Probabilistic image processing by means of the Bethe approximation for the Q-Ising model

2003 J. Phys. A: Math. Gen. 36 11023

(http://iopscience.iop.org/0305-4470/36/43/025)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) 11023–11035 PII: S0305-4470(03)61624-0

Probabilistic image processing by means of the Bethe approximation for the Q-Ising model

Kazuyuki Tanaka[1], **Jun-ichi Inoue**[2] **and D M Titterington**[3]

[1] Graduate School of Information Sciences, Tohoku University, Aramaki-aza-aoba 09, Aoba-ku, Sendai 980-8579, Japan
[2] Complex Systems Engineering, Graduate School of Engineering, Hokkaido University, N13-W8, Kita-ku, Sapporo 060-8628, Japan
[3] Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK

E-mail: kazu@statp.is.tohoku.ac.jp

Received 31 March 2003, in final form 1 August 2003
Published 15 October 2003
Online at stacks.iop.org/JPhysA/36/11023

Abstract

The framework of Bayesian image restoration for multi-valued images by means of the Q-Ising model with nearest-neighbour interactions is presented. Hyperparameters in the probabilistic model are determined so as to maximize the marginal likelihood. A practical algorithm is described for multi-valued image restoration based on the Bethe approximation. The algorithm corresponds to loopy belief propagation in artificial intelligence. We conclude that, in real world grey-level images, the Q-Ising model can give us good results.

PACS numbers: 02.50−r, 02.50.Cw, 02.50.Tt, 05.20.−y, 05.50.+q, 75.10.Nr, 89.70.+c

1. Introduction

Statistical-mechanical methods are applicable to probabilistic information processing [1]. Conversely, many problems in probabilistic information processing have created research ideas in statistical mechanics. Some of these ideas have been revived as new paradigms for powerful information technologies in the interdisciplinary studies of statistical mechanics and computer science since many notions that are well known in the context of spin glass problems are still not so familiar in the context of computer science. A number of new subject areas that have not been considered by physicists exploit the theory of spin glass problems or other problems in statistical mechanics. Probabilistic image restoration is one of these examples. On the basis of Bayes' formula and with an *a priori* probability distribution assumed for the original image, one creates a probabilistic image restoration in the form of an *a posteriori* probability distribution for the original image when the degraded image is given.

The *a priori* distribution is often assumed to correspond to a classical spin system, represented by an Ising model, a Gaussian model and so on [2]. Moreover, practical algorithms are constructed by using simulated annealing [3], the mean-field approximation [4, 5], the cluster variation method [6] and so on.

One statistical-mechanical method that has been widely applied to many problems in computer science is that of advanced mean-field theory [7]. The mean-field approximation and the Bethe approximation can be formulated by means of the variational principle for the minimization of approximate free energies. Recently, some computer scientists have become interested in such variational approaches for advanced mean-field theory. It has been suggested that the extremum conditions with respect to the effective fields for the approximate free energy in the Bethe approximation are equivalent to simultaneous fixed-point equations associated with belief propagation in the context of probabilistic inference in artificial intelligence [8].

In Bayesian image restoration, we have to determine some model parameters, such as an interaction parameter and so on. These model parameters are referred to as hyperparameters in statistics. In practice it is common for the hyperparameters to be determined so as to maximize a marginal likelihood [2, 4, 5], which is expressed in terms of free energies of the *a priori* probabilistic model and the *a posteriori* probabilistic model. Tanaka has applied the Bethe approximation to the estimation of hyperparameters by maximizing the marginal likelihood [2, 9]. In particular, in [2], the Potts model was adopted as an *a priori* probability distribution, and the results of the Bethe approximation were compared with those using mean-field approximations for hyperparameter estimation in the context of multi-level image restoration. Tanaka [2] concluded that the Bethe approximation can improve the quality of image restoration over that achieved by mean-field approximation. Patterns generated by Monte Carlo simulations from the Potts model have many spatially flat parts, but real world images are characterized more by spatially smooth parts than by spatially flat parts. The Q-Ising model is a classical spin system which can generate images with many spatially smooth parts [10]. Here, Q is the number of possible states in the model. For the case of $Q = 3$, the Q-Ising model is a special case of the Blum–Emery–Griffiths model, which has complicated critical phenomena [11].

In the present paper, we investigate Bayesian image restoration with hyperparameters estimated by maximizing the marginal likelihood when we adopt a Q-Ising model as the *a priori* probability distribution. The maximization of the marginal likelihood is achieved by using the Bethe approximation. In section 2, we explain how to construct the *a posteriori* probability distributions when the *a priori* probability distribution is assumed to be a Q-Ising model with only nearest-neighbour interactions. In section 3, we give the basic framework of hyperparameter estimation based on maximization of the marginal likelihood. In section 4, we summarize the simultaneous equations for marginal probability distributions in any classical spin system with spatially non-uniform external fields and nearest-neighbour interactions in the Bethe approximation and explain how to calculate the optimal values of the hyperparameters numerically. In section 5, we give some numerical experiments. Conclusions are given in section 6.

2. Bayesian image analysis based on the Q-Ising model

We consider an image on a square lattice $\Omega \equiv \{i\}$ such that each pixel takes one of the grey-levels $\boldsymbol{Q} = \{0, 1, 2, \ldots, Q-1\}$, with 0 and $Q-1$ corresponding to black and white, respectively. The intensities at pixel i in the original image and the degraded image are regarded as random variables 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 actual original image and the degraded image are denoted by $\boldsymbol{f} = \{f_i\}$ and $\boldsymbol{g} = \{g_i\}$, respectively.

The probability that the original image is $\boldsymbol{f}$, $\Pr\{\boldsymbol{F} = \boldsymbol{f}\}$, is called the *a priori* probability of the image. In the Bayes formula, the *a posteriori* probability $\Pr\{\boldsymbol{F} = \boldsymbol{f} | \boldsymbol{G} = \boldsymbol{g}\}$, that the original image is $\boldsymbol{f}$ when the given degraded image is $\boldsymbol{g}$, is expressed as

$$\Pr\{\boldsymbol{F} = \boldsymbol{f} | \boldsymbol{G} = \boldsymbol{g}\} = \frac{\Pr\{\boldsymbol{G} = \boldsymbol{g} \big| \boldsymbol{F} = \boldsymbol{f}\} \Pr\{\boldsymbol{F} = \boldsymbol{f}\}}{\sum_z \Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{z}\} \Pr\{\boldsymbol{F} = \boldsymbol{z}\}} \tag{1}$$

where the summation $\sum_z$ is taken over all possible configurations of images $\boldsymbol{z} = \{z_i | i \in \Omega\}$. The probability $\Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{f}\}$ is the conditional probability that the degraded image is $\boldsymbol{g}$ when the original image is $\boldsymbol{f}$ and describes the degradation process.

In the present paper, it 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 with the same probability p, independently of the other pixels; the probability that the intensity is unchanged is therefore $1-(Q-1)p$, which constrains p to lie in the range $0 \leqslant p \leqslant 1/Q$. The conditional probability distribution associated with the degradation process when the original image is $\boldsymbol{f}$ is

$$\begin{aligned} \Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{f}\} &= \Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{f}, p\} \\ &= \prod_{i \in \Omega} \big(p\big(1 - \delta_{f_i, g_i}\big) + \big(1 - (Q-1)p\big)\delta_{f_i, g_i}\big) \end{aligned} \tag{2}$$

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

$$\Pr\{\boldsymbol{F} = \boldsymbol{f}\} = \Pr\{\boldsymbol{F} = \boldsymbol{f} | \alpha\} = \frac{\prod_{ij \in B} \exp\left(-\frac{1}{2}\alpha(f_i - f_j)^2\right)}{\sum_z \prod_{ij \in B} \exp\left(-\frac{1}{2}\alpha(z_i - z_j)^2\right)} \tag{3}$$

where B is the set of all the nearest-neighbour pairs of pixels on the square lattice Ω. By substituting equations (2) and (3) into equation (1), we obtain

$$\Pr\{\boldsymbol{F} = \boldsymbol{f} | \boldsymbol{G} = \boldsymbol{g}, \alpha, p\} = \frac{1}{Z(\alpha, p)} \left(\prod_{i \in \Omega} \exp\big(\beta(p)\delta_{f_i, g_i}\big)\right) \left(\prod_{ij \in B} \exp\left(-\frac{1}{2}\alpha(f_i - f_j)^2\right)\right) \tag{4}$$

where $Z(\alpha, p)$ is a normalization constant and $\beta(p)$ is defined by

$$\beta(p) \equiv \ln\left(\frac{1 - Qp + p}{p}\right). \tag{5}$$

3. Hyperparameter estimation

In the maximum marginal likelihood estimation approach, the hyperparameters α and p are determined so as to maximize the marginal likelihood $\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}$, where

$$\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\} \equiv \sum_z \Pr\{\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{F} = \boldsymbol{z}, p\} \Pr\{\boldsymbol{F} = \boldsymbol{z} | \alpha\}. \tag{6}$$

We denote the maximizers of the marginal likelihood $\Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}$ by $\hat{\alpha}$ and $\hat{p}$. Thus

$$(\hat{\alpha}, \hat{p}) = \arg\max_{(\alpha, p)} \Pr\{\boldsymbol{G} = \boldsymbol{g} | \alpha, p\}. \tag{7}$$

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

$$\sum_{i\in\Omega}\sum_{z_i\in\boldsymbol{Q}}\bigl(1-\delta_{z_i,g_i}\bigr)\Pr\{F_i=z_i|\boldsymbol{G}=\boldsymbol{g},\hat{\alpha},\hat{p}\}=(Q-1)p \tag{8}$$

$$\sum_{ij\in B}\sum_{z_i\in\boldsymbol{Q}}\sum_{z_j\in\boldsymbol{Q}}(z_i-z_j)^2\Pr\{F_i=z_i,F_j=z_j|\boldsymbol{G}=\boldsymbol{g},\hat{\alpha},\hat{p}\}$$
$$=\sum_{ij\in B}\sum_{z_i\in\boldsymbol{Q}}\sum_{z_j\in\boldsymbol{Q}}(z_i-z_j)^2\Pr\{F_i=z_i,F_j=z_j|\hat{\alpha}\}. \tag{9}$$

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

$$\Pr\{F_i=f_i|\boldsymbol{G}=\boldsymbol{g},\alpha,p\}\equiv\sum_{z}\delta_{f_i,z_i}\Pr\{\boldsymbol{F}=\boldsymbol{z}|\boldsymbol{G}=\boldsymbol{g},\alpha,p\} \tag{10}$$

$$\Pr\{F_i=f_i,F_j=f_j|\boldsymbol{G}=\boldsymbol{g},\alpha,p\}\equiv\sum_{z}\delta_{f_i,z_i}\delta_{f_j,z_j}\Pr\{\boldsymbol{F}=\boldsymbol{z}|\boldsymbol{G}=\boldsymbol{g},\alpha,p\} \tag{11}$$

$$\Pr\{F_i=f_i,F_j=f_j|\alpha\}\equiv\sum_{z}\delta_{f_i,z_i}\delta_{f_j,z_j}\Pr\{\boldsymbol{F}=\boldsymbol{z}|\alpha\}. \tag{12}$$

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

$$\hat{f}_i=\arg\max_{f_i\in\boldsymbol{Q}}\Pr\{F_i=f_i|\boldsymbol{G}=\boldsymbol{g},\alpha,p\}. \tag{13}$$

This way of producing a restored image is called maximum posterior marginal estimation [12, 13].

4. Bethe approximation for classical spin systems

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 $\Pr\{\boldsymbol{F}=\boldsymbol{f}|\alpha\}$.

Next we summarize the simultaneous equations to be satisfied by marginal probability distributions, namely

$$\rho_i(f_i)=\sum_{z}\delta_{f_i,z_i}\rho(\boldsymbol{z}) \tag{14}$$

$$\rho_{ij}(f_i,f_j)=\sum_{z}\delta_{f_i,z_i}\delta_{f_j,z_j}\rho(\boldsymbol{z}) \tag{15}$$

in the Bethe approximation for a probability distribution $\rho(\boldsymbol{f})$ defined by

$$\rho(\boldsymbol{f})=\frac{\prod_{i\in\Omega}\psi_i(f_i)\prod_{ij\in B}\phi_{ij}(f_i,f_j)}{\sum_{z}\prod_{i\in\Omega}\psi_i(z_i)\prod_{ij\in B}\phi_{ij}(z_i,z_j)}. \tag{16}$$

Here $\phi_{ij}(\xi,\xi')$ and $\psi_i(\xi)$ are always positive for any values of $\xi\in\boldsymbol{Q}$ and $\xi'\in\boldsymbol{Q}$. The detailed derivation is similar to that given in [2, 9] and here we merely state the results of the variational calculation in the Bethe approximation. In the Bethe approximation, the

simultaneous equations for the sets of marginal probabilities $\{\rho_i(\xi)|i \in \Omega, \xi \in \boldsymbol{Q}\}$ and $\{\rho_{ij}(\xi, \xi')|ij \in B, \xi \in \boldsymbol{Q}, \xi' \in \boldsymbol{Q}\}$ are given by

$$\rho_i(\xi) = \frac{\psi_i(\xi)\prod_{k\in \boldsymbol{c}_i}\mu_{k\to i}(\xi)}{\sum_{\zeta\in\boldsymbol{Q}}\psi_i(\zeta)\prod_{k\in \boldsymbol{c}_i}\mu_{k\to i}(\zeta)} \tag{17}$$

$$\rho_{ij}(\xi, \xi') = \frac{\psi_i(\xi)\phi_{ij}(\xi, \xi')\psi_j(\xi')\prod_{k\in \boldsymbol{c}_i\setminus j}\mu_{k\to i}(\xi)\prod_{l\in \boldsymbol{c}_j\setminus i}\mu_{l\to j}(\xi')}{\sum_{\zeta\in\boldsymbol{Q}}\sum_{\zeta'\in\boldsymbol{Q}}\psi_i(\zeta)\phi_{ij}(\zeta, \zeta')\psi_j(\zeta')\prod_{k\in \boldsymbol{c}_i\setminus j}\mu_{k\to i}(\zeta)\prod_{l\in \boldsymbol{c}_j\setminus i}\mu_{l\to j}(\zeta')} \tag{18}$$

$$\mu_{j\to i}(\xi) = \frac{\sum_{\zeta\in\boldsymbol{Q}}\phi_{ij}(\xi, \zeta)\psi_j(\zeta)\prod_{k\in \boldsymbol{c}_j\setminus i}\mu_{k\to j}(\zeta)}{\sum_{\zeta'\in\boldsymbol{Q}}\sum_{\zeta\in\boldsymbol{Q}}\phi_{ij}(\zeta', \zeta)\psi_j(\zeta)\prod_{k\in \boldsymbol{c}_j\setminus i}\mu_{k\to j}(\zeta)}. \tag{19}$$

Here $\boldsymbol{c}_i \equiv \{j|ij \in B\}$ is the set of all the nearest-neighbour pixels of i. By setting $\psi_i(\xi) = \exp(\beta(p)\delta_{\xi,g_i})$ and $\phi_{ij}(\xi, \xi') = \exp\bigl(-\frac{1}{2}\alpha(\xi - \xi')^2\bigr)$, we obtain the marginal probabilities $\Pr\{F_i = \xi|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ and $\Pr\{F_i = \xi, F_j = \xi'|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ as $\rho_i(\xi)$ and $\rho_{ij}(\xi, \xi')$, respectively. By setting $\psi_i(\xi) = 1$ and $\phi_{ij}(\xi, \xi') = \exp\bigl(-\frac{1}{2}\alpha(\xi - \xi')^2\bigr)$, we obtain the marginal probabilities $\Pr\{F_i = \xi, F_j = \xi'|\alpha\}$ as $\rho_{ij}(\xi, \xi')$. Though these forms may not be so familiar to some physicists, $\ln(\mu_{i\to j}(\xi))$ corresponds to the effective field in the conventional Bethe approximation. In probabilistic inference, the quantity $\mu_{i\to j}(\xi)$ is called a message propagated from i to j.

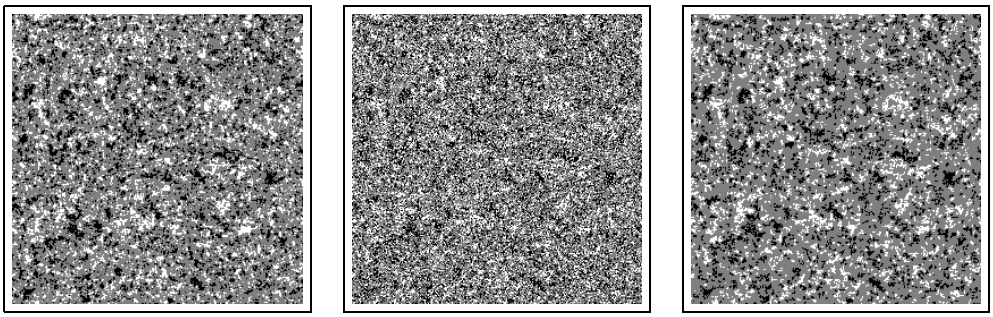
Equations (19) have the form of fixed-point equations for the messages $\mu_{i\to j}(\xi)$. In practical numerical calculations, we solve the simultaneous equations (19) by using iterative methods. For various values of the hyperparameters α and p, we obtain the marginal probability distributions $\Pr\{F_i = f_i|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$, $\Pr\{F_i = f_i, F_j = f_j|\boldsymbol{G} = \boldsymbol{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j|\alpha\}$ and search for the optimal set of values, $(\hat{\alpha}, \hat{p})$, that satisfy equations (9) and (8) numerically.

5. Numerical experiments

In this section, we report some numerical experiments. The optimal set of values of the hyperparameters, $(\hat{\alpha}, \hat{p})$, is determined by means of maximum marginal likelihood estimation and the Bethe approximation given in sections 3 and 4.

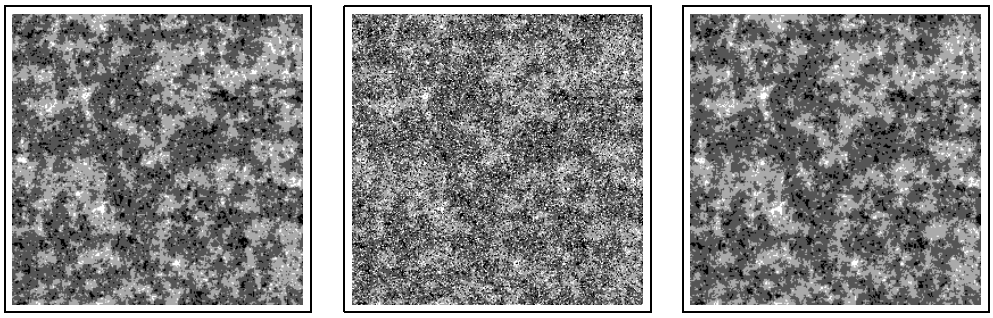
To evaluate restoration performance quantitatively, 20 original images $\boldsymbol{f}$ are simulated from the *a priori* probability distribution (3) for the Q-Ising model. We produce a degraded image $\boldsymbol{g}$ from each original image $\boldsymbol{f}$ by means of the degradation process (2) for $(Q - 1)p = 0.3$. By applying the iterative algorithm of the Bethe approximation to each degraded image $\boldsymbol{g}$, we obtain estimates of the hyperparameters $\hat{p}$ and $\hat{\alpha}$ and the restored image $\hat{\boldsymbol{f}}$ for each degraded image $\boldsymbol{g}$. For each of the cases $Q = 3$ and $Q = 4$, one of the numerical experiments is shown in figures 1 and 2, respectively. We give, in figures 3 and 4, $\beta(p)$- and α-dependences of the left-hand side and the right-hand side of equations (8) and (9) corresponding to the image restorations shown in figures 1 and 2, respectively. Moreover, in order to show the achievement of the maximization of marginal likelihood, we give, in figures 5 and 6, also $\beta(p)$- and α-dependences of $\ln \Pr\{\boldsymbol{G} = \boldsymbol{g}|\alpha, p\}$ corresponding to the image restorations shown in figures 1 and 2, respectively. From the 20 degraded images $\boldsymbol{g}$ and the corresponding restored images $\hat{\boldsymbol{f}}$, we calculate the 95% confidence intervals for the hyperparameters, $\hat{p}$ and $\hat{\alpha}$, and the values of the Hamming distance $d(\boldsymbol{f}, \hat{\boldsymbol{f}})$,

$$d(\boldsymbol{f}, \hat{\boldsymbol{f}}) \equiv \frac{1}{|\Omega|}\sum_{i\in\Omega}\bigl(1 - \delta_{f_i,\hat{f}_i}\bigr) \tag{20}$$



(*a*) (*b*) (*c*)

Figure 1. Image restoration based on the Q-Ising model ($Q = 3$). The original image $\boldsymbol{f}$ is generated from the *a priori* probability distribution (3). (*a*) Original image $\boldsymbol{f}$ ($\alpha = 0.65$). (*b*) Degraded image $\boldsymbol{g}$ ($(Q-1)p = 0.3$). (*c*) Restored image $\hat{\boldsymbol{f}}$ obtained by the Bethe approximation.



(*a*) (*b*) (*c*)

Figure 2. Image restoration based on the Q-Ising model ($Q = 4$). The original image $\boldsymbol{f}$ is generated by the *a priori* probability distribution (3). (*a*) Original image $\boldsymbol{f}$ ($\alpha = 0.75$). (*b*) Degraded image $\boldsymbol{g}$ ($(Q-1)p = 0.3$). (*c*) Restored image $\hat{\boldsymbol{f}}$ obtained by the Bethe approximation.

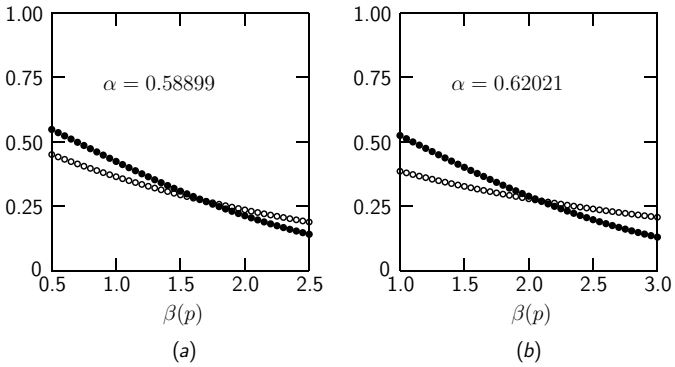


Figure 3. $\beta(p)$-dependence of the left-hand side and the right-hand side of equation (8) corresponding to the image restorations shown in figures 1 and 2. The left-hand side and the right-hand side are shown as open circles and solid circles, respectively. (*a*) $Q = 3$. (*b*) $Q = 4$.

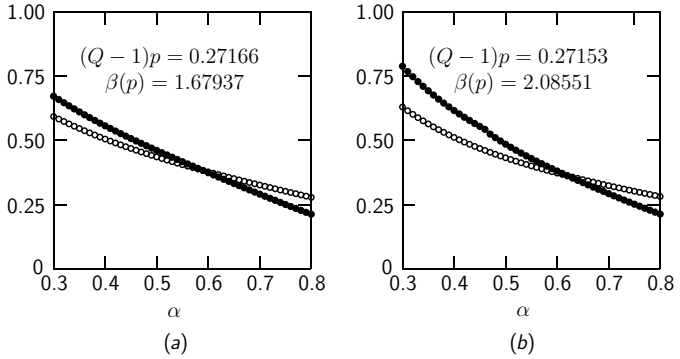


Figure 4. α-dependence of the left-hand side and the right-hand side of equation (9) corresponding to the image restorations shown in figures 1 and 2. The left-hand side and the right-hand side are shown as open circles and solid circles, respectively. (*a*) $Q = 3$. (*b*) $Q = 4$.

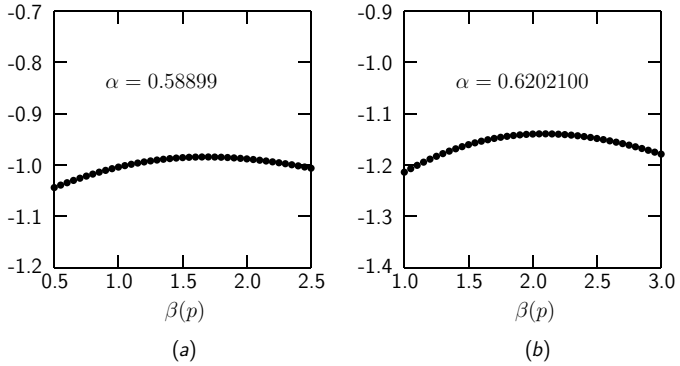


Figure 5. $\beta(p)$-dependence of the logarithm of marginal likelihood per pixel, $\frac{1}{|\Omega|}\ln \Pr\{\boldsymbol{G} = \boldsymbol{g}|\alpha, p\}$, corresponding to the image restorations shown in figures 1 and 2. (*a*) $Q = 3$. (*b*) $Q = 4$.

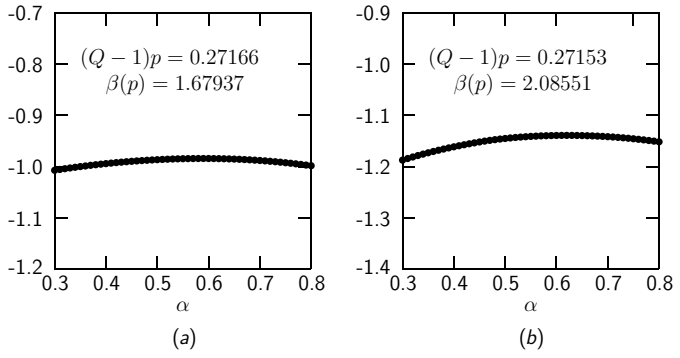


Figure 6. α-dependence of the logarithm of marginal likelihood per pixel, $\frac{1}{|\Omega|}\ln \Pr\{\boldsymbol{G} = \boldsymbol{g}|\alpha, p\}$, corresponding to the image restorations shown in figures 1 and 2. (*a*) $Q = 3$. (*b*) $Q = 4$.

and the improvement in the signal to noise ratio, Δ_{SNR},

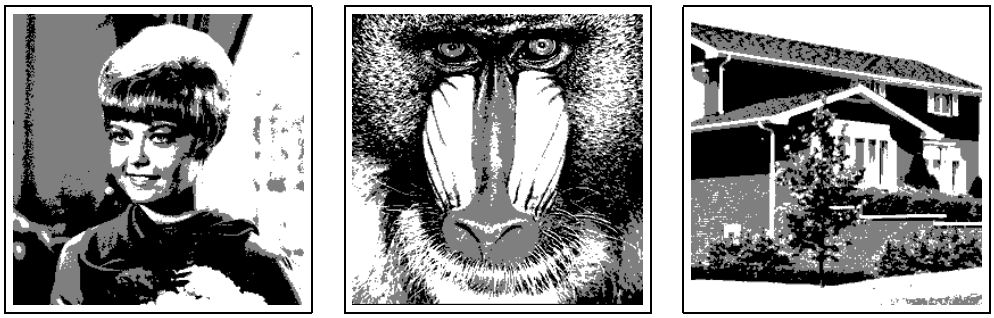
$$\Delta_{\mathrm{SNR}} \equiv 10 \log_{10} \left(\frac{\sum_{i\in\Omega}(g_i - f_i)^2}{\sum_{i\in\Omega}(\hat{f}_i - f_i)^2} \right) \ (\mathrm{dB}) \tag{21}$$

Table 1. Approximate 95% confidence intervals for the hyperparameters, p and α, and the values of $d(\boldsymbol{f}, \hat{\boldsymbol{f}})$ and Δ_{SNR} obtained for some degraded images $\boldsymbol{g}$, which are produced for $(Q-1)p = 0.3$ from 20 original images $\boldsymbol{f}$. The 20 original images $\boldsymbol{f}$ are generated by Monte Carlo simulations from the *a priori* probability distributions (3) for the Q-Ising models. The hyperparameters are estimated by applying the Bethe approximation to maximum marginal likelihood estimation.

	$Q = 3$	$Q = 4$
α	0.65	0.75
$(Q-1)p$	0.30	0.30
$\hat{\alpha}$	[0.586 32 ± 0.001 367]	[0.620 58 ± 0.001 314]
$(Q-1)\hat{p}$	[0.272 01 ± 0.001 011]	[0.269 61 ± 0.000 748]
$d(\boldsymbol{f}, \boldsymbol{g})$	[0.298 02 ± 0.000 421]	[0.298 06 ± 0.000 422]
$d(\boldsymbol{f}, \hat{\boldsymbol{f}})$	[0.222 41 ± 0.000 854]	[0.169 57 ± 0.000 565]
Δ_{SNR} (dB)	[0.265 96 ± 0.052 205]	[4.618 42 ± 0.032 449]

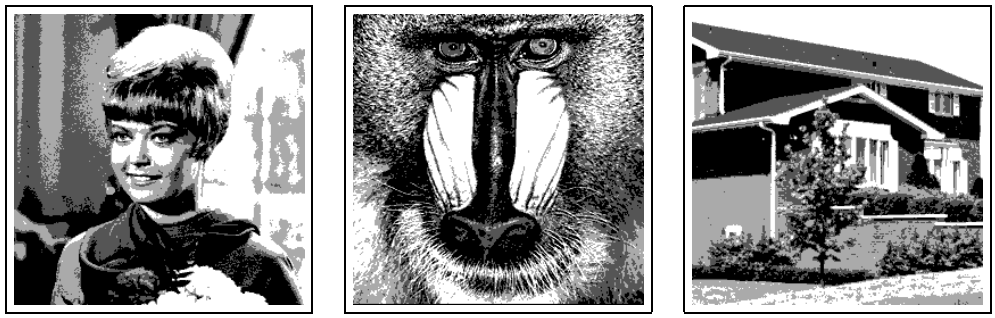
in decibels (dB). The definition of the signal to noise ratio $R_{\mathrm{SNR}}(\boldsymbol{f}, \boldsymbol{g})$ of the original image $\boldsymbol{f}$ and the degraded image $\boldsymbol{g}$ is given by $R_{\mathrm{SNR}}(\boldsymbol{f}, \boldsymbol{g}) \equiv 10 \log_{10}\left(\frac{\text{Variance of signal } \boldsymbol{f}}{\text{Variance of noise in } \boldsymbol{g}}\right)$ (dB). The original image $\boldsymbol{f}$ and the difference $\boldsymbol{g} - \boldsymbol{f} = \{g_i - f_i | i \in \Omega\}$ can be regarded as a signal and noise, respectively. The variance of $\boldsymbol{g} - \boldsymbol{f}$ is equal to $\frac{1}{|\Omega|}\sum_{i \in \Omega}(g_i - f_i)^2$. The signal to noise ratio $R_{\mathrm{SNR}}(\boldsymbol{f}, \boldsymbol{g})$ becomes small when the noise included in the degraded image $\boldsymbol{g}$ is large. We can also consider the signal to noise ratio $R_{\mathrm{SNR}}(\boldsymbol{f}, \hat{\boldsymbol{f}})$ in the same way. The improvement in the signal to noise ratio is defined by $\Delta_{\mathrm{SNR}} = R_{\mathrm{SNR}}(\boldsymbol{f}, \hat{\boldsymbol{f}}) - R_{\mathrm{SNR}}(\boldsymbol{f}, \boldsymbol{g})$, which corresponds to the extent to which noise is reduced. For the cases of $Q = 3$ and $Q = 4$, these confidence intervals are given in table 1. Clearly the true values of the hyperparameters, α and p, lie outside the 95% confidence intervals. The phrase '95% confidence' generally means that the true parameter is included in the estimated interval in 95 out of 100 experiments when the underlying approximation is valid. The present results might appear to throw doubt on the accuracy in the Bethe approximation so far as estimating the values of p and α is concerned. The fact that the confidence intervals do not include the true values indicate a bias in estimating the hyperparameters using the Bethe approximation. However, the point estimates of the hyperparameters are quite close to the true values in absolute terms. Further investigation of the impact of these biases will be the subject of future research.

These results show that combining the Bethe approximation with maximum marginal likelihood estimation gives us good results for hyperparameter estimation for the Q-Ising model. In the image $\boldsymbol{f}$ given in figures 1(a), many pixels i have $f_i = 1$ and there are some small-sized clusters of which the neighbouring pixels have the same values as $f_i = 0$ or $f_i = 2$. In the image $\boldsymbol{f}$ given in figure 2(a), many pixels i have $f_i = 1$ or $f_i = 2$ and there are some large-sized clusters of which the neighbouring pixels have the same values as $f_i = 1$ or $f_i = 2$. The images given in figures 1(a) and 2(a) seem to be different from each other. In the probabilistic model given in equation (3), phase transition does not occur for the case of $Q = 3$, though a phase transition does occur for the case of $Q = 4$. From table 1, it is clear that the results for $Q = 3$ are not so good, though those for $Q = 4$ are satisfactory. For the case of $Q = 3$, the images $\boldsymbol{f}$ generated by Monte Carlo simulation from the *a priori* probability distribution (3) have many large regions in which the state of every pixel is $f_i = 1$ and have only small regions in which the state of every pixel is $f_i = 0$ or $f_i = 2$. Many large regions consist only of state 1. It is difficult to recover such small regions by using our probabilistic model. We expect that, if Q is odd, the generated images $\boldsymbol{f}$ have similar properties to the case of $Q = 3$. On the other hand, the generated images $\boldsymbol{f}$ have many large regions in which the state of every pixel is $f_i = 1$ or $f_i = 2$ and these patterns in the images $\boldsymbol{f}$ consist of two states



(*a*) (*b*) (*c*)

Figure 7. Artificial 3-valued images $\boldsymbol{f}$ generated from the 256-valued standard images 'girl', 'mandrill' and 'home', obtained by thresholding ($Q = 3$). (*a*) Girl. (*b*) Mandrill. (*c*) Home.



(*a*) (*b*) (*c*)

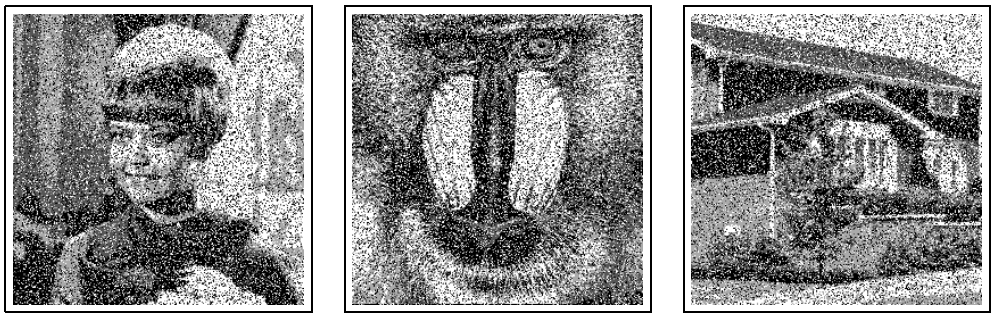
Figure 8. Artificial 4-valued images $\boldsymbol{f}$ generated from the 256-valued standard images 'girl', 'mandrill' and 'home', obtained by thresholding ($Q = 4$). (*a*) Girl. (*b*) Mandrill. (*c*) Home.

(*a*) (*b*) (*c*)

Figure 9. Degraded images $\boldsymbol{g}$ which are produced for $(Q-1)p = 0.3$ from the original images $\boldsymbol{f}$ given in figure 7 ($Q = 3$). (*a*) Girl. (*b*) Mandrill. (*c*) Home.

1 and 2 for the case of $Q = 4$. We expect that, if Q is even, the generated images $\boldsymbol{f}$ will have similar properties to the case of $Q = 4$.

We then performed numerical experiments based on the artificial images in figures 7 and 8. The artificial images $\boldsymbol{f}$ are generated from the 256-valued standard images 'girl' and 'mandrill' by thresholding. Degraded images $\boldsymbol{g}$, produced from the original images $\boldsymbol{f}$ with

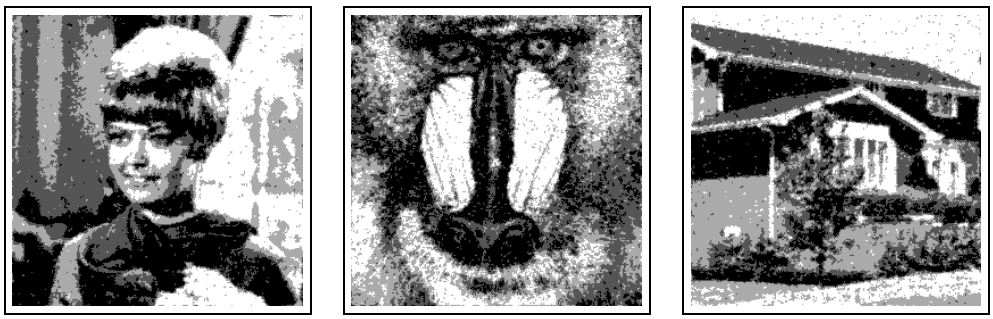


(*a*) (*b*) (*c*)

Figure 10. Degraded images $\boldsymbol{g}$ which are produced for $(Q-1)p = 0.3$ from the original images $\boldsymbol{f}$ given in figure 8 ($Q = 4$). (*a*) Girl. (*b*) Mandrill. (*c*) Home.

(*a*) (*b*) (*c*)

Figure 11. Restoration of the degraded images $\boldsymbol{g}$ given in figure 9 based on the Q-Ising model ($Q = 3$). (*a*) Girl. (*b*) Mandrill. (*c*) Home.



(*a*) (*b*) (*c*)

Figure 12. Restoration of the degraded images $\boldsymbol{g}$ given in figure 10 based on the Q-Ising model ($Q = 4$). (*a*) Girl. (*b*) Mandrill. (*c*) Home.

$(Q-1)p = 0.3$, are shown in figures 9 and 10. The image restorations created by means of the iterative algorithm using the Bethe approximation for the Q-Ising model are shown in figures 11 and 12. We give in table 2 the estimates, $\hat{p}$ and $\hat{\alpha}$, of the hyperparameters, and the values of the Hamming distance $d(\boldsymbol{f}, \hat{\boldsymbol{f}})$ as well as the improvement in the signal to noise ratio, Δ_{SNR} (dB). The image 'girl' has many spatially smooth parts, whereas the image

Table 2. The estimates $\hat{p}$ and $\hat{\alpha}$ of the hyperparameters and the values of $d(\boldsymbol{f}, \hat{\boldsymbol{f}})$ and Δ_{SNR} obtained for some degraded images $\boldsymbol{g}$, which are produced for $(Q-1)p = 0.3$ from the original images $\boldsymbol{f}$ given in figure 7 ($Q = 3$) and figure 8 ($Q = 4$). The hyperparameters are estimated by applying the Bethe approximation to maximum marginal likelihood estimation.

Original image	$\hat{p}$	$\hat{\alpha}$	$d(\boldsymbol{f}, \hat{\boldsymbol{f}})$	Δ_{SNR} (dB)
		(*a*) $Q = 3$		
Girl	0.296 86	0.873 67	0.066 94	8.801 03
Mandrill	0.297 94	0.724 26	0.194 96	3.560 09
Home	0.297 94	0.788 34	0.115 59	6.091 29
		(*b*) $Q = 4$		
Girl	0.280 52	0.787 99	0.098 07	8.892 41
Mandrill	0.343 97	0.623 62	0.240 57	4.081 05
Home	0.282 65	0.727 36	0.124 86	7.314 26

'home' has many spatially flat parts and the image 'mandrill' has many spatially changing parts. However, the Q-Ising model gives us good results for the cases of both $Q = 3$ and $Q = 4$.

6. Concluding remarks

In the present paper, we have investigated probabilistic image restoration when the Q-Ising model is adopted as the *a priori* probability distribution. The practical algorithm was constructed by using the Bethe approximation and the hyperparameters are determined so as to maximize the marginal likelihood for each degraded image. We conclude that, for many real world grey-level images, the Q-Ising model can provide good results. In numerical experiments based on the original images given in figure 7, the estimated values of the hyperparameter α are about 0.873 67, 0.724 26 and 0.788 34, as shown in table 2(a). For such values of α, original images $\boldsymbol{f}$ generated by Monte Carlo simulations from the *a priori* probability distribution (3) for the Q-Ising model will typically have most pixels in state 1. However, the Q-Ising model does give us good results for real-world images when we use such large values of α. We conclude that image restorations obtained by using probabilistic models are not always dictated by the most likely configuration corresponding to the *a priori* probability distribution.

Another important conclusion is that the Bethe approximation is clearly applicable to probabilistic image restoration including the estimation of hyperparameters. Moreover, the basic procedure in the algorithm can be reduced to dealing with very simple equations (17)–(19). This procedure corresponds to loopy belief propagation in artificial intelligence [8]. In the numerical experiments in the present paper, we have not applied an expectation–maximization procedure to the maximization of the marginal likelihood, although this is often done. Zhang [4] has succeeded in applying an advanced mean-field method within the expectation-maximization procedure in the context of probabilistic image restoration. Detailed investigation of an algorithm constructed by combining loopy belief propagation with the expectation-maximization algorithm is a problem for future research.

The present authors employed the Bethe approximation in maximizing the marginal likelihood $\Pr\{\boldsymbol{G} = \boldsymbol{g}|\alpha, p\}$ of hyperparameters α and p. In the Bethe approximation, the approximate free energy for every probabilistic model in equation (16) is given by

$$\mathcal{F}_{\mathrm{Bethe}}[\{\rho_i, \rho_{ij} | i \in \Omega, ij \in B\}] \equiv \sum_{i \in \Omega} \mathcal{F}_i[\rho_i] + \sum_{ij \in B} (\mathcal{F}_{ij}[\rho_{ij}] - \mathcal{F}_i[\rho_i] - \mathcal{F}_j[\rho_j]), \tag{22}$$

where

$$\mathcal{F}_i[\rho_i] \equiv \sum_{\zeta \in \boldsymbol{Q}} \rho_i(\zeta) \ln \left(\frac{\rho_i(\zeta)}{\psi_i(\zeta)} \right) \tag{23}$$

$$\mathcal{F}_{ij}[\rho_{ij}] \equiv \sum_{\zeta \in \boldsymbol{Q}} \sum_{\zeta' \in \boldsymbol{Q}} \rho_{ij}(\zeta, \zeta') \ln \left(\frac{\rho_{ij}(\zeta, \zeta')}{\psi_i(\zeta) \phi_{ij}(\zeta, \zeta') \psi_j(\zeta')} \right). \tag{24}$$

The simultaneous fixed-point equations (17)–(19) are equivalent to the extremum conditions of the Bethe free energy $\mathcal{F}_{\mathrm{Bethe}}[\{\rho_i, \rho_{ij} | i \in \Omega, ij \in B\}]$ with respect to the marginal probability distributions $\{\rho_i, \rho_{ij} | i \in \Omega, ij \in B\}$ under the constraints $\sum_{\zeta \in \boldsymbol{Q}} \rho_i(\zeta) = 1$ $(i \in \Omega)$, $\sum_{\zeta \in \boldsymbol{Q}} \sum_{\zeta' \in \boldsymbol{Q}} \rho_{ij}(\zeta, \zeta') = 1$ $(ij \in B)$ and $\rho_i(\xi) = \sum_{\zeta \in \boldsymbol{Q}} \rho_{ij}(\xi, \zeta)$ $(\xi \in \boldsymbol{Q}, j \in \boldsymbol{c}_i, i \in \Omega)$. However, it is known that the Bethe free energy $\mathcal{F}_{\mathrm{Bethe}}[\{\rho_i, \rho_{ij} | i \in \Omega, ij \in B\}]$ does not provide any bounds for the true free energy $\mathcal{F}[\rho] = -\ln\bigl(\sum_{\boldsymbol{z}} \prod_{i \in \Omega} \psi_i(z_i) \prod_{ij \in B} \phi_{ij}(z_i, z_j)\bigr)$, while a mean-field free energy is a bound for the true free energy [7]. Furthermore, in some cases the solution of the simultaneous fixed point equations (17)–(19) corresponds not to a local minimum but to a saddle point of the Bethe free energy [14]. In spite of that, the present scheme provides satisfactory results. The Bethe approximation often gives us poor results for some Ising models with frustration effects [15, 16]. Frustration effects in probabilistic models cause poor results in the Bethe approximation. The present *a priori* probabilistic model (3) corresponds to spatially uniform ferromagnetic interactions and has no frustration effect. The frustration effects in the present *a posteriori* probabilistic model (4) come from only a part of spatially non-uniform external fields $-\beta(p)\delta_{f_i, g_i}$ and is very small at that. This seems to be a qualitative explanation for obtaining satisfactory results in the present scheme in spite of some problems in the Bethe approximation.

Inoue and Carlucci [10] adopted Q-Ising models with infinite-range interactions as *a priori* probability distributions and investigated statistical performance in terms of some distances between the original and the restored images, calculated by using the replica method. In addition, some of the present authors have investigated hyperparameter estimation by applying the replica method to the maximization of the marginal likelihood for the spin-$\frac{1}{2}$ Ising model with infinite-range interactions. In work in progress we are extending the framework to the Q-Ising model with infinite-range interactions. The results will be reported elsewhere [17].

Acknowledgments

The authors are 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.

References

[1] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford: Oxford University Press)
[2] Tanaka K 2002 *J. Phys. A: Math. Gen.* **35** R81
[3] Geman S and Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721
[4] Zhang J 1992 *IEEE Trans. Image Process.* **40** 2570
[5] Pryce J M and Bruce A D 1995 *J. Phys. A: Math. Gen.* **28** 511
[6] Tanaka K and Morita T 1995 *Phys. Lett.* A **203** 122

[7] Opper M and Saad D (ed) 2001 *Advanced Mean Field Methods—Theory and Practice* (Cambridge, MA: MIT Press)
[8] Yedidia J S, Freeman W T and Weiss Y 2001 *Advances in Neural Information Processing Systems 13* ed T Leen, T Dietterich and V Tresp (Cambridge, MA: MIT Press) pp 689–95
[9] Tanaka K 2000 *IEICE Trans.* A **J83-A** 1148 (in Japanese)
Tanaka K 2000 *Electron. Commun. Japan* **85** 50 (Engl. Transl.)
[10] Inoue J and Carlucci D M 2001 *Phys. Rev.* E **64** 036121
[11] Blum M, Emery V J and Griffiths R B 1971 *Phys. Rev.* A **4** 1071
[12] Marroquin J, Mitter S and Poggio T 1987 *J. Am. Stat. Assoc.* **82** 76
[13] Nishimori H and Wong K Y M 1999 *Phys. Rev.* E **60** 132
[14] Heskes T 2003 *Adv. Neural Inf. Process. Syst.* **15** at press
[15] Morita T 1987 *Physica* A **141** 335
[16] Morita T 1988 *Phys. Lett.* A **132** 1
[17] Inoue J and Tanaka K 2003 *J. Phys. A: Math. Gen.* **36** 10997