# LOOPY BELIEF PROPAGATION AND PROBABILISTIC IMAGE PROCESSING

Kazuyuki Tanaka
Graduate School of Information Sciences, Tohoku University
Aramaki-Aza-Aoba 04, Aoba-ku, Sendai 980-8579, Japan
E-mail: kazu@statp.is.tohoku.ac.jp

Jun-ichi Inoue
Graduate School of Engineering, Hokkaido University
H13-W8, Kita-ku, Sapporo 060-8628, Japan
E-mail: j_inoue@complex.eng.hokudai.ac.jp

D. M. Titterington
Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK
E-mail: mike@stats.gla.ac.uk

**Abstract. The hyperparameter estimation in the maximization of the marginal likelihood in the probabilistic image processing is investigated by using the cluster variation method. The algorithms are substantially equivalent to generalized loopy belief propagations.**

## INTRODUCTION

Combining Bayesian approach with Markov random fields is one of powerful methods for probabilistic image processing[1]. It is also known that advanced mean-field methods and other statistical-mechanical methods are applicable to the Bayesian image processing[2, 3].

Advanced mean-field methods have been widely applied to many problems in computer science[4]. The advanced mean-field methods can be formulated by means of the variational principle for the minimization of an approximate Kullback-Leibler divergence or an approximate free energy. We have a cluster variation method, which is sometimes referred to as Kikuchi method[5, 6], as one of the familiar advanced mean-field methods. Probabilistic inference algorithms in artificial intelligence are constructed by applying loopy belief propagations to probabilistic models with graph representations for nodes[8]. It was suggested that, for probabilistic models, the extremum conditions for approximate free energies in some approximations of the cluster variation method is equivalent to simultaneous fixed point equations of the loopy belief propagations and the ordinary loopy belief propagation can be extended to a generalized belief propagation by using the cluster variation method[11, 12].

Tanaka and Morita[7] proposed to construct the probabilistic image processing algorithms in the Bayesian approach with Markov random fields by using the cluster variation method. Weiss[9] also investigated an approximate inference in Markov random fields by means of some loopy belief propagations. Tanaka[10] applied a pair approximation in the cluster variation method to the hyperparameter estimation by means of the maximization of marginal likelihood and concluded that a pair approximation in the cluster variation method can improve the image quality for restoration in contrast to the mean-field approximation.

In the present paper, we investigate the hyperparameter estimation in the maximization of the marginal likelihood in the probabilistic image processing by using the cluster variation method. We adopt a square-cactus approximation as well as a pair approximation in the cluster variation method. The algorithms are substantially equivalent to the generalized loopy belief propagation.

## BAYESIAN IMAGE ANALYSIS BY GRAPHICAL MODEL

We consider a image on a square lattice $\Omega{\equiv}\{i\}$ with $Q$ gray-levels. Each pixel takes one of the gray-levels $\boldsymbol{Q} = \{0, 1, 2, \cdots, Q-1\}$. Each gray-level expresses the intensity of light at a pixel in computer vision and 0 and $Q-1$ is corresponding to the black and the white. A random variable of intensity at each pixel $i$ in the original image and the degraded image are denoted by $F_i$ and $G_i$, respectively. Then the random fields of intensities in the original image and the degraded image are represented by $\boldsymbol{F}{\equiv}\{F_i|i{\in}\Omega\}$ and $\boldsymbol{G}{\equiv}\{G_i|i{\in}\Omega\}$, respectively. The original image and the degraded image are denoted by $\boldsymbol{f} = \{f_i\}$ and $\boldsymbol{g} = \{f_i\}$, respectively. Here $f_i$ and $g_i$ are the intensities at the pixel $i$ in the original image and the degraded image, respectively.

In the present paper, the degradation process is assumed that the degraded image $\boldsymbol{g}$ is generated from the original image $\boldsymbol{f}$ by changing the intensity of each pixel to another intensity the same probability $p$, independently of the other pixels. The conditional probability distribution of the degradation process when the original image is $\boldsymbol{f}$ is given as follows:

$$\Pr\{\boldsymbol{G} = \boldsymbol{g}|\boldsymbol{F} = \boldsymbol{f}, p\} = \prod_{i\in\Omega}\Big(p(1-\delta_{f_i,g_i}) + \big(1-(Q-1)p\big)\delta_{f_i,g_i}\Big), \tag{1}$$

where $\delta_{a,b}$ is the Kronecker delta. Moreover, the *a priori* probability distribution that the original image is $\boldsymbol{f}$ is assumed to be as

$$\Pr\{\boldsymbol{F} = \boldsymbol{f}|\alpha\} = \frac{\displaystyle\prod_{ij\in B}\exp\Big(-\frac{1}{2}\alpha(f_i - f_j)^2\Big)}{\displaystyle\sum_{\boldsymbol{z}}\prod_{ij\in N}\exp\Big(-\frac{1}{2}\alpha(z_i - z_j)^2\Big)}, \tag{2}$$

where $N$ is the set of all the nearest-neighbor pairs of pixels on the square lattice $\Omega$ and the summation $\sum_{\boldsymbol{z}}$ is taken over all possible configurations of images $\boldsymbol{z} = \{z_i | i \in \Omega\}$. By substituting Eqs.(1) and (2) into the Bayes formula:

$$\Pr\{\boldsymbol{F} = \boldsymbol{f} | \boldsymbol{G} = \boldsymbol{g}, p, \alpha\} = \frac{\Pr\{\boldsymbol{G} = \boldsymbol{g} \big| \boldsymbol{F} = \boldsymbol{f}, p\} \Pr\{\boldsymbol{F} = \boldsymbol{f} | \alpha\}}{\sum_{\boldsymbol{z}} \Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{z}, p\} \Pr\{\boldsymbol{F} = \boldsymbol{z} | \alpha\}}, \tag{3}$$

we obtain the *a posteriori* probability distribution $\Pr\{\boldsymbol{F} = \boldsymbol{f} | \boldsymbol{G} = \boldsymbol{g}, p, \alpha\}$.

In the maximum marginal likelihood estimation in statistics, the hyper-parameters $\alpha$ and $p$ are determined so as to maximize the marginal likelihood $\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}$:

$$\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\} \equiv \sum_{\boldsymbol{z}} \Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{z}, p\} \Pr\{\boldsymbol{F} = \boldsymbol{z} | \alpha\}. \tag{4}$$

The maximizers of marginal likelihood $\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}$ are denoted by $\widehat{\alpha}$ and $\widehat{p}$, such that

$$(\widehat{\alpha}, \widehat{p}) = \arg \max_{(\alpha, p)} \Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}. \tag{5}$$

The conditions for an extremum of $\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}$ at $\alpha = \widehat{\alpha}$ and $p = \widehat{p}$ can be reduced to the following simultaneous equations:

$$\begin{aligned} &\sum_{ij \in B} \sum_{z_i \in \boldsymbol{Q}} \sum_{z_j \in \boldsymbol{Q}} (z_i - z_j)^2 \Pr\{F_i = z_i, F_j = z_j | \boldsymbol{G} = \boldsymbol{g}, \widehat{\alpha}, \widehat{p}\} \\ &\qquad = \sum_{ij \in N} \sum_{z_i \in \boldsymbol{Q}} \sum_{z_j \in \boldsymbol{Q}} (z_i - z_j)^2 \Pr\{F_i = z_i, F_j = z_j | \widehat{\alpha}\}, \end{aligned} \tag{6}$$

$$\sum_{\{i | i \in \Omega\}} \sum_{z_i \in \boldsymbol{Q}} \delta_{z_i, g_i} \Pr\{F_i = z_i | \boldsymbol{G} = \boldsymbol{g}, \widehat{\alpha}, \widehat{p}\} = 1 - (Q - 1)p. \tag{7}$$

Here $\Pr\{F_i = f_i | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$, $\Pr\{F_i = f_i, F_j = f_j | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j | \alpha\}$ are marginal probabilities defined by

$$\Pr\{F_i = f_i | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\} \equiv \sum_{\boldsymbol{z}} \delta_{f_i, z_i} \Pr\{\boldsymbol{F} = \boldsymbol{z} | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\}, \tag{8}$$

$$\Pr\{F_i = f_i, F_j = f_j | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\} \equiv \sum_{\boldsymbol{z}} \delta_{f_i, z_i} \delta_{f_j, z_j} \Pr\{\boldsymbol{F} = \boldsymbol{z} | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\}, \tag{9}$$

$$\Pr\{F_i = f_i, F_j = f_j | \alpha\} \equiv \sum_{\boldsymbol{z}} \delta_{f_i, z_i} \delta_{f_j, z_j} \Pr\{\boldsymbol{F} = \boldsymbol{z} | \alpha\}. \tag{10}$$

For the obtained estimates $\widehat{\alpha}$ and $\widehat{p}$, the restored image $\widehat{\boldsymbol{f}} = \{\widehat{f}_i | i \in \Omega\}$ is determined by

$$\widehat{f}_i = \arg \max_{z_i \in \boldsymbol{Q}} \Pr\{F_i = z_i | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\}. \tag{11}$$

The estimation framework for restored image is called maximum posterior marginal estimation.

## LOOPY BELIEF PROPAGATION AND CLUSTER VARIATION METHOD

In the above framework, we have to calculate the marginal probability distributions $\Pr\{F_i = f_i|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ $(i{\in}\Omega)$, $\Pr\{F_i = f_i, F_j = f_j|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ $(ij{\in}B)$ and $\Pr\{F_i = f_i, F_j = f_j|\alpha\}$ $(ij{\in}B)$. Since it is hard to calculate these marginal probability distributions exactly, we apply the Bethe approximation to the above probabilistic models given by $\Pr\{\boldsymbol{F} = \boldsymbol{f}|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ and $\Pr\{\boldsymbol{F} = \boldsymbol{f}|\alpha\}$.

In order to explain the framework of CVM, we should define some notations for clusters. Cluster is a set of nodes. When a node $i$ belongs to a cluster $\gamma$, we call $i$ an element of $\gamma$ and we express it in terms of the notation $i{\in}\gamma$. When all the nodes in a cluster $\gamma'$ belong to a cluster $\gamma$, we call $\gamma'$ a subcluster of $\gamma$. When a cluster $\gamma'$ is a subcluster of a cluster $\gamma$, we use the notation $\gamma'{\leq}\gamma$. We express $\gamma' < \gamma$ when a cluster $\gamma'$ is a proper subcluster of $\gamma$. The set of all the nodes, which are belonging to the cluster $\gamma$ and are not belonging to the cluster $\gamma'$, is denoted by the notation $\gamma\backslash\gamma'$. The notation $\Omega\backslash\gamma$ is the set of all nodes not belonging to the cluster $\gamma$.

First of all, we have to specify a set of basic clusters. Every basic cluster must not be a subcluster of another element in the set of basic clusters. We denote the set of basic clusters by $B$. We consider such a set $C$ of clusters that a cluster is in $C$ if and only if it is a cluster in $B$ or is the cluster of the common nodes of two or more clusters in $B$, excluding the empty cluster $\mathbf{0}$. The set of all the clusters, which are belonging to the set $C$ and are not belonging to the set $B$, is denoted by the notation $C\backslash B$. A set of random variables $f_i$ associated with nodes $i$ belonging to a cluster $\gamma$ is denoted by $\boldsymbol{f}_\gamma{\equiv}\{f_i|i{\in}\gamma\}$.

We consider a probability distribution given by

$$P(\boldsymbol{f}){\equiv}\frac{1}{Z}\prod_{\{\gamma|\gamma\in C\}} W_\gamma(\boldsymbol{f}_\gamma)^{-\mu(\gamma)} \tag{12}$$

where $Z$ is a normalization constant and $\mu(\alpha)$ $(\alpha{\in}C)$ is a Möbius function defined by

$$\mu(\alpha){\equiv}-1-\sum_{\{\gamma|\gamma>\alpha,\ \gamma\in C\}}\mu(\gamma) \qquad (\alpha{\in}C). \tag{13}$$

We wish to compute the marginal probability distribution defined by

$$P_\alpha(\boldsymbol{f}_\alpha){\equiv}\sum_{\boldsymbol{z}}\Big(\prod_{\{i|i\in\alpha\}}\delta_{f_i,z_i}\Big)P(\boldsymbol{z}). \tag{14}$$

It is difficult to obtain the exact values of the marginal probability distribution and we have to employ some approximations.

We introduce a Kullback-Leibler divergence between the probability distribution $P(\boldsymbol{f})$ and a probability distribution $Q(\boldsymbol{f})$, which is defined by

$$\mathcal{D}[Q||P]\equiv\sum_{\boldsymbol{f}}Q(\boldsymbol{f})\ln\Big(\frac{Q(\boldsymbol{f})}{P(\boldsymbol{f})}\Big). \tag{15}$$

In the cluster variation method, the probability distribution $P(\boldsymbol{f})$ is approximately restricted to be the following form:

$$P(\boldsymbol{f})\simeq\prod_{\{\gamma|\gamma\in C\}}Q_{\gamma}(\boldsymbol{f}_{\gamma})^{-\mu(\gamma)}, \tag{16}$$

where $Q_{\alpha}(\boldsymbol{f}_{\alpha})$ is the marginal probability distribution of the probability distribution $Q(\boldsymbol{f})$ and is defined by

$$Q_{\alpha}(\boldsymbol{f}_{\alpha})\equiv\sum_{\boldsymbol{z}}\Big(\prod_{\{i|i\in\alpha\}}\delta_{f_i,z_i}\Big)Q(\boldsymbol{z}). \tag{17}$$

The definition (17) can be rewritten as the following relations among the marginal probability distributions $Q_{\alpha}(\boldsymbol{f}_{\alpha})$:

$$Q_{\alpha}(\boldsymbol{f}_{\alpha})=\sum_{\boldsymbol{z}_{\gamma}}(\prod_{\{i|i\in\alpha\}}\delta_{f_i,z_i})Q_{\gamma}(\boldsymbol{z}_{\gamma}) \quad (\alpha<\gamma,\ \alpha\in C\backslash B,\ \gamma\in C). \tag{18}$$

In the cluster variation method, the above relations are called *reducibilities*.

By using Eqs.(16) and (17), the Kullback-Leibler divergence $\mathcal{D}[Q||P]$ can be reduced to

$$\mathcal{D}[Q||P]=\mathcal{F}[\{Q_{\gamma}|\gamma\in C\}]-\ln(Z), \tag{19}$$

where

$$\mathcal{F}[\{Q_{\gamma}|\gamma\in C\}]\equiv-\sum_{\{\gamma|\gamma\in C\}}\mu(\gamma)\mathcal{D}[Q_{\gamma}||W_{\gamma}]. \tag{20}$$

If we choose functions $W_{\gamma}(\boldsymbol{f}_{\gamma})$ so as to satisfy the normalization conditions, $\mathcal{D}[Q_{\gamma}||W_{\gamma}]$ can be also regarded as the Kullback-Leibler divergence between $Q_{\gamma}(\boldsymbol{f}_{\gamma})$ and $W_{\gamma}(\boldsymbol{f}_{\gamma})$. In the cluster variation method, the approximate marginal probability distributions $P_{\alpha}(\boldsymbol{f}_{\alpha})$ are determined so as to minimize the right-hand side of Eq.(19), the approximate form of the Kullback-Leibler divergence in the cluster variation method, under the normalizations and the reducibilities in the marginal probability distributions as follows:

$$\begin{aligned}\{P_{\alpha}|\alpha\in C\}\simeq\arg\min_{\{Q_{\alpha}|\alpha\in C\}}\Big\{\mathcal{F}[\{Q_{\alpha}|\alpha\in C\}]\Big|\sum_{\boldsymbol{z}_{\alpha}}Q_{\alpha}(\boldsymbol{z}_{\alpha})=1\ (\alpha\in C),\\ Q_{\alpha}(\boldsymbol{f}_{\alpha})=\sum_{\boldsymbol{z}_{\gamma}}(\prod_{\{i|i\in\alpha\}}\delta_{f_i,z_i})Q_{\gamma}(\boldsymbol{z}_{\gamma})\ (\alpha<\gamma\in C)\Big\}.\end{aligned} \tag{21}$$

By introducing the Lagrange multipliers for the normalizations and the reducibilities and by taking the first variation of $\mathcal{F}[\{Q(\gamma)|\gamma\in C\}]$ with respect to marginal probability distributions $Q_\alpha(\boldsymbol{f}_\alpha)$, the approximate forms of the marginal probability distributions $Q_\alpha(\boldsymbol{f}_\alpha)$ can be derived as follows:

$$Q_\alpha(\boldsymbol{f}_\alpha) = \frac{1}{Z_\alpha} W_\alpha(\boldsymbol{f}_\alpha) \prod_{\{\gamma|\gamma\leq\alpha,\ \gamma\in C\backslash B\}} \exp(\lambda_{\gamma,\alpha}(\boldsymbol{f}_\gamma)), \tag{22}$$

where

$$\lambda_{\alpha,\alpha}(\boldsymbol{f}_\alpha) = -\sum_{\{\gamma|\gamma>\alpha,\ \gamma\in C\}} \frac{\mu(\gamma)}{\mu(\alpha)} \lambda_{\alpha,\gamma}(\boldsymbol{f}_\alpha). \tag{23}$$

Here $\lambda_{\alpha,\gamma}$ are Lagrange multipliers for the reducibilities (21) and are determined so as to satisfy Eqs.(21) and (23) with (22).

We choose the sets $B$ and $C$ so as to satisfy $C\backslash B = \Omega$ and consider the cluster variation method for the probabilistic model

$$P(\boldsymbol{f}) = \frac{1}{Z}\Bigg(\prod_{\{i|i\in\Omega\}} W_i(f_i)\Bigg)\Bigg(\prod_{\{\alpha|\alpha\in B\}} W_\alpha(\boldsymbol{f}_\alpha)\Big(\prod_{\{j|j\in\alpha\}} W_j(f_j)^{-1}\Big)\Bigg). \tag{24}$$

Here $W_\alpha(\boldsymbol{f}_\alpha)$ and $W_i(f_i)$ are always positive for any values of $\boldsymbol{f}_\alpha$ and $f_i$. We introduce a set $\boldsymbol{c}_i\equiv\{\gamma|\gamma > i,\ \gamma\in C\}$ for each pixel $i$. By replacing $\lambda_{i,\alpha}(f_i)$ by $M_{\alpha\to i}(f_i)$ as follows:

$$M_{\alpha\to i}(f_i) = \prod_{\gamma\in\boldsymbol{c}_i\backslash\alpha} \exp\Big(\lambda_{i,\alpha}(f_i)\Big), \tag{25}$$

the simultaneous equations for the sets of marginal probabilities $\{P_i(\xi)|i\in\Omega\}$ and $\{P_\alpha(\boldsymbol{f}_\alpha)|\alpha\in B\}$ are given as follows:

$$P_i(f_i) = \frac{W_i(f_i) \displaystyle\prod_{\{\alpha|\alpha\in\boldsymbol{c}_i\}} M_{\alpha\to i}(f_i)}{\displaystyle\sum_{z_i} W_i(z_i) \prod_{\{\alpha|\alpha\in\boldsymbol{c}_i\}} M_{\alpha\to i}(z_i)}, \tag{26}$$

$$P_\alpha(\boldsymbol{f}_\alpha) = \frac{W_\alpha(\boldsymbol{f}_\alpha) \displaystyle\prod_{\{i|i\in\alpha\}} \prod_{\{\gamma|\gamma\in\boldsymbol{c}_i\backslash\alpha\}} M_{\gamma\to i}(f_i)}{\displaystyle\sum_{\boldsymbol{z}_\alpha} W_\alpha(\boldsymbol{z}_\alpha) \prod_{\{i|i\in\alpha\}} \prod_{\{\gamma|\gamma\in\boldsymbol{c}_i\backslash\alpha\}} M_{\gamma\to i}(z_i)}, \tag{27}$$

$$M_{\alpha\to i}(f_i) = \frac{\displaystyle\sum_{\boldsymbol{z}_\alpha} \delta_{f_i,z_i} \Big(\frac{W_\alpha(\boldsymbol{z}_\alpha)}{W_i(z_i)}\Big) \prod_{\{j|j<\alpha\}} \prod_{\{\gamma|\gamma\in\boldsymbol{c}_j\backslash\alpha\}} M_{\gamma\to j}(z_j)}{\displaystyle\sum_{\boldsymbol{z}_\alpha} \Big(\frac{W_\alpha(\boldsymbol{z}_\alpha)}{W_i(z_i)}\Big) \prod_{\{j|j<\alpha\}} \prod_{\{\gamma|\gamma\in\boldsymbol{c}_j\backslash\alpha\}} M_{\gamma\to j}(z_j)}. \tag{28}$$

The marginal probability distribution $P_{ij}(f_i, f_j)$ ($ij{\le}\alpha$, $\alpha{\in}B$) is obtained by

$$P_{ij}(f_i, f_j) = \sum_{\boldsymbol{z}_\alpha} \delta_{z_i,f_i}\delta_{z_j,f_j} P_\alpha(\boldsymbol{z}_\alpha). \tag{29}$$

In the pair approximation and the cactus-square approximation of the cluster variation method, the sets $B$ are chosen as shown in Figs.1(a) and 1(b), respectively. By setting

$$W_i(f_i) = p\big(1-\delta_{f_i,g_i}\big) + \big(1-(Q-1)p\big)\delta_{f_i,g_i}, \tag{30}$$

$$W_\alpha(\boldsymbol{f}_\alpha) = \Big(\prod_{\{i|i\in\Omega\}} W_i(f_i)\Big)\exp\Big(-\frac{1}{2}\alpha \sum_{\{ij|ij\le\alpha,ij\in N\}} (f_i-f_j)^2\Big), \tag{31}$$

the marginal probabilities $\Pr\{F_i = f_i|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ are obtained as $P_i(f_i)$ and $P_{ij}(f_i, f_j)$, respectively. By setting

$$W_i(f_i) = 1, \tag{32}$$

$$W_\alpha(\boldsymbol{f}_\alpha) = \exp\Big(-\frac{1}{2}\alpha \sum_{\{ij|ij\le\alpha,ij\in N\}} (f_i-f_j)^2\Big), \tag{33}$$

we obtain the marginal probabilities $\Pr\{F_i = f_i, F_j = f_j|\alpha\}$ as $P_{ij}(f_i, f_j)$. Though these forms may be not so familiar for some physicists, $\ln(M_{i\to j}(\xi))$ is corresponding to the effective field in the conventional Bethe approximation. In the probabilistic inference, the quantity $M_{i\to j}(\xi)$ is called a message propagated from $i$ to $j$.

Eqs.(28) have forms of fixed point equations for the messages $M_{i\to j}(\xi)$. In practical numerical calculations, we solve the simultaneous equations (28) by using the iterative method in the conventional numerical analysis. Various values of hyperparameters $\alpha$ and $p$, we obtain the marginal probability distributions $\Pr\{F_i = f_i|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$, $\Pr\{F_i = f_i, F_j = f_j|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j|\alpha\}$ and search the optimal set of values, $(\widehat{\alpha}, \widehat{p})$, satisfying Eqs.(6) and (7) numerically.

## NUMERICAL EXPERIMENTS

In this section, we give some numerical experiments. The optimal set of values of hyperparameters, $(\widehat{\alpha}, \widehat{p})$, are determined by means of the maximum marginal likelihood estimation and the cluster variation method. We adopt a pair approximation and a square-cactus approximation in the cluster variation method. The sets $B$ of basic clusters in the pair approximation and in the square-cactus approximation are shown in Fig.1.

We then performed numerical experiments for artificial binary image generated from the 256-valued standard image “Mandrill” by using a thresholded

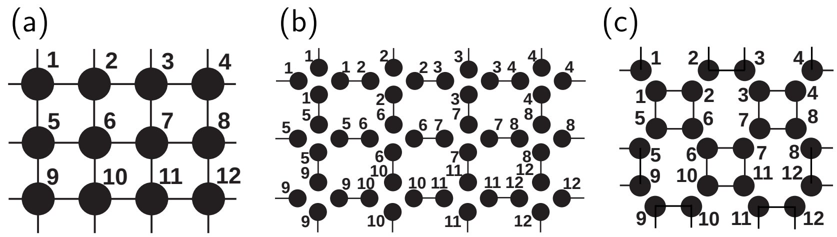

Figure 1: Square lattice and the sets of basic clusters of the pair approximation and the square-cactus approximation in the cluster variation method. (a) Square lattice. (b) Set $B$ of basic clusters in the pair approximation. (c) Set $B$ of basic cluster in the square-cactus approximation.

processing. The image restorations by means of the iterative algorithms of the mean-field approximation, the pair approximation and the square-cactus approximation are shown in Fig. 2. We give in table 1 the estimates of hyperparameters, $\widehat{p}$ and $\widehat{\alpha}$, and the values of the improvement of signal to noise ratio, $\Delta_{\rm SNR}$ (dB):

$$\Delta_{\rm SNR} \equiv 10 \log_{10}\Big( \frac{||\boldsymbol{f}-\boldsymbol{g}||^2}{||\boldsymbol{f}-\widehat{\boldsymbol{f}}||^2} \Big) \text{ (dB)}. \tag{34}$$

As shown in Fig. 1 and Table 1, though the pair approximation and the

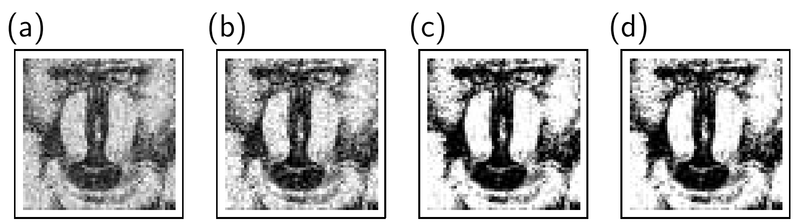

Figure 2: Image restorations for an artificial binary image $\boldsymbol{f}$ generated from the 256-valued standard images "Mandrill" by using a thresholded processing. (a) Degraded image $\boldsymbol{g}$. (b) Restored image $\widehat{\boldsymbol{f}}$ in the mean-field approximation. (c) Restored image $\widehat{\boldsymbol{f}}$ in the pair approximation. (d) Restored image $\widehat{\boldsymbol{f}}$ in the square-cactus approximation.

square-cactus approximation in the cluster variation method can improve the image quality for restoration in contrast to the mean-field approximation, the result of the square-cactus approximation is substantially equal to the one of the pair approximation. Moreover, we adopt as original images twenty binary images which are generated by Monte Carlo simulations in the *a priori* probability distribution (2) for $\alpha = 2.15^{-1}$ and obtain the similar results as the ones in Fig.2 and Table 1.

Table 1: The estimates of hyperparameters, $\widehat{p}$ and $\widehat{\alpha}$, and the values of $\Delta_{\mathrm{SNR}}$ obtained for some degraded images $\boldsymbol{g}$ given in figures 2(a). The hyperparameters are estimated by applying the mean-field approximation the pair approximation and the square-cactus approximation to the maximum marginal likelihood estimation, respectively.

| Apprix. | $\widehat{p}$ | $\widehat{\alpha}$ | $\Delta_{\mathrm{SNR}}$ (dB) |
|---|---|---|---|
| MFA | 0.100666 | 0.250971 | 0.979681 |
| PA | 0.179189 | 0.396343 | 2.304090 |
| SCA | 0.179701 | 0.393583 | 2.297306 |

## CONCLUDING REMARKS

In the present paper, we summarize the framework of the cluster variation method for the probabilistic image processing based on the Bayesian analysis and the maximization of the marginal likelihood. Particularly, we gave a generalized belief propagation algorithm by restricting the set $C\backslash B$ to the one consisting only of single pixels, so that $C\backslash B = \Omega$. Of course, we expect that the obtained results are improved as we adopt larger basic clusters. However, the computational complexity also grows as we adopt larger basic clusters. Most important point is how large basic clusters we should adopt to obtain sufficiently good results. The results in the present paper suggest that the pair approximation can give us good results for the probabilistic model with the interactions between the nearest-neighbor pairs of pixels.

We show the results of the 4-valued image restoration by means of the pair approximation in Fig.3. This result is obtained by assuming Eq.(2) as *a priori* probability distribution. The probabilistic model given in Eq.(2) is referred to as $Q$-state Ising model in the statistical mechanics. We have a $Q$-state Potts model as the other familiar probabilistic model in the statistical mechanics[10]. The $Q$-state Potts model reflect the spatially flatness in images, while the $Q$-state Ising model reflect the spatially smoothness. Now we investigate the hyperparameter estimations in probabilistic image restorations by adopting the $Q$-state Potts model and $Q$-state Ising model as *a priori* probability distribution. One of the results obtained by using the $Q$-state Potts model as *a priori* probability distribution is shown in Fig.3(c). We will report the detailed investigation elsewhere[13].

### Acknowledgements

The author is grateful to Professor T. Horiguchi of the Graduate School of Information Science, Tohoku University, for valuable discussions. This work was partly supported by the Grants-In-Aid (No.13680384 and No.14084203) for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

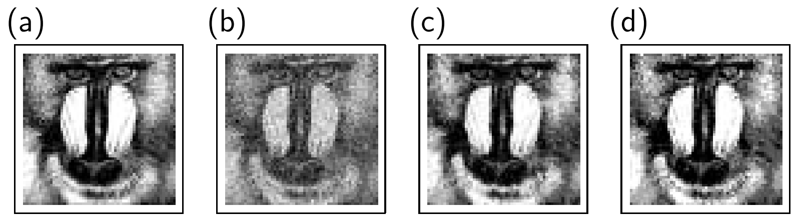


Figure 3: Image restorations for an artificial 4-valued image $\boldsymbol{f}$ ($Q = 4$). (a) Original image $\boldsymbol{f}$. (b) Degraded image $\boldsymbol{g}$. (c) Restored image $\widehat{\boldsymbol{f}}$ in the pair approximation for the *a priori* prbability distribution (2). (d) Restored image $\widehat{\boldsymbol{f}}$ in the pair approximation for the *a priori* prbability distribution given as a $Q$-state Potts model.

## REFERENCES


[1] S. Z. Li, *Markov Random Field Modeling in Computer Vision* (Springer-Verlag, 1995).

[2] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, 2001)

[3] K. Tanaka, "Statistical-mechanical Approach to Image Processing (Topical Review)," Journal of Physics A: Mathematical and General, vol.35, no.37, pp.R81-R150, 2002.

[4] M. Opper and D. Saad (edited), *Advanced Mean Field Methods - Theory and Practice* (MIT Press, 2001).

[5] R. Kikuchi, "A theory of cooperative phenomena," Phys. Rev. vol.81, no.6, pp.988-1003, March 1951.

[6] T. Morita, "Cluster variation method for non-uniform Ising and Heisenberg models and spin-pair correlation function," Prog. Theor. Phys., vol.85, no.2, pp.243-255, Feb. 1991.

[7] K. Tanaka and T. Morita, "Cluster Variation Method and Image Restoration Problem," Physics Letters A, vol.203, no.2-3, pp.122-128, 1995.

[8] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference* (Morgan Kaufmann, 1988).

[9] Y. Weiss, "Correctness of Belief Propagation in Gaussian Graphical Models of Arbitrary Topology," Neural Computation, vol.13, pp.2173-2200, 2001.

[10] K. Tanaka, "Theoretical Study of Hyperparameter Estimation by Maximization of Marginal Likelihood in Image Restoration by means of Cluster Variation Method," Electronics and Communications in Japan, Part 3, vol.85, no.7, pp.50-62, 2002.

[11] J. S. Yedidia, W. T. Freeman and Y. Weiss, "Generalized Belief Propagation," Advances in Neural Information Processing Systems, vol.13, T. Leen, T. Dietterich, V. Tresp (eds), pp.689-695 (Cambridge, MA: MIT Press, 2001).

[12] H. J. Kappen and W. Wiegerinck, "Novel iteration schemes for the cluster variation method," Advances in Neural Information Processing System, vol.14, T. G. Dietterich, S. Becker, and Z. Ghahramani (eds), pp.415-422 (Cambridge, MA: MIT Press, 2002).

[13] K. Tanaka, J. Inoue and D. M. Titterington, "Probabilistic Image Processing by means of Bethe Approximation for Multi-state Classical Spin Systems," to be submitted to J. Phys. A, 2003.