

© 2002 The Author(s). Originally published in *Physical Review E*. Also available on arXiv.

Statistical Mechanics of the Bayesian Image Restoration under Spatially Correlated Noise

Jun Tsuzurugi^{†‡*} and Masato Okada[‡]

[†]*Nara Institute of Science and Technology,
Takayama-cho 8916-5, Ikoma-shi, Nara, 630-0101 Japan*

[‡]*RIKEN Brain Science Institute Hirosawa 2-1,
Wakou-shi, Saitama, 351-0456 Japan*

Abstract

We investigated the use of the Bayesian inference to restore noise-degraded images under conditions of spatially correlated noise. The generative statistical models used for the original image and the noise were assumed to obey multi-dimensional Gaussian distributions whose covariance matrices are translational invariant. We derived an exact description to be used as the expectation for the restored image by the Fourier transformation and restored an image distorted by spatially correlated noise by using a spatially uncorrelated noise model. We found that the resulting hyperparameter estimations for the minimum error and maximal posterior marginal criteria did not coincide when the generative probabilistic model and the model used for restoration were in different classes, while they did coincide when they were in the same class.

PACS numbers: 07.05.Mh, 87.80.Xa, 87.18.Sn

I. INTRODUCTION

Recent research has shown that the Bayesian inference is a useful approach to the image restoration problem [1]-[3]. However, previous research has usually assumed that the superimposed noise is not correlated between pixels [1]-[12]. Statistical mechanics provides a useful method for analyzing the image restoration problem [4],[5]. Still, when we take optical effects into consideration, it seems natural to take spatially correlated noise into account. We have investigated image restoration under the condition of spatially correlated noise. We considered a case in which both the original image and noise obey a Gaussian distribution composed of translational-symmetric matrices. We were able to diagonalize the Gaussian distribution by using Fourier transformation. While previous research was done using spatially uncorrelated noise, we considered the availability of a spatially uncorrelated noise model against spatially correlated noise.

We considered two methods for estimating the hyperparameters. The first is minimizing the mean squared error - the hyperparameters are set so that the mean squared error is minimized. However, minimizing the mean squared error is not applicable to the image restoration problem because we need the original image to practice this method. The second

* juntuzu@brain.riken.go.jp

method is maximizing the marginal likelihood [6] - the hyperparameters are estimated by maximizing the marginal likelihood acquired from only the distorted image. When the noise probabilities for generating and restoring coincidence are the same, the results given by the two methods are the same. When the probabilities are different, we found that they are not.

II. MODEL

The images we see in our daily life are always two-dimensional, although we assume a d -dimensional image for mathematical generalization, where the length of each side is L and the total number of pixels, N , is L^d . The i -th pixel value, ξ_i , of an original image can be written in a multiple Gaussian distribution as

$$P(\boldsymbol{\xi}) = \frac{1}{Z_{\text{prior}}(\beta, h)} \exp[-H(\boldsymbol{\xi}, \beta, h)], \quad (1)$$

$$H(\boldsymbol{\xi}, \beta, h) = \boldsymbol{\xi}^T(\beta G + hI)\boldsymbol{\xi}, \quad (2)$$

where β and h are positive scalar values, G is a matrix, and I is the unit matrix. The partition function, $Z_{\text{prior}}(\beta, h)$, is given by

$$Z_{\text{prior}}(\beta, h) = (2\pi)^{\frac{N}{2}} |\beta G + hI|^{-\frac{1}{2}}. \quad (3)$$

Note that Eq. (2) is called the Hamiltonian in statistical mechanics. In this paper, we argue that matrix G is translational invariant and that element $G_{\mathbf{i}, \mathbf{j}}$ is given by [5]

$$G_{\mathbf{i}, \mathbf{j}} = 2d\delta_{(\mathbf{i}-\mathbf{j}), \mathbf{0}} - \sum_{\boldsymbol{\delta}} \delta_{(\mathbf{i}-\mathbf{j}), -\boldsymbol{\delta}} - \sum_{\boldsymbol{\delta}} \delta_{(\mathbf{i}-\mathbf{j}), \boldsymbol{\delta}}, \quad (4)$$

where the d -dimensional vector $\boldsymbol{\delta}$ is

$$\boldsymbol{\delta} = (1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1). \quad (5)$$

The $\delta_{\mathbf{i}, \mathbf{j}}$ in Eq. (4) is the Kronecker delta, which is assumed to be

$$\delta_{\mathbf{i}, \mathbf{j}} = \begin{cases} 1 & (\mathbf{i} = \mathbf{j}) \\ 0 & (\mathbf{i} \neq \mathbf{j}). \end{cases} \quad (6)$$

Equation (4) means that interactions occur only between nearest-neighbor pixels. If β takes a large value, i.e., the interaction of the first term has a large effect, the neighboring pixels

tend to take the same value. When h is large, the absolute value of the pixels tends to be small. Furthermore, we use a periodic boundary condition,

$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_{i+L}\boldsymbol{\delta}. \quad (7)$$

The distorted image, $\boldsymbol{\tau} = \{\tau_i\}$, is generated according to the following conditional probability, which obeys a Gaussian distribution:

$$P_{\text{out}}(\boldsymbol{\tau}|\boldsymbol{\xi}) = \frac{1}{Z_{\text{noise}}} \exp \left[-\frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\xi})^T R^{-1}(\boldsymbol{\tau} - \boldsymbol{\xi}) \right], \quad (8)$$

$$Z_{\text{noise}}(R) = (2\pi)^{\frac{N}{2}} |R|^{\frac{1}{2}}, \quad (9)$$

$$R_{i,j} = (1-a)b^2 \delta_{i,j} + ab^2 \exp \left[-\frac{d_{i,j}^2}{\kappa^2} \right], \quad (10)$$

where $0 \leq a \leq 1$ and $d_{i,j}$ denotes the distance on a d -dimensional lattice of linear size L with periodic boundary conditions,

$$d_{i,j}^2 = \sum_{\mu=1}^d \min(|i_\mu - j_\mu|, L - |i_\mu - j_\mu|)^2. \quad (11)$$

where $0 \leq a \leq 1$. When $a = 0$ the noise reduces to i.i.d. Gaussian noise with variance b^2 .

$0 \leq a \leq 1$. The noise obeys an identical independent Gaussian distribution with mean 0 and variance b^2 , when $a = 0$.

A general strategy commonly used in image restoration is to apply the Bayes formulation to the posterior probability. Here, $\boldsymbol{\sigma}$ is a restoration image based on the Bayes formulation. When a distorted image, $\boldsymbol{\tau}$, is given, one can use the formulation to calculate the restored image, $\boldsymbol{\sigma}$:

$$P(\boldsymbol{\sigma}|\boldsymbol{\tau}) = \frac{P_{\text{out}}(\boldsymbol{\tau}|\boldsymbol{\sigma})P(\boldsymbol{\sigma})}{\int d\boldsymbol{\sigma} P_{\text{out}}(\boldsymbol{\tau}|\boldsymbol{\sigma})P(\boldsymbol{\sigma})} = \frac{\exp[-H_{\text{eff}}]}{\int d\boldsymbol{\sigma} \exp[-H_{\text{eff}}]}, \quad (12)$$

where

$$H_{\text{eff}} = \boldsymbol{\sigma}^T(\beta G + hI)\boldsymbol{\sigma} + \frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\sigma})^T R^{-1}(\boldsymbol{\tau} - \boldsymbol{\sigma}), \quad (13)$$

is an effective Hamiltonian.

III. THEORY

A. Expectation of the restored image

Since the covariance matrices given by Eqs. (4) and (10) are translational invariant, they can be diagonalized using the discrete Fourier transformation.

The discrete Fourier transformation is

$$\tilde{\xi}_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{j}} \xi_{\mathbf{j}} e^{-i\mathbf{k} \cdot \mathbf{j}}, \quad (14)$$

and the inverse Fourier transformation is

$$\xi_{\mathbf{j}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \tilde{\xi}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{j}}, \quad (15)$$

where i is the imaginary unit, and \mathbf{k} is a d -dimensional vector with the same as \mathbf{j} , and the degree of freedom is L^d .

$$\sum_{\mathbf{j}} = L^d = N, \quad \sum_{\mathbf{k}} = L^d = N. \quad (16)$$

Moreover, each component of \mathbf{k} takes the value

$$0, \frac{2}{L}\pi, \frac{4}{L}\pi, \dots, \frac{2(L-1)}{L}\pi. \quad (17)$$

We can diagonalize Eq. (1) by using the Fourier representation:

$$P(\boldsymbol{\xi}) = \frac{1}{Z_{\text{prior}}(\beta, h)} \exp \left[- \sum_{\mathbf{k}} (\beta \tilde{G}_{\mathbf{k}} + h) \tilde{\xi}_{\mathbf{k}} \tilde{\xi}_{-\mathbf{k}} \right], \quad (18)$$

$$\tilde{G}_{\mathbf{k}} = \sum_{\boldsymbol{\delta}} [2 - 2 \cos(\mathbf{k} \cdot \boldsymbol{\delta})]. \quad (19)$$

Furthermore, we diagonalize $R_{\mathbf{i}, \mathbf{j}}$ in Eq. (10):

$$\tilde{R}_{\mathbf{k}} = (1-a)b^2 + ab^2 \sum_{\mathbf{l}} e^{-\frac{l^2}{\kappa^2}} \cos(\mathbf{k} \cdot \mathbf{l}), \quad (20)$$

where the range of l , which is a component of vector \mathbf{l} , is given by

$$-(L-1) \leq l \leq L-1. \quad (21)$$

We can execute the Fourier transformation on the inside of exp in Eq. (8) in the same way as in Eq. (1).

$$P_{out}(\boldsymbol{\tau}|\boldsymbol{\xi}) = \frac{1}{Z_{noise}} \times \exp \left[-\frac{1}{2} \sum_{\mathbf{k}} \tilde{R}_{\mathbf{k}}^{-1} (\tilde{\boldsymbol{\tau}}_{\mathbf{k}} - \tilde{\boldsymbol{\xi}}_{\mathbf{k}}) (\tilde{\boldsymbol{\tau}}_{-\mathbf{k}} - \tilde{\boldsymbol{\xi}}_{-\mathbf{k}}) \right]. \quad (22)$$

We define $\langle \cdot \rangle$ as denoting the thermal average based on the Bayes formula for the posterior probability $P(\boldsymbol{\sigma}|\boldsymbol{\tau})$ in Eq. (12). The expectation of the restored image, $\sigma_{\mathbf{j}}$, is thus assumed to be

$$\langle \sigma_{\mathbf{j}} \rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \langle \tilde{\sigma}_{\mathbf{k}} \rangle e^{i\mathbf{k} \cdot \mathbf{j}}. \quad (23)$$

Using Eq. (12), the expectation of the Fourier component, $\sigma_{\mathbf{k}}$, can be written as

$$\langle \tilde{\sigma}_{\mathbf{k}} \rangle = \int \prod_{\mathbf{k}'} d\tilde{\sigma}_{\mathbf{k}'} \tilde{\sigma}_{\mathbf{k}} P(\boldsymbol{\sigma}|\boldsymbol{\tau}) = \frac{\int \prod_{\mathbf{k}'} d\tilde{\sigma}_{\mathbf{k}'} \tilde{\sigma}_{\mathbf{k}} e^{-\hat{H}_{eff}}}{\int \prod_{\mathbf{k}'} d\tilde{\sigma}_{\mathbf{k}'} e^{-\hat{H}_{eff}}}, \quad (24)$$

where \hat{H}_{eff} is obtained by applying the Fourier transformation to Eq. (13) as follows.

$$\begin{aligned} \hat{H}_{eff} &= \sum_{\mathbf{k}} (\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h}) \tilde{\sigma}_{\mathbf{k}} \tilde{\sigma}_{-\mathbf{k}} \\ &+ \sum_{\mathbf{k}} \frac{1}{2\hat{R}_{\mathbf{k}}} (\tilde{\boldsymbol{\tau}}_{\mathbf{k}} - \tilde{\sigma}_{\mathbf{k}}) (\tilde{\boldsymbol{\tau}}_{-\mathbf{k}} - \tilde{\sigma}_{-\mathbf{k}}). \end{aligned} \quad (25)$$

β , h , and $\tilde{R}_{\mathbf{k}}$ are unknown adjustable parameters that determine the properties of the original image and noise. One of our major focus here is how to estimate these hyperparameters precisely.

To proceed further, we have to assume some explicit form of the source prior and noise posterior formations, which are used when we restore the image. In this paper, we define both of these probability formations as the same formation except that they have different hyperparameters. We define the hyperparameters that correspond to β , h , and $\tilde{R}_{\mathbf{k}}$ as $\hat{\beta}$, \hat{h} , and $\hat{R}_{\mathbf{k}}$, respectively. If $\hat{R}_{\mathbf{k}}$ is independent of \mathbf{k} , the noise is spatially uncorrelated.

Substituting

$$\hat{A}_{\mathbf{k}} = \hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{R}_{\mathbf{k}}}, \quad \hat{B}_{\mathbf{k}} = \frac{1}{2\hat{R}_{\mathbf{k}}} \quad (26)$$

into Eqs. (24) and (25), we get

$$\langle \tilde{\sigma}_{\mathbf{k}} \rangle = \frac{\hat{B}_{\mathbf{k}}}{\hat{A}_{\mathbf{k}}} \tilde{\tau}_{\mathbf{k}}. \quad (27)$$

Consequently, $\langle \sigma_{\mathbf{j}} \rangle$ in Eq. (23) is given by

$$\langle \sigma_{\mathbf{j}} \rangle = \frac{1}{N} \sum_{\mathbf{k}} \sum_{\mathbf{i}} \frac{\tau_{\mathbf{i}} \frac{1}{2\hat{R}_{\mathbf{k}}} \cos[\mathbf{k} \cdot (\mathbf{j} - \mathbf{i})]}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{R}_{\mathbf{k}}}}. \quad (28)$$

In this paper, we regard Eq. (28) as representing the restored image.

B. Minimization of mean squared error

In this subsection, we estimate hyperparameters $\hat{\beta}$, \hat{h} , and $\hat{R}_{\mathbf{k}}$ by using the criterion defined by Eq. (28) for minimizing the mean squared error between the original and restored images. The expectation of this mean squared error, E_1 , is represented by

$$E_1 = \left\| \frac{1}{N} \sum_{\mathbf{j}} (\xi_{\mathbf{j}} - \langle \sigma_{\mathbf{j}} \rangle)^2 \right\|, \quad (29)$$

where $\|\cdot\|$ denotes the data average of a simultaneous distribution

$$P(\boldsymbol{\tau}, \boldsymbol{\xi}) = P_{out}(\boldsymbol{\tau}|\boldsymbol{\xi})P_s(\boldsymbol{\xi}).$$

Applying the Fourier transformation to Eq. (29), we can derive the next representation.

$$E_1 = \frac{1}{N} \sum_{\mathbf{k}} \left\| (\tilde{\xi}_{\mathbf{k}} - \langle \tilde{\sigma}_{\mathbf{k}} \rangle) (\tilde{\xi}_{-\mathbf{k}} - \langle \tilde{\sigma}_{-\mathbf{k}} \rangle) \right\|. \quad (30)$$

Since $P(\boldsymbol{\tau}, \boldsymbol{\xi})$ is diagonalized by the Fourier transformation, we can easily calculate each \mathbf{k}

as follows,

$$\left\| \left(\tilde{\xi}_{\mathbf{k}} - \langle \tilde{\sigma}_{\mathbf{k}} \rangle \right) \left(\tilde{\xi}_{-\mathbf{k}} - \langle \tilde{\sigma}_{-\mathbf{k}} \rangle \right) \right\| \quad (31)$$

$$\begin{aligned} &= \frac{1}{Z} \int \int d\tilde{\xi}_{\mathbf{k}} d\tilde{\tau}_{\mathbf{k}} \left| \tilde{\xi}_{\mathbf{k}} - \frac{\hat{B}_{\mathbf{k}}}{\hat{A}_{\mathbf{k}}} \tilde{\tau}_{\mathbf{k}} \right|^2 \\ &\quad \times \exp \left[-A_{\mathbf{k}} \left| \tilde{\xi}_{\mathbf{k}} - \frac{\tilde{B}_{\mathbf{k}}}{\tilde{A}_{\mathbf{k}}} \tilde{\tau}_{\mathbf{k}} \right|^2 \right. \\ &\quad \quad \left. - \left(\tilde{B}_{\mathbf{k}} - \frac{\tilde{B}_{\mathbf{k}}^2}{\tilde{A}_{\mathbf{k}}} \right) |\tilde{\tau}_{\mathbf{k}}|^2 \right] \end{aligned} \quad (32)$$

$$= \frac{1}{2\tilde{A}_{\mathbf{k}}} + \left(\frac{\tilde{B}_{\mathbf{k}}}{\tilde{A}_{\mathbf{k}}} - \frac{\hat{B}_{\mathbf{k}}}{\hat{A}_{\mathbf{k}}} \right)^2 \frac{1}{2\tilde{B}_{\mathbf{k}} - \frac{2\tilde{B}_{\mathbf{k}}^2}{\tilde{A}_{\mathbf{k}}}}, \quad (33)$$

where $\tilde{A}_{\mathbf{k}}$, $\tilde{B}_{\mathbf{k}}$, and Z substitute for

$$\tilde{A}_{\mathbf{k}} = \beta\tilde{G}_{\mathbf{k}} + h + \frac{1}{2\tilde{R}_{\mathbf{k}}}, \quad \tilde{B}_{\mathbf{k}} = \frac{1}{2\tilde{R}_{\mathbf{k}}}, \quad (34)$$

$$\begin{aligned} Z = \int \int d\tilde{\xi}_{\mathbf{k}} d\tilde{\tau}_{\mathbf{k}} \exp \left[-\tilde{A}_{\mathbf{k}} \left| \tilde{\xi}_{\mathbf{k}} - \frac{\tilde{B}_{\mathbf{k}}}{\tilde{A}_{\mathbf{k}}} \tilde{\tau}_{\mathbf{k}} \right|^2 \right. \\ \left. - \left(\tilde{B}_{\mathbf{k}} - \frac{\tilde{B}_{\mathbf{k}}^2}{\tilde{A}_{\mathbf{k}}} \right) |\tilde{\tau}_{\mathbf{k}}|^2 \right]. \end{aligned} \quad (35)$$

Therefore, the mean squared error is represented by

$$\begin{aligned} E_1 &= \frac{1}{2N} \sum_{\mathbf{k}} \left[\frac{1}{\beta\tilde{G}_{\mathbf{k}} + h + \frac{1}{2\tilde{R}_{\mathbf{k}}}} \right. \\ &\quad \left. + \left(\frac{\frac{1}{2\tilde{R}_{\mathbf{k}}}}{\beta\tilde{G}_{\mathbf{k}} + h + \frac{1}{2\tilde{R}_{\mathbf{k}}}} - \frac{\frac{1}{2\hat{R}_{\mathbf{k}}}}{\hat{\beta}\tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{R}_{\mathbf{k}}}} \right)^2 \right. \\ &\quad \left. \times \frac{\beta\tilde{G}_{\mathbf{k}} + h + \frac{1}{2\tilde{R}_{\mathbf{k}}}}{\frac{1}{2\tilde{R}_{\mathbf{k}}} (\beta\tilde{G}_{\mathbf{k}} + h)} \right]. \end{aligned} \quad (36)$$

When the conditions

$$\hat{\beta} = \beta, \quad \hat{h} = h, \quad \text{and} \quad \hat{R}_{\mathbf{k}} = \tilde{R}_{\mathbf{k}}, \quad (37)$$

hold, the mean squared error takes the following minimum:

$$E_{1_{min}} = \frac{1}{2N} \sum_{\mathbf{k}} \left[\frac{1}{\beta \tilde{G}_{\mathbf{k}} + h + \frac{1}{2\tilde{R}_{\mathbf{k}}}} \right], \quad (38)$$

More precisely, when the ratio $\hat{\beta} : \hat{h} : \hat{R}_{\mathbf{k}}$ is equal to the ratio $\beta : h : \tilde{R}_{\mathbf{k}}$, E_1 is the minimum given by Eq. (38). Note that this equation represents the limit of the restoration.

Likewise, we can derive the mean squared error, E_2 , between the original image and the distorted image as being represented by

$$E_2 = \left\| \frac{1}{N} \sum_{\mathbf{j}} (\tilde{\xi}_{\mathbf{j}} - \tilde{\tau}_{\mathbf{j}})^2 \right\| \quad (39)$$

$$= \frac{1}{N} \sum_{\mathbf{k}} \tilde{R}_{\mathbf{k}} = \frac{1}{N} \sum_{\mathbf{i}} \tilde{R}_{\mathbf{i},\mathbf{i}} = b^2 \quad (40)$$

E_2 depends only on the diagonal element of the noise covariance matrix.

C. Maximization of the marginal likelihood

The minimized mean squared error criterion generally cannot be used since the unknown original image itself is needed to evaluate the squared error. If we already had the original image, there would be no need to restore the distorted image. Hence, in this paper, we argue for using the maximization of the marginal likelihood.

As shown in Eqs. (1) and (8), we already know the source prior and noise probabilities. Using both probabilities, we can derive $P(\boldsymbol{\tau})$:

$$P(\boldsymbol{\tau}) = \int d\mathbf{S} P_{\text{out}}(\boldsymbol{\tau}|\mathbf{S}; \hat{R}) P(\mathbf{S}; \hat{\beta}, \hat{h}). \quad (41)$$

$P(\boldsymbol{\tau})$ is called the marginal likelihood.

Maximization of the marginal likelihood is used to set the hyperparameters, $\hat{\beta}$, \hat{h} , and \hat{R} , to use to obtain the maximum value of $P(\boldsymbol{\tau})$ for distorted image $\boldsymbol{\tau}$.

With Eqs. (3) and (9), the marginal likelihood is given by

$$P(\boldsymbol{\tau}) = \frac{Z_{\text{posterior}}(\hat{\beta}, \hat{h}, \hat{R})}{Z_{\text{noise}}(\hat{R}) Z_{\text{prior}}(\hat{\beta}, \hat{h})}, \quad (42)$$

where

$$Z_{\text{posterior}}(\hat{\beta}, \hat{h}, \hat{R}) = \prod_{\mathbf{k}} \int d\tilde{S}_{\mathbf{k}} \exp(-\hat{H}_{eff}) \quad (43)$$

$$= \pi^{\frac{N}{2}} \prod_{\mathbf{k}} \frac{1}{\sqrt{\hat{A}_{\mathbf{k}}}} \exp \left[-\left(\hat{B}_{\mathbf{k}} - \frac{\hat{B}_{\mathbf{k}}^2}{\hat{A}_{\mathbf{k}}} \right) |\tilde{\tau}_{\mathbf{k}}|^2 \right]. \quad (44)$$

Because \ln is a monotonically increasing function that maximizes the logarithmic marginal likelihood, maximizing $\ln P(\boldsymbol{\tau})$ is equivalent to maximizing $P(\boldsymbol{\tau})$.

The logarithmic marginal likelihood $\ln P(\boldsymbol{\tau})$ can be written as,

$$\begin{aligned} \ln(P(\boldsymbol{\tau})) &= \ln \left(\frac{Z_{\text{posterior}}(\hat{\beta}, \hat{h}, \hat{R})}{Z_{\text{noise}}(\hat{R}) Z_{\text{prior}}(\hat{\beta}, \hat{h})} \right) \\ &= -\frac{1}{2} \sum_{\mathbf{k}} \ln \left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{R}_{\mathbf{k}}} \right) - \frac{N}{2} \ln(2\pi) \\ &\quad - \frac{1}{2} \sum_{\mathbf{k}} \ln(\hat{R}_{\mathbf{k}}) + \frac{1}{2} \sum_{\mathbf{k}} \ln(\hat{\beta} G_{\mathbf{k}} + \hat{h}) \\ &\quad - \sum_{\mathbf{k}} \frac{(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h}) \frac{1}{2\hat{R}_{\mathbf{k}}}}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{R}_{\mathbf{k}}}} |\tilde{\tau}_{\mathbf{k}}|^2. \end{aligned} \quad (45)$$

IV. RESULT

In this section, we consider a method to restore an image distorted by spatially correlated noise by means of the spatially uncorrelated noise model. To do this, we replace R as one of the hyperparameters:

$$\hat{R}_{\mathbf{i}, \mathbf{j}} = \hat{r} \delta_{\mathbf{i}, \mathbf{j}}, \quad \hat{R}_{\mathbf{k}} = \hat{r}. \quad (46)$$

We can now write Eq. (28) as

$$\langle \sigma_{\mathbf{j}}^{app} \rangle = \frac{1}{N} \sum_{\mathbf{k}} \sum_{\mathbf{i}} \frac{\tau_{\mathbf{i}} \frac{1}{2r} \cos[\mathbf{k} \cdot (\mathbf{j} - \mathbf{i})]}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}}}. \quad (47)$$

Figure 1 shows the analytical and simulated mean squared errors between the original image and each restored image, Fig. 2 shows the estimated \hat{r} , and Fig. 3 show the estimated parameters \hat{h} and $\hat{\beta}$. We used 100 artificially generated two-dimensional $N = 16^2$ images with hyperparameters $\beta = 1.0$, $h = 1.0$, $b = 0.75$, and $\kappa = 3.0$. The horizontal axes of these

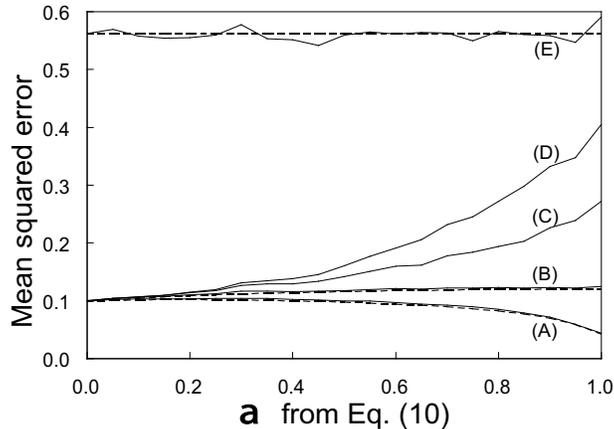


FIG. 1. Mean squared error between original and each restored image. Horizontal axis denotes a from Eq. (10); vertical axis denotes mean squared error. (A) Optimum decode given by Eq. (38). (B) Restored using criterion of minimum restored error. (C) Restored using \hat{r} estimated by maximizing marginal likelihood. (D) Restored using $\hat{\beta}, \hat{h}$, and \hat{r} , all estimated by maximizing marginal likelihood. (E) Error before decoding given by Eq. (40).

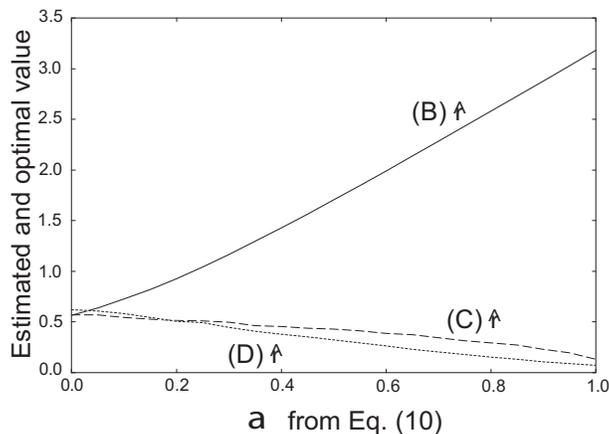


FIG. 2. Estimated \hat{r} . Horizontal axis denotes a from Eq. (10); vertical axis denotes average estimated \hat{r} obtained by maximizing marginal likelihood or optimal \hat{r} obtained using criterion of minimum restored error. Solid line (B) denotes \hat{r} used in Fig. 1(B), dashed line (C) denotes \hat{r} used in Fig. 1(C), and dotted line (D) denotes \hat{r} used in Fig. 1(D).

three figures denote a from Eq. (10); a represents the magnitude of the spatial correlation of the noise ($a = 0$ implies no spatial correlation).

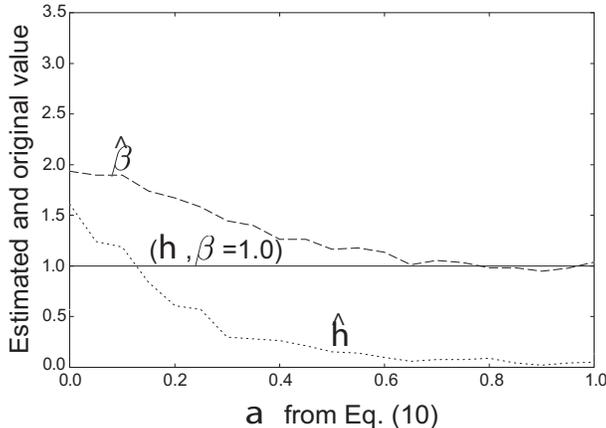


FIG. 3. Estimated \hat{h} (dotted line) and $\hat{\beta}$ (dashed line). Both values were averaged with respect to 100 artificially generated images. Horizontal axis denotes a from Eq. (10); vertical axis denotes average estimated \hat{h} or $\hat{\beta}$ or actual h or β . Solid line denotes actual $h(= \beta = 1.0)$.

A. Minimization of mean squared error

Curve (E) in Fig. 1 represents error E_2 given by Eq. (40), that is, the mean squared error between the original and distorted images. (The dashed line represents the analytical error, and the solid line represents the simulated results.) Because the numbers of pixels ($N = 16^2$) and samples (100) were finite, the simulated results have some fluctuation. This fluctuation tends to asymptotically disappear as the numbers are increased.

Curve (A) in Fig. 1 represents minimum error $E_{1_{min}}$ given by Eq. (38), that is, the mean squared error between the original and distorted images restored using the spatially correlated noise model with the best parameters. Hence, $E_{1_{min}}$ implies the limits of image restoration-no further error reduction is possible. Curve (B) in Fig. 1 represents the error in an image restored using the uncorrelated noise model. The restoration was done using hyperparameter \hat{r} from Eq. (29), which was acquired minimizing the mean squared error. (The dashed line represents the analytical results, and the solid line represents the simulated results.) Curve (B) in Fig. 2 shows the optimal values of \hat{r} by minimizing the mean squared error.

Note that when $a = 0.0$, curve (A) in Fig. 1 is equal to curve (B). Curve (B) in Fig. 1 shows the optimum performance when using the minimization of the mean squared error as the criterion. However, as mentioned, the hyperparameters cannot be estimated in practice by minimizing the mean squared error. Therefore, we will discuss how the error can be

reduced by maximizing the logarithmic marginal likelihood in the next subsection.

B. Maximization of the marginal likelihood

Substituting Eq. (45) for Eq. (46), we get

$$\begin{aligned} \ln(P_{app}(\boldsymbol{\tau})) &= -\frac{1}{2} \sum_{\mathbf{k}} \ln \left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}} \right) - \frac{N}{2} \ln(2\pi) \\ &\quad - \frac{N}{2} \ln(\hat{r}) + \frac{1}{2} \sum_{\mathbf{k}} \ln \left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} \right) \\ &\quad - \sum_{\mathbf{k}} \frac{\left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} \right)^{\frac{1}{2\hat{r}}}}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}}} |\tilde{\tau}_{\mathbf{k}}|^2. \end{aligned} \quad (48)$$

We can derive the conditions necessary for the extremal $P_{app}(\boldsymbol{\tau})$, and we can represent them using the following nonlinear simultaneous equations.

$$\begin{aligned} \frac{1}{\hat{\beta}} &= \left(\sum_{\mathbf{k}} \frac{\tilde{G}_{\mathbf{k}}}{\tilde{G}_{\mathbf{k}} + \frac{\hat{h}}{\hat{\beta}}} \right)^{-1} \left[\sum_{\mathbf{k}} \frac{\tilde{G}_{\mathbf{k}}}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}}} \right. \\ &\quad \left. + \sum_{\mathbf{k}} |\tilde{\tau}_{\mathbf{k}}|^2 \frac{\tilde{G}_{\mathbf{k}}}{2\hat{r}^2 \left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}} \right)^2} \right], \end{aligned} \quad (49)$$

$$\begin{aligned} \hat{r} &= \frac{1}{2N} \sum_{\mathbf{k}} \frac{\tilde{G}_{\mathbf{k}}}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}}} \\ &\quad + \frac{1}{N} \sum_{\mathbf{k}} |\tilde{\tau}_{\mathbf{k}}|^2 \frac{(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h})^2}{\left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}} \right)^2}, \end{aligned} \quad (50)$$

$$\begin{aligned} \frac{1}{\hat{h}} &= \left(\sum_{\mathbf{k}} \frac{1}{1 + \frac{\hat{\beta} \tilde{G}_{\mathbf{k}}}{\hat{h}}} \right)^{-1} \left[\sum_{\mathbf{k}} \frac{1}{\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}}} \right. \\ &\quad \left. + \sum_{\mathbf{k}} |\tilde{\tau}_{\mathbf{k}}|^2 \frac{1}{2\hat{r}^2 \left(\hat{\beta} \tilde{G}_{\mathbf{k}} + \hat{h} + \frac{1}{2\hat{r}} \right)^2} \right]. \end{aligned} \quad (51)$$

We can then determine the convergence point by means of an iterative calculation.

We next determined the unknown hyperparameter, \hat{r} , by maximizing the marginal likelihood, assuming that $\hat{\beta} = \beta$ and $\hat{h} = h$. Thus, we restricted ourselves to the case where

the generation model of the image was already known, as was β and h . Curve (C) in Fig. 1 shows the mean squared error when the image was restored using the estimated \hat{r} . Curve (C) in Fig. 2 shows the estimated \hat{r} .

Note that the difference between curve (B) in Fig. 2, the limit of the restoration, and curve (C), obtained by minimizing the mean squared error increased with the noise correlation parameter, a . Therefore, the error obtained by maximizing the marginal likelihood (curve (C) in Fig. 1) increases with the noise correlation parameter, a . This implies that without knowing the noise model, estimating the hyperparameters by maximizing the marginal likelihood does not work well. That is, the conventional uncorrelated noise model cannot cope with spatially correlated noise.

To confirm this, we used maximization of the marginal likelihood to estimate $\hat{\beta}$ and \hat{h} -the hyperparameters of the image-generation probability-as well as \hat{r} . Curve (D) in Fig. 1 shows the mean squared error estimated using hyperparameters $\hat{\beta}$, \hat{h} , and \hat{r} . Curve (D) in Fig. 2 shows the obtained \hat{r} , and Fig. 3 shows the obtained $\hat{\beta}$ and \hat{h} . The error shown by curve (D) in Fig. 1 is much greater than that shown by curve (C).

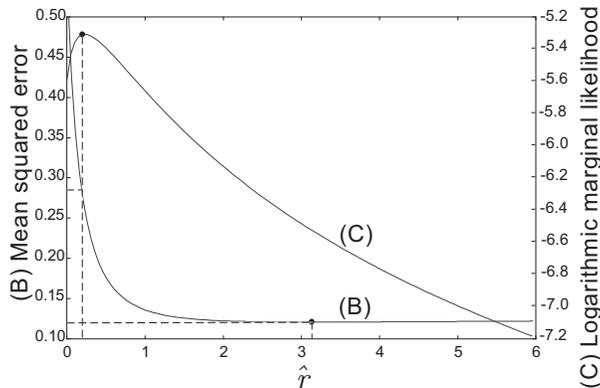


FIG. 4. Mean squared error and marginal likelihood. Horizontal axis denotes hyperparameter \hat{r} ; left vertical axis denotes mean squared error (B) given by Eq.(36); right vertical axis denotes logarithmic marginal likelihood (C) given by Eq.(48). The parameters were $a = 1.0$, $\beta = 1.0$, $h = 1.0$, and $\kappa = 3.0$. Those used for (B) correspond to those for curve (B) in Fig. 1 when $a = 1.0$. Likewise, those used for (C) correspond to those for curve (C) in Fig. 1(C) when $a = 1.0$.

Next, we demonstrate that the logarithmic marginal likelihood, curve (C) in Fig. 1, does not have a local maximum and that the iteratively obtained solution of Eq. (50) corresponds to the global maximum. Curve (B) in Fig. 4 shows the mean squared error, E_1 , given by

Eq.(36), and curve (C) shows the logarithmic marginal likelihood, $\ln(P_{app}(\boldsymbol{\tau}))$, given by Eq. (48). The horizontal axis denotes \hat{r} . Curve (B) shows that E_1 takes a minimum value, $E_1 = 0.12$, at $\hat{r} = 3.2$. The parameters used for curve (B) correspond to those for curve (B) in Fig. 1 when $a = 1.0$. Curve (C) shows that that the logarithmic marginal likelihood takes a maximum value, $E_1 = 0.28$, at $\hat{r} = 0.2$. The parameters used for curve (C) correspond to those for curve (C) in Fig. 1 when $a = 1.0$. The important point is that the logarithmic marginal likelihood, curve (C) in Fig. 4, does not have a local maximum.

C. Sample of images

Next, we show the practical significance of these differences for a typical set of artificial images. Figure 5(a) shows an original image generated using hyperparameters $\beta = 0.5$ and

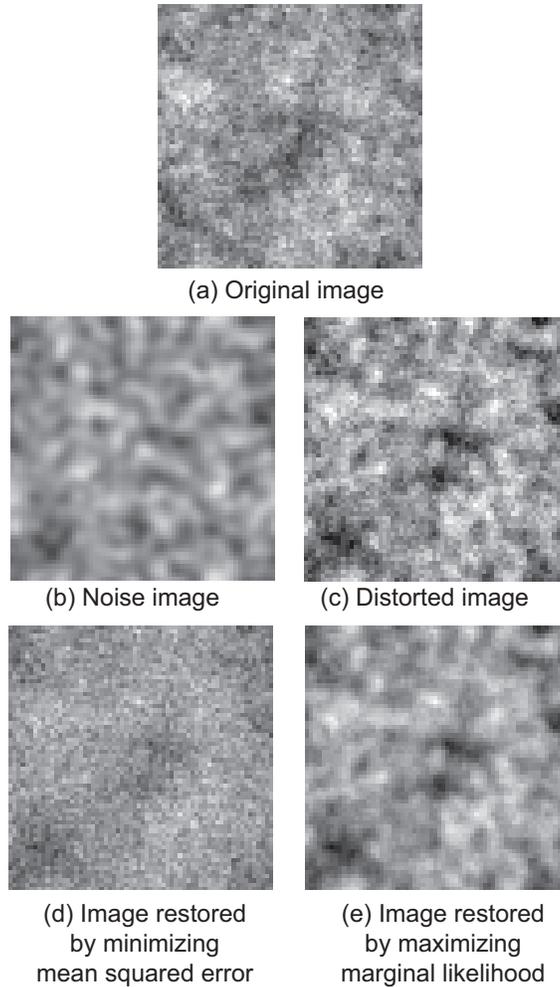


FIG. 5. Artificial images. $N = 64^2$.

$h = 1.0^{-4}$. Figure 5(c) shows the same image after it was distorted by the noise shown in Fig. 5(b) ($a = 1.0, b = 0.75$, and $\kappa = 3.0$). The mean squared error, E_2 , between the original and distorted images was 0.57. Figure 5(d) shows the image after it was restored using the complete restoration model: the generation model and noise model were consistent with the original models, and optimum values were used for the hyperparameters. When $a = 1.0$, curve (A) in Fig. 1 corresponds to Fig. 5(d), and the restored mean squared error, E_1 , is 0.27. Figure 5(e) shows the image after it was restored using an uncorrelated noise model with hyperparameters obtained by maximizing the marginal likelihood. When $a = 1.0$, curve (D) in Fig. 1 corresponds to Fig. 5(e), and the restored mean squared error is 0.45. The image in Fig. 5(d) resembles the original one, Fig. 5(a), while the one in Fig. 5(e) is similar to the distorted one in Fig. 5(c) and looks slightly out of focus.

Furthermore, we show the results when natural images were used. Figure 6 shows the natural images and a table of the estimated parameters. We used $b = 0.75$ and $\kappa = 3.0$ for the noise parameters and three values for a : 0.0, 0.5, and 1.0. The three images on the left are distorted, and those on the right are the same images restored using $\hat{\beta}$, \hat{h} , and \hat{r} estimated using the maximized marginal likelihood. In this simulation as well, E_1 increased with the noise correlation parameter, a . This tendency is reflected in the three restored images. For $a = 1.0$, the lower right image, the noise looks like a stain that was not completely removed.

D. Discussion

Here, we consider why line (A) in Fig. 1 tends to decrease as a approaches 1.0. This phenomenon appears to be universal, and also occurs with another parameter set as shown in Fig. 7(ii). For reference, we also show line (i) which is based on the parameter set that was used in Fig. 1(A) with only N changed to $N = 64^2$. (Note that line (A) is independent of N).

For the original image, we used the Gaussian model represented by Eq. (1), which has neighboring interactions on each pixel. For the noise image, we used the Gaussian model represented by Eq. (8), in which the interaction is Gaussian functionally decreased. We expected the difference between these two models to become conspicuously large near $a = 1.0$, and this difference between the models would enable successful restoration near $a = 1.0$. Figure 8 shows the typical images about Fig. 7(ii). The image was most difficult

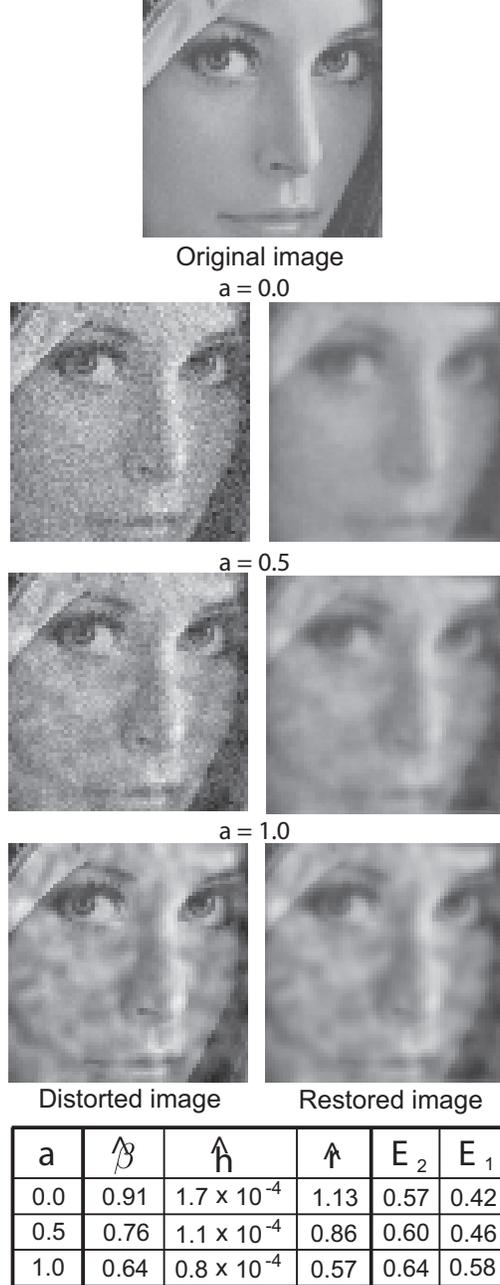


FIG. 6. Natural images. $N = 64^2$.

to restore in Fig. 8 when $a = 0.55$, as shown in Fig. 7(ii). In this case, the original and the noise images were visually similar; more precisely, the correlation lengths of the two images were similar. Thus, to successfully restore a distorted image, the character of the original image should be different from that of the noise image.

Here, we summarize our explanation of why line (A) in Fig. 1 decreases near $a = 1.0$. As a approaches 1.0, the mean squared error decreases because the difference in character,

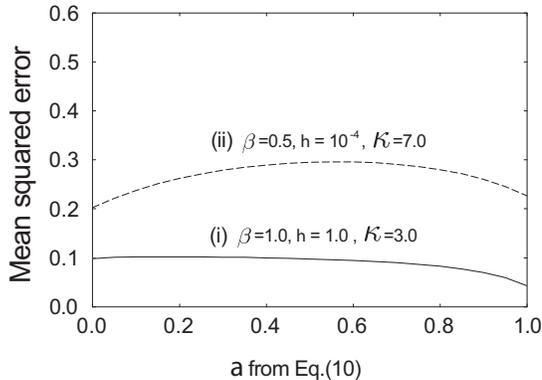


FIG. 7. Mean squared error between original image and restored image. The horizontal axis denotes a from Eq. (10), and the vertical axis denotes the mean squared error. As a criterion, the optimal decode given by Eq. (38) is used. The parameters $b = 0.75$, and $N = 64^2$ are fixed. For the solid line (i), the other parameters were set as $\beta = 1.0$, $h = 1.0$, and $\kappa = 3.0$. For the dashed line (ii), the other parameters were set as $\beta = 0.5$, $h = 10^{-4}$, and $\kappa = 7.0$.

especially the correlation length, between the two images enables easier restoration. Conversely, the mean squared error is greater at $a = 0.55$ on line (ii) in Fig. 7 because the correlation-length similarity between the images makes it difficult to distinguish between them, and that also makes it difficult to restore the distorted image. To ensure successful restoration, the correlation length of the original image should be different from that of the noise image, which is the case when $a = 0.00$ and $a = 1.00$ in Fig. 8.

V. CONCLUSION

We investigated the use of the Bayesian inference to restore images under conditions of spatially correlated noise. We assumed that both the original image and the noise obeyed a Gaussian distribution composed of translational symmetric matrices. We used the Fourier transformation to diagonalize the covariance matrices, which enabled us to apply various forms of statistical analysis. We obtained the expected value of a restored image and obtained the optimal analytical hyperparameters by minimizing the mean squared error and used these hyperparameters to determine the generation and noise models. Furthermore, we discussed whether the conventional spatially uncorrelated noise model could cope with the spatially correlated noise or not. We used two methods to estimate the hyperparameters:

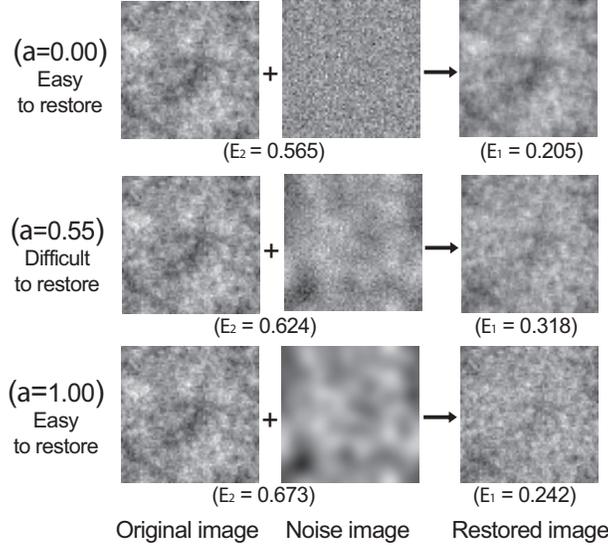


FIG. 8. Typical images about Fig. 7 (ii). $N = 64^2$. Parameters for the original images were $\beta = 0.5$ and $h = 10^{-4}$. The three original prepared images were the same. Parameters for noise images were $\beta = 0.5, h = 10^{-4}, \kappa = 7.0$, and $b = 0.75$. Only a was changed. Note that when $a = 0.55$, the original and noise images were visually similar, and the restoration quality was lower in this case.

minimizing the mean squared error and maximizing the marginal likelihood. The difference between the errors obtained using the two methods increased with the noise correlation parameter. The restoration error was larger when we used the hyperparameters obtained using the maximization of the marginal likelihood method. Thus, the conventional spatially uncorrelated noise model could not cope with the spatially correlated noise.

-
- [1] R. Morina, *IEEE Trans. Image Process.*, **8**, 231 (1999).
 - [2] R. Morina and K. Katsaggelos, *IEEE Trans. Image Process.*, **8**, 231 (1999).
 - [3] J. Inoue and K. Tanaka, *Phys. Rev. E*, **65**,016125-1 (2002).
 - [4] H. Nishimori and K. Y. M .Wong, *Phys. Rev. E*, **60**, 132 (1999).
 - [5] H. Nishimori, *BUSSEI KENKYU. Jpn.*, **73**, 850 (2000).
 - [6] K. Tanaka and J. Inoue, *IEICE Trans. Inf. & Syst*, **E84-D**, 546 (2002).
 - [7] H. Derin, H. Elliott, R. Cristi, and D. Geman, *IEEE Trans. Pattern Anal. & Machine Intell.*, **6**, 707 (1984).

- [8] S. Geman and D. Geman, *IEEE Trans. Pattern Anal. & Machine Intell.*, **6**, 721 (1984).
- [9] J. Inoue, *Phys. Rev. E*, **60**, 2547 (1999).
- [10] J. Inoue, *Phys. Rev. E*, **63** 046114 (2001).
- [11] J. Inoue, *Phys. Rev. E*, **64**, 036121 (2001).
- [12] J. Inoue and K. Tanaka, *Neural Netw. Signal Process.* XI, 383 (2001).