特集論文

Maximum Marginal Likelihood Estimation and Constrained Optimization in Image Restoration

Kazuyuki Tanaka

Department of Computer and Mathematical Sciences, Graduate School of Information Sciences, Tohoku University
 kazu@statp.is.tohoku.ac.jp

keywords: Markov random fields, image restoration, mean-field approximation, Bayes statistics, maximum likelihood estimation, evidence, cluster variation method

Summary

In image restorations, two different mathematical frameworks for massive probabilistic models are given in the standpoint of Bayes statistics. One of them is formulated by assuming that *a priori* probability distribution has a form of Gibbs micro-canonical distribution and can be reduced to a constrained optimization. In this framework, we have to know a quantity estimated from the original image though we do not have to know the configuration of the original image. We give a new method to estimate the length of boundaries between areas with different grey-levels in the original image only from the degraded image in grey-level image restorations. When the length of boundaries in the original image is estimated from the given degraded image, we can construct a probabilistic model for image restorations by means of Bayes formula. This framework can be regarded as probabilistic information processing at zero temperature. The other framework is formulated by assuming the *a priori* probability distribution has a form of Gibbs canonical distribution. In this framework, some hyperparameters are determined so as to maximize a marginal likelihood. In this paper, it is assumed that the *a priori* probability distribution is a Potts model, which is one of familiar statistical-mechanical models. In this assumption, the logarithm of marginal likelihood can be expressed in terms of free energies of some probabilistic models and hence this framework can be regarded as probabilistic information processing at finite temperature. The practical algorithms in the constrained optimization and the maximum marginal likelihood estimation are given by means of cluster variation method, which is a statistical-mechanical approximation with high accuracy for massive probabilistic models. We compare them with each others in some numerical experiments for grey-level image restorations.

1. Introduction

In statistical method, many probabilistic models for image processing are constructed as *a posteriori* probability distributions by using the Bayes formula [Derin 84]. *A priori* information of natural images is taken to the *a posteriori* probability distribution through *a priori* probability distribution in the Bayes formula. The *a priori* probability distribution is often given in a form of Gibbs canonical distribution. Particularly, Markov random fields, in which the state of a pixel is dependent only on the states of its nearest neighbour pixels [Derin 84], is the most popular random fields of *a priori* probability distribution. The Markov random field (MRF) model, which is a probabilistic model with Markov random fields, is a very interesting model also in the statistical-mechanical point of view. One of difficulties in the statistical method for image processing is the determination of hyperparameters in the *a priori* probability distribution.

In statistics, Pryce and Bruce [Pryce 95], and Zhou, Leahy and Qi [Zhou 97] applied a maximum marginal likelihood (MML) estimation to the hyperparameter determination in MRF models. In the MML estimation, the marginal likelihood is defined by marginalizing a joint probability for the original and the degraded image with respect to original images and can be regarded as a likelihood function for the hyperparameters when the degraded image is given. If the *a priori* probability distribution for the original image is assumed to be a Gibbs canonical distribution, the temperature plays a role of a hyperparameter in the MML estimation. The MML estimation can be regarded as probabilistic information processing at finite temperature.

In the standpoint of statistical mechanics, Morita and the present author proposed a new method [Morita 96]. In the method, we estimate some properties of the original image to be recovered from a single given degraded image with the aid of some *a priori* knowl-

edge, and then seek an image which is nearest to the degraded image among the images satisfying the estimated properties. The problem is formulated as a constrained optimization (CO). We can derive an energy function by introducing some Lagrange multipliers to ensure the constraint conditions. These Lagrange multipliers are determined so as to satisfy the corresponding constraint conditions from among the minimum-energy configurations searched for various values of Lagrange multipliers. Temperature is just a control parameter in order to apply an annealing procedure to the search of minimum-energy configurations. In such a point of view, the mathematical framework constructed by using the CO can be regarded as probabilistic information processing at zero temperature.

The main purpose of the present paper is to clarify the mathematical frameworks of the CO [Morita 96] and the MML estimation [Pryce 95, Zhou 97] in the standpoint of statistical mechanics. In the CO, we choose the length of boundaries between areas with different grey-levels in images as the most important quantity that characterizes the original image and give a new method to estimate the length of boundaries in the original image only from the given degraded image. In the MML estimation, we adopt a Potts model [Wu 82] as *a priori* probability distribution for original images. We apply the cluster variation method (CVM) [Tanaka 95] to construct practical algorithms for the CO and the MML estimation. The CVM can give us not only the optimal configuration but also some statistical quantities in a massive probabilistic model with high accuracy. ^{*1} Recently, the present author investigated the MML estimation for hyperparameters in the MRF model by means of the CVM and compared some results in the CVM with the ones in the MFA [Tanaka 00]. It was shown that the accuracy of the MFA is not enough to obtain the estimates of hyperparameters and we can improve the framework of the MML estimation by using the CVM instead of the MFA. The difficulty in Bayes statistics is how we calculate some statistical quantities in the massive probabilistic model, because the total number of all possible configuration is $\mathcal{O}(\exp(N))$ where N is the system size of the massive

probabilistic model. We need the computational time of $\mathcal{O}(\exp(N))$ 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 $\mathcal{O}(N)$. In the case of MRF model on a square lattice, the total number of unknown parameters in the CVM proposed in [Tanaka 95, Tanaka 00] 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.

The present paper is organized in the following way. In Sec. 2, we explain about the construction of a massive probabilistic model for grey-level image restorations by means of the Bayes formula. In Sec. 3, we extend a formulation for binary image restorations in Ref. [Morita 96] to the case of grey-level image restorations. The formulation in Ref. [Morita 96] is based on the CO and we impose a constraint condition as *a priori* information for an original image. In the formulation in Ref. [Morita 96], we have to know the length of boundaries between areas with different grey-levels in the original image. In Sec. 3, we gave a method to estimate the length of boundaries in the original image only from the given degraded image for grey-level image restorations.^{*2} In Sec. 4, we explain about the MML estimation and give a set of deterministic equations for hyperparameters in terms of some marginal probability distributions of the *a posteriori* and the *a priori* probability distribution. In Sec. 5, we give approximate deterministic equations for the marginal probability distributions of a massive probabilistic model expressed in a form of Gibbs canonical distribution, by using the CVM. By applying the approximate method to the massive probabilistic models given in Secs. 3 and 4, we can construct approximate algorithms for grey-level image restorations by means of the CO and the MML estimation. Some numerical experiments by using the CO and the MML estimation are shown in Sec. 6. Concluding remarks are given in Sec. 7.

2. Bayes Statistics

In this section, we give the mathematical framework

^{*1} The CVM is an extension of mean-field approximation. The mean-field approximation is one of familiar statistical-mechanical techniques for massive probabilistic models. Pryce and Bruce [Pryce 95], and Zhou, Leathy and Qi [Zhou 97] obtained the estimates of hyperparameters in the MML estimation by using the mean-field approximation.

^{*2} This method has given only for binary image restorations in Ref. [Morita 96].

of image restoration by means of probabilistic models in Bayes statistics.

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

$$\Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*\} = \frac{\Pr\{\mathbf{Y} = \mathbf{y}^* | \mathbf{X} = \mathbf{x}\} \Pr\{\mathbf{X} = \mathbf{x}\}}{\sum_{\mathbf{z}} \Pr\{\mathbf{Y} = \mathbf{y}^* | \mathbf{X} = \mathbf{z}\} \Pr\{\mathbf{X} = \mathbf{z}\}}, \quad (1)$$

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

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*\}. \quad (2)$$

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

$$\Pr\{X_{i,j} = n | \mathbf{Y} = \mathbf{y}^*\} = \sum_{\mathbf{z}} \delta(z_{i,j}, n) \Pr\{\mathbf{X} = \mathbf{z} | \mathbf{Y} = \mathbf{y}^*\}, \quad (3)$$

so that we have

$$\hat{x}_{i,j} = \arg \max_{n \in \mathbf{Q}} \Pr\{X_{i,j} = n | \mathbf{Y} = \mathbf{y}^*\} \quad ((i, j) \in \mathbf{I}). \quad (4)$$

Here $\delta(a, b)$ represents Kronecker's delta δ_{ab} . This criteria is called maximum *posterior* marginal (MPM) estimation [Marroquin 87, Nishimori 99].

In order to estimate those properties only from knowledge of the given degraded image $\mathbf{y}^*$, we adopt two assumptions.

Assumption (i). *We have a given degraded image $\mathbf{y}^*$ which is obtained from the original image $\mathbf{x}^*$ 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 $\Pr\{\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}^*\}$ is given by

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

where

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

The conditional probability distribution can be rewritten as

$$\Pr\{\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}^*, p^*\} = \frac{\exp(-\beta(p^*)d(\mathbf{x}^*, \mathbf{y}))}{\sum_{\mathbf{z}} \exp(-\beta(p^*)d(\mathbf{x}^*, \mathbf{z}))}, \quad (7)$$

where

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

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

The summation $\sum_{(i,j) \in \mathbf{I}}$ 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).

3. Constrained Optimization and Gibbs Micro-canonical Distribution

3.1 Constrained Optimization

In this subsection, we explain the CO given in Ref. [Morita 96].

We adopt not only the assumption (i) but also the following assumption:

Assumption (ii). *Even though the original image $\mathbf{x}^*$ itself is unknown, we assume that we know the quantity $\sigma_2(\mathbf{x}^*)$ defined by*

$$\sigma_2(\mathbf{x}^*) = \sum_{ij \in \mathbf{B}} \left(2 - \delta(x_{i,j}^*, x_{i+1,j}^*) - \delta(x_{i,j}^*, x_{i,j+1}^*) \right). \quad (10)$$

where $\mathbf{B}$ is the set of nearest-neighbour pairs of pixels. The quantity $\sigma_2(\mathbf{x}^*)$ is the total number of nearest-neighbour pair of pixels which have different state to each other. We show that the MAP estimation, which is constructed under the assumptions (i) and (ii), can be reduced to the CO by using the Bayes formula (1). In Ref. [Morita 96], it was shown that the value of $\sigma_2(\mathbf{x}^*)$ defined by Eq. (10) can be estimated only from the given degraded image $\mathbf{y}^*$ under an assumption for the original image, in the case of binary image restorations. We will give the corresponding method of estimating the value of $\sigma_2(\mathbf{x}^*)$ only from a single given degraded image $\mathbf{y}^*$ in the case of grey-level image restorations in the next subsection.

The assumption (ii) requires that the restored image $\hat{\mathbf{x}} = \{\hat{x}_{i,j} | (i,j) \in \mathbf{I}\}$ satisfies the following condition:

$$\sigma_2(\hat{\mathbf{x}}) = \sigma_2(\mathbf{x}^*). \quad (11)$$

When this is the only requirement for the restored image $\hat{\mathbf{x}}$, the best choice for $\Pr\{\mathbf{X} = \mathbf{x}\}$ is

$$\Pr\{\mathbf{X} = \mathbf{x}\} \equiv \frac{\delta(\sigma_2(\mathbf{x}), \sigma_2(\mathbf{x}^*))}{\sum_{\mathbf{z}} \delta(\sigma_2(\mathbf{z}), \sigma_2(\mathbf{x}^*))}. \quad (12)$$

We remark that the probability distribution given in Eq. (12), can be regarded as a Gibbs micro-canonical distribution in statistical mechanics. The image $\mathbf{x} = \hat{\mathbf{x}}$, which makes the *a posteriori* probability distribution $\Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*\}$ given by Eq. (1) maximum, is given by the one which makes the likelihood function $\Pr\{\mathbf{Y} = \mathbf{y}^* | \mathbf{X} = \mathbf{x}\}$ maximum under the condition (11). Since $\beta(p^*)$ given in Eq. (9) is always positive, the MAP estimation given in Eqs. (1), (2), (7), (8), (10) and (12) is reduced to the following constrained optimization:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ d(\mathbf{x}, \mathbf{y}^*) \mid \sigma_2(\mathbf{x}) = \sigma_2(\mathbf{x}^*) \right\}. \quad (13)$$

We introduce a Lagrange multiplier J for the condition $\sigma_2(\mathbf{x}) = \sigma_2(\mathbf{x}^*)$ and rewrite the conditional optimization (13) as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ d(\mathbf{x}, \mathbf{y}^*) + J(\sigma_2(\mathbf{x}) - \sigma_2(\mathbf{x}^*)) \right\}. \quad (14)$$

The optimal configuration $\hat{\mathbf{x}}$ can be obtained as the minimum-energy configuration which minimizes the energy function $H(\mathbf{x} | \mathbf{y}^*, J)$ given by

$$H(\mathbf{x} | \mathbf{y}^*, J) \equiv - \sum_{(i,j) \in \mathbf{I}} \left(\delta(x_{i,j}, y_{i,j}^*) + J\delta(x_{i,j}, x_{i+1,j}) + J\delta(x_{i,j}, x_{i,j+1}) \right). \quad (15)$$

We determine an optimal value of J by enforcing the constraint condition (11). The optimal value of Lagrange multiplier J is denoted by $\hat{J}$.

In the CO, we search the configuration $\mathbf{x} = \hat{\mathbf{x}}$ which maximizes the likelihood function $\Pr\{\mathbf{Y} = \mathbf{y}^* | \mathbf{X} = \mathbf{x}\}$ among the set of configurations $\mathbf{x}$ satisfying the constraint (11). Because the inequality $\beta(p^*) > 0$ is always valid for any p^* , we do not need to determine the value of p^* in Eq. (7). All the quantity that we need in the CO is only quantity $\sigma_2(\mathbf{x}^*)$ for the true original image $\mathbf{x}^*$.

In order to employ the annealing procedure in the search of the optimal configuration $\hat{\mathbf{x}}(J)$, we introduce the temperature T and the Gibbs canonical distribution $\rho(\mathbf{x} | \mathbf{y}^*, T, J)$ as follows:

$$\rho(\mathbf{x} | \mathbf{y}^*, T, J) \equiv \frac{\exp(-H(\mathbf{x} | \mathbf{y}^*, J)/T)}{\sum_{\mathbf{z}} \exp(-H(\mathbf{z} | \mathbf{y}^*, J)/T)}. \quad (16)$$

From the form of the Gibbs canonical distribution given in Eq. (16), it is obvious that the configuration maximizing the probability distribution $\rho(\mathbf{x} | \mathbf{y}^*, T, J)$ does not depend on the value of T and that Eq. (14) can be rewritten as

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \rho(\mathbf{x} | \mathbf{y}^*, T, \hat{J}) \quad (T > 0). \quad (17)$$

This fact means that we obtain the optimal configuration $\hat{\mathbf{x}}$ in terms of Eq. (17) at the optimal value of Lagrange multiplier, $\hat{J}$.

In the CO, we construct the energy function by using the Bayes formula under the assumptions (i) and (ii). While the Lagrange multiplier J is introduced to ensure the constrained condition (11), the temperature T is introduced only in the purpose of employing the annealing procedure in the search of the minimum-energy configuration in the energy function $H(\mathbf{x} | \mathbf{y}^*, J)$ given in Eq. (15).

Now, we describe about the relationship between the Gibbs micro-canonical distribution given in (12) and the Gibbs canonical distribution in the similar argument of Ref. [Tanaka 99]. We consider the Laplace transform of a function $\mathcal{N}(\tau) \equiv \sum_{\mathbf{z}} \delta(\sigma_2(\mathbf{z}) - \tau)$ as follows:

$$\nu(\kappa) \equiv \int_0^{+\infty} \exp(-\kappa\tau) \sum_{\mathbf{z}} \delta(\sigma_2(\mathbf{z}) - \tau) d\tau$$

$$= \sum_{\mathbf{z}} \exp(-\kappa \sigma_2(\mathbf{z})), \quad (18)$$

where $\delta(a)$ is a delta function. By using the Laplace inversion formula, we have the following equality:

$$\begin{aligned} & \sum_{\mathbf{z}} \delta(\sigma_2(\mathbf{z}) - \sigma_2(\mathbf{x}^*)) \\ &= \lim_{\lambda \rightarrow +\infty} \frac{1}{2\pi i} \int_{-i\lambda}^{+i\lambda} \exp(\kappa \sigma_2(\mathbf{x}^*)) \nu(\kappa) d\kappa \\ &= \lim_{\lambda \rightarrow +\infty} \frac{1}{2\pi i} \int_{-i\lambda}^{+i\lambda} \exp(\sigma_2(\mathbf{x}^*) g(\kappa)) d\kappa, \end{aligned} \quad (19)$$

where

$$g(\kappa) \equiv \kappa + \frac{1}{\sigma_2(\mathbf{x}^*)} \ln \left(\sum_{\mathbf{z}} \exp(-\kappa \sigma_2(\mathbf{z})) \right). \quad (20)$$

Since the quantity $\sigma_2(\mathbf{x}^*)$ is the number of boundary length between different darkness in the original image $\mathbf{x}^*$, we have $\sigma_2(\mathbf{x}^*) = \mathcal{O}(|\mathbf{I}|)$. The quantity $-\ln \left(\sum_{\mathbf{z}} \exp(-\kappa \sigma_2(\mathbf{z})) \right)$ is a free energy of Potts model on a square lattice and is $\mathcal{O}(|\mathbf{I}|)$ [Wu 82]. Hence we have

$$\frac{1}{\sigma_2(\mathbf{x}^*)} \ln \left(\sum_{\mathbf{z}} \exp(\kappa \sigma_2(\mathbf{z})) \right) = \mathcal{O}(1). \quad (21)$$

In the limit of $|\mathbf{I}| \rightarrow +\infty$, we obtain the following equality by means of the saddle point approximation:

$$\begin{aligned} & \sum_{\mathbf{z}} \delta(\sigma_2(\mathbf{z}) - \sigma_2(\mathbf{x}^*)) \\ & \sim \sum_{\mathbf{z}} \exp \left(\hat{\kappa} (\sigma_2(\mathbf{x}^*) - \sigma_2(\mathbf{z})) \right), \end{aligned} \quad (22)$$

where $\hat{\kappa}$ is determined so as to satisfy the following equation:

$$\frac{\sum_{\mathbf{z}} \sigma_2(\mathbf{z}) \exp(-\hat{\kappa} \sigma_2(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-\hat{\kappa} \sigma_2(\mathbf{z}))} = \sigma_2(\mathbf{x}^*). \quad (23)$$

This equality means that the Gibbs micro-canonical distribution given in Eq. (12) is equivalent to the Gibbs canonical distribution given as follows:

$$\Pr\{\mathbf{X} = \mathbf{x} | \hat{\kappa}\} \equiv \frac{\exp(-\hat{\kappa} \sigma_2(\mathbf{x}))}{\sum_{\mathbf{z}} \exp(-\hat{\kappa} \sigma_2(\mathbf{z}))}. \quad (24)$$

Moreover, we remark also that the deterministic equation (23) for hyperparameter $\hat{\kappa}$ is equivalent to the following equation:

$$\hat{\kappa} = \arg \max_{\kappa} \Pr\{\mathbf{X} = \mathbf{x}^* | \kappa\}. \quad (25)$$

The equation (25) means that we regard the probability distribution $\Pr\{\mathbf{X} = \mathbf{x}^* | \kappa\}$ as a likelihood function

of hyperparameter κ when the true original image is $\mathbf{x}^*$. In Sec. 6, some numerical experiments for the formulation in the present section will be given in a standpoint of the Gibbs micro-canonical distribution (12) and the constrained optimization (13).

3.2 Estimation of $\sigma_2(\mathbf{x}^*)$

In this subsection, we give the estimation of $d(\mathbf{x}^*, \mathbf{y}^*)$ and $\sigma_2(\mathbf{x}^*)$ based on the assumptions (i), (ii) and the following assumption:

Assumption (iii). *The configuration of a lattice site (i, j) in the original image $\mathbf{x}^*$ should take the same value as the configuration of either the nearest neighbour $(i-1, j)$ or $(i+1, j)$ in the same row as well as the configuration of either $(i, j-1)$ or $(i, j+1)$ in the same column, such that we have*

$$\begin{aligned} & \left(\delta(x_{i,j}^*, x_{i-1,j}^*) + \delta(x_{i,j}^*, x_{i+1,j}^*) \right) \\ & \times \left(\delta(x_{i,j}^*, x_{i,j-1}^*) + \delta(x_{i,j}^*, x_{i,j+1}^*) \right) \neq 0 \\ & ((i, j) \in \mathbf{I}). \end{aligned} \quad (26)$$

These assumptions are applicable to the mathematical framework for the MRF of the image with any grades. However the assumption (iii) is too strong to apply the framework to the practical image with 256 grades because this assumption emphasizes just flatness and the assumption for practical images should include not only flatness but also smoothness. On the other hand, the flatness is the most important assumption for original images in image restoration. In order to clarify the basic mathematical framework of MRF model, we restrict the assumption for original images only to flatness. This is the reason why we adopt such a strong assumption for original images.

In order to express the above assumption (iii) in terms of probability theory, we introduce two probabilities $P_x(n, n', n'')$ and $P_y(n, n', n'')$ which are defined by

$$\begin{aligned} P_x(n, n', n'') & \equiv \frac{1}{2|\mathbf{I}|} \\ & \times \sum_{(i,j) \in \mathbf{I}} \left(\Pr\{X_{i-1,j} = n, X_{i,j} = n', X_{i+1,j} = n''\} \right. \\ & \quad \left. + \Pr\{X_{i,j-1} = n, X_{i,j} = n', X_{i,j+1} = n''\} \right) \\ & (n, n', n'' \in \mathbf{Q}), \end{aligned} \quad (27)$$

and

$$\begin{aligned} P_y(n, n', n'') & \equiv \frac{1}{2|\mathbf{I}|} \\ & \times \sum_{(i,j) \in \mathbf{I}} \left(\Pr\{Y_{i-1,j} = n, Y_{i,j} = n', Y_{i+1,j} = n''\} \right. \end{aligned}$$

$$+ \Pr\{Y_{i,j-1} = n, Y_{i,j} = n', Y_{i,j+1} = n''\} \\ (n, n', n'' \in \mathbf{Q}), \quad (28)$$

in the original and the degraded image, respectively. By assumption (ii), $P_x(n, n', n'')$ and $P_y(n, n', n'')$ are related by

$$P_y(n, n', n'') = \sum_{m \in \mathbf{Q}} \sum_{m' \in \mathbf{Q}} \sum_{m'' \in \mathbf{Q}} P_{yx}(n|m, p^*) \\ \times P_{yx}(n'|m', p^*) P_{yx}(n''|m'', p^*) P_x(m, m', m'') \\ (n, n', n'' \in \mathbf{Q}), \quad (29)$$

where $P_{yx}(n|m, p^*)$ is defined by Eq. (6). In order to estimate the value of $\sigma_2(\mathbf{x}^*)$ only from the given degraded image, we adopt the assumption (iii). The assumption (iii) states that

$$P_x(n, n', n'') = 0, \quad (n \neq n', n'' \neq n'). \quad (30)$$

Moreover, $P_x(n, n', n'')$ should have the following symmetry:

$$P_x(n, n', n'') = P_x(n'', n', n). \quad (31)$$

In terms of $P_x(n, n', n'')$ and $P_{yx}(n|m, p^*)$, the values of $\bar{\sigma}_2(\mathbf{x}^*) \equiv \sigma_2(\mathbf{x}^*)/(2|\mathbf{I}|)$ and $\bar{d}(\mathbf{x}^*, \mathbf{y}^*) \equiv d(\mathbf{x}^*, \mathbf{y}^*)/|\mathbf{I}|$ are approximately expressed as the following quantities $\hat{\sigma}_{2x}$ and $\hat{d}_{xy}$, respectively:

$$\hat{\sigma}_{2x} \equiv \sum_{m \in \mathbf{Q}} \sum_{m' \in \mathbf{Q}} \sum_{m'' \in \mathbf{Q}} P_x(m, m', m'') \\ \times (1 - \delta(m, m')), \quad (32)$$

and

$$\hat{d}_{xy} \equiv \sum_{n \in \mathbf{Q}} \sum_{m \in \mathbf{Q}} P_{yx}(n|m, p^*) (1 - \delta(n, m)). \quad (33)$$

By substituting Eq. (6) into Eq. (33), the quantity $\hat{d}_{xy}$ is expressed in terms of p^* as follows:

$$\hat{d}_{xy} = (q-1)p^*. \quad (34)$$

From Eqs. (29), (30), (31) and (34), the deterministic equations for $\hat{\sigma}_{2x}$ and $\hat{d}_{xy}$ are given as follows:

$$\hat{\sigma}_{3y} = \hat{d}_{xy} \left(2 - \frac{q}{q-1} \hat{d}_{xy} \right) \\ + 2 \left(1 - \frac{q}{q-1} \hat{d}_{xy} \right)^2 \hat{\sigma}_{2x}, \quad (35)$$

$$\hat{\sigma}_{2y} = \hat{d}_{xy} \left(2 - \frac{q}{q-1} \hat{d}_{xy} \right) \\ + \left(1 - \frac{q}{q-1} \hat{d}_{xy} \right)^2 \hat{\sigma}_{2x}. \quad (36)$$

Here $\hat{\sigma}_{2y}$ and $\hat{\sigma}_{3y}$ are defined by

$$\hat{\sigma}_{2y} \equiv \sum_{m \in \mathbf{Q}} \sum_{m' \in \mathbf{Q}} \sum_{m'' \in \mathbf{Q}} P_y(m, m', m'') \\ \times (1 - \delta(m, m')), \quad (37)$$

$$\hat{\sigma}_{3y} \equiv \sum_{m \in \mathbf{Q}} \sum_{m' \in \mathbf{Q}} \sum_{m'' \in \mathbf{Q}} P_y(m, m', m'') \\ \times (1 - \delta(m, m'')). \quad (38)$$

The detailed explanation for derivation of Eqs. (35) and (36) is given in Appendix A. From the spatial averages in the single given degraded image $\mathbf{y}^*$, these values are approximately estimated as the following quantities $\hat{\sigma}_{2y}$ and $\hat{\sigma}_{3y}$:

$$\hat{\sigma}_{2y} = \frac{1}{2|\mathbf{I}|} \sigma_2(\mathbf{y}^*), \quad (39)$$

$$\hat{\sigma}_{3y} = \frac{1}{2|\mathbf{I}|} \sum_{(i,j) \in \mathbf{I}} \left(2 - \delta(y_{i,j}^*, y_{i+2,j}^*) \right. \\ \left. - \delta(y_{i,j}^*, y_{i,j+2}^*) \right). \quad (40)$$

Thus, the left-hand sides of Eqs. (35) and (36) can be calculated from the spatial average in the given degraded image $\mathbf{y}^*$. By solving the simultaneous equations (35) and (36) with respect of $\hat{\sigma}_{2x}$ and $\hat{d}_{xy}$, we can estimate the quantities $\sigma_2(\mathbf{x}^*)$ and $d(\mathbf{x}^*, \mathbf{y}^*)$. The quantities $\hat{\sigma}_{2x}$ and $\hat{d}_{xy}$ are regarded as the estimated values for $\sigma_2(\mathbf{x}^*)/(2|\mathbf{I}|)$ and $d(\mathbf{x}^*, \mathbf{y}^*)/|\mathbf{I}|$. We adopt the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$ as the value of $\sigma_2(\mathbf{x}^*)$ in the constrained optimization (13).

4. Maximum Marginal Likelihood Estimation and Gibbs Canonical Distribution

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 MML estimation [Zhou 97, Pryce 95].

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 (iv). *The a priori probability distribution, that the original image is $\mathbf{x}$, is given by*

$$\Pr\{\mathbf{X} = \mathbf{x}\} = \Pr\{\mathbf{X} = \mathbf{x} | K^*\} \\ \equiv \frac{\exp(-K^* \sigma_2(\mathbf{x}))}{\sum_{\mathbf{z}} \exp(-K^* \sigma_2(\mathbf{z}))}. \quad (41)$$

This assumption means that the original image is one of images that are produced with large probability in the a priori probability distribution (41).

We remark that the *a priori* probability distribution given by Eq. (41) can be regarded as a Gibbs canonical distribution in statistical mechanics. By substituting Eqs. (7) and (41) into Eq. (1), the *a posteriori* probability distribution is given by

$$\begin{aligned} \Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*\} &= \Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*, p^*, K^*\} \\ &= \frac{\exp(-\beta(p^*)d(\mathbf{x}, \mathbf{y}^*) - K^*\sigma_2(\mathbf{x}))}{\sum_{\mathbf{z}} \exp(-\beta(p^*)d(\mathbf{z}, \mathbf{y}^*) - K^*\sigma_2(\mathbf{z}))}. \end{aligned} \quad (42)$$

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

In the standpoint of the MML estimation, the estimates of hyperparameters, $\hat{p}$ and $\hat{K}$, are determined so as to maximize a marginal likelihood function

$$\begin{aligned} \Pr\{\mathbf{Y} = \mathbf{y}^* | p, K\} \\ \equiv \sum_{\mathbf{z}} \Pr\{\mathbf{Y} = \mathbf{y}^* | \mathbf{X} = \mathbf{z}, p\} \Pr\{\mathbf{X} = \mathbf{z} | K\}, \end{aligned} \quad (43)$$

so that we have

$$(\hat{p}, \hat{K}) = \underset{(p, K)}{\operatorname{argmax}} \Pr\{\mathbf{Y} = \mathbf{y}^* | p, K\}. \quad (44)$$

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

$$\begin{aligned} \sum_{(i,j) \in \mathbf{I}} \sum_{n \in \mathbf{Q}} \delta(y_{i,j}^*, n) \Pr\{X_{i,j} = n | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\} \\ = (q-1)\hat{p}, \end{aligned} \quad (45)$$

$$\begin{aligned} \sum_{(i,j) \in \mathbf{I}} \sum_{n \in \mathbf{Q}} \sum_{m \in \mathbf{Q}} \delta(n, m) \\ \times \left(\Pr\{X_{i,j} = n, X_{i+1,j} = m | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\} \right. \\ \left. + \Pr\{X_{i,j} = n, X_{i,j+1} = m | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\} \right) \\ = \sum_{(i,j) \in \mathbf{I}} \sum_{n \in \mathbf{Q}} \sum_{m \in \mathbf{Q}} \delta(n, m) \\ \times \left(\Pr\{X_{i,j} = n, X_{i+1,j} = m | \hat{K}\} \right. \\ \left. + \Pr\{X_{i,j} = n, X_{i,j+1} = m | \hat{K}\} \right), \end{aligned} \quad (46)$$

where

$$\begin{aligned} \Pr\{X_{i,j} = n | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\} \\ \equiv \sum_{\mathbf{x}} \delta(x_{i,j}, n) \Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\}, \end{aligned}$$

(47)

$$\begin{aligned} \Pr\{X_{i,j} = n, X_{i+1,j} = n' | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\} \\ \equiv \sum_{\mathbf{x}} \delta(x_{i,j}, n) \delta(x_{i+1,j}, n') \Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\}, \end{aligned} \quad (48)$$

$$\begin{aligned} \Pr\{X_{i,j} = n, X_{i,j+1} = n' | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\} \\ \equiv \sum_{\mathbf{x}} \delta(x_{i,j}, n) \delta(x_{i,j+1}, n') \Pr\{\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}^*, \hat{p}, \hat{K}\}, \end{aligned} \quad (49)$$

$$\begin{aligned} \Pr\{X_{i,j} = n, X_{i+1,j} = n' | \hat{K}\} \\ \equiv \sum_{\mathbf{x}} \delta(x_{i,j}, n) \delta(x_{i+1,j}, n') \Pr\{\mathbf{X} = \mathbf{x} | \hat{K}\}, \end{aligned} \quad (50)$$

$$\begin{aligned} \Pr\{X_{i,j} = n, X_{i,j+1} = n' | \hat{K}\} \\ \equiv \sum_{\mathbf{x}} \delta(x_{i,j}, n) \delta(x_{i,j+1}, n') \Pr\{\mathbf{X} = \mathbf{x} | \hat{K}\}. \end{aligned} \quad (51)$$

For the estimates of hyperparameters, $\hat{p}$ and $\hat{K}$, the restored image $\hat{\mathbf{x}}$ in the MML estimation is given as follows:

$$\begin{aligned} \hat{x}_{i,j} &= \arg \max_{n \in \mathbf{Q}} \Pr\{X_{i,j} = n | \mathbf{y}^*, \hat{p}, \hat{K}\} \\ &((i, j) \in \mathbf{I}). \end{aligned} \quad (52)$$

In this framework, we have to calculate the marginal probability distributions for the massive probabilistic model. An approximate method for calculating these marginal probability distributions for such a massive probabilistic model will be given as the CVM in Sec. 5.

5. Cluster Variation Method

In this section, we give the general frameworks of the CVM [Tanaka 95] 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:

$$\rho(\mathbf{x}) = \frac{\exp(-E(\mathbf{x}))}{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))}, \quad (53)$$

where

$$\begin{aligned} E(\mathbf{x}) &\equiv \sum_{(i,j) \in \mathbf{I}} \left(g_{i,j}(x_{i,j}) \right. \\ &\quad \left. + e(x_{i,j}, x_{i+1,j}) + e(x_{i,j}, x_{i,j+1}) \right). \end{aligned} \quad (54)$$

The Gibbs canonical distribution $\rho(\mathbf{x})$ satisfies the variational principle for the minimization of the free

energy defined by

$$\mathcal{F}[\rho] \equiv \sum_{\mathbf{z}} \rho(\mathbf{x}) \left(E(\mathbf{x}) + \ln(\rho(\mathbf{x})) \right). \quad (55)$$

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

$$\rho_{i,j}(n) \equiv \sum_{\mathbf{x}} \delta(n, x_{i,j}) \rho(\mathbf{x}), \quad (56)$$

and two-dimensional marginal probability distributions:

$$\rho_{i,j}^{i+1,j}(n, n') \equiv \sum_{\mathbf{x}} \delta(n, x_{i,j}) \delta(n', x_{i+1,j}) \rho(\mathbf{x}), \quad (57)$$

$$\rho_{i,j}^{i,j+1}(n, n') \equiv \sum_{\mathbf{x}} \delta(n, x_{i,j}) \delta(n', x_{i,j+1}) \rho(\mathbf{x}). \quad (58)$$

The marginal probability distributions satisfy the normalization conditions:

$$\sum_{n \in \mathbf{Q}} \rho_{i,j}(n) = 1, \quad (59)$$

$$\sum_{n \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} \rho_{i,j}^{i+1,j}(n, n') = \sum_{n \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} \rho_{i,j}^{i,j+1}(n, n') = 1, \quad (60)$$

and the reducibility conditions:

$$\begin{aligned} \rho_{i,j}(n) &= \sum_{n' \in \mathbf{Q}} \rho_{i,j}^{i+1,j}(n, n') = \sum_{n' \in \mathbf{Q}} \rho_{i-1,j}^{i,j}(n', n) \\ &= \sum_{n' \in \mathbf{Q}} \rho_{i,j}^{i,j+1}(n, n') = \sum_{n' \in \mathbf{Q}} \rho_{i,j-1}^{i,j}(n', n). \end{aligned} \quad (61)$$

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:

$$\begin{aligned} \rho(\mathbf{x}) &\simeq \prod_{(i,j) \in \mathbf{I}} \rho_{i,j}(x_{i,j}) \left(\frac{\rho_{i,j}^{i+1,j}(x_{i,j}, x_{i+1,j})}{\rho_{i,j}(x_{i,j}) \rho_{i+1,j}(x_{i+1,j})} \right) \\ &\quad \times \left(\frac{\rho_{i,j}^{i,j+1}(x_{i,j}, x_{i,j+1})}{\rho_{i,j}(x_{i,j}) \rho_{i,j+1}(x_{i,j+1})} \right). \end{aligned} \quad (62)$$

By applying Eqs. (54), (56), (57), (58) and (62) to Eq. (55), the free energy $\mathcal{F}[\rho]$ can be expressed in terms of the marginal probability distributions as follows:

$$\begin{aligned} \mathcal{F}[\rho] &\simeq \mathcal{F}[\{\rho_{i,j}, \rho_{i,j}^{i+1,j}, \rho_{i,j}^{i,j+1}\}] \\ &\equiv \sum_{(i,j)} \left\{ \sum_{n \in \mathbf{Q}} g_{i,j}(n) \rho_{i,j}(n) \right. \\ &\quad + \sum_{n \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} e(n, n') \left(\rho_{i,j}^{i+1,j}(n, n') + \rho_{i,j}^{i,j+1}(n, n') \right) \\ &\quad + S_{i,j} + (S_{i,j}^{i+1,j} - S_{i,j} - S_{i+1,j}) \\ &\quad \left. + (S_{i,j}^{i,j+1} - S_{i,j} - S_{i,j+1}) \right\}, \end{aligned} \quad (63)$$

where

$$S_{i,j}^{i+1,j} \equiv - \sum_{n \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} \rho_{i,j}^{i+1,j}(n, n') \ln(\rho_{i,j}^{i+1,j}(n, n')), \quad (64)$$

$$S_{i,j}^{i,j+1} \equiv - \sum_{n \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} \rho_{i,j}^{i,j+1}(n, n') \ln(\rho_{i,j}^{i,j+1}(n, n')). \quad (65)$$

By taking the first variation of $\mathcal{F}[\{\rho_{i,j}, \rho_{i,j}^{i+1,j}, \rho_{i,j}^{i,j+1}\}]$ 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 (59) and (60), and the reducibility conditions (61), the set of deterministic equations for the marginal probability distributions in the CVM is obtained as follows:

$$\rho_{i,j}(n) = \frac{\exp(-E_{i,j}(n))}{\sum_{m \in \mathbf{Q}} \exp(-E_{i,j}(m))} \quad ((i,j) \in \mathbf{I}), \quad (66)$$

$$\rho_{i,j}^{i+1,j}(m, n) = \frac{\exp(-E_{i,j}^{i+1,j}(m, n))}{\sum_{m' \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} \exp(-E_{i,j}^{i+1,j}(m', n'))} \quad ((i,j) \in \mathbf{I}), \quad (67)$$

$$\rho_{i,j}^{i,j+1}(m, n) = \frac{\exp(-E_{i,j}^{i,j+1}(m, n))}{\sum_{m' \in \mathbf{Q}} \sum_{n' \in \mathbf{Q}} \exp(-E_{i,j}^{i,j+1}(m', n'))} \quad ((i,j) \in \mathbf{I}), \quad (68)$$

$$\begin{aligned} E_{i,j}(n) &\equiv g_{i,j}(n) + \lambda_{i,j}^{i+1,j}(n) + \lambda_{i,j}^{i-1,j}(n) \\ &\quad + \lambda_{i,j}^{i,j+1}(n) + \lambda_{i,j}^{i,j-1}(n), \end{aligned} \quad (69)$$

$$\begin{aligned} E_{i,j}^{i+1,j}(m, n) &\equiv g_{i,j}(m) + g_{i+1,j}(n) + e(m, n) \\ &\quad + \lambda_{i,j}^{i-1,j}(m) + \lambda_{i,j}^{i,j-1}(m) + \lambda_{i,j}^{i,j+1}(m) \\ &\quad + \lambda_{i+1,j}^{i+2,j}(m) + \lambda_{i+1,j}^{i+1,j-1}(m) + \lambda_{i+1,j}^{i+1,j+1}(m), \end{aligned} \quad (70)$$

$$\begin{aligned} E_{i,j}^{i,j+1}(m, n) &\equiv g_{i,j}(m) + g_{i,j+1}(n) + e(m, n) \\ &\quad + \lambda_{i,j}^{i-1,j}(m) + \lambda_{i,j}^{i,j-1}(m) + \lambda_{i,j}^{i,j+1}(m) \\ &\quad + \lambda_{i+1,j}^{i+2,j}(m) + \lambda_{i+1,j}^{i+1,j-1}(m) + \lambda_{i+1,j}^{i+1,j+1}(m), \end{aligned} \quad (71)$$

$$\begin{aligned} \lambda_{i,j}^{i+1,j}(n) &= -g_{i,j}(n) \\ &\quad - \lambda_{i,j}^{i-1,j}(n) - \lambda_{i,j}^{i,j+1}(n) - \lambda_{i,j}^{i,j-1}(n) \\ &\quad - \ln \left(\sum_{m \in \mathbf{Q}} \exp(-E_{i,j}(m)) \right) \\ &\quad - \ln \left(\sum_{m \in \mathbf{Q}} \rho_{i,j}^{i+1,j}(n, m) \right) \quad (n \in \mathbf{Q}), \end{aligned} \quad (72)$$

$$\begin{aligned} \lambda_{i+1,j}^{i,j}(n) &= -g_{i+1,j}(n) \\ &\quad - \lambda_{i+1,j}^{i+2,j}(n) - \lambda_{i+1,j}^{i+1,j+1}(n) - \lambda_{i+1,j}^{i+1,j-1}(n) \\ &\quad - \ln \left(\sum_{m \in \mathbf{Q}} \exp(-E_{i+1,j}(m)) \right) \end{aligned}$$

$$-\ln\left(\sum_{m \in \mathbf{Q}} \rho_{i,j}^{i+1,j}(n,m)\right) \quad (n \in \mathbf{Q}), \quad (73)$$

$$\begin{aligned} \lambda_{i,j}^{i,j+1}(n) &= -g_{i,j}(n) \\ &\quad - \lambda_{i,j}^{i,j-1}(n) - \lambda_{i,j}^{i+1,j}(n) - \lambda_{i,j}^{i-1,j}(n) \\ &\quad - \ln\left(\sum_{m \in \mathbf{Q}} \exp(-E_{i,j}(m))\right) \\ &\quad - \ln\left(\sum_{m \in \mathbf{Q}} \rho_{i,j}^{i,j+1}(n,m)\right) \quad (n \in \mathbf{Q}), \end{aligned} \quad (74)$$

$$\begin{aligned} \lambda_{i,j+1}^{i,j}(n) &= -g_{i,j+1}(n) \\ &\quad - \lambda_{i,j+1}^{i,j+2}(n) - \lambda_{i,j+1}^{i+1,j+1}(n) - \lambda_{i,j+1}^{i-1,j+1}(n) \\ &\quad - \ln\left(\sum_{m \in \mathbf{Q}} \exp(-E_{i,j+1}(m))\right) \\ &\quad - \ln\left(\sum_{m \in \mathbf{Q}} \rho_{i,j+1}^{i,j}(n,m)\right) \quad (n \in \mathbf{Q}). \end{aligned} \quad (75)$$

The detailed calculations and explanations have been given in Ref. [Tanaka 95]. By applying these formulations to the CO and the MML estimation, 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 CO and the MML estimation. In the CO, we have to estimate the restored images $\hat{\mathbf{x}}$ from Eq. (17) for various values of hyperparameter J . For a sufficiently small positive value of T , Eq. (17) can be approximately reduced to the following equation:

$$\hat{x}_{i,j} = \arg \max_{n \in \mathbf{Q}} \rho_{i,j}(n|\mathbf{y}^*, T, J) \quad ((i,j) \in \mathbf{I}), \quad (76)$$

where

$$\begin{aligned} \rho_{i,j}(n|\mathbf{y}^*, T, J) &\equiv \sum_{\mathbf{x}} \delta(n, x_{i,j}) \rho(\mathbf{x}|\mathbf{y}^*, T, J) \\ &\quad (n \in \mathbf{Q}, (i,j) \in \mathbf{I}). \end{aligned} \quad (77)$$

In order to calculate $\rho_{i,j}(n|\mathbf{y}^*, T, J)$, we set that

$$g_{i,j}(n) \equiv -\frac{1}{T} \delta(n, y_{i,j}^*), \quad e(n, n') \equiv -\frac{J}{T} \delta(n, n'). \quad (78)$$

Approximate values of $\rho_{i,j}(n|\mathbf{y}^*, T, J)$ ($n \in \mathbf{Q}, (i,j) \in \mathbf{I}$) by means of the CVM are approximately obtained as follows:

$$\rho_{i,j}(n|\mathbf{y}^*, T, J) = \rho_{i,j}(n), \quad (79)$$

by solving In the MML estimation, we have to calculate the marginal probability distributions for $\Pr\{\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}^*, p, K\}$ and $\Pr\{\mathbf{X} = \mathbf{x}|K\}$ for various values of hyperparameters p and K . In order to calculate the marginal probability distributions for $\Pr\{\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}^*, p, K\}$, we set that

$$g_{i,j}(n) \equiv -\beta(p) \delta(n, y_{i,j}^*), \quad e(n, n') \equiv -K \delta(n, n'). \quad (80)$$

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

$$\Pr\{X_{i,j} = n|\mathbf{Y} = \mathbf{y}^*, p, K\} = \rho_{i,j}(n), \quad (81)$$

$$\begin{aligned} \Pr\{X_{i,j} = n, X_{i+1,j} = n'|\mathbf{Y} = \mathbf{y}^*, p, K\} \\ = \rho_{i,j}^{i+1,j}(n, n'), \end{aligned} \quad (82)$$

$$\begin{aligned} \Pr\{X_{i,j} = n, X_{i,j+1} = n'|\mathbf{Y} = \mathbf{y}^*, p, K\} \\ = \rho_{i,j}^{i,j+1}(n, n'), \end{aligned} \quad (83)$$

by solving Eqs. (66)-(75) for various values of p and K .

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

$$g_{i,j}(n) \equiv 0, \quad e(n, n') \equiv -K \delta(n, n'). \quad (84)$$

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

$$\Pr\{X_{i,j} = n, X_{i+1,j} = n'|K\} = \rho_{i,j}^{i+1,j}(n, n'), \quad (85)$$

$$\Pr\{X_{i,j} = n, X_{i,j+1} = n'|K\} = \rho_{i,j}^{i,j+1}(n, n'), \quad (86)$$

by solving Eqs. (66)-(75) for various values of p and K .

6. Numerical Experiments

In this section, some numerical experiments are given for the original images $\mathbf{x}^*$ in Figure 1(a), Figure 2(a), Figure 3(a) and Figure 4(a). By setting $(q-1)p^* = 0.2$, the degraded images $\mathbf{y}^*$ are produced from the original images $\mathbf{x}^*$ in the degradation process given in Eqs. (5) and (6). The degraded images $\mathbf{y}^*$ are shown in Figure 1(b), Figure 2(b), Figure 3(b) and Figure 4(b). We compare the restored image $\hat{\mathbf{x}}$ obtained by using the CO in Sec. 3 with the one by the MML estimation in Sec. 4. In both the CO and the MML estimation, the practical algorithms are constructed by means of the CVM given in Sec. 5. We remark that the main difference between the CO and the MML estimation is in the assumption for the *a priori* probability distribution. In the CO given in Sec. 3, we assume that the *a priori* probability distribution is given as the Gibbs micro-canonical distribution in Eq. (12) with Eq. (10). In the MML estimation given in Sec. 4, we adopt the Gibbs canonical distribution given in Eq. (41) with Eq. (10) as the *a priori* probability distribution.

In the CO, the optimal configurations $\hat{\mathbf{x}}$ are obtained from Eq. (76) for various values of Lagrange

multiplier J by means of the CVM. From the obtained optimal configurations for various values of J , the optimal value $\hat{J}$ and the restored image $\hat{\mathbf{x}}$ are chosen so as to satisfy the constrained condition $\sigma_2(\hat{\mathbf{x}}) = \sigma_2(\mathbf{x}^*)$. In all numerical experiments for the CO of the present paper, we use the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$, which is obtained by using simultaneous equations (35) and (36) only from the degraded image $\mathbf{y}^*$, as the value of $\sigma_2(\mathbf{x}^*)$ in the constraint (11). The true values of $\sigma_2(\mathbf{x}^*)/(2|\mathbf{I}|)$ and the estimates of $\hat{\sigma}_{2x}$ obtained from the given degraded images $\mathbf{y}^*$ by using Eqs. (35) and (36) are shown in Table 1. The optimal values of Lagrange multipliers, $\hat{J}$, which are estimated from the given degraded images $\mathbf{y}^*$ by means of the CO, are shown in Table 2. The restored images $\hat{\mathbf{x}}$ obtained by using the CO are shown in Figure 1(c), Figure 2(c), Figure 3(c) and Figure 4(c). The values of $d(\hat{\mathbf{x}}, \mathbf{x}^*)/|\mathbf{I}|$ and

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

for the restored images $\hat{\mathbf{x}}$ obtained by means of the CO, are also shown in Table 2. Here, Δ_{SNR} is an improvement of signal-to-noise ratio.

In the MML estimation, the right-hand side in Eq. (46) and the left-hand sides in Eq. (45) and Eq. (46) 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 (45) and (46) for $\Pr\{\mathbf{Y} = \mathbf{y}^* | p, K\}$. At the estimates of hyperparameters, $\hat{p}$ and $\hat{K}$, the restored image $\hat{\mathbf{x}}$ are obtained from Eq. (52). The estimates of hyperparameters, $\hat{p}$ and $\hat{K}$, which are estimated from the given degraded images $\mathbf{y}^*$ by means of the MML estimation, are shown in Table 3. The restored images $\hat{\mathbf{x}}$ obtained by using the MML estimation are shown in Figure 1(d), Figure 2(d), Figure 3(d) and Figure 4(d). The values of $d(\hat{\mathbf{x}}, \mathbf{x}^*)/|\mathbf{I}|$ and Δ_{SNR} for the restored images $\hat{\mathbf{x}}$ obtained by means of the MML estimation are also shown in Table 3.

Before closing this section, we remark again that, all restored images in Figure 1(c), Figure 2(c), Figure 3(c) and Figure 4(c) and all values of $\hat{J}$, $d(\hat{\mathbf{x}}, \mathbf{x}^*)/|\mathbf{I}|$ and Δ_{SNR} in Table 2 are obtained by adopting the values of $2|\mathbf{I}|\hat{\sigma}_{2x}$ given in Table 1 as the values of $\sigma_2(\mathbf{x}^*)$ in the constraint condition (11) of the CO. The values of $\hat{\sigma}_{2x}$ given in Table 1 are estimated only from the degraded images $\mathbf{y}^*$ by using Eqs. (35) and (36) in Sec. 3.

Fig. 1 Binary image restorations by means of the constrained optimization (CO) and the maximum marginal likelihood (MML) estimation ($|\mathbf{I}| = 256 \times 256$). (a) Original image $\mathbf{x}^*$ ($q = 2$). (b) Degraded image $\mathbf{y}^*$ ($(q-1)p^* = 0.2$). (c) Restored image $\hat{\mathbf{x}}$ by the CO given in Sec. 3. We remark that the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$ given in table Table 1 is used as the value of $\sigma_2(\mathbf{x}^*)$ in the constraint (11) in the CO. (d) Restored image $\hat{\mathbf{x}}$ by the MML estimation given in Sec. 4.

7. Concluding Remarks

In this paper, we give some comparisons the CO with the MML estimation not only for binary image restorations but also for grey-level ones. We extend the framework of the CO given only for binary image restorations in Ref. [Morita 96] to grey-level image restorations. The most important point of the extension is how the length of boundaries $\sigma_2(\mathbf{x}^*)$ between areas with different grey-levels in the original image $\mathbf{x}^*$ is estimated from only the given degraded image $\mathbf{y}^*$. In Ref. [Morita 96], the estimation method was given only in the case of binary image restoration. In this paper, we extend the estimation method of $\sigma_2(\mathbf{x}^*)$ for binary images to the one for grey-level images. In the image restoration by the CO, we have to search the optimal solution of the massive optimization problem (14). In order to search the optimal solution for various values of Lagrange multiplier J , we introduce a Gibbs distribution (16) and construct two annealing algorithms for the optimization by means of the CVM. On the other hand, we have to calculate some statistical quantities in the massive probabilistic models given by Eqs. (41) and (42), in the MML esti-

Fig. 2 Binary image restorations by means of the constrained optimization (CO) and the maximum marginal likelihood (MML) estimation ($|\mathbf{I}| = 256 \times 256$). (a) Original image $\mathbf{x}^*$ ($q = 2$). (b) Degraded image $\mathbf{y}^*$ ($((q-1)p^* = 0.2)$). (c) Restored image $\hat{\mathbf{x}}$ by the CO given in Sec. 3. We remark that the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$ given in table Table 1 is used as the value of $\sigma_2(\mathbf{x}^*)$ in the constraint (11) in the CO. (d) Restored image $\hat{\mathbf{x}}$ by the MML estimation given in Sec. 4.

mation. We gave the algorithm also by means of the CVM in order to calculate some statistical quantities for various values of hyperparameters p and K .

The image restoration algorithm in the CO can be constructed from only the given degraded image under the assumptions (i), (ii) and (iii). The assumption (iii) favor that an original image has a lot of flat region. For original images that satisfy the assumption (iii) at most of all pixels, we obtain good restored images. On the other hand, if the original image does not satisfy the assumption (iii) at a lot of pixels, we obtain not so good results. The reason is that the assumption (iii) is too strongly forced on the image restorations in the CO.

As another hyperparamter estimation method in MRF models, we have the MML estimation [Pryce 95, Zhou 97], which is one of familiar techniques in statistics. In the present paper, we explained the framework of the MML estimation under the assumptions (i), (ii) and (iv). If we compare the results obtained by using the CO with those by the MML estimation for the degraded images given in Figure 1(b) and Figure 3(b), the restored images $\hat{\mathbf{x}}$ in the CO are better than those in the MML estimation. The rea-

Fig. 3 Image restorations by means of the constrained optimization (CO) and the maximum marginal likelihood (MML) estimation ($|\mathbf{I}| = 256 \times 256$). (a) Original image $\mathbf{x}^*$ ($q = 4$). (b) Degraded image $\mathbf{y}^*$ ($((q-1)p^* = 0.2)$). (c) Restored image $\hat{\mathbf{x}}$ by the CO given in Sec. 3. We remark that the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$ given in table Table 1 is used as the value of $\sigma_2(\mathbf{x}^*)$ in the constraint (11) in the CO. (d) Restored image $\hat{\mathbf{x}}$ by the MML estimation given in Sec. 4.

son is that the assumption (iii) satisfies at most of all pixels in the original images given in Figure 1(a) and Figure 3(a). On the other hand, the original images given in Figure 2(a) and Figure 4(a) include a lot of patterns which do not satisfy the assumption (iii). In the image restorations for the degraded images given in Figure 2(b) and Figure 4(b), the restored images $\hat{\mathbf{x}}$ obtained by means of the MML estimation are better than those by the CO. Moreover, the results obtained from Figure 1(b) and Figure 3(b), by using the MML are also not so bad as shown in Figure 1(d) and Figure 3(d). It is shown that the MML estimation can give us a good results in the image restorations not only for the original images which satisfy the assumption (iii) but also for the original images which always not satisfy the assumption (iii) perfectly.

◇ Appendix ◇

1. Derivation of Eqs. (35) and (36).

The marginal probabilities $P_\nu^{(1)}(m)$, $P_\nu^{(2)}(m, m')$ and $P_\nu^{(3)}(m, m')$ ($\nu = x, y$, $m, m' \in \mathbf{Q}$) are defined by the

Fig. 4 Image restorations by means of the constrained optimization (CO) and the maximum marginal likelihood (MML) estimation ($|\mathbf{I}| = 256 \times 256$). (a) Original image $\mathbf{x}^*$ ($q = 4$). (b) Degraded image $\mathbf{y}^*$ ($((q-1)p^* = 0.2)$). (c) Restored image $\hat{\mathbf{x}}$ by the CO given in Sec. 3. We remark that the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$ given in table Table 1 is used as the value of $\sigma_2(\mathbf{x}^*)$ in the constraint (11) in the CO. (d) Restored image $\hat{\mathbf{x}}$ by the MML estimation given in Sec. 4.

following reducibilities:

$$P_\nu^{(1)}(m) = \sum_{m' \in \mathbf{Q}} \sum_{m'' \in \mathbf{Q}} P_\nu(m, m', m''), \quad (1)$$

$$P_\nu^{(2)}(m, m') = \sum_{m'' \in \mathbf{Q}} P_\nu(m, m', m''), \quad (2)$$

$$P_\nu^{(3)}(m, m') = \sum_{m'' \in \mathbf{Q}} P_\nu(m, m'', m'). \quad (3)$$

By these definitions, $\hat{\sigma}_{2x}$, $\hat{\sigma}_{2y}$ and $\hat{\sigma}_{3y}$ are given by

$$\hat{\sigma}_{2x} = 1 - \sum_{m \in \mathbf{Q}} P_x^{(2)}(m, m), \quad (4)$$

$$\hat{\sigma}_{2y} = 1 - \sum_{n \in \mathbf{Q}} P_y^{(2)}(n, n), \quad (5)$$

$$\hat{\sigma}_{3y} = 1 - \sum_{n \in \mathbf{Q}} P_y^{(3)}(n, n). \quad (6)$$

Substituting Eq. (6) into Eq. (29), we obtain

$$\begin{aligned} P_y(n, n', n'') &= p^{*3} \\ &+ \left(P_x^{(1)}(n) + P_x^{(1)}(n') + P_x^{(1)}(n'') \right) p^{*2} (1 - qp^*) \\ &+ \left(P_x^{(2)}(n, n') + P_x^{(2)}(n', n'') \right. \\ &\quad \left. + P_x^{(3)}(n, n'') \right) p^* (1 - qp^*)^2 \\ &+ P_x(n, n', n'') (1 - qp^*)^3. \end{aligned} \quad (7)$$

Table 1 Values of $\sigma_2(\mathbf{x}^*)/(2|\mathbf{I}|)$, which are estimated from the original images $\mathbf{x}^*$ given in Figure 1(a), Figure 2(a), Figure 3(a) and Figure 4(a), and values of $\hat{\sigma}_{2x}$, which are obtained by using Eqs. (35) and (36) only from the degraded images $\mathbf{y}^*$ given in Figure 1(b), Figure 2(b), Figure 3(b) and Figure 4(b).

(a) $q = 2$.

$\mathbf{x}^*$	$\mathbf{y}^*$	$\sigma_2(\mathbf{x}^*)/(2 \mathbf{I})$	$\hat{\sigma}_{2x}$
Fig. 1a	Fig. 1b	0.0629	0.0620
Fig. 2a	Fig. 2b	0.1700	0.0624

(b) $q = 4$.

$\mathbf{x}^*$	$\mathbf{y}^*$	$\sigma_2(\mathbf{x}^*)/(2 \mathbf{I})$	$\hat{\sigma}_{2x}$
Fig. 3a	Fig. 3b	0.1004	0.0993
Fig. 4a	Fig. 4b	0.3689	0.1383

Table 2 Values of $\hat{J}$, $d(\mathbf{x}^*, \hat{\mathbf{x}})/|\mathbf{I}|$ and Δ_{SNR} (dB) in the CO given in Sec. 3. We remark that the value of $2|\mathbf{I}|\hat{\sigma}_{2x}$ given in Table 1 is used as the value of $\sigma_2(\mathbf{x}^*)$ in the constraint (11).

(a) $q = 2$.

Fig.	$\hat{J}$	$d(\mathbf{x}^*, \hat{\mathbf{x}})/ \mathbf{I} $	Δ_{SNR} (dB)
1c	0.6081	0.04106	6.8523
2c	0.5793	0.12437	2.0393

(b) $q = 4$.

Fig.	$\hat{J}$	$d(\mathbf{x}^*, \hat{\mathbf{x}})/ \mathbf{I} $	Δ_{SNR} (dB)
3c	0.5750	0.0320	10.353
4c	0.5696	0.2375	2.4278

Table 3 Values of $\hat{p}$, $\hat{K}$, $d(\mathbf{x}^*, \hat{\mathbf{x}})$ and Δ_{SNR} (dB) in the restored images $\hat{\mathbf{x}}$ which is obtained by using the MML estimation given in Sec. 4.

(a) $q = 2$.

Fig.	$(q-1)\hat{p}$	$\hat{K}$	$d(\mathbf{x}^*, \hat{\mathbf{x}})/ \mathbf{I} $	Δ_{SNR} (dB)
1d	0.1331	0.7825	0.0844	3.7218
2d	0.1720	0.7653	0.1176	2.2832

(b) $q = 4$.

Fig.	$(q-1)\hat{p}$	$\hat{K}$	$d(\mathbf{x}^*, \hat{\mathbf{x}})/ \mathbf{I} $	Δ_{SNR} (dB)
3d	0.1295	1.1018	0.0732	4.3300
4d	0.2371	1.1022	0.1693	2.9534

By using the definitions (2) and (3), we obtain

$$\begin{aligned} P_y^{(3)}(n, n) &= p^{*2} + 2P_x^{(1)}(n)p^*(1 - qp^*) \\ &\quad + P_x^{(3)}(n, n)(1 - qp^*)^2, \end{aligned} \quad (8)$$

$$\begin{aligned} P_y^{(2)}(n, n) &= p^{*2} + 2P_x^{(1)}(n)p^*(1 - qp^*) \\ &\quad + P_x^{(2)}(n, n)(1 - qp^*)^2. \end{aligned} \quad (9)$$

If we set $m \neq m'$ in Eq. (2), we obtain

$$P_x^{(2)}(m, m') = P_x(m, m', m') = P_x(m', m, m'), \quad (10)$$

because of Eqs. (30) and (31). Because of the relations (30) and (10), we obtain

$$P_x^{(3)}(m, m') = 2P_x^{(2)}(m, m') \quad (m \neq m'). \quad (11)$$

By summing Eqs. (8) and (9) with respect to y and by using Eq. (11) and relations (4)-(6), (33) and (34), we obtain Eqs. (35) and (36).

◇ References ◇

- [Derin 84] H. Derin, H. Elliott, R. Cristi and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.6, no.6, p.707-720, 1984.
- [Pryce 95] J. M. Pryce and A. D. Bruce, "Statistical mechanics of image restoration," Journal of Physics A, vol.28, no.3, pp.511-532, 1995.
- [Zhou 97] Z. Zhou, R. M. Leahy and J. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," IEEE Trans. Image Processing, vol.6, no.6, pp.844-861, 1997.
- [Morita 96] T. Morita and K. Tanaka, "Determination of parameters in an image recovery of statistical-mechanical means," Physica A, vol.223, pp.244-262, 1996.
- [Tanaka 99] T. Tanaka, "Mean-field approximation and maximum likelihood estimation," Technical Report of IE-ICE, NC98-173, pp.155-162, March 1999 (in Japanese).
- [Wu 82] F. Y. Wu, "The Potts model," Reviews of Modern Physics, vol.54, no.1, pp.235-268, 1982.
- [Tanaka 95] K. Tanaka and T. Morita, "Cluster variation method and image restoration problem," Physics Letters A, vol.203, no.2-3, pp.122-128, 1995.
- [Marroquin 87] J. Marroquin, S. Mitter and T. Poggio, "Probabilistic solution of ill-posed problems in computer vision," Journal of the American Statistical Association, vol.82, no.397, pp.76-89, 1987.
- [Nishimori 99] H. Nishimori and K. Y. M. Wong, "Statistical mechanics of image restoration and error-correcting codes," Physical Review E, vol.60, no.1, pp.132-144, 1999.
- [Tanaka 00] K. Tanaka, "Theoretical Study of Hyperparameter Estimation by Maximization of Marginal Likelihood in Image Restoration by means of Cluster Variation Method," Trans. Inst. Electron. Inf. Commun. Eng. A, vol. J83-A, no.10, pp.1148-1160, 2000 (in Japanese).
- [Tanaka 97] K. Tanaka and T. Morita, "Statistical-mechanical iterative algorithm for image restoration of a grey level image," Trans. Inst. Electron. Inf. Commun. Eng. A, vol. J80-A, no.6, pp.1033-1037, 1997 (in Japanese).

〔担当委員：新田克己〕

Received October 6, 2000.

Acknowledgments

The author is grateful to Professor D. M. Titterton of the Department of Statistics, University of Glasgow, Professor T. Morita, Department of Computer Science, College of Engineering, Nihon University and Professor T. Horiguchi of the Department of Computer and Mathematical Sciences, Graduate School of Information Science, Tohoku University.

Author's Profile

Tanaka, Kazuyuki (Member)

Received his B.E. and Ph.D degrees in electronic engineering from Tohoku University in 1984 and 1989, respectively. In 1989, he joined Fac. of Eng. of Tohoku University as research associate. He joined Graduate School of Information Sciences of the same university, in 1993. He joined Muroran Institute of Technology as associate professor in 1994. Since April 1999, he has been an associate professor at Graduate School of Information Sciences of Tohoku University. He has been doing research in the areas of probabilistic information processing. He is a member of Institute of Electronics, Information and Communication Engineerings, Society of Instrument and Control Engineerings, Physical Society of Japan and Information Processing Society of Japan.