, *Mean Field Models for Spin Glass. Volume I: Basic Examples* (Springer, 2011)
80. M. Talagrand, *Mean Field Models for Spin Glass. Volume II: Advanced Replica-Symmetry and Low Temperature* (Springer, 2011)
81. K. Tanaka, S. Kataoka, M. Yasuda, M. Ohzeki, Inverse renormalization group transformation in Bayesian image segmentations. *J. Phys. Soc. Jpn.* **84**(4) (2015). Article ID 045001. <https://doi.org/10.7566/JPSJ.84.045001>
82. D.A. Harville, *Matrix algebra from a statistician's perspective* (Springer, 1997)
83. S. Suzuki, J. Inoue, B.K. Chakrabarti, *Quantum Ising Phases and Transitions in Transverse Ising Models*. Lecture Notes in Physics Book, vol. 862 (Springer, 2013)
84. A. Dutta, G. Aeppli, B.K. Chakrabarti, U. Divakaran, T.F. Rosenbaum, D. Sen, *Quantum Phase Transitions in Transverse Field Spin Models: From Statistical Physics to Quantum Information* (Springer, 2015)
85. S.G. Ovchinnikov, V.V. Val'kov, *Hubbard Operators in the Theory of Strongly Correlated Electrons* (Imperial College Press, 2004). https://doi.org/10.1142/9781860945977_0001
86. H. Tasaki, *Physics and Mathematics of Quantum Many-Body Systems (Graduate Texts in Physics)* (Springer, 2020)

87. M. Suzuki, Relationship between d -dimensional quantal spin systems and $(d + 1)$ -dimensional Ising systems—equivalence, critical exponents and systematic approximants of the partition function and spin correlations. *Prog. Theor. Phys.* **56**(5), 1454–1469 (1976). <https://doi.org/10.1143/PTP.56.1454>
88. M. Suzuki, Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics. *J. Math. Phys.* **26**(4), 601–612 (1985). <https://doi.org/10.1063/1.526596>
89. J. Gubernatis, N. Kawashima, P. Werner, *Quantum Monte Carlo Methods: Algorithms for Lattice Models* (Cambridge University Press, 2016)
90. T. Morita, Cluster variation method of cooperative phenomena and its generalization II. Quantum statistics. *J. Phys. Soc. Jpn.* **12**(10), 1060–1063 (1957). <https://doi.org/10.1143/JPSJ.12.1060>
91. R. Kubo, The spin-wave theory as a variational method and Its application to antiferromagnetism. *Rev. Modern Phys.* **25**(1), 344–351 (1953). <https://doi.org/10.1103/RevModPhys.25.344>
92. T. Morita, An approximation scheme of the cluster variation method for quantum lattice gases. *Prog. Theor. Phys.* **92**(6), 1081–1093 (1994). <https://doi.org/10.1143/ptp/92.6.1081>
93. T. Morita, A Bose lattice gas equivalent to Heisenberg model and its QVM study. *J. Phys. Soc. Jpn.* **64**(4), 1211–1216 (1995). <https://doi.org/10.1143/JPSJ.64.1211>
94. R. Miyazaki, H. Nishimori, G. Ortiz, Real-space renormalization group for the transverse-field Ising model in two and three dimensions. *Phys. Rev. E* **83**(5) (2011). Article ID 051103. <https://doi.org/10.1103/PhysRevE.83.051103>
95. M. Opper, O. Winther, Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling. *Phys. Rev. E* **64**(5) (2001). Article ID 056131 <https://doi.org/10.1103/PhysRevE.64.056131>
96. M. Opper, O. Winther, Expectation consistent approximate inference. *J. Mach. Learn. Res.* **6**(73), 2177–2204 (2005). <https://doi.org/10.5555/1046920.1194917>
97. M. Yasuda, C. Takahashi, K. Tanaka, Perturbative interpretation of adaptive Thouless-Anderson-Palmer free energy. *J. Phys. Soc. Jpn.* **85**(7) (2016). Article ID 075001. <https://doi.org/10.7566/JPSJ.85.075001>
98. C. Takahashi, M. Yasuda, K. Tanaka, Adaptive Thouless-Anderson-Palmer equation for higher-order Markov random fields. *J. Phys. Soc. Jpn.* **89**(6) (2020). Article ID 064007. <https://doi.org/10.7566/JPSJ.89.064007>
99. K. Tanaka, M. Nakamura, S. Kataoka, M. Ohzeki, M. Yasuda, Momentum-space renormalization group transformation in Bayesian image modeling by Gaussian graphical model. *J. Phys. Soc. Jpn.* **87**(8), 1–2 (2018). Article ID 085001. <https://doi.org/10.7566/JPSJ.87.085001>
100. K. Tanaka, M. Ohzeki, M. Yasuda, Sublinear computational time modeling by momentum-space renormalization group theory in statistical machine learning procedures. *Rev. Socionetwork Strat.* **13**(2), 281–306 (2019). <https://doi.org/10.1007/s12626-019-00053-1>
101. J. Inoue, Deterministic flows of order-parameters in stochastic processes of quantum Monte Carlo method. *J. Phys. Conf. Ser.* **233** (2010). Article ID 012010. <https://doi.org/10.1088/1742-6596/233/1/012010>
102. J. Inoue, Pattern-recalling processes in quantum Hopfield networks far from saturation. *J. Phys. Conf. Ser.* **297** (2011). Article ID 012012. <https://doi.org/10.1088/1742-6596/297/1/012012>
103. M. Ohzeki, S. Okada, M. Terabe, S. Taguchi, Optimization of neural networks via finite-value quantum fluctuations. *Sci. Rep.* **8** (2018). Article ID 9950. <https://doi.org/10.1038/s41598-018-28212-4>
104. S. Arai, M. Ohzeki, K. Tanaka, Dynamics of order parameters of nonstoquastic Hamiltonians in the adaptive quantum Monte Carlo method. *Phys. Rev. E* **99**(3) (2019). Article ID 032120. <https://doi.org/10.1103/PhysRevE.99.032120>
105. S. Arai, M. Ohzeki, K. Tanaka, Teacher-student learning for a binary perceptron with quantum fluctuations. *J. Phys. Soc. Jpn.* **90**(7) (2021). Article ID 074002. <https://doi.org/10.7566/JPSJ.90.074002>

106. S. Arai, M. Ohzeki, K. Tanaka, Mean field analysis of reverse annealing for code-division multiple-access multiuser detection. *Phys. Rev. Res.* **3**(3) (2021). Article ID 033006. <https://doi.org/10.1103/PhysRevResearch.3.033006>
107. K. Tanaka, T. Horiguchi, T. Morita, Critical indices for the two-dimensional Ising model with nearest-neighbor and next-nearest-neighbor interactions. II. Strip cluster approximation. *Phys. A Stat. Mech. Appl.* **192**(4), 647–664 (1993). [https://doi.org/10.1016/0378-4371\(93\)90114-J](https://doi.org/10.1016/0378-4371(93)90114-J)
108. M.B. Hastings, Quantum belief propagation: an algorithm for thermal quantum systems. *Phys. Rev. B* **76**(20) (2007). Article ID 201102(R). <https://doi.org/10.1103/PhysRevB.76.201102>
109. M.S. Leifer, D. Poulin, Quantum graphical models and belief propagation. *Ann. Phys.* **323**(8), 1899–1946 (2008). <https://doi.org/10.1016/j.aop.2007.10.001>
110. F. Krzakala, A. Rosso, G. Semerjian, F. Zamponi, Path-integral representation for quantum spin models: application to the quantum cavity method and Monte Carlo simulations. *Phys. Rev. B* **78**(13) (2008). Article ID 134428. <https://doi.org/10.1103/PhysRevB.78.134428>
111. M. Ohzeki, Message-passing algorithm of quantum annealing with nonstoquastic Hamiltonian. *J. Phys. Soc. Jpn.* **88**(6) (2019). Article ID 061005. <https://doi.org/10.7566/JPSJ.88.061005>
112. N. Nishimura, K. Tanahashi, K. Sugauma, M.J. Miyama, M. Ohzeki, Item listing optimization for E-commerce websites based on diversity. *Front. Comput. Sci.* **1** (2019). Article ID 2. <https://doi.org/10.3389/fcomp.2019.00002>
113. S. Okada, M. Ohzeki, M. Terabe, S. Taguchi, Improving solutions by embedding larger sub-problems in a D-wave quantum annealer. *Sci. Rep.* **9** (2019). Article ID 2098
114. M. Ohzeki, Breaking limitation of quantum annealer in solving optimization problems under constraints. *Sci. Rep.* **10** (2020). Article ID 3126
115. A.S. Koshikawa, M. Ohzeki, T. Kadowaki, K. Tanaka, Benchmark test of black-box optimization using D-wave quantum annealer. *J. Phys. Soc. Jpn.* **90**(6) (2021). Article ID 064001. <https://doi.org/10.7566/JPSJ.90.064001>
116. T. Sato, M. Ohzeki, K. Tanaka, Assessment of image generation by quantum annealer. *Sci. Rep.* **11** (2021). Article ID 13523

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

