Statistical-mechanical Approaches to Probabilistic Image Processing
— Markov Random Field Model and Mean Field Theory —

Kazuyuki Tanaka

Department of Computer and Mathematical Sciences, Graduate School of Information Sciences, Tohoku University, Aramaki-aza-aoba 04, Aoba-ku, Sendai 980-8579, Japan

Summary: Hyperparameters that are treated in statistical methods of image restorations are determined by maximizing an evidence (marginal likelihood) in a standpoint of maximum likelihood estimation. A powerful algorithm for the maximization of the evidence in a Markov random field model can be constructed by using a cluster variation method, which is a familiar methodology for formulating a high-level approximation in massive probabilistic models. The cluster variation method can give us better accuracy than the mean field approximation. We compare the results which are obtained by using the cluster variation method with the ones by the mean-field approximation in some numerical experiments. The cluster variation method can provide us the values of hyperparameters with high accuracy. A part of the present investigations have already reported in Refs. [17] and [18].

1. Introduction.

Probabilistic information processing by means of graphical models is one of powerful methods in soft computing and machine learning. Their practical applications are expected[1, 2]. The framework of probabilistic information processing is often constructed by assuming some a priori informations as probability distributions and by using the Bayes formula. Markov random field (MRF) models are very useful a priori probabilities in the probabilistic image processing[3, 4]. In the MRF’s, the state of a pixel is dependent only on the states of its nearest neighbour pixels [5, 6]. In image processing, we have to treat a massive probabilistic model and hardly calculate some statistical quantities in such a model, exactly. Many authors employed a mean-field approximation (MFA) [7, 8, 9], to calculate approximate values of some statistical quantities. The MFA is one of familiar techniques for massive probabilistic models in the statistical mechanics. Moreover, the MRF model is very close to some statistical-mechanical models, e.g. spin glass models[10]. It is expected that the nature of the mechanism of image processing by means of the MRF models is investigated in the statistical-mechanical point of view.

The major difficulty in the image processing by using the MRF model is how the hyperparameters are determined. In familiar statistical approaches, the hyperparameters are estimated by using the evidence framework. It is hard exactly to calculate the evidence in such a massive probabilistic model, because the total number of all possible configuration is exponential order. It is necessary to apply something approximation to the calculation of evidence. In the statistical-mechanical point of view, some investigations for the evidence frameworks of the MRF models have already been done[11, 12], by employing the MFA. Many computer scientists pay attention to the MFA as a key application for massive probabilistic models because an exponential order of computatial complexity can be reduced to a polynomial order by the
MFA. However, it is known that the MFA is the first-level approximation, and the accuracy of the MFA is not sufficient. We have a cluster variation method (CVM) as an extended version of approximation in the MFA and was applied to the material sciences[13]. Morita reformulated the CVM by using a Möbius[14, 15]. The author applied the CVM to the Bayesian image restorations[16, 17, 18, 19].

The main purpose of this paper is to study the accuracy in the CVM and the MFA in image restorations by means of the MRF model and the evidence framework. A part of the present investigations have already done in Refs. [17] and [18]. In Sec. 2, the evidence framework for the MRF model in image restorations are summarized. In Sec. 3, the basic framework of the MFA and the CVM are formulated for the MRF model given as a form of Gibbs distribution. In Secs. 4 and 5, some numerical experiments and some concluding remarks are presented.

2. Bayesian Statistics and Image Processing.

In this section, we explain the framework of image restoration by means of probabilistic models in Bayes statistics.

We consider a $q$-valued image on a square lattice $I = \{(i, j)\}$, meaning that each pixel can take one of $q$ states. The lattice is assumed to consist of $|I|$ pixels and to satisfy the periodic boundary conditions, so that the lattice is on a torus. The true original and given degraded image are represented by $x = \{x_{i,j} | (i, j) \in I\}$ and $y = \{y_{i,j} | (i, j) \in I\}$, respectively. Each variable, $x_{i,j}$ and $y_{i,j}$, takes a value from $Q = \{0, 1, \ldots, q - 1\}$. In recovering the original image $x$ from the given degraded image $y$, we use some a priori properties of the original image $x$. We express the set of random variables representing the original and the degraded image by $X = \{X_{i,j} | (i, j) \in I\}$ and $Y = \{Y_{i,j} | (i, j) \in I\}$, respectively. In the Bayes formula, the a posteriori probability distribution $Pr\{X = x | Y = y\}$, that the original image is $x$ when the given degraded image is $y$, is expressed as

$$Pr\{X = x | Y = y\} = \frac{Pr\{Y = y | X = x\}Pr\{X = x\}}{\sum_z Pr\{Y = y | X = z\}Pr\{X = z\}},$$

where the summation $\sum_z$ takes over all possible images $z = \{z_{i,j} | (i, j) \in I\}$. $Pr\{X = x\}$ is the a priori probability distribution that the original image is $x$ and $Pr\{Y = y | X = x\}$ is the conditional probability distribution that the degraded image is $y$ when the original image is $x$. In the Bayesian statistics, the conditional probability distribution $Pr\{Y = y | X = x\}$ is referred to as a likelihood function for the original image $x$ when the degraded image is $y$. The restored image $\hat{x} = \{\hat{x}_{i,j} | (i, j) \in I\}$ can be determined so as to maximize the a posteriori probability distribution as follows:

$$\hat{x} = \arg \max_z Pr\{X = z | Y = y\}. \tag{2}$$

This criteria is called maximum a posteriori (MAP) estimation. On the other hand, the restored image $\hat{x}$ can be determined also by maximizing the posterior marginal probability distribution for each pixel defined by

$$Pr\{X_{i,j} = n | Y = y\} = \sum_z \delta(z_{i,j}, n)Pr\{X = z | Y = y\}, \tag{3}$$

so that we have

$$\hat{x}_{i,j} = \arg \max_{n \in Q} Pr\{X_{i,j} = n | Y = y\} \quad ((i, j) \in I). \tag{4}$$
Here $\delta(a, b)$ represents Kronecker's delta $\delta_{ab}$. This criteria is called maximum posterior marginal (MPM) estimation \[20\], \[21\].

In order to estimate those properties only from knowledge of the given degraded image $y$, we adopt two assumptions.

**Assumption (i).** We have a given degraded image $y$ which is obtained from the original image $x$ by changing the state of each pixel to another state by the same probability $p^*$, independently of the other pixels. Here $p^*$ is assumed to be less than $1/q$.

By assumption (i), the conditional probability distribution $Pr\{Y = y|X = x\}$ is given by

$$Pr\{Y = y|X = x\} = Pr\{Y = y|X = x, p^*\} \equiv \prod_{(i,j) \in I} P_{yx}(y_{i,j}|x_{i,j}, p^*), \quad (5)$$

where

$$P_{yx}(n|m, p^*) \equiv (1 - \delta(n, m))p^* + (1 - (q - 1)p^*)\delta(n, m). \quad (6)$$

The conditional probability distribution can be rewritten as

$$Pr\{Y = y|X = x, p^*\} = \frac{\exp(-\beta(p^*)d(x, y))}{(q - 1 + \exp(-\beta(p^*))^{\Pi}}, \quad (7)$$

where

$$d(x, y) \equiv \sum_{(i,j) \in I} (1 - \delta(x_{i,j}, y_{i,j})), \quad (8)$$

$$\beta(p^*) \equiv \ln\left(\frac{1 - (q - 1)p^*}{p^*}\right). \quad (9)$$

The summation $\sum_{(i,j) \in I}$ is taken over all the pixels $(i, j)$. In the present paper, we treat only the case that $\beta(p^*)$ is positive since $p^* < 1/q$. This means that, in all the formulations given in this paper, we adopt the Gibbs canonical distribution as the probability distribution for the degradation process, as shown in Eq. (7).

In this section, we assume that the a priori probability distribution has a form of Gibbs canonical distribution and give the mathematical framework for the determination of hyperparameters by means of the evidence framework\[12\], \[11\].

First, we explain about the formulation for a posteriori probability distribution as a Gibbs canonical distribution. We can construct the a posteriori probability distribution by using the Bayes formula under the assumptions (i) and the following assumptions:

**Assumption (ii).** The a priori probability distribution, that the original image is $x$, is given by

$$Pr\{X = x\} = Pr\{X = x|K^*\} \equiv \frac{\exp(-K^*\sigma_2(x))}{\sum_x \exp(-K^*\sigma_2(z))}. \quad (10)$$

This assumption means that the original image is one of images that are produced with large probability in the a priori probability distribution (10).
We remark that the \textit{a priori} probability distribution given by Eq. (10) can be regarded as a Gibbs canonical distribution in statistical mechanics. The probabilistic model given in Eq. (10) is called \textit{q}-state Potts model in the statistical mechanics[22]. In the case of \( q = 2 \), this model is called Ising model[23] By substituting Eqs. (7) and (10) into Eq. (1), the \textit{a posteriori} probability distribution is given by

\[
\Pr\{X = x|Y = y\} = \Pr\{X = x|Y = y, p^*, K^*\}
\]

\[
= \frac{\exp(-\beta(p^*)d(x, y) - K^*\sigma_2(x))}{\sum_{z}\exp(-\beta(p^*)d(z, y) - K^*\sigma_2(z))}.
\]

\[
(11)
\]

For the values of \( p^* \) and \( K^* \), the restored image \( \hat{x} \) is given by maximizing the \textit{posterior} marginal probability distribution \( \Pr\{X_{i,j} = n|Y = y, p^*, K^*\} \) defined by Eq. (3).

In the standpoint of the evidence framework, the estimates of hyperparameters, \( \hat{p} \) and \( \hat{K} \), are determined so as to maximize an evidence

\[
\Pr\{Y = y|p, K\} \equiv \sum_{z} \Pr\{Y = y|X = z, p\} \Pr\{X = z|K\},
\]

\[
(12)
\]

with respect to the hyperparameters, \( p \) and \( K \), so that we have

\[
(\hat{p}, \hat{K}) = \text{argmax}_{(p, K)} \Pr\{Y = y|p, K\}.
\]

\[
(13)
\]

The restored image \( \hat{x} \) for the estimates of hyperparameters, \( \hat{p} \) and \( \hat{K} \), can be obtained from Eq. (4). This quantity \( \Pr\{Y = y|p, K\} \) is called evidence or marginal likelihood. From the extremum conditions for \( \Pr\{Y = y|p, K\} \), the deterministic equations for \( \hat{p} \) and \( \hat{K} \) are reduces to the following equations:

\[
\sum_{(i,j) \in \mathcal{I} \in \mathcal{Q}} \sum \delta(y_{i,j}, n)\Pr\{X_{i,j} = n|Y = y, \hat{p}, \hat{K}\} = (q - 1)\hat{p},
\]

\[
(14)
\]

\[
\sum_{(i,j) \in \mathcal{I} \in \mathcal{Q}m \in \mathcal{Q}} \sum \delta(n, m)\bigg(\Pr\{X_{i,j} = n, X_{i+1,j} = m|Y = y, \hat{p}, \hat{K}\}
\]

\[
+ \Pr\{X_{i,j} = n, X_{i,j+1} = m|Y = y, \hat{p}, \hat{K}\}\bigg)
\]

\[
= \sum_{(i,j) \in \mathcal{I} \in \mathcal{Q}m \in \mathcal{Q}} \sum \delta(n, m)\bigg(\Pr\{X_{i,j} = n, X_{i+1,j} = m|\hat{K}\}
\]

\[
+ \Pr\{X_{i,j} = n, X_{i,j+1} = m|\hat{K}\}\bigg),
\]

(15)

where

\[
\Pr\{X_{i,j} = n|Y = y, \hat{p}, \hat{K}\} \equiv \sum_{z} \delta(z_{i,j}, n)\Pr\{X = z|Y = y, \hat{p}, \hat{K}\},
\]

(16)

\[
\Pr\{X_{i,j} = n, X_{i+1,j} = n'|Y = y, \hat{p}, \hat{K}\}
\]

\[
\equiv \sum_{z} \delta(z_{i,j}, n)\delta(z_{i+1,j}, n')\Pr\{X = z|Y = y, \hat{p}, \hat{K}\},
\]

(17)

\[
\Pr\{X_{i,j} = n, X_{i,j+1} = n'|Y = y, \hat{p}, \hat{K}\}
\]

\[
\equiv \sum_{z} \delta(z_{i,j}, n)\delta(z_{i,j+1}, n')\Pr\{X = z|Y = y, \hat{p}, \hat{K}\},
\]

(18)

\[
\Pr\{X_{i,j} = n, X_{i+1,j} = n'|\hat{K}\}
\]

\[
\equiv \sum_{z} \delta(z_{i,j}, n)\delta(z_{i+1,j}, n')\Pr\{X = z|\hat{K}\},
\]

(19)

\[
\Pr\{X_{i,j} = n, X_{i,j+1} = n'|\hat{K}\}
\]

\[
\equiv \sum_{z} \delta(z_{i,j}, n)\delta(z_{i,j+1}, n')\Pr\{X = z|\hat{K}\}.
\]

(20)
For the estimates of hyperparameters, \( \hat{p} \) and \( \hat{K} \), the restored image \( \hat{x} \) in the evidence framework is given as follows:

\[
\hat{x}_{i,j} = \arg \max_{n \in Q} \Pr\{X_{i,j} = n | y, \hat{p}, \hat{K}\} \quad (i,j) \in \mathbf{I}.
\]  \hspace{1cm} (21)

In this framework, we have to calculate the marginal probability distributions for the massive probabilistic model. Some approximate methods for calculating these marginal probability distributions for such a massive probabilistic model will be given as the MFA and the CVM in Sec. 3.

### 3. Mean Field Theory and Cluster Variation Method.

In this section, we give the general frameworks of the CVM [16] to calculate some statistical quantities given in the above sections.

In order to determine the optimal values of hyperparameters and the restored images, we have to apply an approximation to the massive probabilistic model which is expressed in the form of Gibbs canonical distribution as follows:

\[
\rho(x) = \frac{\exp(-E(x))}{\sum_z \exp(-E(z))},
\]  \hspace{1cm} (22)

where

\[
E(x) = \sum_{(i,j) \in \mathbf{I}} \left( g_{i,j}(x_{i,j}) + e(x_{i,j}, x_{i+1,j}) + e(x_{i,j}, x_{i+1,j}) \right).
\]  \hspace{1cm} (23)

The Gibbs canonical distribution \( \rho(x) \) satisfies the variational principle for the minimization of the free energy defined by

\[
\mathcal{F}[\rho] = \sum_z \rho(z) \left( E(z) + \ln(\rho(z)) \right).
\]  \hspace{1cm} (24)

In order to explain about the MFA, we have to introduce a one-dimensional marginal probability distribution:

\[
\rho_{i,j}(n) = \sum_z \delta(n, z_{i,j}) \rho(z).
\]  \hspace{1cm} (25)

The marginal probability distributions satisfy the normalization conditions:

\[
\sum_{n \in Q} \rho_{i,j}(n) = 1.
\]  \hspace{1cm} (26)

In the MFA, the probability distribution of massive probabilistic model is approximately expressed in terms of the marginal probability distributions as follows:

\[
\rho(x) \approx \prod_{(i,j) \in \mathbf{I}} \rho_{i,j}(x_{i,j}).
\]  \hspace{1cm} (27)

By applying Eqs. (23), (25), and (27) to Eq. (24), the free energy \( \mathcal{F}[\rho] \) can be expressed in terms of the marginal probability distributions as follows:

\[
\mathcal{F}[\rho] \approx \mathcal{F}[\{\rho_{i,j}\}]
\equiv \sum_{(i,j) \in \mathbf{I}} \rho_{i,j}(n) \{ g_{i,j}(n) + \sum_{n' \in Q} e(n, n') \left( \rho_{i+1,j}(n') + \rho_{i,j+1}(n') \right) \} + \sum_{(i,j)} S_{i,j},
\]  \hspace{1cm} (28)
where
\[ S_{i,j} = - \sum_{n \in Q} \rho_{i,j}(n) \ln(\rho_{i,j}(n)), \tag{29} \]

By taking the first variation of \( F[\{\rho_{i,j}\}] \) with respect to the marginal probability distributions \( \rho_{i,j}(n) \) under the normalization conditions (26), the set of deterministic equations for the marginal probability distributions in the MFA is obtained as follows:

\[ \rho_{i,j}(n) = \frac{\exp\{-E_{i,j}^{\text{MF}}(n)\}}{\sum_{m \in Q} \exp\{-E_{i,j}^{\text{MF}}(m)\}} \quad (n \in Q, (i,j) \in \mathbf{I}), \tag{30} \]

\[ E_{i,j}^{\text{MF}}(n) = g_{i,j}(n) + \sum_{n' \in Q} e(n,n') \left( \rho_{i+1,j}(n') + \rho_{i-1,j}(n') + \rho_{i,j+1}(n') + \rho_{i,j-1}(n') \right). \tag{31} \]

In order to explain about the CVM, we have to introduce two-dimensional marginal probability distributions:

\[ \rho_{i,j}^{i+1,j}(n,n') = \sum_{z} \delta(n,z_{i,j}) \delta(n',z_{i+1,j}) \rho(z), \tag{32} \]

\[ \rho_{i,j}^{i,j+1}(n,n') = \sum_{z} \delta(n,z_{i,j}) \delta(n',z_{i,j+1}) \rho(z). \tag{33} \]

The two-dimensional marginal probability distributions satisfy the normalization conditions:

\[ \sum_{n \in Q} \sum_{n' \in Q} \rho_{i,j}^{i+1,j}(n,n') = \sum_{n \in Q} \sum_{n' \in Q} \rho_{i,j}^{i,j+1}(n,n') = 1, \tag{34} \]

and the reducibility conditions:

\[ \rho_{i,j}(n) = \sum_{n' \in Q} \rho_{i,j}^{i+1,j}(n,n') = \sum_{n' \in Q} \rho_{i,j}^{i,j+1}(n',n) = \sum_{n' \in Q} \rho_{i,j}^{i,j+1}(n,n') = \sum_{n' \in Q} \rho_{i,j-1}(n',n). \tag{35} \]

In the pair approximation of the CVM, the probability distribution of massive probabilistic model is approximately expressed in terms of the marginal probability distributions as follows:

\[ \rho(x) \approx \prod_{(i,j) \in \mathbf{I}} \rho_{i,j}(x_{i,j}) \left( \frac{\rho_{i,j}^{i+1,j}(x_{i,j},x_{i+1,j})}{\rho_{i,j}(x_{i,j}) \rho_{i+1,j}(x_{i+1,j})} \right) \left( \frac{\rho_{i,j}^{i,j+1}(x_{i,j},x_{i,j+1})}{\rho_{i,j}(x_{i,j}) \rho_{i,j+1}(x_{i,j+1})} \right). \tag{36} \]

By applying Eqs. (23), (25), (32), (33) and (36) to Eq. (24), the free energy \( F[\rho] \) can be expressed in terms of the marginal probability distributions as follows:

\[ F[\rho] \approx F[\{\rho_{i,j}, \rho_{i,j}^{i+1,j}, \rho_{i,j}^{i,j+1}\}] \equiv \sum_{(i,j)} \left\{ \sum_{n \in Q} q_{i,j}(n) \rho_{i,j}(n) + \sum_{n \in Q} \sum_{n' \in Q} e(n,n') \left( \rho_{i,j}^{i+1,j}(n,n') + \rho_{i,j}^{i,j+1}(n,n') \right) \right. 

\[ + S_{i,j} + (S_{i,j}^{i+1,j} - S_{i,j} - S_{i+1,j}) + (S_{i,j}^{i,j+1} - S_{i,j} - S_{i,j+1}) \}, \tag{37} \]

where

\[ S_{i,j}^{i+1,j} = - \sum_{n \in Q} \sum_{n' \in Q} \rho_{i,j}^{i+1,j}(n,n') \ln(\rho_{i,j}^{i+1,j}(n,n')), \tag{38} \]
By taking the first variation of $F\{\rho_{i,j}, \rho_{i,j}^{i+1,j}, \rho_{i,j}^{i,j+1}\}$ with respect to the marginal probability distributions $\rho_{i,j}(n)$, $\rho_{i,j}^{i+1,j}(n,n')$ and $\rho_{i,j}^{i,j+1}(n,n')$ under the normalization conditions (26) and (34), and the reducibility conditions (35), the set of deterministic equations for the marginal probability distributions in the CVM is obtained as follows:

\[
S_{i,j}^{i+1,j} \equiv - \sum_{n' \in Q} \sum_{n \in Q} \rho_{i,j}^{i+1,j}(n,n') \ln(\rho_{i,j}^{i+1,j}(n,n')).
\] (39)

\[
\rho_{i,j}(n) = \frac{\exp(-E_{i,j}(n))}{\sum_{m \in Q} \exp(-E_{i,j}(m))} (i,j) \in I,
\] (40)

\[
\rho_{i,j}^{i+1,j}(m,n) = \frac{\exp(-E_{i,j}^{i+1,j}(m,n))}{\sum_{m' \in Q} \sum_{n' \in Q} \exp(-E_{i,j}^{i+1,j}(m',n'))} (i,j) \in I,
\] (41)

\[
\rho_{i,j}^{i,j+1}(m,n) = \frac{\exp(-E_{i,j}^{i,j+1}(m,n))}{\sum_{m' \in Q} \sum_{n' \in Q} \exp(-E_{i,j}^{i,j+1}(m',n'))} (i,j) \in I,
\] (42)

\[
E_{i,j}(n) \equiv g_{i,j}(n) + \lambda_{i,j}^{i+1,j}(n) + \lambda_{i,j}^{i,j-1}(n) + \lambda_{i,j}^{i,j+1}(n) + \lambda_{i,j}^{i-1,j}(n),
\] (43)

\[
E_{i,j}^{i+1,j}(m,n) \equiv g_{i,j}(m) + g_{i+1,j}(n) + e(m,n) + \lambda_{i,j}^{i+1,j-1}(m) + \lambda_{i,j}^{i,j+1-1}(m) + \lambda_{i,j}^{i+1,j+1}(m) + \lambda_{i,j}^{i-1,j+1}(m) + \lambda_{i,j}^{i+2,j}(m) + \lambda_{i,j}^{i+1,j}(m) + \lambda_{i,j}^{i+1,j+1}(m) + \lambda_{i,j}^{i+1,j+1}(m),
\] (44)

\[
\lambda_{i,j}^{i+1,j}(n) = -g_{i,j}(n) - \lambda_{i,j}^{i-1,j}(n) - \lambda_{i,j}^{i,j+1}(n) - \lambda_{i,j}^{i,j+1}(n) - \lambda_{i,j}^{i,j+1}(n) - \ln\left(\sum_{m \in Q} \exp(-E_{i,j}(m))\right) - \ln\left(\sum_{m \in Q} \rho_{i,j}^{i+1,j}(m,n)\right),
\] (46)

\[
\lambda_{i,j}^{i+1,j}(n) = -g_{i+1,j}(n) - \lambda_{i+1,j}^{i+1,j}(n) - \lambda_{i+1,j}^{i+1,j+1}(n) - \lambda_{i,j}^{i,j+1-1}(n) - \ln\left(\sum_{m \in Q} \exp(-E_{i+1,j}(m))\right) - \ln\left(\sum_{m \in Q} \rho_{i,j}^{i+1,j}(m,n)\right),
\] (47)

\[
\lambda_{i,j}^{i,j+1}(n) = -g_{i,j}(n) - \lambda_{i,j}^{i,j+1}(n) - \lambda_{i,j}^{i,j+1}(n) - \lambda_{i,j}^{i,j+1}(n) - \ln\left(\sum_{m \in Q} \exp(-E_{i,j}(m))\right) - \ln\left(\sum_{m \in Q} \rho_{i,j}^{i,j+1}(m,n)\right),
\] (48)

\[
\lambda_{i,j}^{i,j+1}(n) = -g_{i,j+1}(n) - \lambda_{i,j+1}^{i,j+1}(n) - \lambda_{i,j+1}^{i,j+1}(n) - \lambda_{i,j+1}^{i,j+1}(n) - \ln\left(\sum_{m \in Q} \exp(-E_{i,j+1}(m))\right) - \ln\left(\sum_{m \in Q} \rho_{i,j}^{i,j+1}(m,n)\right),
\] (49)

The detailed calculations and explanations have been given in Ref. [16]. By applying these formulations to the evidence framework, we obtain the optimal hyperparameters and the restored images with high accuracy.

Before closing this section, we explain when the CVM are used in the evidence framework. In the evidence framework, we have to calculate the marginal probability distributions for $\Pr\{X = x|Y = y, p, K\}$ and $\Pr\{X = x|K\}$ for various values of hyperparameters $p$ and $K$. In order to calculate the marginal probability distributions for $\Pr\{X = x|Y = y, p, K\}$, we set that

\[
g_{i,j}(n) \equiv -\beta(p)\delta(n,y_i^*), \quad e(n,n') \equiv -K\delta(n,n').
\] (50)

Approximate values of the marginal probability distributions by means of the CVM are approximately obtained as follows:

\[
\Pr\{X_{i,j} = n|Y = y^*, p, K\} = \rho_{i,j}(n),
\] (51)
Pr\{X_{i,j} = n, X_{i+1,j} = n' | Y = y^*, p, K\} = \rho_{i,j}^{i+1,j}(n, n'), \quad (52)
Pr\{X_{i,j} = n, X_{i,j+1} = n' | Y = y^*, p, K\} = \rho_{i,j}^{i,j+1}(n, n'), \quad (53)

by solving Eqs. (40)-(49) for various values of \(p\) and \(K\).

In order to calculate the marginal probability distributions for \(Pr\{X = x | K\}\), we set that

\[ g_{i,j}(n) \equiv 0, \quad e(n, n') \equiv -K\delta(n, n'). \quad (54) \]

Approximate values of the marginal probability distributions by means of the CVM are approximately obtained as follows:

\[ Pr\{X_{i,j} = n, X_{i+1,j} = n' | K\} = \rho_{i,j}^{i+1,j}(n, n'), \quad (55) \]
\[ Pr\{X_{i,j} = n, X_{i,j+1} = n' | K\} = \rho_{i,j}^{i,j+1}(n, n'), \quad (56) \]

by solving Eqs. (40)-(49) for various values of \(p\) and \(K\).


In this section, some numerical experiments are given for the original images \(x\) in Fig. 1(a) \((q = 2)\) and Fig. 2(a) \((q = 4)\). By setting \((q - 1)p^* = 0.2\), the degraded images \(y\) are produced from the original images \(x\) in the degradation process given in Eqs. (5) and (6). The degraded images \(y\) are shown in Fig. 1(b) and Fig. 2(b). We compare the restored image \(\hat{x}\) obtained by using the CVM with the one by the MFA.

In the evidence framework, the right-hand side in Eq. (15) and the left-hand sides in Eq. (14) and Eq. (15) are calculated by means of the CVM for various values of \(p\) and \(K\). From these obtained quantities, the estimates of hyperparameters, \(\hat{p}\) and \(\hat{K}\), are chosen so as to satisfy the extremum conditions (14) and (15) for \(Pr\{Y = y | p, K\}\). At the estimates of hyperparameters, \(\hat{p}\) and \(\hat{K}\), the restored image \(\hat{x}\) are obtained from Eq. (21). The estimates of hyperparameters, \(\hat{p}\) and \(\hat{K}\), which are estimated from the given degraded images \(y\) by means of the evidence framework, are shown in Table 1. The restored images \(\hat{x}\) obtained by applying the MFA to the evidence framework are shown in Fig. 1(c) and Fig. 2(c). The restored images \(\hat{x}\) obtained by applying the CVM to the evidence framework are shown in Fig. 1(d) and Fig. 2(d). The values of \(d(\hat{x}, x) / |I|\) and the improvement of signal to noise ratio defined by

\[ \Delta_{SNR} \equiv 10 \log_{10} \left( \frac{\sum_{(i,j) \in I} (x^*_{i,j} - y^*_{i,j})^2}{\sum_{(i,j) \in I} (x^*_{i,j} - \hat{x}_{i,j})^2} \right) \text{ (dB)} \]

for the restored images \(\hat{x}\) obtained by means of the evidence framework are also shown in Table 1.

5. Concluding Remarks.

In this paper, we gave the formulations of the MFA and the CVM in image restorations by means of the MRF model and the evidence framework. Some numerical experiments showed that the results in image restorations are much affected from the accuracy of the approximation adopted in calculating the evidence of the MRF model.

The CVM is a methodology for the construction of an approximation for the massive probabilistic model whose probability distribution is given in a form of Gibbs canonical distribution. In the present paper, we adopt the pair approximation in the
Figure 1: Comparison between the cluster variation method (CVM) and the mean field approximation (MFA) in the binary image restorations by means of the evidence framework ($|I| = 256 \times 256$). (a) Original image $x^*$ ($q = 2$). (b) Degraded image $y^*$ ($(q-1)p^* = 0.2$). (c) Restored image $\hat{x}$ by the MFA. (d) Restored image $\hat{x}$ by the CVM.

Figure 2: Comparison between the cluster variation method (CVM) and the mean field approximation (MFA) in the image restorations by means of the evidence framework ($|I| = 256 \times 256$). (a) Original image $x^*$ ($q = 2$). (b) Degraded image $y^*$ ($(q-1)p^* = 0.2$). (c) Restored image $\hat{x}$ by the MFA. (d) Restored image $\hat{x}$ by the CVM.

CVM as shown in Eq. (36). In the pair approximation, it can be regarded that the set of all nearest-neighbor pair of pixels is adopted as basic clusters in the CVM. Here, a cluster means a set of pixels and the size of cluster is the number of pixels constructing the cluster. We can constructed the more high-level approximation in the CVM by adopted the bigger clusters as basic clusters Recently, some extensions of the mean field approximation have proposed for probabilistic information processing in graphical models[2], for example, a deterministic algorithm in Bayesian image restorations[24], brief propagation algorithm in error correcting codes[25], machine learning algorithm in neural network system[26], and so on. These extensions are also based on the concept of the CVM. It is future problem that the algorithms be means of high-level approximations of the CVM are constructed for some practical applications in the massive probabilistic information processing.

The difficulty in Bayes statistics is how we calculate some statistical quantities in massive probabilistic models, because the total number of all possible configuration is $O(\exp(N))$ where $N$ is the system size. We need the computational time of $O(\exp(N))$ in calculating the statistical quantities exactly. By introducing the MFA or the CVM, the calculations of the statistical quantities can be approximately reduced to the problem to solve simultaneous non-linear equations in which the total number of unknown parameters is $O(N)$. In the case of MRF model on a square
Table 1: The estimates of hyperparameters, $\hat{p}$, $\hat{K}$, and the values of $d(x,y)/|I|$, $d(x,\hat{x})/|I|$ and $\Delta_{SNR}$ obtained for some degraded images $y$, which are produced for the case of $(q-1)p^* = 0.2$ from the original image $x$ given in Fig. 1a ($q = 2$) and Fig. 2a ($q = 4$). The hyperparameters are estimated by applying the mean field approximation (MFA) and the cluster variation method (CVM) to the evidence framework.

| $\hat{x}$ | Approx. | $(q-1)\hat{p}$ | $\hat{K}$ | $d(x,y)/|I|$ | $d(x,\hat{x})/|I|$ | $\Delta_{SNR}$ (dB) |
|-----------|---------|----------------|------------|---------------|----------------|------------------|
| Fig. 1c   | MFA     | 0.0700         | 0.6073     | 0.1989        | 0.1989         | 0                |
| Fig. 1d   | CVM     | 0.1468         | 0.8033     | 0.1989        | 0.0694         | 4.5751           |
| Fig. 2c   | MFA     | 0.1079         | 1.0018     | 0.1989        | 0.1235         | 2.3092           |
| Fig. 2d   | CVM     | 0.1549         | 1.1019     | 0.1989        | 0.0753         | 5.3193           |

lattice, the total number of unknown parameters in the pair approximation of the CVM is about four times of the one in the MFA. The computational time in some numerical experiments by means of the CVM is also about four times of the one by means of the MFA.

We have an expectation-maximization (EM) procedure as a powerful one fastly for calculating the estimates of hyperparameters in the evidence framework[27]. In fact, Zhang proposed a new algorithm that is constructed for estimating the hyperparameters in the MRF model for segmentations by using the MFA[28]. It is interesting to construct an EM algorithm by means of the CVM. It is also future problem.

Acknowledgment
The author is grateful to Professor J. Inoue, Graduate School of Engineering, Hokkaido University, Professor D. M. Titterington of the Department of Statistics, University of Glasgow, and Professor T. Horiguchi of the Department of Computer and Mathematical Sciences, Graduate School of Information Science, Tohoku University.

References


Kazuyuki Tanaka
Associate Professor
Department of Computer and Mathematical Sciences
Graduate School of Information Sciences, Tohoku University
Aramaki-aza-aoba 04, Aoba-ku, Sendai 980-8579, Japan
Phone: +81-22-217-5899
FAX: +81-22-217-5851
E-mail: kazu@statp.is.tohoku.ac.jp.