Statistical trajectory of an approximate EM algorithm for probabilistic image processing

2007 J. Phys. A: Math. Theor. 40 11285

(http://iopscience.iop.org/1751-8121/40/37/007)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 11285–11300 doi:10.1088/1751-8113/40/37/007

Statistical trajectory of an approximate EM algorithm for probabilistic image processing

Kazuyuki Tanaka[1] **and D M Titterington**[2]

[1] Graduate School of Information Sciences, Tohoku University, Aramaki-aza-aoba 6-3-09, Aoba-ku, Sendai 980-8579, Japan

[2] Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK

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

Received 4 June 2007, in final form 30 July 2007
Published 29 August 2007
Online at stacks.iop.org/JPhysA/40/11285

Abstract

We calculate analytically a statistical average of trajectories of an approximate expectation-maximization (EM) algorithm with generalized belief propagation (GBP) and a Gaussian graphical model for the estimation of hyperparameters from observable data in probabilistic image processing. A statistical average with respect to observed data corresponds to a configuration average for the random-field Ising model in spin glass theory. In the present paper, hyperparameters which correspond to interactions and external fields of spin systems are estimated by an approximate EM algorithm. A practical algorithm is described for gray-level image restoration based on a Gaussian graphical model and GBP. The GBP approach corresponds to the cluster variation method in statistical mechanics. Our main result in the present paper is to obtain the statistical average of the trajectory in the approximate EM algorithm by using loopy belief propagation and GBP with respect to degraded images generated from a probability density function with true values of hyperparameters. The statistical average of the trajectory can be expressed in terms of recursion formulas derived from some analytical calculations.

PACS numbers: 02.50.−r, 02.50.Cw, 02.50.Tt, 05.20.−y, 05.50.+q, 75.10.Nr, 87.19.Dd, 89.70.+c

1. Introduction

Belief propagation and its related algorithms have been used to carry out statistical inference for computational models on an arbitrary graph with a large number of nodes and there are links to work in computer science [1–4]. In particular it is known that loopy belief propagation (LBP) and generalized belief propagation (GBP) are equivalent respectively to Bethe approximations and the cluster variation method in statistical mechanics [5–9]. One of the successful applications is to probabilistic image processing [10–12].

One of the difficult problems in probabilistic image processing is how to estimate hyperparameters from observed data, for example in the form of degraded images in the context of image restoration. One familiar statistical approach is to maximize a marginal likelihood. By using LBP, we can construct approximate frameworks for estimating the hyperparameters from the observed data alone [11]. In probabilistic image processing based on Gaussian graphical models, Tanaka *et al* [13] have investigated the accuracy of hyperparameter estimation by comparing approximate values of the marginal likelihood for LBP with the exact values. Tanaka *et al* [14] have investigated the accuracy of hyperparameter estimation by comparing approximate values of the marginal likelihood for the generalized belief propagation with the exact values, numerically.

As another statistical-mechanical approach to computer science, we have statistical performance estimation by using an aspect of spin glass theory [15]. Statistical performance estimation corresponds to a configuration average for random spin systems. A statistical performance estimation scheme has been derived by applying the replica method to image restoration, although the computational model has been defined on a complete graph [18]. It is expected that they have similar properties to those of approaches to image restoration on a square lattice. One familiar statistical technique for estimating hyperparameters is the expectation-maximization (EM) algorithm [4]. A statistical average of trajectories corresponding to the EM algorithm with respect to observed data has been investigated for some probabilistic computational models on a complete graph by using the replica method [19, 20].

Some authors have investigated the accuracy of certain statistical quantities and the convergence of algorithms in LBP [21–25]. However, statistical averages of trajectories of approximate EM algorithms in LBP and GBP have not been investigated yet.

The main purpose of the present paper is to calculate the statistical average of the trajectories for hyperparameter estimation by using LBP and GBP analytically and to compare the statistical average with that obtained by exact calculation. We obtain the statistical average of the trajectory in the approximate EM algorithm by using LBP and GBP with respect to degraded images generating from a probability density function with true values of hyperparameters. The statistical average of the trajectory can be expressed in terms of recursion formulas derived from some analytical calculations as our main results in the present paper. In section 2, we explain the basic framework of the EM algorithm for Bayesian image analysis. In section 3, we present an EM algorithm for image processing based on a Gaussian graphical model, making use of the multi-dimensional Gaussian integral formula. In section 4, we present the EM algorithm for GBP in the Gaussian graphical model and describe some numerical experiments. In section 5, we derive the statistical average of the trajectories of the EM algorithm in the Gaussian graphical model on a square lattice and compare the results with those based on the exact calculations given in section 3. In section 6, we provide some concluding remarks.

2. EM algorithm in probabilistic image processing

In this section, we present a framework for probabilistic image restoration for gray-level images by using a Gaussian graphical model as an *a priori* probabilistic model in the Bayesian framework. We also develop the EM algorithm for hyperparameter estimation in probabilistic image restoration.

In computer vision, images are typically defined on a set of pixels arranged on a rectangular lattice $\Omega \equiv \{i|i = 1, 2, \ldots, |\Omega|\}$ in the two-dimensional xy-plane, where $|\Omega|$ is the total number of pixels[3]. The spatial coordinate vector of the ith pixel is denoted by $\boldsymbol{r}_i$. We consider

[3] We denote the number of whole elements belonging to a set A by $|A|$.

that there is a link ij between every nearest-neighbor pair of pixels i and j, and the set of all the links is denoted by $\mathcal{N}$.

At each pixel, the intensity of light is represented by an integer or a real number. Here, we consider an image on a rectangular lattice Ω such that the intensity at each pixel takes a real value in the range $(-\infty, +\infty)$. A monochrome digital image is then expressed as a two-dimensional light intensity function f_i, where f_i is proportional to the brightness of the image at pixel i. The rectangular lattice Ω is assumed to have periodic boundary conditions in both x- and y-directions.

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, and the random fields of intensities in the original image and the degraded image are represented by $\boldsymbol{F} \equiv (F_1, F_2, \ldots, F_{|\Omega|})^{\mathrm{T}}$ and $\boldsymbol{G} \equiv (G_1, G_2, \ldots, G_{|\Omega|})^{\mathrm{T}}$, respectively. The actual original image and the degraded image actually observed are denoted by $\boldsymbol{f} = (f_1, f_2, \ldots, f_{|\Omega|})^{\mathrm{T}}$ and $\boldsymbol{g} = (g_1, g_2, \ldots, g_{|\Omega|})^{\mathrm{T}}$, respectively.

In the present paper, it is assumed that the degraded image $\boldsymbol{g}$ is generated from the original image $\boldsymbol{f}$ by the addition of white Gaussian noise with mean 0 and variance σ^2, so that

$$\mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\boldsymbol{F} = \boldsymbol{f}, \sigma) \equiv \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{|\Omega|}{2}} \prod_{i\in\Omega} \exp\left(-\frac{1}{2\sigma^2}(f_i - g_i)^2\right). \tag{1}$$

Moreover, the *a priori* probability density for the original image is assumed to be

$$\mathcal{P}(\boldsymbol{F} = \boldsymbol{f}|\alpha) \equiv \frac{1}{\mathcal{Z}_{\mathrm{PR}}(\alpha)} \prod_{ij\in\mathcal{N}} \exp\left(-\frac{1}{2}\alpha(f_i - f_j)^2\right), \tag{2}$$

which represents local spatial correlation among the pixel intensities. The denominator $\mathcal{Z}_{\mathrm{PR}}(\alpha)$ in equation (2) is a normalization constant. By substituting equations (1) and (2) into the Bayes formula, we obtain

$$\begin{aligned}\mathcal{P}(\boldsymbol{F} = \boldsymbol{f}|\boldsymbol{G} = \boldsymbol{g}, \alpha, \sigma) &= \frac{\mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\boldsymbol{F} = \boldsymbol{f}, \sigma)\mathcal{P}(\boldsymbol{F} = \boldsymbol{f}|\alpha)}{\int \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\boldsymbol{F} = \boldsymbol{z}, \sigma)\mathcal{P}(\boldsymbol{F} = \boldsymbol{z}|\alpha)\,\mathrm{d}\boldsymbol{z}} \\ &= \frac{1}{\mathcal{Z}_{\mathrm{PO}}(\boldsymbol{g}, \alpha, \sigma)} \exp\left(-\frac{1}{2\sigma^2}\sum_{i\in\Omega}(f_i - g_i)^2 - \frac{1}{2}\alpha\sum_{ij\in\mathcal{N}}(f_i - f_j)^2\right),\end{aligned} \tag{3}$$

where $\int \mathrm{d}\boldsymbol{z} \equiv \int_{-\infty}^{+\infty} \mathrm{d}z_1 \int_{-\infty}^{+\infty} \mathrm{d}z_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}z_{|\Omega|}$. The denominator $\mathcal{Z}_{\mathrm{PO}}(\boldsymbol{g}, \alpha, \sigma)$ in equation (3) is a normalization constant.

In the maximum likelihood approach, values for the hyperparameters α and σ are determined so as to maximize the marginal likelihood $\mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha, \sigma)$, where

$$\mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha, \sigma) \equiv \int \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\boldsymbol{F} = \boldsymbol{z}, \sigma)\mathcal{P}(\boldsymbol{F} = \boldsymbol{z}|\alpha)\,\mathrm{d}\boldsymbol{z}. \tag{4}$$

We denote these maximizers by $\widehat{\alpha}$ and $\widehat{\sigma}$:

$$(\widehat{\alpha}, \widehat{\sigma}) = \arg\max_{(\alpha,\sigma)} \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha, \sigma). \tag{5}$$

Given the estimates $\widehat{\alpha}$ and $\widehat{\sigma}$, the restored image $\widehat{\boldsymbol{f}} = (\widehat{f}_1, \widehat{f}_2, \ldots, \widehat{f}_{|\Omega|})^{\mathrm{T}}$ is determined by

$$\widehat{\boldsymbol{f}} = \boldsymbol{h}(\boldsymbol{g}, \widehat{\alpha}, \widehat{\sigma}) \equiv \int \boldsymbol{z}\mathcal{P}(\boldsymbol{F} = \boldsymbol{z}|\boldsymbol{G} = \boldsymbol{g}, \widehat{\alpha}, \widehat{\sigma})\,\mathrm{d}\boldsymbol{z}. \tag{6}$$

This way of producing a restored image is called maximum posterior mean estimation. In this case, it also provides the 'mode' because of the Gaussian nature of the posterior.

The maximization of the marginal likelihood in equation (5) can be achieved by means of the EM algorithm. If we define

$$Q(\alpha, \sigma|\alpha', \sigma', \boldsymbol{g}) \equiv \int \mathcal{P}(\boldsymbol{F} = \boldsymbol{z}|\boldsymbol{G} = \boldsymbol{g}, \alpha', \sigma') \ln(\mathcal{P}(\boldsymbol{F} = \boldsymbol{z}, \boldsymbol{G} = \boldsymbol{g}|\alpha, \sigma))\,\mathrm{d}\boldsymbol{z}, \tag{7}$$

the EM algorithm is summarized as follows.

EM algorithm

Step 1: Set $(\alpha(0), \sigma(0))$ and $t \leftarrow 0$.
Step 2: Iterate the following EM step until convergence:

$$(\alpha(t+1), \sigma(t+1)) \leftarrow \arg\max_{(x,y)} Q(x, y|\alpha(t), \sigma(t), \boldsymbol{g}) \tag{8}$$

and $t \leftarrow t + 1$.
The Q-function has the following relationship with the marginal likelihood:

$$\left[\frac{\partial}{\partial\theta} Q(x, y|\alpha, \sigma, \boldsymbol{g})\right]_{x=\alpha, y=\sigma} = \left[\frac{\partial}{\partial\theta} \ln\, \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|x, y)\right]_{x=\alpha, y=\sigma}, \qquad (\theta = x, y). \tag{9}$$

It then follows that the above EM algorithm leads to one of the extreme points of the marginal likelihood $\mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha, \sigma)$ with respect to (α, σ). Equation (8) can be replaced by the update rule in which $(\alpha(t+1), \sigma(t+1))$ is determined so as to be the extreme point of $Q(x, y|\alpha(t), \sigma(t), \boldsymbol{g})$ with respect to x and y, i.e. such that

$$\left[\frac{\partial}{\partial\theta} Q(x, y|\alpha(t), \sigma(t), \boldsymbol{g})\right]_{x=\alpha(t+1), y=\sigma(t+1)} = 0, \qquad (\theta = x, y). \tag{10}$$

By substituting equations (1) and (2) into equation (10) with equation (7), we can reduce the update rule in the EM algorithm to the following equations:

$$\sum_{ij\in\mathcal{N}} \langle (F_i - F_j)^2 | \alpha(t+1)\rangle = \sum_{ij\in\mathcal{N}} \langle (F_i - F_j)^2 | \boldsymbol{g}, \alpha(t), \sigma(t)\rangle, \tag{11}$$

$$\sigma(t+1)^2 = \frac{1}{|\Omega|}\sum_{i\in\Omega} \langle (F_i - g_i)^2 | \boldsymbol{g}, \alpha(t), \sigma(t)\rangle. \tag{12}$$

Here $\langle (F_i - F_j)^2|\boldsymbol{g}, \alpha(t), \sigma(t)\rangle$ and $\langle (F_i - g_i)^2|\boldsymbol{g}, \alpha(t), \sigma(t)\rangle$ are the expectation values of $(F_i - F_j)^2$ and $(F_i - g_i)^2$ with respect to $\mathcal{P}(\boldsymbol{F} = \boldsymbol{f}|\boldsymbol{G} = \boldsymbol{g}, \alpha, \sigma)$, respectively, while $\langle (F_i - F_j)^2|\alpha(t+1)\rangle$ is the expectation value of $(F_i - F_j)^2$ with respect to $\mathcal{P}(\boldsymbol{F} = \boldsymbol{f}|\alpha(t+1))$.

Let us suppose that degraded images $\boldsymbol{g}$ are generated according to $\mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*)$ for the true values of hyperparameters $\alpha = \alpha^*$ and $\sigma = \sigma^*$. Now we introduce the statistical average of the Q-function, defined by

$$\mathcal{Q}(\alpha, \sigma|\alpha(t), \sigma(t)) \equiv \int Q(\alpha, \sigma|\alpha(t), \sigma(t), \boldsymbol{g}) \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*)\,\mathrm{d}\boldsymbol{g}. \tag{13}$$

The statistical average of the EM algorithm (8) can be written as

$$(\alpha(t+1), \sigma(t+1)) \leftarrow \arg\max_{\alpha,\sigma} \mathcal{Q}(\alpha, \sigma|\alpha(t), \sigma(t)). \tag{14}$$

By differentiating $\mathcal{Q}(\alpha, \sigma|\alpha(t), \sigma(t))$ with respect to α and σ, we can reduce equation (14) to

$$\sum_{ij\in\mathcal{N}} \langle (F_i - F_j)^2 | \alpha(t+1)\rangle = \sum_{ij\in\mathcal{N}} \int \langle (F_i - F_j)^2 | \boldsymbol{g}, \alpha(t), \sigma(t)\rangle \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*)\,\mathrm{d}\boldsymbol{g}, \tag{15}$$

$$\sigma(t+1)^2 = \frac{1}{|\Omega|}\sum_{i\in\Omega} \int \langle (F_i - g_i)^2 | \boldsymbol{g}, \alpha(t), \sigma(t)\rangle \mathcal{P}(\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*)\,\mathrm{d}\boldsymbol{g}. \tag{16}$$

Equations (15) and (16) provide the statistical average of the trajectories in the EM algorithm with respect to the probability density function $\mathcal{P}(\boldsymbol{G}=\boldsymbol{g}|\alpha^*,\sigma^*)$. For example, suppose that we generate 20 degraded images $\boldsymbol{g}^{(1)},\boldsymbol{g}^{(2)},\ldots,\boldsymbol{g}^{(20)}$ from $\mathcal{P}(\boldsymbol{G}=\boldsymbol{g}|\alpha^*,\sigma^*)$ and calculate the sequence $\{(\alpha^{(k)}(t),\sigma^{(k)}(t))|t=0,1,2,\ldots\}$ for each degraded image $\boldsymbol{g}^{(k)}$. Then we can estimate the statistical average of the trajectory in the EM algorithm by the sample average of the sequences obtained by applying the EM algorithm to each degraded image $\boldsymbol{g}^{(k)}$. Equations (15) and (16) can be regarded as update rules for calculating the statistical average of the EM trajectory with respect to the infinite number of degraded images generated from the probability density function $\mathcal{P}(\boldsymbol{G}=\boldsymbol{g}|\alpha^*,\sigma^*)$.

3. Exact EM algorithm for Gaussian graphical model

Some statistical quantities associated with the Gaussian graphical model mentioned in the previous section can be calculated exactly by means of the multi-dimensional Gaussian integral formula. In this section, we give an explicit algorithm for calculating the statistical average of the trajectory $(\alpha(t),\sigma(t))$ in the EM algorithm, as defined in equations (15) and (16). The basic scheme underlying the present section has been given in [11, 26].

By using the multi-dimensional Gaussian integral formula, we obtain the exact expression of the restored image $\widehat{\boldsymbol{f}}=\boldsymbol{h}(\boldsymbol{g},\widehat{\alpha},\widehat{\sigma})$ as

$$\widehat{\boldsymbol{f}}=\boldsymbol{h}(\boldsymbol{g},\widehat{\alpha},\widehat{\sigma})=(\boldsymbol{I}+\widehat{\alpha}\widehat{\sigma}^2\boldsymbol{C})^{-1}\boldsymbol{g}. \tag{17}$$

Here $\boldsymbol{I}$ is the $|\Omega|\times|\Omega|$ unit matrix and $\boldsymbol{C}$ is the $|\Omega|\times|\Omega|$ matrix whose (i,j)-component $\langle i|\boldsymbol{C}|j\rangle$ is defined by

$$\langle i|\boldsymbol{C}|j\rangle\equiv\begin{cases}4 & (i=j)\\ -1 & (ij\in\mathcal{N})\\ 0 & (\text{otherwise}),\end{cases}\qquad (i,j\in\Omega). \tag{18}$$

Furthermore, the expectation values $\langle(F_i-F_j)^2|\alpha(t+1)\rangle$, $\sum_{ij\in\mathcal{N}}\langle(F_i-F_j)^2|\boldsymbol{g},\alpha(t),\sigma(t)\rangle$ and $\langle(F_i-g_i)^2|\boldsymbol{g},\alpha(t),\sigma(t)\rangle$ can be derived in closed form in terms of the matrices $\boldsymbol{I}$ and $\boldsymbol{C}$ and the vector $\boldsymbol{g}$ by using the multi-dimensional Gaussian integral formula. As a result, the update rules (11) and (12) in the EM algorithm can be written as

$$\alpha(t+1)^{-1}=\frac{1}{|\Omega|}\mathrm{Tr}(\sigma(t)^2\boldsymbol{C}(\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})+\frac{1}{|\Omega|}\boldsymbol{g}^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1}\boldsymbol{g}, \tag{19}$$

$$\sigma(t+1)^2=\frac{1}{|\Omega|}\mathrm{Tr}(\sigma(t)^2(\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})+\frac{1}{|\Omega|}\boldsymbol{g}^{\mathrm{T}}\alpha(t)^2\sigma(t)^4(\boldsymbol{C}(\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})^2\boldsymbol{g}. \tag{20}$$

If we average the right-hand sides of equations (19) and (20) with respect to the probability density function $\mathcal{P}(\boldsymbol{G}=\boldsymbol{g}|\alpha^*,\sigma^*)$, the update rules (15) and (16) are as follows:

$$\begin{aligned}\alpha(t+1)^{-1}={}&\frac{1}{|\Omega|}\mathrm{Tr}(\sigma(t)^2\boldsymbol{C}(\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})\\&-\frac{1}{|\Omega|}\mathrm{Tr}\left(\frac{1}{\alpha^*}\right)(\boldsymbol{I}+\alpha^*\sigma^{*2}\boldsymbol{C})((\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})^2,\end{aligned} \tag{21}$$

$$\begin{aligned}\sigma(t+1)^2={}&\frac{1}{|\Omega|}\mathrm{Tr}(\sigma(t)^2(\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})\\&-\frac{1}{|\Omega|}\mathrm{Tr}\left(\frac{\alpha(t)^2\sigma(t)^4}{\alpha^*}\right)\boldsymbol{C}(\boldsymbol{I}+\alpha^*\sigma^{*2}\boldsymbol{C})((\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1})^2.\end{aligned} \tag{22}$$

It is obvious that equations (21) and (22) are satisfied by $\alpha(t) = \alpha(t+1) = \alpha^*$ and $\sigma(t) = \sigma(t+1) = \sigma^*$.

4. GBP for Gaussian graphical model

In this section, we summarize the theoretical structure of GBP. We explain the framework for probabilistic models defined by the following probability density function:

$$p(\boldsymbol{f}|\boldsymbol{g},\alpha,\beta) = \frac{1}{\mathcal{Z}(\boldsymbol{g},\alpha,\beta)} \exp\left(-\frac{1}{2}\beta\sum_{i\in\Omega}(f_i-g_i)^2 - \frac{1}{2}\alpha\sum_{ij\in\mathcal{N}}(f_i-f_j)^2\right), \tag{23}$$

where

$$\mathcal{Z}(\boldsymbol{g},\alpha,\beta) \equiv \int \exp\left(-\frac{1}{2}\beta\sum_{i\in\Omega}(z_i-g_i)^2 - \frac{1}{2}\alpha\sum_{ij\in\mathcal{N}}(z_i-z_j)^2\right)\mathrm{d}\boldsymbol{z}. \tag{24}$$

If we set $\beta = 0$ and $\beta = 1/\sigma^2$, the probability density function (23) reduces to equations (2) and (3).

We introduce the Kullback–Leibler (KL) divergence defined by

$$\mathcal{D}[q\|p] \equiv \int \mathrm{d}\boldsymbol{z}\, q(\boldsymbol{z}) \ln\left(\frac{q(\boldsymbol{z})}{p(\boldsymbol{z})}\right). \tag{25}$$

If we substitute equation (23) into equation (25), the KL divergence can be expressed in terms of the average $m_i = \int z_i q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$, the variance $V_{ii} = \int (z_i - m_i)^2 q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$ and the covariance $V_{ij} = \int (z_i - m_i)(z_j - m_j) q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$ as follows:

$$\mathcal{D}[q\|p] = \ln(\mathcal{Z}(\boldsymbol{g},\alpha,\beta)) + \mathcal{F}[q], \tag{26}$$

where

$$\mathcal{F}[q] \equiv \frac{1}{2}\beta\sum_{i\in\Omega}(V_{ii} + (m_i - g_i)^2) + \frac{1}{2}\alpha\sum_{ij\in\mathcal{N}}((V_{ii} - 2V_{ij} + V_{jj} + (m_i - m_j)^2) + \int q(\boldsymbol{z})\ln(q(\boldsymbol{z}))\,\mathrm{d}\boldsymbol{z}. \tag{27}$$

In order to explain the framework of GBP, we should define some notation for clusters. In the present paper, a set of pixels is called cluster. When a pixel i belongs to a cluster γ, we call i an element of γ and express it in terms of the notation $i \in \gamma$. When all the nodes in a cluster γ' belong to a cluster γ and γ is not equal to γ', we call γ' a subcluster of γ and use the notation $\gamma' \subset \gamma$.

Now we 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 $\mathcal{B}$. We consider a set $\mathcal{C}$ of clusters such that a cluster is in $\mathcal{C}$ if and only if it is a cluster in $\mathcal{B}$ or is the cluster of the common nodes of two or more clusters in $\mathcal{B}$.

Suppose that $|\gamma| = n$, and that the pixels $\gamma_1, \gamma_2, \ldots, \gamma_n$ belonging to γ are ordered so that $\gamma_1 < \gamma_2 < \cdots < \gamma_n$. Let the n-dimensional vectors $\boldsymbol{f}_\gamma$ and $\boldsymbol{z}_\gamma$ be defined by $\boldsymbol{f}_\gamma \equiv \left(f_{\gamma_1}, f_{\gamma_2}, \ldots, f_{\gamma_n}\right)^{\mathrm{T}}$ and $\boldsymbol{z}_\gamma \equiv \left(z_{\gamma_1}, z_{\gamma_2}, \ldots, z_{\gamma_n}\right)^{\mathrm{T}}$, respectively. We introduce the marginal probability density function $Q_\gamma(\boldsymbol{f}_\gamma)$ defined by

$$q_\gamma(\boldsymbol{f}_\gamma) \equiv \int \left(\prod_{i\in\gamma}\delta(f_i - z_i)\right) q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} \qquad (\gamma \in \mathcal{C}). \tag{28}$$

For the purposes of GBP, we assume that the trial function q is restricted to the following factorized form:

$$q(\boldsymbol{f}) = \prod_{\gamma\in\mathcal{C}} Q_\gamma(\boldsymbol{f}_\gamma)^{-\mu(\gamma)}, \tag{29}$$

where $\mu(\gamma)$ is a Möbius function defined by

$$\mu(\gamma) \equiv -1 - \sum_{\{\gamma'|\gamma'\supset\gamma,\gamma'\in\mathcal{C}\}} \mu(\gamma') \qquad (\gamma \in \mathcal{C}) \tag{30}$$

Moreover, we assume that $q_\gamma(\boldsymbol{f}_\gamma)$ is the $|\gamma|$-dimensional Gaussian density given by

$$q_\gamma(\boldsymbol{f}_\gamma) = \frac{1}{\sqrt{(2\pi)^{|\gamma|}\det(\boldsymbol{V}_\gamma)}} \exp\left(-\frac{1}{2}(\boldsymbol{f}_\gamma - \boldsymbol{m}_\gamma)^{\mathrm{T}}\boldsymbol{V}_\gamma{}^{-1}(\boldsymbol{f}_\gamma - \boldsymbol{m}_\gamma)\right). \tag{31}$$

In the case of $|\gamma| = n$, the average vector $\boldsymbol{m}_\gamma$ and the covariance matrix $\boldsymbol{V}_\gamma$ are defined by

$$\boldsymbol{m}_\gamma = \begin{pmatrix} m_{\gamma_1} \\ m_{\gamma_2} \\ \vdots \\ m_{\gamma_n} \end{pmatrix}, \qquad \boldsymbol{V}_\gamma = \begin{pmatrix} V_{\gamma_1,\gamma_1} & V_{\gamma_1,\gamma_2} & \cdots & V_{\gamma_1,\gamma_n} \\ V_{\gamma_2,\gamma_1} & V_{\gamma_2,\gamma_2} & \cdots & V_{\gamma_2,\gamma_n} \\ \vdots & \vdots & \ddots & \vdots \\ V_{\gamma_n,\gamma_1} & V_{\gamma_n,\gamma_2} & \cdots & V_{\gamma_n,\gamma_n} \end{pmatrix}. \tag{32}$$

From equations (29) and (31), the free energy (27) can be expressed in terms of the averages m_i and variances V_{ii} for all the pixels ($i \in \Omega$), and the covariances V_{ij} for all the nearest-neighbor pairs of pixels ($ij \in \mathcal{N}$) as follows:

$$\begin{aligned}\mathcal{F}[q] = \mathcal{F}[\boldsymbol{m}, \boldsymbol{V}] \equiv{}& \frac{1}{2}\beta\sum_{i\in\Omega}(V_{ii} + (m_i - g_i)^2) + \frac{1}{2}\alpha\sum_{ij\in\mathcal{N}}(V_{ii} - 2V_{ij} + V_{jj} + (m_i - m_j)^2) \\ &+\sum_{\gamma\in\mathcal{C}}\mu(\gamma)\left(1 + \frac{1}{2}\ln((2\pi)^{|\gamma|}\det\boldsymbol{V}_\gamma)\right).\end{aligned} \tag{33}$$

The extremum conditions of $\mathcal{F}[\boldsymbol{m}, \boldsymbol{V}]$ with respect to the average vector $\boldsymbol{m}$ and the covariance matrix $\boldsymbol{V}$ are as follows:

$$\beta(m_i - g_i) + \alpha\sum_{j\in\mathcal{N}_i}(m_i - m_j) = 0 \tag{34}$$

$$\beta + 4\alpha + \sum_{\{\gamma|\gamma\in\mathcal{C},i\in\gamma\}} \mu(\gamma)\left(\boldsymbol{V}_\gamma^{-1}\right)_{ii} = 0 \qquad (i \in \Omega) \tag{35}$$

$$-\alpha + \sum_{\{\gamma|\gamma\in\mathcal{C},i\in\gamma,j\in\gamma\}} \mu(\gamma)\left(\boldsymbol{V}_\gamma^{-1}\right)_{ij} = 0 \qquad (ij \in \mathcal{N}) \tag{36}$$

$$\sum_{\{\gamma|\gamma\in\mathcal{C},i\in\gamma,j\in\gamma\}} \mu(\gamma)\left(\boldsymbol{V}_\gamma^{-1}\right)_{ij} = 0 \qquad (i, j \in \Omega, ij\notin\mathcal{N}). \tag{37}$$

For each (α, β), we can solve equations (34)–(37) numerically and obtain $\boldsymbol{m}$ and $\boldsymbol{V}_\gamma$ $(\gamma \in \mathcal{C})$.

5. Hyperparameter estimation by the EM algorithm and GBP

In this section, we apply GBP to the EM algorithm for the Gaussian graphical model when a degraded image $\boldsymbol{g}$ is given. The explicit algorithm and some numerical experiments involving practical images are also given.

The analytical solution of equation (34) gives

$$\boldsymbol{m} = \boldsymbol{h}(\boldsymbol{g}, \alpha, \sigma) = \begin{pmatrix} h_1(\boldsymbol{g}, \alpha, \sigma) \\ h_2(\boldsymbol{g}, \alpha, \sigma) \\ \vdots \\ h_{|\Omega|}(\boldsymbol{g}, \alpha, \sigma) \end{pmatrix} = \beta(\beta \boldsymbol{I} + \alpha \boldsymbol{C})^{-1}\boldsymbol{g}. \tag{38}$$

In GBP, the restored image $\widehat{\boldsymbol{f}}$ from the posterior probability density function is given by equation (17) [14].

The covariance matrix $\boldsymbol{V}$ is obtained by solving the system of equations (35)–(37) numerically. These equations do not include the vector $\boldsymbol{g}$ and therefore the approximate values of the covariance matrix in GBP do not depend on the vector $\boldsymbol{g}$ or the position of pixel. We can express the variances and the covariances in the prior and the posterior probabilistic models in terms of notation independent of pixels as

$$V_{\mathrm{PR}}^{(0)}(\alpha) \equiv \left\langle {F_i}^2 \middle| \alpha \right\rangle - \langle F_i|\alpha\rangle^2 \qquad (i \in \Omega) \tag{39}$$

$$V_{\mathrm{PR}}^{(1)}(\alpha) \equiv \langle F_i F_j|\alpha\rangle - \langle F_i|\alpha\rangle\langle F_j|\alpha\rangle \qquad (ij \in \mathcal{N}) \tag{40}$$

$$V_{\mathrm{PO}}^{(0)}(\alpha, \sigma) \equiv \left\langle {F_i}^2 \middle| \boldsymbol{g}, \alpha, \sigma \right\rangle - \langle F_i|\boldsymbol{g}, \alpha, \sigma\rangle^2 \qquad (i \in \Omega) \tag{41}$$

$$V_{\mathrm{PO}}^{(1)}(\alpha, \sigma) \equiv \langle F_i F_j|\boldsymbol{g}, \alpha, \sigma\rangle - \langle F_i|\boldsymbol{g}, \alpha, \sigma\rangle\langle F_j|\boldsymbol{g}, \alpha, \sigma\rangle \qquad (ij \in \mathcal{N}). \tag{42}$$

By calculating the expectations of both sides of equations (19) and (20) with respect to $\mathcal{P}\{\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*\}$, we obtain the following update rules for the hyperparameters $\alpha(t)$ and $\sigma(t)$:

$$\begin{aligned} 2V_{\mathrm{PR}}^{(0)}(\alpha(t+1)) - 2V_{\mathrm{PR}}^{(1)}(\alpha(t+1)) &= 2V_{\mathrm{PO}}^{(0)}(\alpha(t), \sigma(t)) - 2V_{\mathrm{PO}}^{(1)}(\alpha(t), \sigma(t)) \\ &\quad + \frac{1}{|\mathcal{N}|}\sum_{ij \in \mathcal{N}} (h_i(\boldsymbol{g}, \alpha(t), \sigma(t)) - h_j(\boldsymbol{g}, \alpha(t), \sigma(t)))^2, \end{aligned} \tag{43}$$

$$\sigma(t+1)^2 = V_{\mathrm{PO}}^{(0)}(\alpha(t), \sigma(t)) + \frac{1}{|\Omega|}\sum_{i \in \Omega} (h_i(\boldsymbol{g}, \alpha(t), \sigma(t)) - g_i)^2. \tag{44}$$

We remark that $V_{\mathrm{PR}}^{(0)}(\alpha)$ and $V_{\mathrm{PR}}^{(1)}(\alpha)$ are the solutions $V_{\mathrm{PR}}^{(0)}(\alpha) = V_{ii}$ $(i \in \Omega)$ and $V_{\mathrm{PR}}^{(1)}(\alpha) = V_{ij}$ $(ij \in \mathcal{N})$ of equations (35)–(37) for $\beta = 0$ and that $V_{\mathrm{PO}}^{(0)}(\alpha, \sigma)$ and $V_{\mathrm{PO}}^{(1)}(\alpha, \sigma)$ are the solutions $V_{\mathrm{PO}}^{(0)}(\alpha, \sigma) = V_{ii}$ $(i \in \Omega)$ and $V_{\mathrm{PO}}^{(1)}(\alpha, \sigma) = V_{ij}$ $(ij \in \mathcal{N})$ of equations (35)–(37) for $\beta = \sigma^{-2}$.

In the Bethe approximation, we set $\mathcal{B} = \mathcal{N}$ and $\mathcal{C} = \Omega \cup \mathcal{B}$, and then the Möbius functions are given as $\mu(i) = 3$ $(i \in \Omega)$, $\mu(ij) = -1$ $(ij \in \mathcal{N})$. In the square approximation of GBP, we set $\mathcal{B} = \{ijkl|ij, jk, kl, li \in \mathcal{N}, i, j, k, l \in \Omega\}$, $\mathcal{C} = \Omega \cup \mathcal{N} \cup \mathcal{B}$ and then the Möbius functions are given as $\mu(i) = -1$ $(i \in \Omega)$, $\mu(ij) = 1$ $(ij \in \mathcal{N})$ and $\mu(ijkl) = -1$ $(ijkl \in \mathcal{B})$. If we apply the system of equations (35)–(37) to the posterior and the prior probabilistic density functions, the deterministic equations for $V_{\mathrm{PR}}^{(0)}(\alpha)$, $V_{\mathrm{PR}}^{(1)}(\alpha)$, $V_{\mathrm{PO}}^{(0)}(\alpha, \sigma)$ and $V_{\mathrm{PO}}^{(1)}(\alpha, \sigma)$ are reduced to

$$\begin{pmatrix} V_{\mathrm{PR}}^{(0)}(\alpha) \\ V_{\mathrm{PR}}^{(1)}(\alpha) \end{pmatrix} = \boldsymbol{\Psi}\left(\begin{pmatrix} V_{\mathrm{PR}}^{(0)}(\alpha) \\ V_{\mathrm{PR}}^{(1)}(\alpha) \end{pmatrix} \middle| \alpha, 0 \right), \tag{45}$$

$$\begin{pmatrix} V_{\mathrm{PO}}^{(0)}(\alpha, \sigma) \\ V_{\mathrm{PO}}^{(1)}(\alpha, \sigma) \end{pmatrix} = \boldsymbol{\Psi}\left(\begin{pmatrix} V_{\mathrm{PO}}^{(0)}(\alpha, \sigma) \\ V_{\mathrm{PO}}^{(1)}(\alpha, \sigma) \end{pmatrix} \middle| \alpha, \sigma^{-2} \right), \tag{46}$$

$$\boldsymbol{\Psi}\left(\begin{pmatrix}\xi\\ \eta\end{pmatrix}\middle|\alpha,\beta\right) \equiv \begin{pmatrix}\frac{3}{4\xi}+\frac{1}{4}\beta+\alpha & -\alpha\\ -\alpha & \frac{3}{4\xi}+\frac{1}{4}\beta+\alpha\end{pmatrix}^{-1}\begin{pmatrix}1\\ 0\end{pmatrix}, \tag{47}$$

for the Bethe approximation, and

$$\boldsymbol{\Psi}\left(\begin{pmatrix}\xi\\ \eta\end{pmatrix}\middle|\alpha,\beta\right) \equiv \begin{pmatrix}1&0&0&0\\0&1&0&0\end{pmatrix}\left(\left(\alpha+\frac{1}{4}\beta-\frac{1}{4\xi}+\frac{\xi}{2(\xi^2-\eta^2)}\right)\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\right.$$
$$\left.-\frac{1}{2}\left(\alpha-\frac{\eta}{\xi^2-\eta^2}\right)\begin{pmatrix}0&1&0&1\\1&0&1&0\\0&1&0&1\\1&0&1&0\end{pmatrix}\right)^{-1}\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \tag{48}$$

for the square approximation, respectively. In the case of the Bethe approximation, the deterministic equations for $V_{\mathrm{PR}}^{(0)}(\alpha)$, $V_{\mathrm{PR}}^{(1)}(\alpha)$ can be reduced to

$$\begin{pmatrix}V_{\mathrm{PR}}^{(0)}(\alpha)\\ V_{\mathrm{PR}}^{(1)}(\alpha)\end{pmatrix} = \frac{1}{8\alpha}\begin{pmatrix}3\\ 1\end{pmatrix}. \tag{49}$$

The explicit procedure for the EM algorithm for GBP is as follows:

EM algorithm for GBP

Step 1: Set initial values $\alpha(0)$, $\sigma(0)$ and set $t \leftarrow 0$.
Step 2: Calculate the vector $\boldsymbol{h}(\boldsymbol{g},\alpha(t),\sigma(t))$ as follows:

$$\boldsymbol{h}(\boldsymbol{g},\alpha(t),\sigma(t)) \leftarrow (\boldsymbol{I}+\alpha(t)\sigma(t)^2\boldsymbol{C})^{-1}\boldsymbol{g}. \tag{50}$$

Step 3: Iterate the following procedure until a and b converge:

$$\begin{pmatrix}a\\ b\end{pmatrix} \leftarrow \boldsymbol{\Psi}\left(\begin{pmatrix}a\\ b\end{pmatrix}\middle|\alpha(t),\sigma(t)^{-2}\right). \tag{51}$$

Step 4: Calculate $\sigma(t+1)$ and c through the following rules:

$$\sigma(t+1) \leftarrow \left(a+\frac{1}{|\Omega|}\sum_{i\in\Omega}(h_i(\boldsymbol{g},\alpha(t),\sigma(t))-g_i)^2\right)^{1/2}, \tag{52}$$

$$c \leftarrow a-b+\frac{1}{2|\mathcal{N}|}\sum_{ij\in\mathcal{N}}(h_i(\boldsymbol{g},\alpha(t),\sigma(t))-h_j(\boldsymbol{g},\alpha(t),\sigma(t)))^2. \tag{53}$$

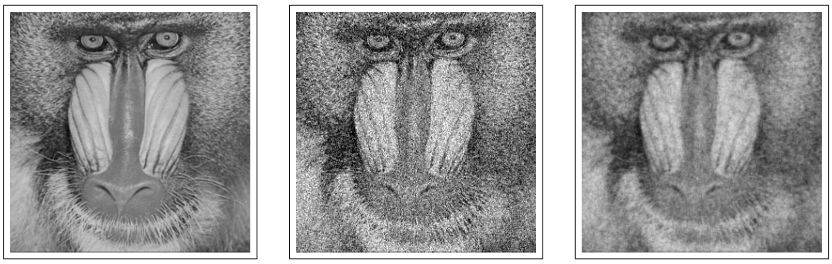
Step 5: Determine $\alpha(t+1)$ so as to satisfy

$$\begin{pmatrix}d\\ d-c\end{pmatrix} = \boldsymbol{\Psi}\left(\begin{pmatrix}d\\ d-c\end{pmatrix}\middle|\alpha(t+1),0\right), \tag{54}$$

and update $t \leftarrow t+1$.
Step 6: Stop if $\alpha(t)$ and $\sigma(t)$ have converged, and go to step 2 otherwise.

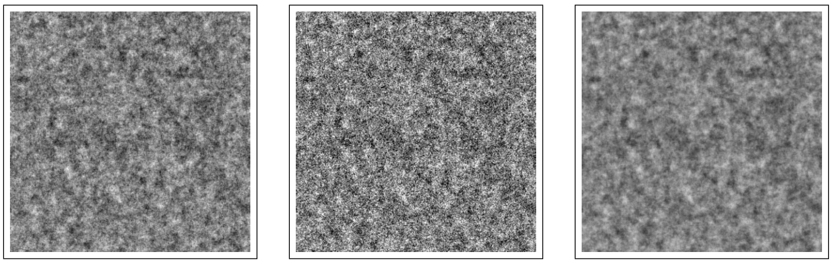
The function $\boldsymbol{\Psi}$ on the right-hand sides of equations (51) and (54) is defined by equations (47) for LBP and by equation (48) for GBP, respectively.

In the present paper, our framework for Bayesian image restoration uses a Gaussian graphical model and considers a continuous random variable for the intensity at each pixel in



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

Figure 1. Image restoration by means of the Gaussian graphical model and the EM algorithm. (*a*) Original image $\boldsymbol{f}$. (*b*) Degraded image $\boldsymbol{g}$ ($\sigma = 40$). (*c*) Restored image $\widehat{\boldsymbol{f}}$ based on generalized belief propagation (GBP) ($\widehat{\alpha} = 0.000\,713$, $\widehat{\sigma} = 37.610$).



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

Figure 2. Image restoration by means of the Gaussian graphical model and the EM algorithm. (*a*) Original image $\boldsymbol{f}$. (*b*) Degraded image $\boldsymbol{g}$ ($\sigma = 40$). (*c*) Restored image $\widehat{\boldsymbol{f}}$ based on generalized belief propagation (GBP) ($\widehat{\alpha} = 0.000\,716$, $\widehat{\sigma} = 39.268$).

the original and the degraded images. However, in practical images in computer vision, the intensity of light at each pixel is represented by an integer chosen from the set $\{0, 1, \ldots, 255\}$. In our numerical experiments, we apply the framework established in the previous sections to images consisting of integers $\{0, 1, \ldots, 255\}$. Instead of equation (6), we use

$$\widehat{f}_i \equiv \arg \min_{n=0,1,\ldots,255} \left(n - \int z_i \mathcal{P}(\boldsymbol{F} = \boldsymbol{z} | \boldsymbol{G} = \boldsymbol{g}, \widehat{\alpha}, \widehat{\sigma})\, \mathrm{d}\boldsymbol{z} \right)^2. \tag{55}$$

In our numerical experiments, we adopt the two images in figures 1(*a*) and 2(*a*) as original images $\boldsymbol{f}$. Figure 1(*a*) shows a standard image. Figure 2(*a*) is an image generated by sampling by the Markov chain Monte Carlo method from the *a priori* probability density function (2). Degraded images in figures 1(*a*) and 2(*a*) are generated by corrupting the original images by adding white Gaussian noise with mean 0 and variance 40^2. We apply the EM algorithms associated with LBP and GBP to the degraded images $\boldsymbol{g}$ shown in figures 1(*b*) and 2(*b*).

For the degraded images in figures 1(*b*) and 2(*b*), the trajectories of $(\sigma(t), \alpha(t))$ $(t = 0, 1, 2, \ldots)$ in the EM algorithms with LBP and GBP are given in figures 3 and 4, respectively. The initial values $\sigma(0)$ and $\alpha(0)$ in the EM algorithm were 100 and 0.000 10, respectively. The small solid circle in both panel (*a*) and panel (*b*) is the trajectory for the exact solution obtained from the update rules (19) and (20) for $(\alpha(t), \sigma(t))$. The restored images obtained by applying the EM algorithm for GBP to the degraded images $\boldsymbol{g}$ in figures 1(*b*) and 2(*b*) are shown in figures 1(*c*) and 2(*c*), respectively.

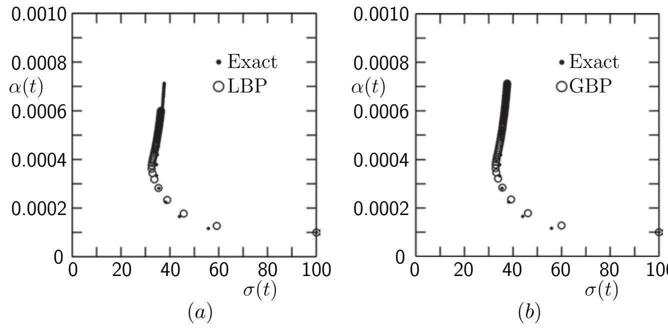


Figure 3. Values of $(\sigma(t), \alpha(t))$ $(t = 0, 1, 2, \ldots)$ in the EM algorithms associated with the degraded image in figure 1(*b*). The initial values $\sigma(0)$ and $\alpha(0)$ in the EM algorithm are 100 and 0.000 10, respectively. (*a*) The open circle is for loopy belief propagation (LBP). (*b*) The open circle is for generalized belief propagation (GBP). The small solid circle in both (*a*) and (*b*) is for the exact solution obtained by using the update rules (19) and (20) for $(\alpha(t), \sigma(t))$.

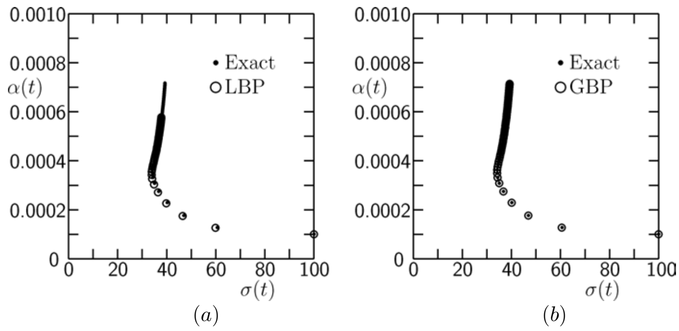


Figure 4. Values of $(\sigma(t), \alpha(t))$ $(t = 0, 1, 2, \ldots)$ in the EM algorithms associated with the degraded image in figure 2(*b*). The initial values $\sigma(0)$ and $\alpha(0)$ are 100 and 0.000 10, respectively. (*a*) The open circle is for loopy belief propagation (LBP). (*b*) The open circle is for generalized belief propagation (GBP). The small solid circle in both (*a*) and (*b*) is for the exact solution obtained by using the update rules (19) and (20) for $(\alpha(t), \sigma(t))$.

6. Statistical trajectory of EM algorithm for GBP

In this section, we give the general procedure for calculating the statistical trajectory for the EM algorithm in probabilistic image processing based on the Gaussian graphical model. The statistical trajectory corresponds to the average of equations (43) and (44) with respect to degraded images $\boldsymbol{g}$. Now we assume that degraded images $\boldsymbol{g}$ are generated according to $\mathcal{P}\{\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*\}$.

By calculating expectations of both sides of equations (19) and (20) with respect to $\mathcal{P}\{\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*\}$, we obtain the following update rules for the hyperparameters $\alpha(t)$ and $\sigma(t)$:

$$2V_{\mathrm{PR}}^{(0)}(\alpha(t+1)) - 2V_{\mathrm{PR}}^{(1)}(\alpha(t+1)) = 2V_{\mathrm{PO}}^{(0)}(\alpha(t), \sigma(t)) - 2V_{\mathrm{PO}}^{(1)}(\alpha(t), \sigma(t)) + \frac{1}{|\mathcal{N}|}\sum_{ij\in\mathcal{N}}\int (h_i(\boldsymbol{g}, \alpha(t), \sigma(t)) - h_j(\boldsymbol{g}, \alpha(t), \sigma(t)))^2 \mathcal{P}\{\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*\}\,\mathrm{d}\boldsymbol{g}, \tag{56}$$

$$\sigma(t+1)^2 = V_{\mathrm{PO}}^{(0)}(\alpha(t), \sigma(t)) + \frac{1}{|\Omega|}\sum_{i\in\Omega}\int (h_i(\boldsymbol{g}, \alpha(t), \sigma(t)) - g_i)^2 \mathcal{P}\{\boldsymbol{G} = \boldsymbol{g}|\alpha^*, \sigma^*\}\,\mathrm{d}\boldsymbol{g}. \tag{57}$$

By substituting equation (17) into equations (56) and (57), we obtain

$$V_{\mathrm{PR}}^{(0)}(\alpha(t+1)) - V_{\mathrm{PR}}^{(1)}(\alpha(t+1)) = V_{\mathrm{PO}}^{(0)}(\alpha(t), \sigma(t)) - V_{\mathrm{PO}}^{(1)}(\alpha(t), \sigma(t)) + \frac{1}{|\mathcal{N}|}\mathrm{Tr}\,\frac{\boldsymbol{I} + \alpha^*\sigma^{*2}\boldsymbol{C}}{\alpha^*(\boldsymbol{I} + \alpha\sigma^2\boldsymbol{C})^2}, \tag{58}$$

$$\sigma(t+1)^2 = V_{\mathrm{PO}}^{(0)}(\alpha(t), \sigma(t)) + \frac{1}{|\Omega|}\mathrm{Tr}\,\frac{\alpha^2(\boldsymbol{I} + \alpha^*\sigma{*}^2\boldsymbol{C})}{\alpha^*(\boldsymbol{I} + \alpha\sigma^2\boldsymbol{C})^2}. \tag{59}$$

We remark that $V_{\mathrm{PR}}^{(0)}(\alpha)$ and $V_{\mathrm{PR}}^{(1)}(\alpha)$ are the solutions $V_{\mathrm{PR}}^{(0)}(\alpha) = V_{ii}$ $(i \in \Omega)$ and $V_{\mathrm{PR}}^{(1)}(\alpha) = V_{ij}$ $(ij \in \mathcal{N})$ of equations (35)–(37) for $\beta = 0$ and that $V_{\mathrm{PO}}^{(0)}(\alpha, \sigma)$ and $V_{\mathrm{PO}}^{(1)}(\alpha, \sigma)$ are the solutions $V_{\mathrm{PO}}^{(0)}(\alpha, \sigma) = V_{ii}$ $(i \in \Omega)$ and $V_{\mathrm{PO}}^{(1)}(\alpha, \sigma) = V_{ij}$ $(ij \in \mathcal{N})$ of equations (35)–(37) for $\beta = \sigma^{-2}$.

From the deterministic equations (45)–(48) for $V_{\mathrm{PR}}^{(0)}(\alpha)$, $V_{\mathrm{PO}}^{(0)}(\alpha, \sigma)$, $V_{\mathrm{PR}}^{(1)}(\alpha)$ and $V_{\mathrm{PO}}^{(1)}(\alpha, \sigma)$, and the update rule of the EM algorithm (58) and (59), the procedure for estimating the statistical trajectory in the EM algorithm in the GBP is summarized as follows.

Statistical trajectory estimation algorithm of EM algorithm for GBP

Step 1: Set initial values $\alpha(0)$, $\sigma(0)$ and set $t \leftarrow 0$.
Step 2: Iterate the following procedure until a and b converge:

$$\begin{pmatrix} a \\ b \end{pmatrix} \leftarrow \boldsymbol{\Psi}\left(\begin{pmatrix} a \\ b \end{pmatrix}\middle|\,\alpha(t), \sigma(t)^{-2}\right). \tag{60}$$

Step 3: Calculate $\sigma(t+1)$ and c through the following rules:

$$\sigma(t+1) \leftarrow \left(a + \frac{1}{|\Omega|}\mathrm{Tr}\,\frac{\alpha(t)^2(\boldsymbol{I} + \alpha^*\sigma^{*2}\boldsymbol{C})}{\alpha^*(\boldsymbol{I} + \alpha(t)\sigma(t)^2\boldsymbol{C})^2}\right)^{1/2}, \tag{61}$$

$$c \leftarrow a - b + \frac{1}{2|\mathcal{N}|}\mathrm{Tr}\,\frac{\boldsymbol{I} + \alpha^*\sigma^{*2}\boldsymbol{C}}{\alpha^*(\boldsymbol{I} + \alpha(t)\sigma(t)^2\boldsymbol{C})^2}. \tag{62}$$

Step 4: Determine $\alpha(t+1)$ so as to satisfy

$$\begin{pmatrix} d \\ d - c \end{pmatrix} = \boldsymbol{\Psi}\left(\begin{pmatrix} d \\ d - c \end{pmatrix}\middle|\,\alpha(t+1), 0\right), \tag{63}$$

and update $t \leftarrow t + 1$.
Step 5: Stop if $\alpha(t)$ and $\sigma(t)$ have converged, and go to step 2 otherwise.

The function $\boldsymbol{\Psi}$ on the right-hand sides of equations (60) and (61) is defined by equations (47) for LBP and by equation (48) for GBP, respectively.

We give the statistical trajectories in the EM algorithms for LBP, GBP and the exact solution in table 1. The statistical trajectories for LBP and GBP are obtained by the procedure described above in this section, while that for the exact solution is calculated by means of the update rules (15) and (16). The true values (σ^*, α^*) for hyperparameters σ and α are $(40, 0.000\,70)$. In figures 5 and 6, the initial values $(\sigma(0), \alpha(0))$ in the EM algorithm

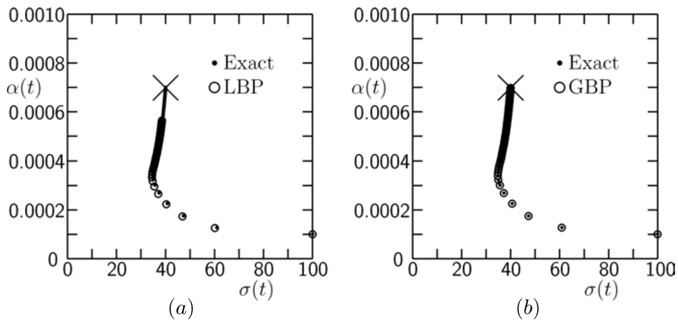


Figure 5. Statistical averages of $(\sigma(t), \alpha(t))$ $(t = 0, 1, 2, \ldots)$ in the EM algorithms. The true values σ^* and α^* for hyperparameters σ and α are 40 and 0.000 70, respectively. The initial values $\sigma(0)$ and $\alpha(0)$ in the EM algorithm are 100 and 0.000 10, respectively. (*a*) The open circle is for loopy belief propagation (LBP). (*b*) The open circle is for generalized belief propagation (GBP). The small solid circle in both (*a*) and (*b*) is for the exact solution.

Table 1. Statistical averages of $(\sigma(t), \alpha(t))$ $(t = 0, 1, 2, \ldots)$ in the EM algorithms for the exact solution, loopy belief propagation (LBP) and generalized belief propagation (GBP). The true values σ^* and α^* for hyperparameters σ and α are 40 and 0.000 70, respectively. (a) $\sigma(0) = 100, \alpha(0) = 0.000\,10$. (b) $\sigma(0) = 100, \alpha(0) = 0.000\,90$.

	Exact		LBP		GBP	
t	$\alpha(t)$	$\sigma(t)$	$\alpha(t)$	$\sigma(t)$	$\alpha(t)$	$\sigma(t)$
			(a)			
0	0.000 1000	100.00	0.000 1000	100.00	0.000 1000	100.00
1	0.000 1274	60.88	0.000 1260	60.18	0.000 1274	60.87
2	0.000 1748	47.31	0.000 1729	46.90	0.000 1748	47.31
5	0.000 3008	35.64	0.000 2961	35.39	0.000 3007	35.64
10	0.000 3749	35.24	0.000 3641	34.89	0.000 3749	35.24
20	0.000 4646	37.23	0.000 4364	36.61	0.000 4645	37.23
50	0.000 6213	39.31	0.000 5357	38.20	0.000 6207	39.30
100	0.000 6885	39.91	0.000 5620	38.52	0.000 6870	39.90
200	0.000 6998	40.00	0.000 5640	38.54	0.000 6979	39.98
500	0.000 7000	40.00	0.000 5640	38.54	0.000 6980	39.98
			(b)			
0	0.000 9000	100.00	0.000 9000	100.00	0.000 9000	100.00
1	0.000 9289	49.69	0.000 9190	48.62	0.000 9280	49.59
2	0.000 9520	43.59	0.000 9286	42.99	0.000 9506	43.55
5	0.000 9376	41.46	0.000 8833	40.94	0.000 9356	41.44
10	0.000 8941	41.20	0.000 8044	40.48	0.000 8917	41.19
20	0.000 8295	40.86	0.000 7013	39.81	0.000 8267	40.84
50	0.000 7393	40.29	0.000 5910	38.84	0.000 7367	40.27
100	0.000 7056	40.04	0.000 5659	38.56	0.000 7035	40.03
200	0.000 7001	40.00	0.000 5640	38.54	0.000 6983	39.98
500	0.000 7000	40.00	0.000 5640	38.54	0.000 6983	39.98

are (100, 0.000 10) and (100, 0.000 90), respectively. The statistical trajectories in the EM algorithms for LBP and GBP in table 1 are shown as the open circles in figures 5 and 6, whereas the statistical trajectory for the exact solution in table 1 is given by the small solid circles.

In table 1 and figures 5 and 6, we see that the trajectories for the approximate EM algorithms for both LBP and GBP are very close to that for the exact EM algorithm. However,

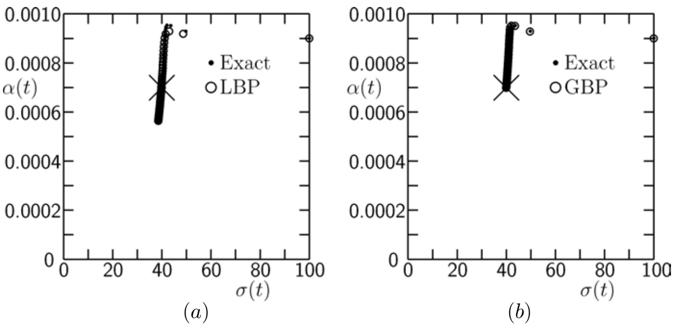


Figure 6. Statistical averages of $(\sigma(t), \alpha(t))$ $(t = 0, 1, 2, \ldots)$ in the EM algorithms. The true values σ^* and α^* for hyperparameters σ and α are 40 and 0.000 70, respectively. The initial values $\sigma(0)$ and $\alpha(0)$ in the EM algorithm are 100 and 0.000 90, respectively. (a) The open circle is for loopy belief propagation (LBP). (b) The open circle is for generalized belief propagation (GBP). The small solid circle in both (a) and (b) is for the exact solution.

the termination points for $(\alpha(t), \sigma(t))$ for both of the approximate EM algorithms are different from the true point (α^*, σ^*) in the hyperparameter space, whereas the exact EM algorithm can reach the true point. Moreover, we see that the termination point of the approximate EM algorithm of GBP is closer to the true point than is the termination point corresponding to LBP in table 1 and then we conclude that GBP gives us better estimation of the hyperparameter by means of the EM algorithm than the does LBP based on the analysis of the statistical trajectory as well as on some numerical experiments described in figures 1–4 in section 5.

7. Concluding remarks

In the present paper, we have extended the EM algorithm for conventional LBP to that for GBP and have described some numerical experiments. The prior probability model has been assumed to correspond to a Gaussian graphical model. Additive white Gaussian noise has been adopted as the degradation process. We have derived recursion formulas for calculating the averages of the trajectories in the EM algorithms for GBP with respect to degraded images generated from the assumed prior probabilities and degradation process. Our main result in the present paper was the derivation of the statistical average of the trajectory as recursion formulas analytically. The results have been compared with the average for the exact solutions obtained by using the multi-dimensional Gaussian integral formula.

The present results have shown that the analytical results of the EM algorithms in section 6 are consistent with the numerical experiments in section 5 for GBP as well as for the exact solution. Moreover, we see that results of GBP and the exact solution are very close to each other for several initial steps for the analytical results as well as for our numerical experiments for degraded images generated according to the assumed prior probability density function and degradation process.

The approximate EM algorithm by means of GBP cannot reach the true point of hyperparameter space (α^*, σ^*). The terminated point of the EM algorithm is closer to the true point than those of LBP. The similar situations are found also in some numerical experiments in section 5. Thus we see that the hyperparameter estimation of GBP is better than LBP from the analysis of the statistical trajectory. We conclude that LBP and GBP differ in terms of the accuracy of estimation of the hyperparameter. On the other hand, in terms of speed, the difference between them is not substantial. The computational time needed for convergence of the approximate EM algorithm by means of GBP is usually at most double that for version based on LBP in the image processing context.

The EM algorithm for estimating the hyperparameters originally included the calculation of statistical quantities for posterior and prior probabilities. In the present paper, we treat only Gaussian graphical models and expressions for the appropriate statistical quantities can be derived analytically. Hence we have constructed the exact EM algorithm and can analyze the accuracy of the approximate EM algorithms of LBP and GBP. However, posterior and prior probabilistic models used in practical image processing are generally intractable and it is usually hard to calculate the corresponding statistical quantities exactly. In such cases, we do have to employ LBP or GBP as approximate algorithms for calculating statistical quantities for intractable probabilistic models.

Moreover, when we apply GBP and LBP to intractable posterior and prior probabilistic models, we have to consider trajectories and convergence not only for the approximate EM algorithm but also for GBP and LBP themselves. In our approximate EM algorithms, GBP and LBP correspond to inner loops and the EM procedure corresponds to an outer loop. Thus we have to consider the convergence both of LBP or GBP and of the EM procedure. It is one of our future research problems to improve the iterative procedures of the inner and outer loops in approximate EM algorithms.

We have investigated the analytical estimation of the statistical averages of the trajectories in the approximate and the exact EM algorithms only for the square lattice because of our particular interest in probabilistic image processing. Moreover, we treated only the case in which a real value in the range $(-\infty, +\infty)$ is taken as the intensity at each pixel. Practical digital images often take integer numbers as the intensity. For some such cases, Inoue and Tanaka have analyzed the statistical average of trajectories in the EM algorithm on complete graphs by using the replica method [19, 20]. We expect that it is possible to carry out similar analysis to that in [19, 20] for probabilistic models on sparse graphs with a finite coordination number. This is left for future research.

Belief propagation has been applied to many problems in computer science. For several problems in computer science, we have to consider probabilistic models on other types of graph. In particular, some researchers are interested in information processing on complex networks [27–29]. It is of interest to do similar analysis to that in the present paper in the context of Gaussian graphical models on complex networks. This also is left for future research.

Acknowledgments

The authors are grateful to Miss Marika Minagawa of the Graduate School of Information Sciences, Tohoku University, for valuable discussions. This work was partly supported by the Grants-In-Aid (no 18079002) for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Frey B J 1998 *Graphical Models for Machine Learning and Digital Communication* (Cambridge, MA: MIT Press)
[2] Jordan M I (ed) 1999 *Learning in Graphical Models* (Cambridge, MA: MIT Press)
[3] Mézard M, Parisi G and Zecchina R 2002 *Science* **297** 812
[4] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)
[5] Kabashima Y and Saad D 1998 *Europhys. Lett.* **44** 668

[6] Opper M and Saad D (ed) 2001 *Advanced Mean Field Methods—Theory and Practice* (Cambridge, MA: MIT Press)
[7] Yedidia J S, Freeman W T and Weiss Y 2001 *Advances in Neural Information Processing Systems* vol 13 (Cambridge, MA: MIT Press) p 689
[8] Yedidia J S, Freeman W T and Weiss Y 2005 *IEEE Trans. Inform. Theory* **51** 2282
[9] Pelizzola A 2005 *J. Phys. A: Math. Gen.* **38** R309 (topical review)
[10] Freeman W T, Jones T R and Pasztor E C 2002 *IEEE Comput. Graph. Appl.* **22** 56
[11] Tanaka K 2002 *J. Phys. A: Math. Gen.* **35** R81 (topical review)
[12] Willsky A S 2002 *Proc. IEEE* **90** 1396
[13] Tanaka K, Shouno H, Okada M and Titterington D M 2004 *J. Phys. A: Math. Gen.* **37** 8675
[14] Tanaka K 2005 *IEICE Trans. Inform. Syst.* J **88-D-II** 2368
[15] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing—An Introduction* (Oxford: Oxford University Press)
[16] Kabashima Y and Saad D 2004 *J. Phys. A: Gen. Phys.* **37** R1 (topical review)
[17] Martin O C, Monasson R and Zecchina R 2001 *Theor. Comput. Sci.* **265** 3
[18] Nishimori H and Wong K Y M 1999 *Phys. Rev.* E **60** 132
[19] Inoue J and Tanaka K 2002 *Phys. Rev.* E **65** 016125
[20] Inoue J and Tanaka K 2003 *J. Phys. A: Math. Gen.* **36** 10997
[21] Weiss Y 2000 *Neural Comput.* **12** 1
[22] Weiss Y and Freeman W T 2001 *Neural Comput.* **13** 2173
[23] Kschischang F R, Frey R J and Loeliger H-A 2001 *IEEE Trans. Inform. Theory* **47** 498
[24] Ikeda S, Tanaka T and Amari S 2004 *Neural Comput.* **16** 1779
[25] Heskes T 2004 *Neural Comput.* **16** 2379
[26] Tanaka K and Inoue J 2002 *IEICE Trans. Inform. Syst.* **E-85D** 546
[27] Mooij J M and Kappen H J 2005 *Advances in Neural Information Processing Systems* vol 17 (Cambridge, MA: MIT Press) p 945
[28] Ohkubo J, Yasuda M and Tanaka K 2005 *Phys. Rev.* E **72** 046135
[29] Wemmenhove B, Nikoletopoulos T and Hatchett J P L 2005 *J. Stat. Mech.* **11** 11007