博士論文（要約）

Bayesian Modeling of Lattice System Measurement for Latent Parameter Estimation

（潜在パラメータ推定を目的とする格子系観測のベイジアンモデリング）

坂本　浩隆

Abstract

Natural science and engineering have developed complementary. Natural science gives to engineering the fundamental principles and the target, and engineering gives some techniques to acquire new paradigms. Microscopes, telescopes and accelerators that are developed by cutting-edge measurement technologies show some new pictures in extreme situations. These technologies are profitable for natural science, however, are also burdensome in some case. For example, in material science, the photomicrography have the significant roles for elucidating structures and functions of materials, but the cost of analysis is so high; researchers who analyze materials, are pressed by work for processing the photos[1]. Besides, experimental equipment requiring enormous cost is developed without sufficient experimental design including an estimation of the results and its confidence. The large-scale experimental equipment, such as high-intensity Fermi-chopper spectrometers[2], have produced a large amount of data which has not yet been fully analyzed. In those case, measurement techniques and measurement process is the bottleneck for appropriate frameworks on natural science: repeating the operations that are proposing a hypothesis and verifying it. We need the engineering for science to solve this bottlenecks. In this thesis, we focus on two aims: the latent parameter estimation using measurements and the model-based experimental design for estimation.

Measurement in natural science is a process to obtain image data. The smoothness and local structures of the image data mirror those continuities and local interactions which targets of natural science have. The measured data are mapped on space (and time) plane, and this process called imaging, for instance, scanning tunneling microscopy images obtained by sampling values on the target surface and then mapped those values [3]. The sampling processes are divided into the case of mapping to real-space or Fourier-space such as observation by using CCD in digital cameras or neutron scattering, respectively.

The motivation of observation in natural science is identifying the model of the target observed. To estimate latent parameter of image data is one of our aims, and necessary for verifying the hypothesis, and this identification. We treat the estimation of latent model parameters and evaluation of confidence of it by using image data. The confidence of estimation is useful for the experimental design which is the other one of our aims.

We focus on quantification of image data and design the observation systems by using Bayesian inference. Our targets are two types of image observation; observing the image sampled and mapped to real-space or Fourier-space. In the case both type of observation,

we modeled what we observed as the lattice, such as the MRFs and crystal systems. Our aim of this thesis is to estimate the interaction parameters of the lattice from image data mapping to real-space or Fourier-space and to evaluate of confidence of the estimation. Besides, we aim the evaluation the confidence and uniqueness of estimates to design the measurement system and processing flow of the image data.

The work in Chapter 3, 5, and 4 study the estimation of the model parameter which is the interaction parameter of the lattice graph by using the image mapped to real-space. In Chapter 3 and 4, we focus on the latent parameter estimation and evaluate the confidence of image data. To begin with, we proposed the theoretical calculation of the distribution of the model parameters of GMRF for the fundamental analysis of the quantification of image data, and this method to calculate is used in the study of the Chapter 3. Next, we show the inevitable effects on observation and estimation of the model parameters in Chapter 4. In Chapter 5, we focus on the experiment design. We show the appropriate processing for the estimation of the model parameters.

The observation mapping to Fourier-space uses the diffraction of waves or particles such as electromagnetic waves or neutron. The recent target of observation by using such diffractions is various materials containing bacteria and polymers, though were some crystals in the beginning [4, 5]. It is the purpose of the measurements for the studies of crystals to elucidate the crystal structure: the atomic arrangement and strength of interaction between the atoms. This interaction can be measured by watching the dispersion relation which reflects the vibrations of the crystals, so the dispersion relations of lattice vibrations and phonons which are quantized lattice vibration has been measured actively. In this thesis, we treat the dispersion relation of lattice vibration.

The work in Chapter 6 studies the estimation of the latent model parameter which is the interaction parameter of the crystal lattice by using the image mapped to Fourier-space. The image mapped to Fourier-space is obtained the spectral data. We use the Bayesian framework to estimate the latent parameters using high-dimensional spectral data. We introduce the previous works of the Bayesian spectral analysis in Sect. 2.2. Our novelty is to estimate the model parameters by using the dispersion relation spectra for analyzing a crystal based on the Bayesian framework. In Chapter 6, we propose the two methods to estimate the model parameters and compare both methods. Besides, we suggest the better procedure to sampling Fourier space on the estimation. In a conventional method, the experimenters visualize sample point in the Fourier-space which has high symmetricity. We can evaluate the sampling method in the viewpoint of the confidence of Bayesian inference. As a consequence, we contribute in this chapter to both aims: to latent parameter estimation of the image mapped to Fourier-space and to suggest the better design of the measurement on Fourier-space.

Acknowledgements

This thesis summarized the work during Ph. D. course of the Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, The University of Tokyo. Many people have really helped me in my research life.

First of all, I would like to express my sincere gratitude to my supervisor Prof. Masato OKADA, who guides and support the work. He gave me constant encouragement and offered valuable advice not only on the studies but also on my life planning.

I would like to give heartfelt thanks to my mentor Dr. Yoshinori NAKANISHI for his enormous and persistent help. He taught many things about studying and writing to me so patiently.

I would like to express my full appreciation Prof. Taka-hisa ARIMA and Dr. Kenji NAGATA for their significant contribution to the work. They made lots of time for discussion and gave many insightful comments.

I also would like to thank other members of my Ph. D. committee, Prof. Akira EJIRI, Prof. Takehiko SASAKI, and Prof. Issei SATO for giving me useful comments on my draft of this thesis.

I would like to thank all the members of the lablatory greatly. Especially I would like to thank Prof. Yonghao YUE, Dr. Yoichi MOTOTAKE, and Dr. Yasuhiko IGARASHI. They gave a lot of advice and suggestion for this thesis. I would like to thank Mr. Yoshinori NAGANO, Mr. Shun KATAKAMI, Mr. Kensuke MUTO, Mr. Hirosato ITOH, and Mr. Kentaro NAGASAWA for their help and comment on this thesis. I would like to thank Dr. Satoru TOKUDA, Dr. Shin MURATA, and Kenji TANAKA for their constant encouragement on the laboratory life. I would like to thank secretaries of OKADA lab. Ms. Ayako FUKUYA, Ms. Chiaki SAKAKIBARA, and Ms. Yoshie IZUMI for daily support for office procedures.

Finally, I would like to express my cordial gratitude to my family for their moral support and warm encouragement.

Contents

Chapter 1

Introduction

Natural science and engineering have developed complementary. Natural science gives to engineering fundamental principles and its target, and engineering gives some techniques to acquire new paradigms. In particular, some measurement techniques such as microscopes, telescopes, and accelerators that are developed by cutting-edge engineering technologies show some new pictures in extreme situations which we could not have watched. These technologies are profitable for natural science, however, are also burdensome in some case. A large amount of measurement data requires the cost of the measurement and analysis. These cost can be bottle-necks for a scientific process: hypothesis and verifying.

Our goal is to treat the measured data and hypothesis in the same framework and give some feedback for the measurement process refinement to solve the bottlenecks on measurement and analysis step. We utilize Bayesian frameworks on this thesis for treating the measured data and physical hypothesis. This thesis focuses on some data that is obtained by measuring values on lattice systems and mapping these values on real or Fourier space. We called the data that has such types, image. We set two targets for this thesis; estimation latent physical parameters of the measured system and feedback to measurement design from the results by using Bayesian modeling because of the algorithmic analysis of measured data and refinement measurement process.

Firstly, this chapter introduces our issues on hypothesis and measurement steps in Sect. 1.1. Secondly, Sect. 1.2 describes the benefit of the Bayesian modeling framework for the scientific process. Thirdly, Sect. 1.3 define our target image data. Lastly, Sect. 1.4 describes the focus of this thesis and the contributions.

1.1 Measurement on Scientific Method

The novel pictures obtained by new measurement method provide the novel scientific paradigms. One of the most famous episodes of the paradigm changes triggered by measurement is that of the discovery of desoxyribonucleic acid (DNA) structure and its

function. Linus Pauling, Francis Crick, and James D. Watson hypothesized that DNA has a helical structure and implied DNA has the self-copying ability. Rosalind Franklin gave the evidence for the hypothesis by measuring the crystalized DNA using X-ray diffraction techniques. The discovery of DNA changed bioscience including genetics completely and created biotechnologies. This episode of the discovery constructed by two-part: proposing a hypothesis and measurement for supporting it. These steps are the elements of a scientific investigation. Repeating these steps enhances the reliability of the scientific fact.

Recent years, the development of the technology of measurement gives a large number of novel pictures. These technologies are profitable for natural science, however, are also burdensome in some case. For example, in material science, the photomicrography have the significant roles for elucidating structures and functions of materials, but the cost of analysis is so high; researchers who analyze materials, are pressed by work for processing the photos[1]. Besides, experimental equipment requiring enormous cost is developed without sufficient experimental design including an estimation of the results and its confidence. The large-scale experimental equipment, such as high-intensity Fermi-chopper spectrometers[2], have produced a large amount of data and its analysis by researchers cannot catch up. Japan Proton Accelerator Research Complex (J-PARC), possess high-intensity Fermi-chopper spectrometers. In recent years, the neutron scattering experiments in J-PARC have produced numerous pieces of four-dimensional event data and researchers have had difficulties in utilizing the whole data which can not treat by human eyes. In those case, measurement techniques and measurement processes are the bottlenecks for appropriate frameworks on natural science: repeating the processes that are proposing a hypothesis and verifying it. We need the engineering for science to solve this bottlenecks.

The reason why the cost of analyzing such data is so high is that the data has a complex structure and they require prior knowledge for analysis. This prior knowledge corresponds to the hypothesis on the scientific method. For instance, the X-ray diffraction pattern obtained from crystalized DNA needs the prior knowledge which DNA has a helical structure for understanding. We need the framework to treat data with prior knowledge flexibly for solving the bottlenecks. The flexible algorithm to utilize the physical knowledge can relax an attributability for reducing the cost to analyze a large-scale data and overcoming the arbitrariness on data analysis. Prior knowledge helps us to analyze the low-quality observed data such as noisy or few sample size. Evaluating the quality of analyzed data by the algorithm can improve the way of measurement.

Our goal is to establish a flexible framework for analyzing measurement data with prior knowledge and utilizing the results of the analysis to improve data processing and measurement method. In this thesis, we approach to the goal by modeling of the measurement. We need a framework for flexible method to model the measurement by using prior knowledge of a measurement target.

The motivation of measurement in natural science is identifying the model of the target observed. We classify the modeling into the following three levels:

Level 1 Model selection from the candidate models of the target.

Level 2 Latent parameter estimation of the selected model.

Level 3 Confidence evaluation for the model and data.

In this thesis, we treat level 2 and 3. That is, we treat the estimation of model (latent) parameters and evaluation of confidence of the estimation by using image data. The confidence obtained in level 3 is useful for the experimental design which is the other one of our aims. We need to obtain the model of not only the target observed but also measurement process in order to improve the way of measurement.

1.2 Bayesian Modeling

The Bayesian framework can perform all levels of modeling of measurement data. The inference based on the Bayesian framework can estimate the model parameters and evaluate the confidence of the values, in the case that we assume the model by which can generate data[6]. We assume the forward model which generate data, involving the condition of the observation and preprocessing.

Let us consider the data generating process, where the probabilistic event x which is parametrized by a occurs, and then we obtain the data y by observation. We formulate this model by the probability distribution of the process to generate the event $P(x|a)$, and the distribution of the process of observation $P(y|x)$. The conditional probability distribution of the model parameter a given data y is $P(a|y)$ calculated by Bayes' theorem and marginalization of probability as follows:

$$\begin{aligned} P(a|y) &= \frac{P(a,y)}{P(y)} \\ &= \int dx \frac{P(a,x,y)}{P(y)} \\ &= \int dx \frac{P(y|x)P(x|a)P(a)}{P(y)}. \end{aligned}$$

The probability $P(a|y)$ is what we want to know, called the posterior distribution. We can treat the expected value of this posterior as the estimates, and the shape of the posterior represents the confidence and singularity of the estimation in the case that we use the candidate model and data[7, 8, 9]. In this thesis, we analyze the posterior to evaluate the appropriateness of the measurement systems, preprocessing before the model parameter estimation, and the amount and quality of the given data.

It is essential to check this appropriateness before the experiments performed and to give the feedback to some conditions for designs of the experiments and data processing. For instance of some conditions, the noise and resolution that the data have. The number

and the performance of the sensors must be limited. Therefore we obtain only deteriorated measurement data inevitably. Hence, we have to evaluate whether the data has sufficient quality or not for our aim, such as the estimation of the model parameters. Experimenters sometimes preprocess the obtained data to try to improve the data quality, such as the data averaging. They obtaining estimates from original data and then use as the input data for the other estimation. We can give the evaluation based on the Bayesian inference for these influences which are deterioration and preprocessing of image data.

Bayes' theorem is very simple from the point of mathematical view, but it is valuable for natural science. We describe the related work of Bayesian computation for model parameter estimation in Sec. 2.3.

1.3 Measurement Data of Lattice Systems

Measurement in natural science is a process to obtain image data. The smoothness and local structures of the image data mirror those continuities and local interactions which targets of natural science have. The measured data are mapped on space (and time) plane, and this process called imaging, for instance, scanning tunneling microscopy images obtained by sampling values on the target surface and then mapped those values [3]. The sampling processes are divided into the case of mapping to real-space or Fourier-space such as observation by using CCD in digital cameras or neutron scattering, respectively.

The observation mapping to real-space can be modeled by using a graphical model. The sampling value obtained from real-space mapped to the node and the correlation expressed by the edges of the graph. For instance, let us consider a simple case which is an observation by using the sensors arranged on a square lattice such as CCD and assumed that their interactions worked in very close. In this case, we can model the image in the graphical model like square lattice, where the nearest neighbor interaction approximates the correlation. On this modeling process, we have to determine the graph structure reflected the existing of the correlations and the strength of interactions.

The observation mapping to Fourier-space uses the diffraction of electromagnetic waves or neutron. The recent target of observation by using such diffractions is various materials containing bacteria and polymers. In the beginning, the targets were some crystals [4, 5]. The purpose of the measurements for the study of a crystal to elucidate the crystal structure. The crystal structure is identified from the atomic arrangement and strength of interaction between the atoms [10, 11]. This interaction can be measured by watching the dispersion relation which reflects the vibrations of the crystals, so the dispersion relations of lattice vibrations and phonons which are quantized lattice vibration has been measured actively [12, 13]. In this thesis, we treat the dispersion relation of a lattice vibration.

1.4 Our Focus and Contribution

We focus on latent parameter estimation of image data and measurement design by using Bayesian inference. Our targets are two types of image observation; observing the image sampled and mapped to real-space or Fourier-space. In the case both type of observation, we modeled what we observed as the lattice, such as the MRFs and crystal systems. Our aim of this thesis is to estimate the interaction parameters of the lattice from image data mapping to real-space or Fourier-space and to evaluate of confidence of the estimation. Besides, we aim the evaluation the confidence of estimates to design the measurement system and processing flow of the image data.

In all chapters of the thesis, we use the minimal model to treat the problems. The computation of the posterior distribution requires a considerable cost, especially in image analysis. Therefore, many techniques for approximation are developed to reduce the cost such as variational Bayes [14]. However, those techniques change the posterior shapes, and then it becomes difficult to divide the influence of what we want to evaluate and the approximation. For this reason, we choose the simple model whose posterior can be calculated analytically, otherwise the sampling method to obtain the strict posterior shape, in each chapter of this thesis.

We construct this thesis as follows (as shown in Fig. 1.1). In Chapter 2, we summarize the related work to this thesis. We introduce some previous work of the Bayesian inference and its application for spectral analysis, and applications and analysis of the MRF models by using statistical mechanics. The work studied in this thesis at the Chapter 3, 5, 4, 6 have been already published in or planned to submit to journals.

The work in Chapter 3, 4, and 5 study the estimation of the latent parameter which is the interaction parameter of the lattice graph by using the image mapped to real-space. In Chapter 3 and 4, we focus on the latent parameter estimation of image data. To begin with, we proposed the theoretical calculation of the distribution of the latent parameters of GMRF for the fundamental analysis of the quantification of image data, and this method to calculate is used in the study of the Chapter 5. We show the inevitable effects on observation and estimation of the model parameters. In Chapter 5, we focus on the experiment design. We show the appropriate processing for the estimation of the model parameters. We describe the summaries of these chapters.

Theory of Distribution Estimation of the Model Parameters in GMRF [15]

In Chapter 3, we show the exact analysis of a Gaussian-MRF(GMRF) model. The GMRFs are models of gray-scale images. In this chapter, we calculate the posterior distribution of the latent model parameters that are interaction parameter of the lattice model: smoothness of the image and the noise strength. We investigated the performance of distribution estimation of latent model parameters in Markov random field models proposed by Y. Nakanishi-Ohno *et al.* [16] when used to evaluate the confidence of data. We analytically calculated the configurational average with respect to data, of the negative logarithm of the posterior distribution

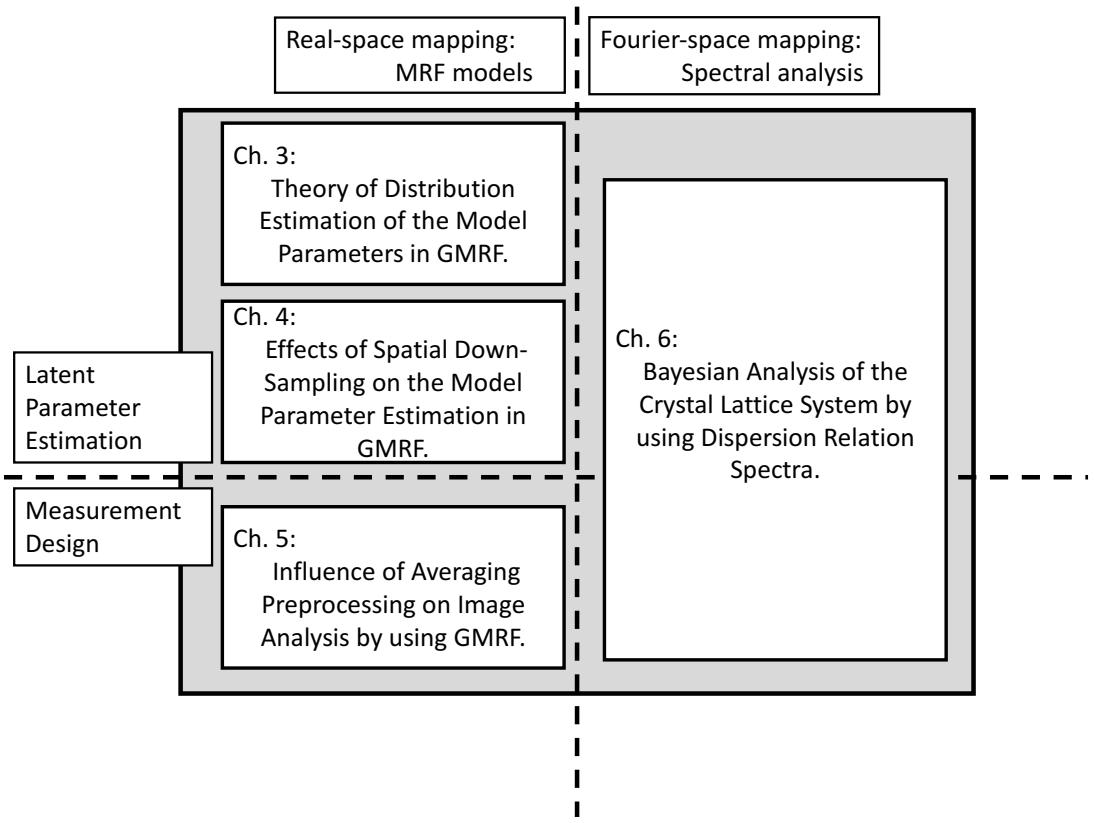

Figure 1.1: Structure of this thesis.

which is defined by each image data. The negative logarithm of the posterior distribution is called free energy based on an analogy with statistical mechanics, so the configurational average is the expected value of this free energy. This configurational average of free energy shrinks as the amount of data increases. Our results theoretically confirm the numerical results from that previous study. This calculation is the basis of the quantification of the image data mapping in real-space because the calculated distribution presents the theoretical limit of the confidence of the latent parameter estimation by using image data where the model form has been known.

The original content of this work is found in Ref. [15].

Effects of Spatial Down-sampling on Model Parameter Estimation in GMRF

In Chapter 4, we show the effects of down-sampling on the posterior distribution estimation of the model parameters of GMRF. We exhibit the effect on the latent parameter estimation by the down-sampling which is the essential and inevitable

process for imaging techniques since spatially continuous objects are recorded as discrete data. In this study, first, we formulate a generative model and a cognitive one of imaging processes with down-sampling by Markov random field models. Next, by applying Bayes' theorem to the cognitive model, we explain a method of estimating latent parameters in the model. Then, we conduct numerical simulations to examine the Bayesian posterior distribution and indicate that down-sampling causes inherent biases. Finally, we discuss the biases focusing on the relation between the generative and cognitive models. It is necessary for considering the quantification of image data to evaluate the effect of spatial down-sampling.

We plan to submit this work to a journal.

Influence of Averaging Preprocessing on Image Analysis by using GMRF [17]

Chapter 5 describes our investigations into the influence of averaging preprocessing on the performance of image analysis. Averaging preprocessing involves a trade-off: image averaging is often undertaken to reduce noise while the number of image data available for image analysis is decreased. We formulated a process of generating image data by using a GMRF model to achieve image analysis tasks such as image restoration and latent parameter estimation by a Bayesian approach. According to the notions of Bayesian inference, posterior distributions were analyzed to evaluate the influence of averaging. There are three main results. First, we found that the performance of image restoration with a predetermined value for latent parameters is invariant regardless of whether we average the images. We then found that the performance of hyper-parameter estimation deteriorates due to averaging. Our analysis of the negative logarithm of the posterior probability, which is called free energy based on an analogy with statistical mechanics, indicated that the confidence of hyper-parameter estimation remains higher without averaging. Finally, we found that when we estimate the latent parameters from the data, the performance of image restoration worsens as we reduce the images by averaging. We conclude that averaging adversely affects the performance of image analysis through latent parameter estimation. This work contributes to the design of measurements by clarifying the adverse effects of the preprocessing. We can treat the averaging as the basic preprocessing for image data, and the result suggests that preprocessing make the performance of model parameter estimation to deteriorate.

The original content of this work is found in Ref. [17].

The work in Chapter 6 study the estimation of the latent parameter which is the interaction parameter of the crystal lattice by using the image mapped to Fourier-space. The image mapped to Fourier-space is obtained the spectral data. We use the Bayesian framework to estimate the model parameters where the spectra are treated as input data. We introduce the previous work of the Bayesian spectral analysis in Sec. 2.3. Our novelty comparing with the previous work is to estimate the model parameters by

using the dispersion relation spectra for analyzing some crystals based on the Bayesian framework. In Chapter 6, we proposed the two methods to estimate the latent parameters and discuss the better method. Besides, we suggest the better way to sampling Fourier space on that estimation. In a conventional method, the experimenters visualize sample point in the Fourier-space which has high symmetricity [10]. However, the results of the estimates indicate that the sampling from a not symmetrical point is better rather than symmetric. As a consequence, we contribute in this chapter to both aims: to modeling of the image mapped to Fourier-space and to suggest the better design of the measurement on Fourier-space.

We describe the summary of the chapter as follows:

Bayesian Analysis of Crystal Lattice System Using Dispersion Relation Spectra

In Chapter 6, we propose two methods to estimate the model parameters which denote the interaction between the atom of the crystal lattice by using dispersion relation spectra that mirror the lattice vibration. One of the methods is the indirect method: the method estimates the model parameters via the estimation of the center of peaks of spectra, which has the same steps of the conventional method by the assessment in visual. The other one of the method is the direct method: the method estimates the model parameters by using whole data without preprocessing which is spectral deconvolution. We show the weak case of the previous spectral deconvolution [8], and we propose the method to be able to handle this case in the indirect method. We compare the results of both methods, and we show the advantages of the direct method. One of our contributions is to show the better method than the method which has the step to analysis as the same to the conventional method on latent parameter estimation of the image mapped in Fourier-space. The other one of our contribution is to propose the method to evaluate the better procedure to measure the Fourier space. The degenerating of the spectra of dispersion relation is used to find the symmetricity of the target crystal by visualizing the symmetric sampling points. However, it is not clear the better sampling procedure of the points in Fourier space. We propose the method for evaluation of the confidence by using Bayesian inference, and demonstrate the design of sampling procedures.

We plan to submit this work to a journal.

Chapter 2

Related Work

In this Chapter, we described related work of this thesis. Firstly, we introduce Markov random field (MRF) models. MRFs are the probabilistic model for the image processing tasks, and the models can represent the continuity on real-space mapping images. We use this model to treat the real-space mapping images. Secondly, we describe a spectral data analysis. The spectral data is a data type for the measurement mapped to the Fourier-space. Spectral deconvolution or spectral function fitting is the typical analysis of the spectral data. We utilize Bayesian spectral deconvolution to analyze spectra. Lastly, we overview the Bayesian computation to estimate hyper parameters such as the model parameters of the target of the measurement. We need to evaluate the confidence of the estimated hyper parameters by using estimate the posterior distribution of model parameters.

2.1 Markov Random Field Model

In this section, we introduce a summary of the development of Markov random field models. In particular, we focus on the relationship with the lattice system and Ising model in statistical mechanics. Then we introduce some applications in natural science data processing.

The word "Markov random field model" was introduced by Levy [18]. Levy discuss the spatial, probabilistic model like Brownian motion, that the values of each point is a continuous value. And this model is used for image processing by Woods [11] and Besag [19]. These MRF models are Gaussian MRF. That is, they modeled the gray-scale image as the value of each node is continuous. And then, the MRF models for digital image processing are proposed by Geman *et al.* [20] and Derin *et al.* [21] formulate the image pixels as the model of lattice systems of atoms, and they use the terminologies of statistical mechanics. Let us consider the original image whose pixels $\mathbf{x} = (x_1, x_2, \ldots, x_N), x_i \in \{0, 1, 2, \ldots, L-1\}$, where L is the number of the gradation of a pixel, and N is the number of the pixels. The probability of the existence of the original image is obtained

by the Gibbs distribution as the following distribution:

$$\pi(\mathbf{x}) = \frac{1}{Z} e^{-U(\mathbf{x})/T}, \tag{2.1}$$

where T is the constant denotes the temperture of this system, $U(\mathbf{x})$ is the energy function, and Z is the normalized constant denoted as

$$Z = \sum_{\mathbf{x}} e^{-U(\mathbf{x})/T}. \tag{2.2}$$

Z is called partition function, it plays important role in the statistical mechanics. Geman *et al.* give the general form of $U(\mathbf{x})$ as

$$U(\mathbf{x}) = \sum V_{i,j}(x_{i,j}) + \sum V_{(i,j),(i+1,j)}(x_{i,j}, x_{i+1,j}) + \sum V_{(i,j),(i,j+1)}(x_{i,j}, x_{i,j+1}). \tag{2.3}$$

For instance, when we assume the isotropic binary image ($L = 2$) and the image has the continuous state, the energy function is obtained by

$$U(\mathbf{x}) = \alpha \sum x_{i,j} + \beta \left(\sum x_{i,j} x_{i+1,j} + \sum x_{i,j} x_{i,j+1} \right). \tag{2.4}$$

By using these formulas, we can restore the image noise which is the typical image processing task. The filtering process for the image restoration is based on the smoothness of the image [22]. The smoothness: a property that the neighbor pixel values are close is known feature of the spins of ferromagnetic materials [22, 23]. Actually, the Ising model which is the simplest model of magnetization has the same formulation to the MRF model of the binary image. The spin system is the model to represent the magnetization which is the total value of the magnetization of each spin. Let us consider the case of the simplest Ising spin system. We assume that the value of spin S_i is $+1$ or -1, where $i = 1, \ldots, N$ is the index of spin-lattice point, and N is the number of atoms. For instance of the lattice, we assume the square lattice as shown in Fig. 2.1. The Hamiltonian of this spin system is defined by

$$H = -J \sum_{<i,j> \in B} S_i S_j - h \sum_{i=1}^{N} S_i, \tag{2.5}$$

where the J denotes interaction whose value is positive in the case of a ferromagnetic system, h denotes external field whose value is 0 when the magnetic field does not exist, and j is the index of the lattice. Let B be the set of interaction worked, and we assume it is the set of nearest neighbor spin. In Eq. (2.5), the more spins faces the same direction, the smaller value the Hamiltonian has. In statistical mechanics, the thermal average is calculated by using the following probability distribution:

$$P(\mathbf{S}) = \frac{e^{-\beta H}}{\sum_{\mathbf{S}} e^{-\beta H}}. \tag{2.6}$$

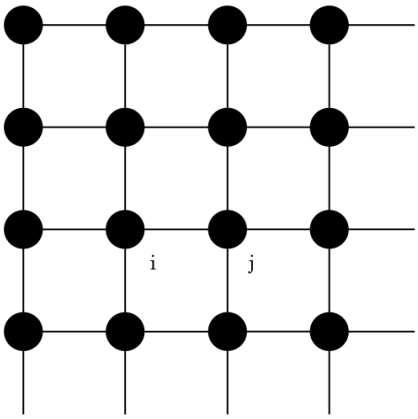

Figure 2.1: Sketch of two dimensional square lattice.

It can be clearly seen that the binary MRF model is equivalent to the model of the Ising spin model by Eqs. (2.1), (2.2), (2.4), (2.5), and (2.6). It has been pointed out that the Bayesian inference procedure is the equivalent framework of the statistical mechanics of spin systems [24, 23]. By using Bayesian frameworks, some image processing method is proposed such as image restoration [20, 21, 25, 26] and image segmentation [27, 28, 29]. On the other hand, by using replica method of the statistical mechanics, Nishimori *et al.* analyze the performance of image restoration of the binary MRF model [30]. Consequently, estimation of the interaction parameters of the image pixels and the lattice spin systems can be performed by using the same framework.

The interaction parameters of MRF models are important meaningful parameters for natural science. Nakanishi *et al.* shows that the interaction parameters of GMRF correspond to the diffusion coefficient of the diffusion equations [16]. They proposed the estimation method of the model parameter distribution of GMRF. Simmons *et al.* suggest that the interaction parameters mirror the wetting behavior which corresponds to the surface tension on the mesoscale structure of the target material [31].

2.2 Spectral Analysis

In this thesis, we treat the spectral data analysis by using Bayesian inference. We describe the previous work in this section.

In an experiment using some diffraction, experimenters obtain the intensity distribution by watching the diffraction patterns of X-ray, light, or neutron beams [4]. This intensity distribution is coupled with the distribution of the scatterers by Fourier transforms. In an analysis of scatterers, the structure factor $|F_S(\mathbf{q})|^2$

$$I_s(\mathbf{q}) \propto |F_S(\mathbf{q})|^2$$

is calculated, where the diffraction vector $\mathbf{q} = \mathbf{s}_0 - \mathbf{s}$, $\mathbf{s}$ is the incident wave, and $\mathbf{s}_0$ is the diffraction wave. Then, the scattering power distribution $f(\mathbf{r})$ is defined by

$$f(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d\mathbf{q} F(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}},$$

which is the distribution of scatterers such as electrons or nucleus. The $F(\mathbf{q})$ has a magnitude $|F(\mathbf{q})|$ and a phase $\phi(\mathbf{q})$ related by the equation

$$F(\mathbf{q}) = |F(\mathbf{q})|\, e^{i\phi(\mathbf{q})}.$$

For that reason that we can obtain only this structure factor $|F_S(\mathbf{q})|^2$ without the phase information, we cannot calculate $f(\mathbf{r})$ directly. This difficulty on the calculation of the scattering power distribution is called a phase retrieval problem. The analysis of scatterers is used the two conventional methods [4]. One is the model-based method. The model-based method has two-step. The first step is the calculation of $|\hat{F}_S(\mathbf{q})|^2$ by using an assumed model and model parameters, and the second step is comparing with calculated $|\hat{F}_S(\mathbf{q})|^2$ and observed $|F_S(\mathbf{q})|^2$. The experimenter repeats the above two steps with different parameters and models until they are satisfied. In most cases, the comparisons of experimenters are made based on a subjective visual assessment. Therefore the assessment may be arbitrary. The other one is the statistical method. In the statistical method, the steps are almost the same as the model-based method, but they compare the autocorrelation functions obtained by using $|F_S(\mathbf{q})|^2$ and an assumed model. What we are concerned here is that the experimenter cannot confirm the uniqueness and confidence of the model obtained by using both methods. The Bayesian inference is the method to solve those problems.

The diffraction patterns is obtained as the spectral data. The spectral data is formulated by the radial basis function networks and it is defined by

$$f(x) = \sum_{k}^{K} a_k \Phi_k(x; \theta) \tag{2.7}$$

where $\Phi_k(x)$ is the radial basis function and θ is the parameters of $\Phi_k(x)$ such as the location of centers or bandwidth of peaks. One of the purposes of the spectral analysis is to deconvolute to basis functions from the spectral data. This task is called spectral deconvolution. In some previous study, the Bayesian spectral deconvolution algorithms are proposed[32, 33, 34, 8].

For example, we describe the case that the spectra consisted by Gaussian basis and

Gaussian error. The likelihood $P(\mathbf{y}|\theta, \mathbf{a}, \sigma_{noise})$

$$\begin{aligned} P(\mathbf{y}|\theta, \mathbf{a}, \sigma_{\mathbf{noise}}) &= \prod_i \mathcal{N}\left(f(x_i;\theta,\mathbf{a}), \sigma^2_{\text{noise}}\right) \\ &= \prod_i \left[\frac{1}{\sqrt{2\pi\sigma_{\text{noise}}}}\right] \exp\left[-\frac{N}{\sigma^2_{\text{noise}}} E(f(x_i;\theta,\mathbf{a}))\right], \end{aligned} \tag{2.8}$$

$$E(f(x_i;\theta,\mathbf{a})) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - f(x_i;\theta,\mathbf{a})\right)^2 \tag{2.9}$$

is defined by the model of the Gaussian observation error, where $\mathbf{a} = \{a_k\}$, $y_i, x_i, (i = 1, \cdots N)$ are the data and its observed point, σ_{noise} is the noise magnitude. When the data set $D = (\mathbf{x}, \mathbf{y})$ is given, we can obtain the posterior $P(\theta|D)$ is obtained

$$\begin{aligned} P(\theta|D) &= \frac{P(D|\theta)P(\theta)}{P(D)} \\ &\propto \exp\left[-\frac{N}{\sigma^2_{\text{noise}}} E(f(x_i;\theta,\mathbf{a}))\right] P(\theta) \end{aligned} \tag{2.10}$$

by Bayes ' theorem. Previous studies estimate the posterior distributions by using Markov Chain Monte Carlo (MCMC) sampling methods such as reverse-jump Monte Carlo (RJMC)[35] or replica exchange Monte Carlo method (REMC)[35]. We described the replica exchange Monte Carlo algorithm in Sect. 2.3.2. Some algorithms can estimate the parameters of the spectrum and select the number of peaks K by sampling K[33, 34] or by optimizing free energy[8]

$$F(K;D) = -\log \int d\theta P(\theta|D) \tag{2.11}$$

which is minus of logarithm of the partition function.

The spectral deconvolution is practical, therefore developed for natural science ,especially physics and chemistry [36, 37, 38, 39]. The spectral deconvolution is utilized to estimate the target model parameters (more "deep" parameters). Murata *et al.* estimate the parameters of the autoregressive model which is a probabilistic model for time series data [40]. In this study, they proposed the method to obtain the parameters of latent dynamics by using time-dependent spectra. Kasai *et al.* proposed the method to deconvolute the spectrum and obtain the target information which is to determine the amino acid [41].

2.3 Computation of Posterior Distribution

In this section, we describe the Bayesian computation for the posterior distribution estimation of latent parameters. Let us consider the latent parameter a and the

measurement target x occurs, and then we obtain the data y by observation. We formulate the probability distribution of the process to generate the event $P(x|a)$, and the distribution of the process of observation $P(y|x)$. The distribution $P(x|a)$ which corresponds to the hypothesis of the event is the prior distribution of x. The probability $P(a|y)$ is what we want to know, called the posterior distribution. To estimate latent parameters of the model is calculated by integration with respect to the visible random variable such as the image pixel values, as follows

$$\begin{aligned} P(a|y) &= \frac{P(a,y)}{P(y)} \\ &= \int dx \frac{P(a,x,y)}{P(y)} \\ &= \int dx \frac{P(y|x)P(x|a)P(a)}{P(y)}. \end{aligned}$$

This integration can be costly for computation in the case that the dimension of the visible variables is high. We can calculate analytically the problems whose prior distribution of the latent parameter is integrable. For example, Gaussian Markov random field models whose posterior can be analytically calculated are the special case. In general, we chose the computational method in two way: variational Bayesian learning or Sampling methods. The variational Bayesian learning is the method to calculate approximately the posterior whose distribution form is limited. We outline the variational inference for the latent parameter estimation in Sect. 2.3.1. On the other hand, the sampling method computes the posterior distribution by using Monte Carlo integration with the sample sequence obtained by unnormalized posterior. We describe the MCMC methods which are the typical sampling method in Sect. 2.3.2. We used MCMC sampling method for the computation of the posterior in Chapter 6.

2.3.1 Variational Inference

Variational inference is the method to calculate approximately the intractable integration. In this section, we mainly focus on the procedure for posterior distribution estimation. Let us consider the parameter w and data D, the posterior $P(w|D)$ is defined by Bayes' theorem as follows

$$\begin{aligned} P(w|y) &= \frac{P(y|w)P(w)}{P(y)} \\ &= \frac{P(y|w)P(w)}{\int dw P(y|w)P(w)}, \end{aligned} \tag{2.12}$$

where, $P(D|w)$ is the probability distribution of observation and $P(w)$ is the prior.

In this method, we fixed the form of the posterior

$$P(w|y) \sim Q(w;\theta) \tag{2.13}$$

as the approximation, where θ is hyperparameters for instance mean and variance of the distribution. We chose $Q(w;\theta)$ what we can calculate easily such as a Gaussian and a conditionally independent distribution with respect to multivariate random variables. The distribution estimation problem is simplified to estimate θ, therefore, the distribution form has been fixed. The hyperparameter θ is obtained commonly by minimization of KL-divergence

$$D_{KL}(Q||P) \equiv \sum_w Q(w) \ln \frac{Q(w)}{Q(w|y)}. \tag{2.14}$$

The variational inference is a flexible method for computation. The accuracy of the approximation depends on the way of calculation of the KL-divergence and the function type of the approximated posterior. Mean field approximation assumes that the random variables independent as follows:

$$P(w_1, w_2) \sim P(w_1)P(w_2) \tag{2.15}$$

This approximation is used for various Bayesian learning algorithms. Although those algorithms are practical, sometimes changes the shape of the approximated posterior substantially.

2.3.2 Sampling method by Markov Chain Monte Carlo

It is necessary to acquire the posteriors and to compute the partition functions. The integration of those partition functions requires an efficient and accurate numerical calculation method. We perform these integrations by sampling method. The probability that X is in $[a_X, b_X]$ of the random variable X is expressed as

$$\begin{aligned} \pi(a_X \leq X \leq b_X) &= \int_{a_X}^{b_X} dX\, \pi(X) && (2.16)\\ &= E\left[a_X \leq X \leq b_X\right]. && (2.17) \end{aligned}$$

When we have a sequence of variables $x_i (i = 1, 2, \ldots N)$, this expectation can be approximated by using Monte Carlo integration defined as

$$\begin{aligned} E\left[a_X \leq X \leq b_X\right] &\sim \frac{1}{M} \sum_i^M R(x_i; a_X, b_X), && (2.18)\\ x_i &\text{ i.i.d } \nu(X), && (2.19)\\ \pi(X) &\propto \nu(X), && (2.20) \end{aligned}$$

where

$$R(x_i; a_X, b_X) = \begin{cases} 1 & (a_X \leq \chi_i \leq b_X) \\ 0 & (\text{otherwise}) \end{cases}. \tag{2.21}$$

We need a sampling method to obtain a sequence of random variables following the function $\nu(x)$. Hence, we describe REMC method which is an efficient sampling method for calculation of the posteriors and free energies. REMC algorithm includes a conventional MCMC method, so we first explain Metropolis-Hastings algorithm, which is a basic MCMC method[42, 43, 44]. Actually, the Gibbs sampling method which is a special case of the Metropolis-Hastings algorithm is used frequently[20, 45], and used in this chapter. Metropolis-Hastings algorithm is explained in algorithm 1.

Algorithm 1 Metropolis-Hastings algorithm

Sample the sequence of random variables following the target probability distribution $\pi(X)$. We assume that the value of the function $\nu(X) \propto \pi(X)$ can be calculated where the candidate-generating distribution is symmetric: $g(x_{i+1}|x_i) = g(x_i|x_{i+1})$
Initialization: Choose an arbitrary sample x_0.
for $i \in (1 \ldots, M)$

Generate the sample candidate $\tilde{x}_{i+1}$ following the candidate-generating distribution $g(\tilde{x}_{i+1}|x_i)$.

Calculate $r = \frac{\nu(\tilde{x}_{i+1})}{\nu(x_i)}$.

Generate a random number p following the uniform distribution on $[0, 1]$.

if $p < \min(1, r)$

accept: $x_{i+1} = \tilde{x}_{i+1}$

else

reject: $x_{i+1} = x_i$

The Metropolis-Hastings algorithm obtains a sequence of random variables that converges to the target distribution under satisfactory conditions[46]. One of the condition for this convergence is that the sample can be moved in entire range of the target function defined. However,if the target distribution has local minima, the Metropolis-Hastings algorithm requires a very long time to move between local solutions because the sample is trapped[47]. Simulated annealing is one way of dealing with such local minima[48]. In the simulated annealing algorithm, a target function is expressed as the following Boltzmann distribution:

$$P(\theta|\mathbf{D};\beta) = \frac{\exp\left[-\beta H(\theta, D)\right]}{\int d\theta \exp\left[-\beta H(\theta, D)\right]}, \tag{2.22}$$

where θ is the parameter, and D is the data. An inverse temperature parameter β corresponds to the magnitude of noise. If the temperature is high, the noise increases. The state following the Boltzmann distribution can escape the local minimum by increasing

or decreasing β. Simulated annealing is an effective algorithm for optimizing the target function but not for sampling.

The REMC method is the method for sampling from the joint distribution of the Boltzmann distributions which have the each parameter of inverse temperatures $\beta_l (l = 1, \ldots, L), \beta_L = 1$:

$$P(\theta_1, \ldots, \theta_L) = \prod_{l=1}^{L} \frac{\exp\left[-N\beta_l E(\theta_l)\right] \psi(\theta_l)}{\int d\theta_l \exp\left[-N\beta_l E(\theta_l)\right] \psi(\theta_l)} \tag{2.23}$$

$$= \prod_{l=1}^{L} \zeta(\theta_l; \beta_l). \tag{2.24}$$

The REMC algorithm can prevent samples from being trapped by parallel sampling and exchanging the samples from each Boltzmann distribution[47]. The detailed algorithm of REMC is explained in algorithm 2.

The REMC algorithm is not only effective for dealing with local minima, but also efficient for the parallel computation of free energies or partition functions[49, 50]. The partition function of $\zeta(\theta_l; \beta_l)$ can be written as

$$z(\beta) = \int d\theta \exp\left[-N\beta E(\theta)\right] \psi(\theta) \tag{2.25}$$

when a sequence of inverse temperatures $0 = \beta_1 < \beta_2 < \cdots < \beta_L = 1$ is given, $z(1)$ which is equal to the distribution function of interest, is obtained as

$$z(1) = \frac{z(\beta_L)}{z(\beta_{L-1})} \times \frac{z(\beta_{L-1})}{z(\beta_{L-2})} \cdots \times \frac{z(\beta_2)}{z(\beta_1)} \tag{2.26}$$

$$= \prod_{l=1}^{L-1} \frac{z(\beta_{l+1})}{z(\beta_l)}$$

$$= \prod_{l=1}^{L-1} \frac{\int d\theta \exp\left[-N\beta_{l+1} E(\theta)\right] \psi(\theta)}{\int d\theta \exp\left[-N\beta_l E(\theta)\right] \psi(\theta)}$$

$$= \prod_{l=1}^{L-1} \frac{\int d\theta \exp\left[-N(\beta_{l+1} - \beta_l) E(\theta)\right] \exp\left[-N\beta_l E(\theta)\right] \psi(\theta)}{\int d\theta \exp\left[-N\beta_l E(\theta) \psi(\theta)\right]}$$

$$= \prod_{l=1}^{L-1} \left\langle \exp\left[-N(\beta_{l+1} - \beta_l) E(\theta)\right] \right\rangle_{\zeta(\theta_l; \beta_l)}, \tag{2.27}$$

where $< \cdot >_{\zeta(\theta_l; \beta_l)}$ denotes an expected value with respect to $\zeta(\theta_l; \beta_l)$. This expected value in Eq. (2.27) can be calculated using samples generated by REMC in parallel. The free energy of Bayesian inference is defined by

$$F(\beta) = -\frac{1}{\beta} \ln z(\beta) \tag{2.28}$$

We have to estimate optimal value of β because β represents the unknown noise magnitude of the system. The method to estimate β is proposed by Tokuda *et al.*[51]. Let us consider the minimization of the free energy. The partial derivative of $F(\beta)$ with respect to β is

$$\frac{\partial F}{\partial \beta} = N\left[\langle E(\theta)\rangle_{\beta} - \frac{1}{2\beta}\right] \tag{2.29}$$

by the approach based on the empirical Bayes framework. They show the optimal $\hat{\beta}$ satisfies the following equation:

$$\langle E(\theta)\rangle_{\hat{\beta}} = \frac{1}{2\hat{\beta}}. \tag{2.30}$$

By using Eq. (2.30), we estimate optimal β, and then we use the sample sequence on $\hat{\beta}$ to calculate the expected value.

Algorithm 2 Replica exchange Monte Carlo method

Sample the sequence of random variables following the target joint distribution $\Upsilon(\theta_1, \ldots, \theta_L) = \prod_{l=1}^{L} \upsilon(\theta_l; \beta_l)$.
Initialize: Choose arbitrary initial samples $x_{i=0,l=1,\ldots,L}$
for $i \in (1 \ldots, M)$

Update samples $x_{i,l=1,\ldots,L}$ by an MCMC method such as the Metropolis-Hastings or Gibbs sampling algorithm.

Exchange process between temperatures at $l, l+1$.

Calculate u and v: $u = \min(1, v), v = \frac{q(\theta_{l+1};\beta_l)q(\theta_l;\beta_{l+1})}{q(\theta_l;\beta_l)q(\theta_{l+1};\beta_{l+1})}$

Generate a random number p following the uniform distribution on $[0, 1]$.

if $p < \min(1, r)$

Exchange $x_{i,l}$ and $x_{i,l+1}$.

Chapter 3

Theory of Distribution Estimation of the Model Parameters in GMRF

Markov random fields (MRFs) have been developed in image processing,[20, 25, 27, 52, 28, 53, 29] and their performance is sensitive to hyper-parameters that represent the smoothness of an image and the magnitude of observation noise. [54, 55, 56, 22] A discrete Fourier transform (DFT) is essential for the theoretical analysis of Gaussian MRF.[26, 56, 57] For example, a DFT enables us to show that an image-restoration method based on Gaussian MRF models is equivalent to a Wiener filter. [26] The marginalization required for hyper-parameter estimation of a Gaussian MRF can be precisely carried out because the DFT allows us to diagonalize adjacency matrices. [56, 16]

Researchers in natural science have begun to use MRF models to extract latent structures of image data. [58, 59, 60, 61] A hyper-parameter of an MRF was pointed out to correspond to a diffusion coefficient. [16] Thus, not only do hyper-parameters need to be estimated, but also the confidence of estimates needs to be evaluated. A previous study proposed a method of distribution estimation in which the posterior distribution of hyper-parameters is calculated by utilizing the framework of Bayesian inference. [16] This method allows us to evaluate the confidence of estimates.

In this study, we theoretically confirm the performance of the method of distribution estimation when it is used to evaluate the confidence of estimates. The posterior distributions vary in shape, and the location of the peak depends on the randomness of the sample data. It is true that the typical distribution can be inferred from the sample mean of distributions, but we have to evaluate a typical distribution directly. Therefore, we analytically calculate the configurational average of the negative logarithm of posterior distribution, also known as that of free energy, with respect to data.

3.1 Formulation

We formulate our MRF model according to Nakanishi-Ohno et al.[16]. We focus on one-dimensional images, but the argument is applicable to high-dimensional integer lattices such as a square lattice. Let us denote an original image as $\mathbf{u} = (u_1, u_2, \ldots, u_N) \in \mathbb{R}^N$ and an observed image as $\mathbf{v} = (v_1, v_2, \ldots, v_N) \in \mathbb{R}^N$, where N is the number of pixels.

Each pixel value v_i of an original image is given by

$$v_i = u_i + n_i, \tag{3.1}$$

where n_i represents measurement noise. We assume that $\mathbf{n} = (n_1, n_2, \ldots, n_N) \in \mathbb{R}^N$ are independent and identically distributed random variables that follow a normal distribution, the mean and variance of which are 0 and b, respectively (denoted as $\mathcal{N}(0, b)$). Then, an observed image $\mathbf{v}$ follows a multivariate normal distribution defined by

$$P_1(\mathbf{v}|\mathbf{u}, b) = \frac{1}{Z_1(b)} \exp\left[-\frac{1}{2b}\sum_{i=1}^{N}(v_i - u_i)^2\right], \tag{3.2}$$

where the function Z_1 is called a partition function and defined by

$$Z_1(b) = \int d\mathbf{v} \exp\left[-\frac{1}{2b}\sum_{i=1}^{N}(v_i - u_i)^2\right]. \tag{3.3}$$

Variable b is called a hyper-parameter representing the magnitude of measurement noise. We also consider a more general case in which T images $\{\mathbf{v}^t\}_{t=1}^T = \{\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^T\}$ are obtained independently by observing the same original image $\mathbf{u}$. In such a case, the probability distribution of a set of observed images $\{\mathbf{v}^t\}_{t=1}^T$ is given by the product of probability distributions, Eq. (5.2), as follows:

$$P_T(\{\mathbf{v}^t\}_{t=1}^T|\mathbf{u}, b) = \prod_{t=1}^{T} P_1(\mathbf{v}^t|\mathbf{u}, b). \tag{3.4}$$

We take the smoothness of an original image $\mathbf{u}$ into account to assign a high probability to an image, for which its values of neighboring pixels are close to each other, as follows:

$$P_{\mathrm{pri}}(\mathbf{u}|a) = \frac{1}{Z_{\mathrm{pri}}(a)} \exp\left[-\frac{1}{2a}\sum_{i=1}^{N}(u_{i+1} - u_i)^2\right], \tag{3.5}$$

where the function Z_{pri} is called a partition function and is defined by

$$Z_{\mathrm{pri}}(a) = \int d\mathbf{u} \exp\left[-\frac{1}{2a}\sum_{i=1}^{N}(u_{i+1} - u_j)^2\right]. \tag{3.6}$$

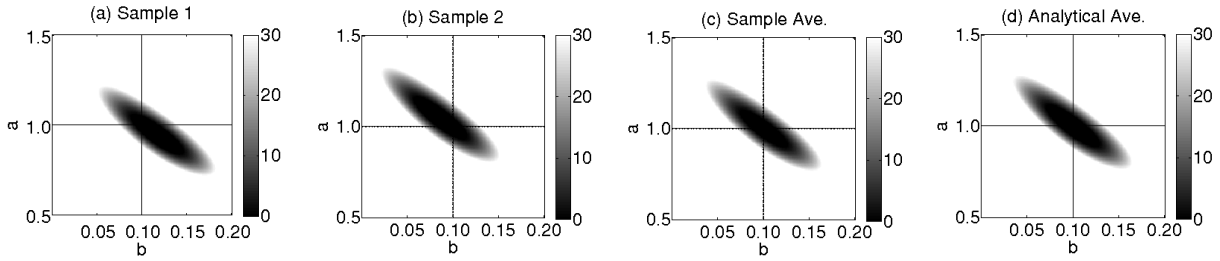

Figure 3.1: Free energies. Figs. (a) and (b) are $F(a, b|\{\mathbf{v}^t\}_{t=1}^T)$ obtained by numerical experiments by using synthetic observed images. Fig. (c) is the arithmetic mean of 1000 trials. Fig. (d) is the configurational average of $\langle F(a, b|\{\mathbf{v}^t\}_{t=1}^T)\rangle$ that is calculated analytically. The images are two dimensional with the number of pixels $N = 128^2$. The number of images is $T = 1$. The true hyper-parameters are $(a_0, b_0) = (1, 0.1)$.

For convenience of analysis, a periodic boundary condition, $u_{N+1} = u_1$, is imposed. Variable a is called a hyper-parameter. The smoothness of an original image increases as this hyper-parameter is set to a small value. This probability function $P_{\rm pri}$, called a prior distribution, corresponds to the diffusion equation, and a is related to the reciprocal number of the diffusion coefficient.

3.2 Analysis

The prior distribution of the hyper-parameters is given by

$$P(a, b) \propto \text{constant}, \tag{3.7}$$

which means that there is no prior knowledge of the hyper-parameters.

Equation (5.4) is rewritten in quadratic form:

$$P_{\rm pri}(\mathbf{u}|a) = \frac{1}{Z_{\rm pri}(a)} \exp\left[-\frac{1}{2a}\mathbf{u}^{\rm T}\Lambda\mathbf{u}\right], \tag{3.8}$$

where the superscript character T represents the transpose. An N-by-N matrix $\mathbf{\Lambda}$, the elements Λ_{ij} of which are 2 if $i = j$, -1 if i is adjacent to j, and 0 otherwise. Matrix $\mathbf{\Lambda}$ can be diagonalized by the DFT because it is a circulant matrix.[26, 56, 57]

By using the DFT defined as

$$\tilde{u}_k = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u_i \exp\left[\frac{2\pi {\rm i}(i \cdot k)}{N}\right], \tag{3.9}$$

$$\tilde{v}_k^t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_i^t \exp\left[\frac{2\pi {\rm i}(i \cdot k)}{N}\right], \tag{3.10}$$

the Fourier representation of the prior probability distribution is given by

$$P_{\rm pri}(\tilde{\mathbf{u}}|a) = \frac{1}{Z_{\rm pri}(a)} \exp\left[-\frac{1}{2a}\tilde{\mathbf{u}}^{\rm T}\tilde{\Lambda}\tilde{\mathbf{u}}\right], \tag{3.11}$$

and each element of $\tilde{\Lambda}$ is calculated as follows:

$$\begin{aligned}\tilde{\Lambda}_{kl} &= \frac{1}{N}\sum_{i,j}\exp\left(\frac{2\pi{\rm i}}{N}li\right)\Lambda_{ij}\exp\left(\frac{2\pi{\rm i}}{N}jk\right), \\ &= \lambda_k\delta_{kl}, \end{aligned}\tag{3.12}$$

where δ_{kl} is the Kronecker delta, and where λ_k is the diagonal elements of $\tilde{\Lambda}$ and is given by

$$\lambda_k = 2 - 2\cos\left(\frac{2\pi k}{N}\right). \tag{3.13}$$

Then, we obtain a diagonalized representation of the prior probability distribution as follows:

$$P_{\rm pri}(\tilde{\mathbf{u}}|a) = \prod_{k=1}^{N}\left\{\sqrt{\frac{\lambda_k}{2\pi a}}\exp\left[-\frac{1}{2a}\lambda_k\left|\tilde{u}_k\right|^2\right]\right\}. \tag{3.14}$$

A probability distribution of a set of observed images in diagonalized form is obtained in a similar way as

$$P_T(\tilde{\mathbf{v}}|\tilde{\mathbf{u}}, b) = \prod_{k=1}^{N}\left\{\left(\sqrt{\frac{1}{2\pi b}}\right)^T\exp\left[-\frac{1}{2b}\sum_{t=1}^{T}\left|\tilde{v}_k^t - \tilde{u}_k\right|^2\right]\right\} \tag{3.15}$$

by using Eqs. (5.3), (3.9), (3.10).

First of all, we focus on image restoration using our MRF models. According to Bayes' theorem, the posterior distribution of an original image given a set of observed images is as follows:

$$P(\mathbf{u}|\{\mathbf{v}^t\}_{t=1}^T, \hat{a}, \hat{b}) = \frac{P_T(\{\mathbf{v}\}_{t=1}^T|\mathbf{u}, \hat{b})P_{\rm pri}(\mathbf{u}|\hat{a})}{P(\{\mathbf{v}^t\}_{t=1}^T|\hat{a}, \hat{b})}, \tag{3.16}$$

where $\hat{a}, \hat{b}$ are estimates of hyper-parameters. By using Eqs. (3.14), (3.15), the Fourier

representation of $P(\mathbf{u}|\{\mathbf{v}^t\}_{t=1}^T, \hat{a}, \hat{b})$ is given by

$$\begin{aligned} & P(\tilde{\mathbf{u}}|\{\tilde{\mathbf{v}}^t\}_{t=1}^T, \hat{a}, \hat{b}) \\ \propto & \prod_{k=1}^N \left\{ \exp\left[-\frac{1}{2\hat{b}} \sum_{t=1}^T \left|\tilde{v}_k^t - \tilde{u}_k\right|^2 - \frac{\lambda_k}{2\hat{a}} |\tilde{u}_k|^2 \right] \right\} \\ = & \prod_{k=1}^N \left\{ \exp\left[-\frac{1}{2} \left(\frac{\lambda_k}{\hat{a}} + \frac{T}{\hat{b}} \right) \right. \right. \\ & \left. \left. \left| \tilde{u}_k - \frac{T}{\hat{b}} \left(\frac{\lambda_k}{\hat{a}} + \frac{T}{\hat{b}} \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tilde{v}_k^t \right|^2 \right] \right\}. \end{aligned} \tag{3.17}$$

Then, we obtain a restored image, by using the maximum *a posteriori* (MAP) estimation, as follows:

$$\tilde{u}_k^{\mathrm{MAP}} = \frac{T}{\hat{b}} \left(\frac{\lambda_k}{\hat{a}} + \frac{T}{\hat{b}} \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tilde{v}_k^t. \tag{3.18}$$

This MAP estimate by utilizing a Gaussian MRF is known to be equivalent to an image obtained by Wiener filter[26].

Next, to evaluate the performance of this image restoration, we will analyze a configurational average of mean squared error (MSE) with respect to data defined by

$$\mathrm{MSE}_{\mathrm{ave}} = \frac{1}{N} \left\langle \sum_{i=1}^N \left(u_i - u_i^{\mathrm{MAP}} \right)^2 \right\rangle_{\mathbf{u},\mathbf{v}|a_0,b_0}, \tag{3.19}$$

where $\langle\cdot\rangle_{\mathbf{u},\mathbf{v}|a_0,b_0}$ is the expected value with respect to $P(\mathbf{u},\mathbf{v}|a_0,b_0) = P_T(\{\mathbf{v}^t\}_{t=1}^T|\mathbf{u},b_0)P_{\mathrm{Pri}}(\mathbf{u}|a_0)$. By using DFT as Eqs. (3.9), (3.10), $\mathrm{MSE}_{\mathrm{ave}}$ is rewritten as follows:

$$\left\langle \sum_{i=1}^N \left(u_i - u_i^{\mathrm{MAP}} \right)^2 \right\rangle_{\mathbf{u},\mathbf{v}|a_0,b_0} = \left\langle \sum_{k=1}^N \left| \tilde{u}_k - \tilde{u}_k^{\mathrm{MAP}} \right|^2 \right\rangle_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}|a_0,b_0} \tag{3.20}$$

By utilizing Eqs. (3.14), (3.15), (3.18), (3.19) and the formula of the Gaussian integral, we obtain $\mathrm{MSE}_{\mathrm{ave}}$ as follows:

$$\begin{aligned} \mathrm{MSE}_{\mathrm{ave}} = & \frac{1}{N} \int d\tilde{\mathbf{u}} d\tilde{\mathbf{v}}^{t=1} \cdots d\tilde{\mathbf{v}}^{t=T} \sum_{k=1}^N \left| \tilde{u}_k - \tilde{u}_k^{\mathrm{MAP}} \right|^2 \\ & P_T(\{\tilde{\mathbf{v}}^t\}_{t=1}^T|\tilde{\mathbf{u}},b_0) P_{\mathrm{Pri}}(\tilde{\mathbf{u}}|a_0) \\ = & \frac{1}{N} \sum_{k=1}^N \left\{ \frac{a_0}{\lambda_k} - 2 \left(\frac{\frac{a_0\hat{a}}{\lambda_k^2}}{\frac{\hat{a}}{\lambda_k} + \frac{\hat{b}}{T}} \right) + \frac{\hat{a}^2}{\lambda_k^2} \frac{\frac{a_0}{\lambda_k} + \frac{b_0}{T}}{\left(\frac{\hat{a}}{\lambda_k} + \frac{\hat{b}}{T} \right)^2} \right\}. \end{aligned} \tag{3.21}$$

We can see that the true hyper-parameters minimize $\mathrm{MSE}_{\mathrm{ave}}$ because the first derivatives of $\mathrm{MSE}_{\mathrm{ave}}$ with respect to $\hat{a}$ and $\hat{b}$ at the point of $(\hat{a}, \hat{b}) = (a_0, b_0)$ are equal to 0. These hyper-parameters correspond to the Nishimori temperature because they maximize the performance of image restoration.[23]

Finally, we will analyze the posterior distribution of the hyper-parameters. According to Bayes' theorem, the posterior distribution is given by

$$P(a, b|\{\mathbf{v}^t\}_{t=1}^T) = \frac{P(\{\mathbf{v}\}_{t=1}^T|a, b)P(a, b)}{P(\{\mathbf{v}^t\}_{t=1}^T)}. \tag{3.22}$$

In our case, the marginal likelihood $P(\{\mathbf{v}\}_{t=1}^T|a, b)$ on the right-hand side of Eq. (3.22) is given by

$$P(\{\mathbf{v}\}_{t=1}^T|a, b) = \int d\mathbf{u} P_T(\{\mathbf{v}^t\}_{t=1}^T|\mathbf{u}, b)P_{\mathrm{Pri}}(\mathbf{u}|a). \tag{3.23}$$

Then, we obtain

$$P(a, b|\{\mathbf{v}^t\}_{t=1}^T) = \frac{\int d\mathbf{u} P_T(\{\mathbf{v}^t\}_{t=1}^T|\mathbf{u}, b)P_{\mathrm{Pri}}(\mathbf{u}|a)P(a, b)}{P(\{\mathbf{v}^t\}_{t=1}^T)}. \tag{3.24}$$

We substitute our model, represented by Eqs. (3.14) and (3.15), into Eq. (3.24) and use the formula of the Gaussian integral to obtain

$$\begin{aligned} P(a, b|\{\tilde{\mathbf{v}}^t\}_{t=1}^T) &\propto \prod_{k=1}^N \left\{ \sqrt{\frac{\lambda_k}{2\pi a}} \left(\sqrt{\frac{1}{2\pi b}} \right)^T \right. \\ &\quad \left. \int d\tilde{u}_k \exp\left[-\frac{1}{2a}\lambda_k \tilde{u}_k^2 - \frac{1}{2b}\sum_{t=1}^T \left|\tilde{v}_k^t - \tilde{u}_k\right|^2 \right] \right\} \\ &\propto \prod_k \left\{ \sqrt{\frac{1}{\frac{a}{\lambda_k} + \frac{b}{T}}} \exp\left[-\frac{1}{2}\frac{1}{\frac{a}{\lambda_k} + \frac{b}{T}}\mu_k^2 \right] \right. \\ &\quad \left. \left(\sqrt{\frac{1}{b}} \exp\left[-\frac{1}{2b}\eta_k \right] \right)^{T-1} \right\}, \end{aligned} \tag{3.25}$$

where $\mu_k = \left|\sum_{t=1}^T \tilde{v}_k^t\right| /T$, $\eta_k = \sum_{t=1}^T |\tilde{v}_k^t - \mu_k|^2/(T-1)$. The variables μ_k and η_k represent the mean and variance, respectively, of the k-th Fourier component with respect to T observed images.

We can calculate the hyper-parameter posterior distribution by substituting synthetic image data into Eq. (5.31). [16] Figures 3.1(a) and (b) show the negative logarithm of

the posterior distribution. Let us call this quantity free energy F based on an analogy with statistical mechanics. We use two-dimensional images whose number of pixels is $N = 128^2$, and the values of the true hyper-parameters are set to $(a_0, b_0) = (1, 0.1)$, in Fig. 3.1 and Fig. 3.2. The number of images is $T = 1$ in Fig. 3.1. Distributions of F vary in location and shape as shown in Figs. 3.1(a) and 3.1(b), but Fig. 3.1(c) suggests that the arithmetic means of trials converge to a certain distribution. In this study, we show the typical F by calculating analytically configurational average with respect to data. As a result, we obtain Fig. 3.1(d), and this analytical configurational average is very similar to the numerical one in Fig. 3.1(c). We will describe the analysis of the configurational average of F in the following paragraphs.

The free energy F is as follows:

$$\begin{aligned} F(a, b|\{\mathbf{v}^t\}_{t=1}^T) &= -\ln P(a, b|\{\mathbf{v}^t\}_{t=1}^T) \\ &= -\frac{1}{2}\sum_k \left\{ \ln \frac{1}{\frac{a}{\lambda_k} + \frac{b}{T}} - \frac{1}{\frac{a}{\lambda_k} + \frac{b}{T}}\mu_k^2 \right. \\ &\quad \left. +(T-1)\left(\ln \frac{1}{b} - \frac{1}{b}\eta_k\right)\right\}. \end{aligned} \tag{3.26}$$

Note that the equal sign between the first and second lines means that both sides are equal up to a constant. We calculate the configurational average of F with respect to data:

$$\begin{aligned} &\langle F(a, b|\{\mathbf{v}^t\}_{t=1}^T)\rangle_{\{\mathbf{v}^t\}_{t=1}^T|a_0,b_0} \\ \equiv &\int \left(\prod_{t=1}^T d\mathbf{v}^t\right) F(a, b|\{\mathbf{v}^t\}_{t=1}^T) P(\{\mathbf{v}^t\}_{t=1}^T|a_0, b_0), \end{aligned} \tag{3.27}$$

where a_0 and b_0 are true values of the hyper-parameters. In previous work [16], the shape of the posterior distribution was evaluated numerically by substituting a few examples of synthetic data $\{\mathbf{v}^t\}_{t=1}^T$ into Eq. (5.31).

We now explain how to calculate this analytically. We substitute Eq. (3.23) into Eq. (3.27) to break down the expectation into two steps:

$$\begin{aligned} &\langle F(a, b|\{\mathbf{v}^t\}_{t=1}^T)\rangle_{\{\mathbf{v^t}\}_{t=1}^T|a_0,b_0} \\ = &\int d\mathbf{u} \left(\prod_{t=1}^T d\mathbf{v}^t\right) F(a, b|\{\mathbf{v}^t\}_{t=1}^T) P_T(\{\mathbf{v}^t\}_{t=1}^T|\mathbf{u}, b_0) P_{\rm pri}(\mathbf{u}|a_0) \\ \equiv &\left\langle \left\langle F(a, b|\{\mathbf{v}^t\}_{t=1}^T)\right\rangle_{\{\mathbf{v}^t\}_{t=1}^T|\mathbf{u},b_0} \right\rangle_{\mathbf{u}|a_0}. \end{aligned} \tag{3.28}$$

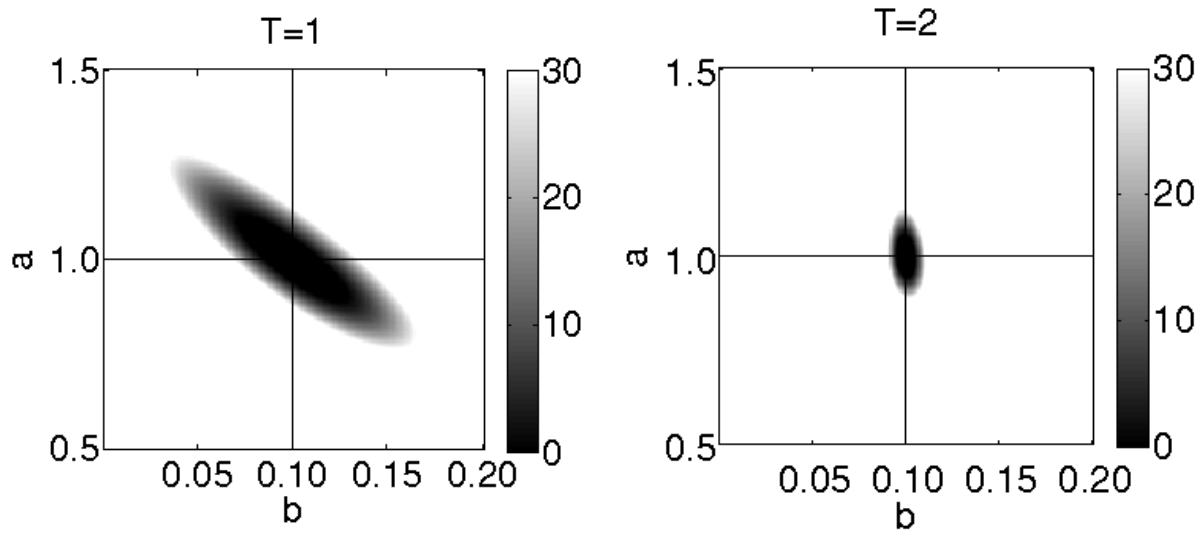

Figure 3.2: Configurational averages of free energy $\langle F(a,b|\{\mathbf{v}^t\}_{t=1}^T)\rangle$. Images are two-dimensional with number of pixels $N = 128^2$. True hyper-parameters are $(a_0, b_0) = (1, 0.1)$. The panels of the left and right columns are when $T = 1, 2$, respectively.

Equation (3.26) is substituted into Eq. (3.28) to obtain

$$\begin{aligned}\langle F(a,b|\{\mathbf{v}^t\}_{t=1}^T)\rangle &= -\frac{1}{2}\sum_k \left\{ \ln \frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}} - \frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}}\langle\langle \mu_k^2\rangle\rangle \right. \\ &\quad \left. +(T-1)\left(\ln\frac{1}{b} - \frac{1}{b}\langle\langle\eta_k\rangle\rangle\right)\right\}, \end{aligned} \tag{3.29}$$

where the subscripts of expectation are omitted. The two expected values $\langle\langle\mu_k^2\rangle\rangle$ and $\langle\langle\eta_k\rangle\rangle$ remain to be calculated.

Let us recall that μ_k and η_k are the sample mean and variance, respectively, of the k-th Fourier component of observed images. We see that variables $\tilde{v}_k^t$ are independent random variables that follow $\mathcal{N}(\tilde{u}_k, b_0)$ and that variables $\tilde{u}_k$ are independent random variables that follow $\mathcal{N}(0, a_0/\lambda_k)$.

Each random variable μ_k follows $\mathcal{N}(\tilde{u}_k, b_0/T)$ because μ_k is the sample mean of $\tilde{v}_k^t$ that follows $\mathcal{N}(\tilde{u}_k, b_0)$. The inner expectation of $\langle\langle\mu_k^2\rangle\rangle$ with respect to $\{v_k^t\}_{t=1}^T|\tilde{u}_k, b_0$ is calculated as follows:

$$\langle\langle\mu_k^2\rangle\rangle = \left\langle|\tilde{u}_k|^2\right\rangle_{\tilde{u}_k|a_0} + \frac{b_0}{T}. \tag{3.30}$$

The calculation of the other expectation $\left\langle|\tilde{u}_k|^2\right\rangle_{\tilde{u}_k|a_0}$ gives us

$$\langle\langle\mu_k^2\rangle\rangle = \frac{a_0}{\lambda_k} + \frac{b_0}{T}, \tag{3.31}$$

using the fact that $\tilde{u}_k$ follows $\mathcal{N}(0, a_0/\lambda_k)$.

Each η_k is the sample variance of $\tilde{v}_k^t$ that follows $\mathcal{N}(\tilde{u}_k, b_0)$. Let us denote

$$\eta_k' = \frac{T-1}{b_0}\eta_k, \tag{3.32}$$

and this random variable η_k' follows the chi-squared distribution with $T-1$ degrees of freedom[62]. Because the mean value of the chi-squared distribution is equal to its degrees of freedom, we obtain

$$\langle\langle \eta_k \rangle\rangle = \frac{b_0}{T-1}\langle\langle \eta_k' \rangle\rangle = b_0. \tag{3.33}$$

As a result, we substitute Eqs. (3.31) and (3.33) into Eq. (3.29) to obtain

$$\begin{aligned}\langle F(a,b|\{\mathbf{v}^t\}_{t=1}^T)\rangle &= -\frac{1}{2}\sum_k \left\{ \ln \frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}} - \frac{\frac{a_0}{\lambda_k}+\frac{b_0}{T}}{\frac{a}{\lambda_k}+\frac{b}{T}} \right. \\ &\quad \left. +(T-1)\left(\ln\frac{1}{b} - \frac{b_0}{b}\right)\right\}. \end{aligned} \tag{3.34}$$

Figure 3.1(d) presents the analytical configurational average of F obtained by Eq. (3.34). Fig. 3.1(c) is similar to Fig. 3.1(d). Our analysis is considered to be valid.

We now focus on the arguments of the minimum of $\langle F(a,b|\{\mathbf{v}^t\}_{t=1}^T)\rangle$. It is shown that the true hyper-parameters minimize $\langle F(a,b|\{\mathbf{v}^t\}_{t=1}^T)\rangle$ because the first derivatives of $\langle F(a,b|\{\mathbf{v}^t\}_{t=1}^T)\rangle$ with respect to a and b at the point of $(a,b) = (a_0, b_0)$ are equal to 0. We confirm a well-known fact that true hyper-parameters minimize this configurational average.

Figure 3.2 presents the analytical configurational average of F. In Fig. 3.2, the left and right panels are when $T = 1, 2$, respectively. We see that the breadth of distribution of F decreases as the amount of data increases from $T = 1$ to $T = 2$. This typical behavior of F corresponds to that of posterior distribution numerically examined by Nakanishi-Ohno et al.[16]. Hence, we theoretically confirm that distribution estimation can be used to evaluate the confidence of data.

We investigated the performance of the distribution estimation method proposed by Nakanishi-Ohno et al.[16]. Our analysis of free energy F validated the fact that the breadth of posterior distribution shrinks as the amount of data increases. In conclusion, we confirm that we can evaluate the confidence of data by analyzing the breadth of F and posterior distribution. It is also important to analyze differences in free energy among original images[63]. This can be carried out by calculating only the inner expectation in the last line of Eq. (3.28) to obtain

$$\begin{aligned}\left\langle F(a,b|\{\tilde{\mathbf{v}}^t\}_{t=1}^T)\right\rangle_{\{\tilde{\mathbf{v}}^t\}_{t=1}^T|\tilde{\mathbf{u}},b_0} &= -\frac{1}{2}\sum_k \left\{ \ln \frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}} - \frac{|\tilde{u}_k|^2+\frac{b_0}{T}}{\frac{a}{\lambda_k}+\frac{b}{T}} \right. \\ &\quad \left. +(T-1)\left(\ln\frac{1}{b} - \frac{b_0}{b}\right)\right\}. \end{aligned} \tag{3.35}$$

It is future work to analyze the data average of the performance of other image-processing such as image inpainting[64].

Chapter 4

Effects of Spatial Down-sampling on Model Parameter Estimation in GMRF

第 4 章は雑誌掲載が予定される内容を含むため，インターネット公表できません.

Chapter 5

Influence of Averaging Preprocessing on Image Analysis by using GMRF

Averaging as preprocessing is often performed on image data, which are typical high-dimensional data. Averaging has two main advantages: an improved signal to noise ratio during observations, and reduced data volume to conserve computational resources, such as the memory space and the computation time spent in postprocessing. However, there is a disadvantage, i.e., averaging can deteriorate the performance of the postprocessing. This is due to a certain amount of irreversible information loss due to averaging. There is a trade-off between the advantages and disadvantages of averaging. Therefore, we have to determine the extent of the signal to noise ratio, data volume, and information loss in the case of averaging image data. This problem is related to how to optimally adjust the frame rate when video data are acquired to enable dynamical systems to be studied. Thus, whether and how averaging should be performed is a common though not trivial problem in the analysis of image and video data in various fields of natural science.

We focused on Bayesian image analysis based on Markov random field (MRF) models. MRF models stochastically express the properties of image data and measurement processes to formulate image data. MRF models have been proposed for Bayesian image analysis[20, 21]. There are many applications in image analysis such as image segmentation and image restoration[25, 27, 52, 28, 53, 29, 64]. Our MRF model has two hyper-parameters which represent the smoothness of the original images and the magnitude of measurement noise. It is important to adjust these hyper-parameters in image analysis. Bayesian inference provides a hyper-parameter estimation method[55, 56, 22]. In addition, our recent research has found that the confidence of hyper-parameter estimation can be evaluated by investigating Bayesian posterior distributions[16, 15]. Such confidence evaluation methods play an important role in natural science because many hyper-parameters are related to physical quantities[16]. Actually, one of the hyper-parameters in our MRF model mathematically corresponds to a diffusion coefficient. This analogy has innovated geophysical research where seismic tomography data are

analyzed by using MRF models in a similar way to ours to reveal geofluid spatial distributions inside the earth[58, 59, 60].

In this study, we examined the influence of averaging preprocessing on image restoration and hyper-parameter estimation. First, we found that if predetermined hyper-parameters are used, the performance of image restoration does not depend on whether averaging is carried out. We then examined the influence of averaging preprocessing on hyper-parameter estimation. We found that the confidence of hyper-parameter estimation decreases as raw image data decreases in volume even though a higher signal to noise ratio is obtained by averaging. We also found that if estimated hyper-parameters are used, the accuracy of image restoration decreases as image data are averaged. Finally, we discuss the trade-off that we found to be involved with averaging on the basis of our results.

The rest of this chapter is organized as follows. Section 2 introduces the MRF model we used to formulate image data and averaging as preprocessing. Section 3 examines the performance of image restoration and hyper-parameter estimation. Section 4 discusses the trade-off involved in averaging and future work. Section 5 summarize the chapter.

5.1 Formulation

This section explains how we formulated a process of probabilistically generating image data by using MRF models to enable image analysis to be discussed based on the Bayesian framework. We then explain how to derive a probability distribution that averaged images follow.

5.1.1 Markov random field models

We formulate an acquisition process of image data from an original image by using MRF models as shown in Fig. 5.1. An item of n-dimensional image data is composed of N pixel values $v_i \in \mathbb{R}$ on an n-dimensional square lattice and is denoted by a vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_N)$. In the case of one-dimensional image data shown in Fig. 5.1(a), the subscript i is an integer from 1 to N. In the case of two-dimensional image data, which is an ordinary case, shown in Fig. 5.1(b), the subscript i is a pair of integers (i_1, i_2) $(1 \leq i_1 \leq N_1,\ 1 \leq i_2 \leq N_2)$ and $N = N_1 \times N_2$ is the number of pixels. In higher-dimensional cases, every pixel is similarly indexed. Each pixel value v_i is acquired by a measurement of the corresponding pixel value $u_i \in \mathbb{R}$ of an original image $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$ represented by the edge between nodes u_i and v_i characterized by a hyper-parameter b. As described later in this subsection, all pairs of neighboring pixels on the lattice can be assumed to take values close to each other, and this relation is represented by the edges characterized by a hyper-parameter a. For simplicity, we explain our formulation using the one-dimensional case hereafter, although the same arguments can be applied to higher-dimensional image data.

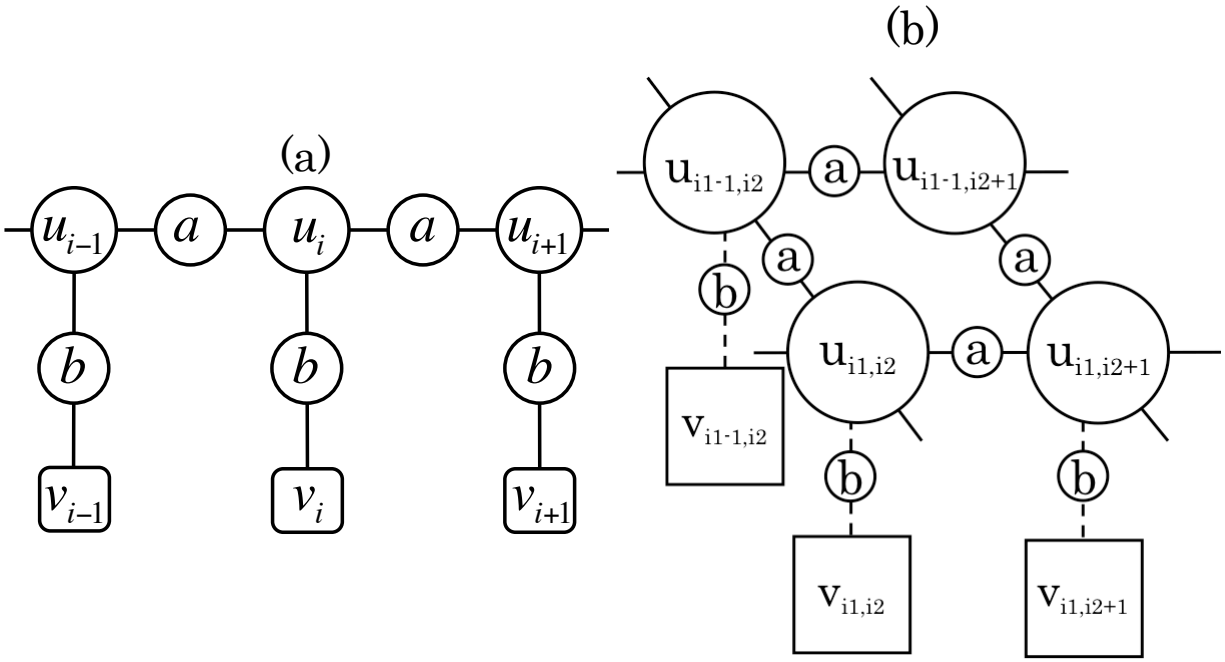

Figure 5.1: Graphical sketches of our MRF model. Variables $\mathbf{u} = \{u_i\}, (i = 1, 2, \ldots, N)$ compose an original image. Variables $\mathbf{v} = \{v_i\}$ compose image data. hyper-parameters a and b represent the smoothness of the original image and the magnitude of measurement noise, respectively. (a) One-dimensional image case. (b) Two-dimensional image case.

First, we explain a simple case where an item of image data is acquired. Let each pixel value v_i of the image data be the sum of the corresponding pixel value u_i of an original image and a noise component n_i as

$$v_i = u_i + n_i. \tag{5.1}$$

When the noise components n_i are i.i.d. random variables, that follow a zero-mean Gaussian distribution with variance b, denoted by $\mathcal{N}(0, b)$, the probability distribution of an item of image data $\mathbf{v}$ is given by

$$P_1(\boldsymbol{v}|\boldsymbol{u}, b) = \left(\frac{1}{\sqrt{2\pi b}}\right)^N \exp\left[-\frac{1}{2b}\sum_{i=1}^{N}(v_i - u_i)^2\right], \tag{5.2}$$

where noise variance b is a hyper-parameter of our MRF model and b is positive.

We next explain the case where more than one item of image data is acquired. Let a set of image data $\{\mathbf{v}^t\}_{t=1}^T = \{\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^T\}$ be acquired by observing the same original image $\mathbf{u}$. When each item of image data $\mathbf{v}^t$ independently follows the probability distribution

in Eq. (5.2), the probability distribution of T items of image data $\{\mathbf{v}^t\}_{t=1}^T$ is given by

$$\begin{aligned} P_T(\{\boldsymbol{v}^t\}_{t=1}^T|\boldsymbol{u},b) &= \prod_{t=1}^{T} P_1(\boldsymbol{v}^t|\boldsymbol{u},b) \\ &= \left(\frac{1}{\sqrt{2\pi b}}\right)^{NT} \exp\left[-\frac{1}{2b}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(v_i^t-u_i\right)^2\right]. \end{aligned} \tag{5.3}$$

We can see that each pixel v_i^t of image data follows $\mathcal{N}(u_i,b)$.

Most natural phenomena targeted as original images have the property of spatial continuity. Then, the prior distribution of an original image is given by

$$P_{\mathrm{pri}}(\boldsymbol{u}|a) = \frac{1}{Z_{\mathrm{pri}}(a)}\exp\left[-\frac{1}{2a}\sum_{i=1}^{N}(u_i-u_{i+1})^2\right], \tag{5.4}$$

where a Gaussian MRF model is used to ensure that neighboring pixels take values close to each other. A periodic boundary condition of $u_{N+1}=u_1$ is imposed for analytical convenience. The function Z_{pri} is called a partition function and is defined by

$$Z_{\mathrm{pri}}(a) = \int d\boldsymbol{u}\exp\left[-\frac{1}{2a}\sum_{i=1}^{N}(u_i-u_{i+1})^2\right]. \tag{5.5}$$

The variable a is called a hyper-parameter of our MRF model. The smoothness of the original image increases as a is set to a smaller value. a represents the variance of the difference in values of neighboring pixels and is positive.

5.1.2 Averaging

We calculated the probability distribution of averaged image data for a set of raw image data to discuss the preprocessing of averaging within the Bayesian framework. For clarity, we introduce a simple case where an item of averaged image data $\boldsymbol{m}^1=(m_1^1,m_2^1,\ldots,m_N^1)$ is computed from two items of raw image data $\{\boldsymbol{v}^1,\boldsymbol{v}^2\}$ as shown in Fig. 5.2(a). Let each pixel m_i^1 of averaged image data be the mean value of corresponding pixels v_i^1 and v_i^2 of raw image data as follows:

$$m_i^1 = \frac{v_i^1+v_i^2}{2}. \tag{5.6}$$

since v_i^1 and v_i^2 follow $\mathcal{N}(u_i,b)$ according to our model, the probability distribution of $\boldsymbol{m}^1$ is given by

$$P_{2\to1}(\boldsymbol{m}^1|\boldsymbol{u},b) = \prod_{i=1}^{N}\frac{\partial}{\partial m_i^1}\iint_{\frac{v_i^1+v_i^2}{2}<m_i^1} dv_i^1dv_i^2\prod_{t=1,2}\sqrt{\frac{1}{2\pi b}}\exp\left[-\frac{1}{2b}(v_i^t-u_i)^2\right], \tag{5.7}$$

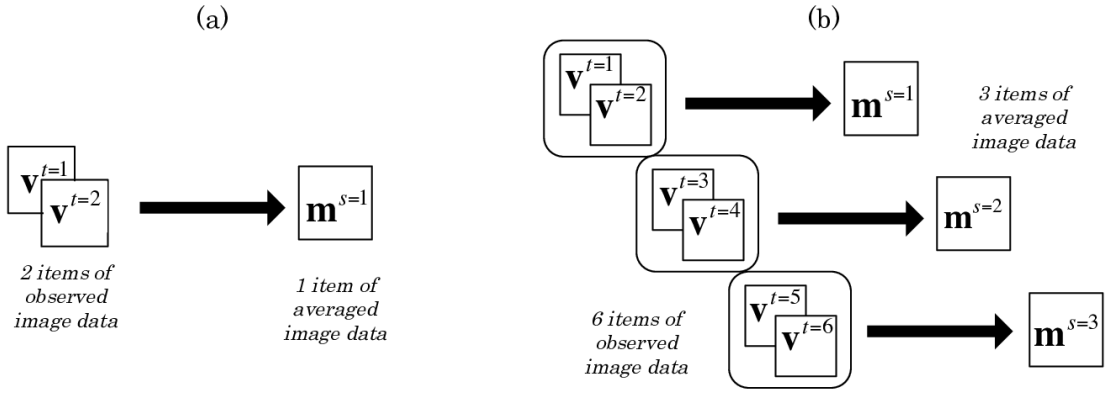

Figure 5.2: Outline of averaging preprocessing. (a) Case where two items of raw image data are preprocessed to obtain one item of averaged image data, viz., $T = 2/S = 1$. (b) Case where six items of raw image data are preprocessed to obtain three items of averaged image data, viz., $T = 6/S = 3$.

where the subscript $2 \to 1$ means that two items of raw image data are averaged into one item of averaged image data. The integral is calculated to obtain

$$P_{2\to1}(\boldsymbol{m}^1|\boldsymbol{u}, b) = \left(\sqrt{\frac{1}{\pi b}}\right)^N \exp\left[-\frac{1}{b}\sum_{i=1}^{N}(m_i^1 - u_i)^2\right]. \tag{5.8}$$

This probability distribution is compared with P_2, the $T = 2$ case of Eq. (5.3), to find the variance of $P_{2\to1}$, which is $b/2$, half that of P_2. This indicates that averaged image data are more precise than raw image data. On the other hand, the number of items of image data available in the main task of image analysis is decreased by averaging preprocessing. Therefore, there is a trade-off and whether averaging should be performed is nontrivial.

We formulate a general case where T given items of raw image data are preprocessed into S items of averaged image data to quantitatively evaluate the influence of averaging. Figure 5.2(b) outlines such a case, where $T = 6$ and $S = 3$. Each pixel m_i^s of averaged image data follows $\mathcal{N}(u_i, bS/T)$ because an item of averaged image data is obtained by calculating the mean of T/S items of raw image data. The probability distribution of a set of averaged image data $\{\boldsymbol{m}^s\}_{s=1}^S$ is given by

$$P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|\boldsymbol{u}, b) = \left(\sqrt{\frac{T/S}{2\pi b}}\right)^{NS} \exp\left[-\frac{T/S}{2b}\sum_{s=1}^{S}\sum_{i=1}^{N}(m_i^s - u_i)^2\right]. \tag{5.9}$$

We can see that the ratio T/S controls the trade-off because it simultaneously represents both the degree of enhanced precision and the degree of decrease in the number of items.

5.2 Image Analysis and Its Performance Evaluation

This section introduces methods of image analysis based on Bayesian inference[6] and explains how we evaluated the influence of averaging on their performance. This study focused on two tasks in image analysis, viz., image restoration and hyper-parameter estimation. It is important to restore image data to the original image, and according to a maximum a posteriori (MAP) framework, image restoration can be formulated as a minimization problem of an energy function. What mattered more to us was hyper-parameter estimation. Our model has two hyper-parameters , which represent the magnitude of measurement noise and the smoothness of original images, the latter of which can be interpreted as a physical quantity, viz., a diffusion coefficient[16]. In general, physical quantities should be assessed not only by point estimation such as MAP estimation but also with the confidence of estimation. Such assessments are possible with a method of estimating posterior distributions that the authors proposed on the basis of a free-energy function, i.e., the negative logarithm of a posterior distribution[16, 15]. These methods of hyper-parameter estimation are also useful for improving the performance of image restoration because the hyper-parameters are included in these energy functions. The framework of hyper-parameter estimation enables us to objectively determine the values of tuning parameters in image analysis using the data, although they are often arbitrarily adjusted.

5.2.1 Image restoration

Posterior distributions play a central role in Bayesian inference. According to Bayes' theorem, the posterior distribution of a restored image given a set of averaged images is given by

$$P_{T\to S}(\boldsymbol{u}|\{\boldsymbol{m}^s\}_{s=1}^S, a_0, b_0) \;=\; \frac{P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|\boldsymbol{u}, b_0)P_{\rm pri}(\boldsymbol{u}|a_0)}{\int d\boldsymbol{u} P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|\boldsymbol{u}, b_0)P_{\rm pri}(\boldsymbol{u}|a_0)}. \tag{5.10}$$

The hyper-parameters are set to the true values $(a, b) = (a_0, b_0)$ in this subsection and hyper-parameter estimation is explained in the next subsection. We substitute Eqs. (5.4) and (5.9) into Eq. (5.10) to obtain

$$P_{T\to S}(\boldsymbol{u}|\{\boldsymbol{m}^s\}_{s=1}^S, a_0, b_0) \;\propto\; e^{-E_{T\to S}(\boldsymbol{u}|\{\boldsymbol{m}^s\}_{s=1}^S, a_0, b_0)}, \tag{5.11}$$

where the energy function, $E_{T\to S}$, is given by

$$E_{T\to S}(\boldsymbol{u}|\{\boldsymbol{m}^s\}_{s=1}^S, a_0, b_0) = \sum_{i=1}^N \left[\frac{T/S}{2b_0} \sum_{s=1}^S \left(m_i^s - u_i\right)^2 + \frac{1}{2a_0}\left(u_i - u_{i+1}\right)^2 \right]. \tag{5.12}$$

The MAP framework, where the posterior distribution is maximized, is equivalent to minimizing the energy function. Thus, we investigate the influence of averaging on the

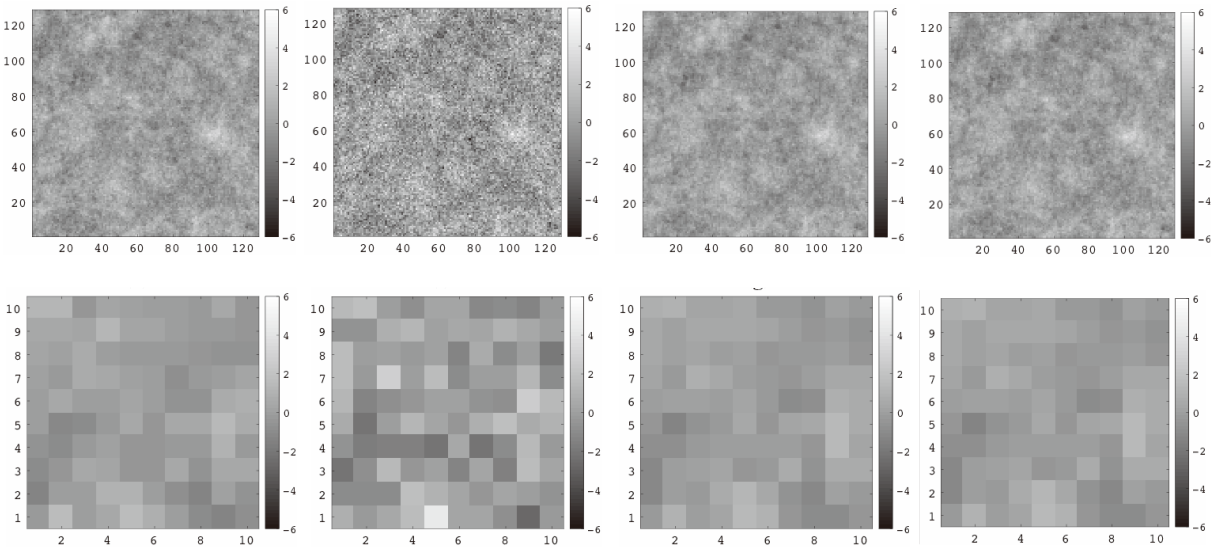

Figure 5.3: Examples of images. The images are two-dimensional with $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2)$. (a) is an original image. (b) is an observed image. (c) is a restored image, where $S = 1$, and (d) is a restored image, where $S = 12$. (e)–(h) Magnified images of the lower-left corner of (a)–(d), respectively.

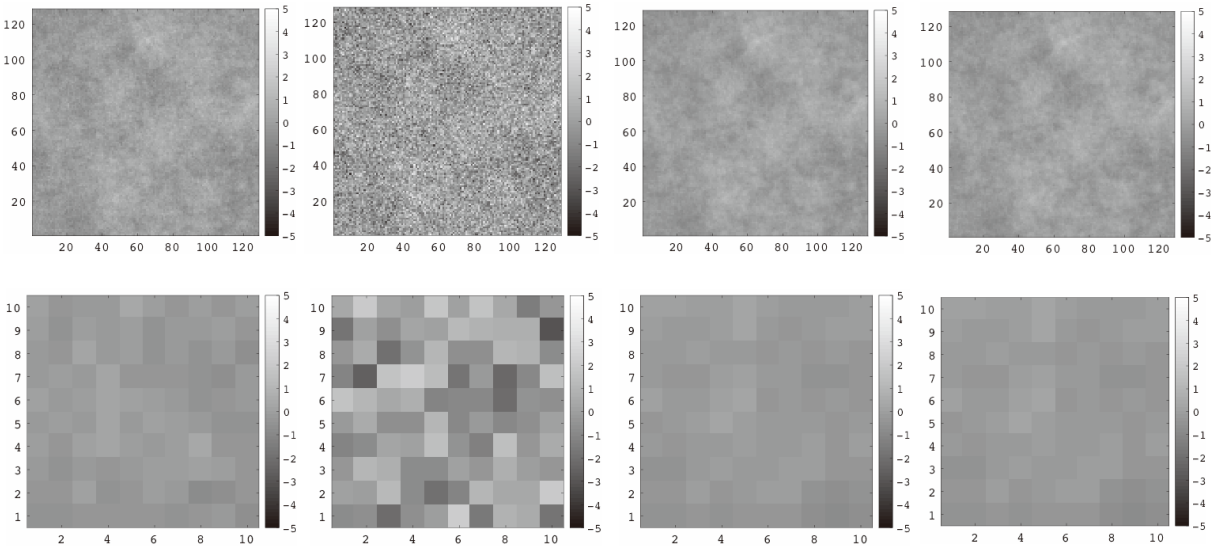

Figure 5.4: Examples of images. This figure is the same as Fig. 5.3 except that the true hyper-parameters are $(a_0, b_0) = (0.2, 1.2)$.

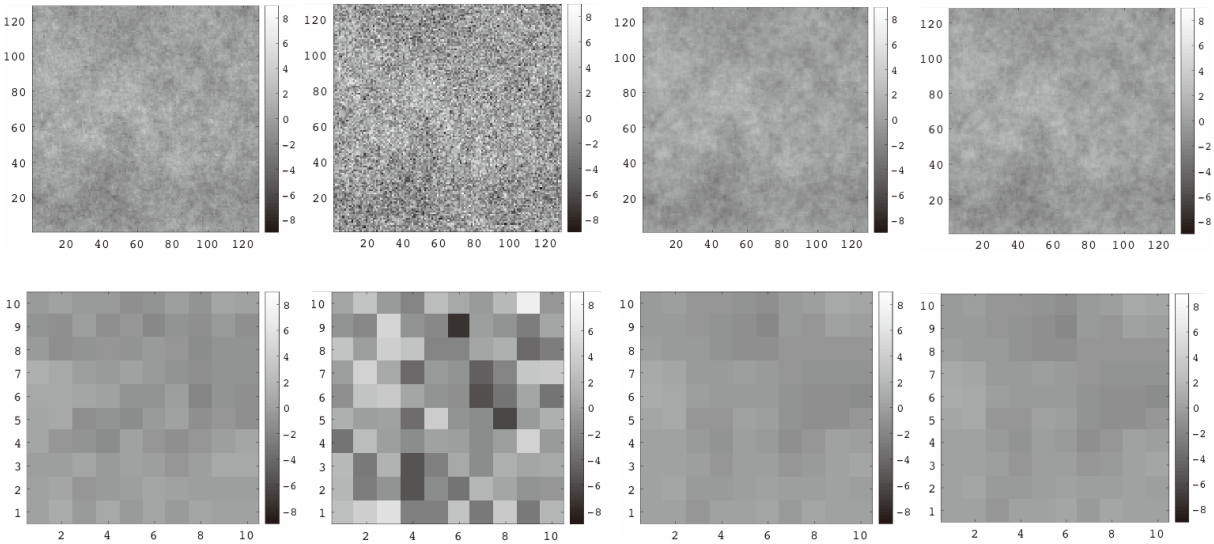

Figure 5.5: Examples of images. This figure is the same as Fig. 5.3 except that the true hyper-parameters are $(a_0, b_0) = (1, 6)$.

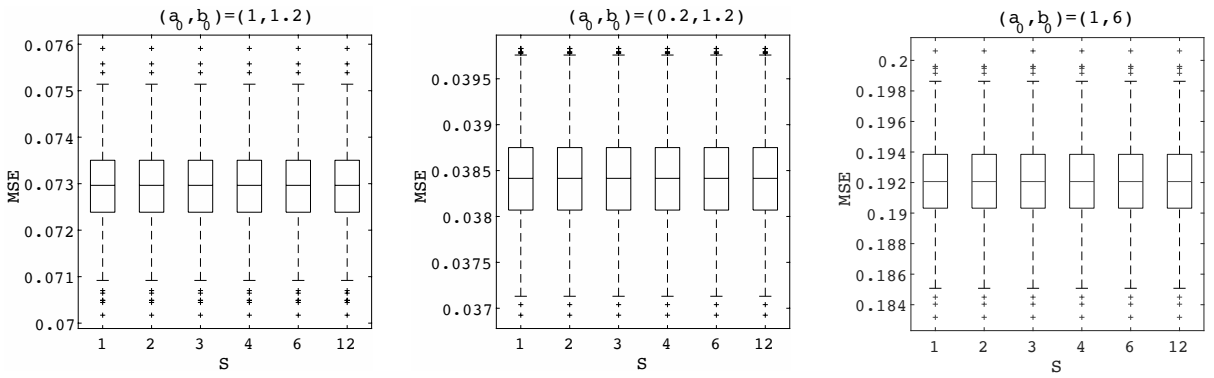

Figure 5.6: Results of 1000 trials of image restoration using true hyper-parameters. The MSE of restored images with respect to original images is shown in the form of a boxplot. The central line in the box is the median and the bottom and top of the box are the first and third quartiles, respectively. The whiskers extending from the bottom and top of the box are 1.5 times as long as the interquartile range. The crosses are outliers. The images are two-dimensional with $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, and $(1, 6)$ from left to right.

energy function to evaluate the performance of image restoration. For simplicity, our analysis begins with the case where the number of raw images is $T = 2$. When two raw images $\boldsymbol{v}^1$ and $\boldsymbol{v}^2$ are averaged to obtain an averaged image $\boldsymbol{m}^1$, in which pixel noise of

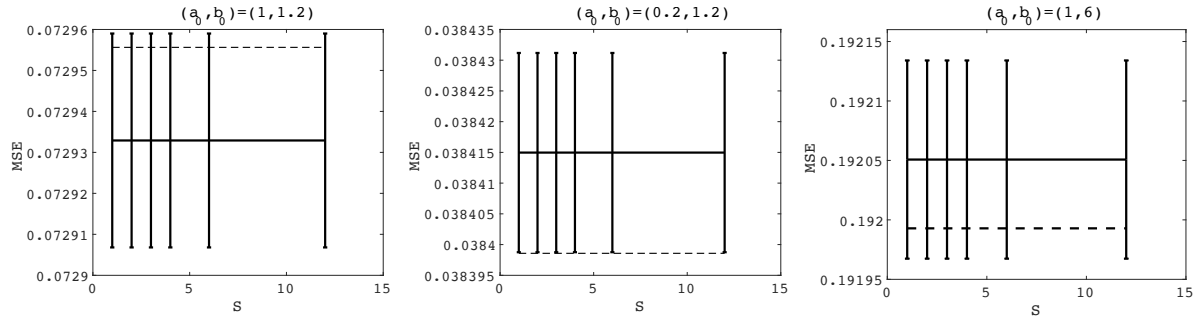

Figure 5.7: Results of 1000 trials of image restoration using true hyper-parameters. The MSE of restored images with respect to original images is plotted against S. The solid line and its error bars represent the mean value and its standard error, respectively. The dotted line represents our theoretical result for the typical value. The images are two-dimensional with $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, and $(1, 6)$ from left to right.

variance $b_0/2$ has been added to each pixel, the energy function $E_{2\to1}$ is given by

$$\begin{aligned} E_{2\to1}(\boldsymbol{u}|\boldsymbol{m}^1, a_0, b_0) &= \sum_{i=1}^{N}\left[\frac{2}{2b_0}\left(m_i^1 - u_i\right)^2 + \frac{1}{2a_0}\left(u_i - u_{i+1}\right)^2\right] \\ &= \sum_{i=1}^{N}\left[\frac{2}{2b_0}\left(\frac{v_i^1 + v_i^2}{2} - u_i\right)^2 + \frac{1}{2a_0}\left(u_i - u_{i+1}\right)^2\right]. \end{aligned} \tag{5.13}$$

When two raw images are used without averaging, the energy function $E_{2\to2}$ is given by

$$\begin{aligned} E_{2\to2}(\boldsymbol{u}|\boldsymbol{v}^1, \boldsymbol{v}^2, a_0, b_0) &= \sum_{i=1}^{N}\left[\frac{1}{2b_0}\left\{\left(v_i^1 - u_i\right)^2 + \left(v_i^2 - u_i\right)^2\right\} + \frac{1}{2a_0}\left(u_i - u_{i+1}\right)^2\right] \\ &= \sum_{i=1}^{N}\left[\frac{2}{2b_0}\left(\frac{v_i^1 + v_i^2}{2} - u_i\right)^2 + \frac{1}{2a_0}\left(u_i - u_{i+1}\right)^2 + \frac{2}{2b_0}\left(\frac{v_i^1 - v_i^2}{2}\right)^2\right]. \end{aligned} \tag{5.14}$$

Since the third term in square brackets on the right of Eq. (5.14) has no component of $\mathbf{u}$, the two energy functions $E_{2\to1}$ and $E_{2\to2}$ are equal up to a constant. Therefore, the same results for image restoration can be expected regardless of whether averaging is performed.

The MAP estimator of the original image is analytically derived from our model, which is based on a Gaussian MRF model with a periodic boundary condition[26, 57, 15]. By

using a discrete Fourier transformation (DFT) defined as

$$\tilde{x}_k = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \exp\left[\frac{2\pi \mathrm{i}(i \cdot k)}{N}\right], \tag{5.15}$$

the Fourier representation of the energy function $\tilde{E}_{T\to S}$ is given by

$$\tilde{E}_{T\to S}(\tilde{\boldsymbol{u}}|\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S, a_0, b_0) \simeq \sum_{k=1}^{N} \frac{1}{2}\left(\frac{T}{b_0} + \frac{\lambda_k}{a_0}\right)\left|\tilde{u}_k - \frac{\frac{T}{b_0}}{\frac{T}{b_0} + \frac{\lambda_k}{a_0}}\tilde{\mu}_k\right|^2, \tag{5.16}$$

where $\lambda_k = 2 - 2\cos\left[\frac{2\pi k}{N}\right]$, $\tilde{\mu}_k = \frac{1}{S}\sum_{s=1}^{S} \tilde{m}_k^s$, and the symbol $\simeq$ means that both sides are equal up to a constant term. It is worth mentioning that all two-body interaction terms such as $(u_i - u_{i+1})^2$ in Eq. (5.4) have been eliminated by the DFT because of the translational symmetry of our model to obtain a one-body representation of the energy function, $\tilde{E}_{T\to S}$. Moreover, it is emphasized that this property is valid in the case of higher-dimensional image data. Because $\tilde{E}_{T\to S}$ is minimized at

$$\tilde{u}_k^{\mathrm{MAP}} = \frac{\frac{T}{b_0}}{\frac{T}{b_0} + \frac{\lambda_k}{a_0}}\tilde{\mu}_k, \tag{5.17}$$

the MAP estimator is given by

$$u_i^{\mathrm{MAP}} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \tilde{u}_k^{\mathrm{MAP}} \exp\left[-\frac{2\pi \mathrm{i}(i \cdot k)}{N}\right] \tag{5.18}$$

with the inverse DFT. Since each $\boldsymbol{m}^s$ is an averaged image of T/S raw image data by definition, μ_k is rewritten as

$$\tilde{\mu}_k \;=\; \frac{1}{T}\sum_{t=1}^{T} \tilde{v}_k^t. \tag{5.19}$$

We can then see that the MAP estimator does not depend on the number of averaged images S and that averaging has no influence on the performance of image restoration.

We carried out numerical experiments to examine the influence of averaging on image restoration. We dealt with a two-dimensional case where the number of pixels was $N = 128^2$ and the number of raw images was $T = 12$. The true hyper-parameter values were set to $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, and $(1, 6)$. Note that the same settings were used in all experiments described in this chapter. The procedure involved three steps. First, an original image was generated by using Eq. (5.4). Second, Gaussian noise was added to the original image to create raw image data in accordance with Eq. (5.3). Finally, the original image was estimated from the image data with or without averaging using the MAP framework. Figs. 5.3–5.5 show some examples of image restoration. The case of

$(a_0, b_0) = (1, 1.2)$ is shown in Fig. 5.3. Fig. 5.3(a) is an original image. Fig. 5.3(b) is raw image data. Fig. 5.3(c) shows the results of image restoration from 12 items of raw image data without averaging, viz., $S = 12$. Fig. 5.3(d) shows the results of image restoration from an item of image data averaged over 12 raw image data, viz., $S = 1$. Figs. 5.3(c) and 5.3(d) were compared to find whether both are the same image. Figs. 5.4 and 5.5 show the cases of $(a_0, b_0) = (0.2, 1.2)$ and $(1, 6)$, respectively, and we can see that the performance of image restoration with the true hyper-parameters is not influenced by averaging preprocessing. Here, let us define mean squared error (MSE) as

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(u_i^{\mathrm{MAP}} - u_i\right)^2, \tag{5.20}$$

to quantitatively evaluate performance. We calculate MSE in cases where $S = 1, 2, 3, 4, 6$, and 12. The results of 1000 trials of numerical experiments are shown in Figs. 5.6 and 5.7. Fig. 5.6 shows boxplots of MSE. Figure 5.7 indicates the mean value of MSE and its standard error to evaluate the typical performance of image restoration. We can see that MSE does not decrease or increase and this supports our argument that averaging does not influence the results of image restoration if the hyper-parameters are predetermined.

It is also important to theoretically analyze the typical performance of image restoration. The expected value of MSE was analytically calculated[15]. The expected value of MSE with respect to image data is defined as

$$\langle \mathrm{MSE} \rangle = \left\langle \frac{1}{N}\sum_{i=1}^{N}\left(u_i^{\mathrm{MAP}} - u_i\right)^2 \right\rangle_{\{\boldsymbol{m}^s\}_{s=1}^S, \boldsymbol{u}|a_0,b_0}, \tag{5.21}$$

where $\langle\cdot\rangle_{\{\boldsymbol{m}^s\}_{s=1}^S, \boldsymbol{u}|a_0,b_0}$ denotes calculating an expected value with respect to

$$P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S, \boldsymbol{u}|a_0, b_0) = P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|\boldsymbol{u}, b_0)P_{\mathrm{pri}}(\boldsymbol{u}|a_0). \tag{5.22}$$

By using the DFT, the expected value is expressed as

$$\langle \mathrm{MSE} \rangle_{\{\boldsymbol{m}^s\}_{s=1}^S, \boldsymbol{u}|a_0,b_0} = \left\langle \frac{1}{N}\sum_{k=1}^{N}\left|\tilde{u}_k^{\mathrm{MAP}} - \tilde{u}_k\right|^2 \right\rangle_{\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S, \tilde{\boldsymbol{u}}|a_0,b_0}. \tag{5.23}$$

The probability distribution with respect to which the expected value is calculated is given by

$$\tilde{P}_{T\to S}(\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S, \tilde{\boldsymbol{u}}|a_0, b_0) = \tilde{P}_{T\to S}(\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S|\tilde{\boldsymbol{u}}, b_0)\tilde{P}_{\mathrm{pri}}(\tilde{\boldsymbol{u}}|a_0), \tag{5.24}$$

where the Fourier representations of the probability distributions on the right-hand side are given by

$$\tilde{P}_{T\to S}(\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S|\tilde{\boldsymbol{u}}, b_0) = \left(\sqrt{\frac{T/S}{2\pi b_0}}\right)^{NS} \exp\left[-\frac{T/S}{2b_0}\sum_{s=1}^{S}\sum_{k=1}^{N}|\tilde{m}_k^s - \tilde{u}_k|^2\right], \tag{5.25}$$

and

$$\tilde{P}_{\rm pri}(\tilde{\boldsymbol{u}}|a_0) = \frac{1}{Z_{\rm pri}(a_0)} \exp\left[-\frac{1}{2a_0}\sum_{k=1}^{N}\lambda_k\left|\tilde{u}_k\right|^2\right]. \tag{5.26}$$

The formula for the Gaussian integral is used to obtain:

$$\langle \mathrm{MSE}\rangle = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{T}{b_0}+\frac{\lambda_k}{a_0}\right)^{-1}. \tag{5.27}$$

Under the same conditions as those in the above simulation, we calculated the expected value of MSE. As seen in Fig. 5.7, the numerical results are considered to be consistent with those of theoretical analysis.

5.2.2 Hyper-parameter estimation

This subsection is devoted to hyper-parameter estimation, which was not discussed in the previous subsection. Hyper-parameter estimation is possible with the Bayesian framework because hyper-parameters can also be regarded as random variables [6]. Posterior distributions play as central a role as ever. According to Bayes' theorem, the posterior distribution of hyper-parameters given a set of averaged images is given by

$$P_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S) \propto P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|a,b)P(a,b), \tag{5.28}$$

where

$$P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|a,b) = \int d\boldsymbol{u} P_{T\to S}(\{\boldsymbol{m}^s\}_{s=1}^S|\boldsymbol{u},b)P_{\rm Pri}(\boldsymbol{u}|a) \tag{5.29}$$

is a marginalized likelihood function. The prior distribution of hyper-parameters $P(a,b)$ is set to a uniform distribution defined in the domain of $a>0$ and $b>0$ because there is no knowledge on the hyper-parameter values beforehand except that they are positive. Thus, the posterior distribution can be identified with the marginalized likelihood function. Marginalization is generally difficult to calculate because of its high-dimensional integral, but in our case, it can be exactly calculated[16]. The DFT is applied to the high-dimensional integral such that it can be broken down into component-wise integrals as

$$P_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S) \propto \prod_{k=1}^{N}\sqrt{\frac{\lambda_k}{2\pi a}}\left(\sqrt{\frac{T/S}{2\pi b}}\right)^S \int d\tilde{u}_k \exp\left[-\frac{T/S}{2b}\sum_{s=1}^{S}|\tilde{m}_k^s-\tilde{u}_k|^2-\frac{\lambda_k}{2a}\left|\tilde{u}_k\right|^2\right]. \tag{5.30}$$

Each of the component-wise integrals can be calculated by using the formula for the Gaussian integral. As a result, we obtain the following exact solution:

$$P_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S) \propto \prod_{k=1}^N \sqrt{\frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}}} \exp\left[-\frac{1}{2}\frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}}|\tilde{\mu}_k|^2\right]\left\{\sqrt{\frac{1}{b}}\exp\left[-\frac{T/S}{2b}\tilde{\eta}_k\right]\right\}^{S-1}, \tag{5.31}$$

where

$$\tilde{\eta}_k = \frac{1}{S-1}\sum_{s=1}^S |\tilde{m}_k^s - \tilde{\mu}_k|^2 . \tag{5.32}$$

We can see from Eq. (5.31) that the posterior distribution depends on S unlike the energy function of image restoration, and the performance of hyper-parameter estimation is influenced by averaging preprocessing.

We begin with the MAP estimation of hyper-parameters. MAP estimation can be carried out by maximizing the posterior distribution with the method of gradient descent. We simulated hyper-parameter estimation with settings similar to those in the numerical experiments on image restoration. We dealt with a two-dimensional case where the number of pixels was $N = 128^2$ and the number of raw images was $T = 12$. The true hyper-parameter values were set to $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, or $(1, 6)$. Figure 5.8 shows the results in the form of a boxplot. The whiskers are shorter as there are more averaged images, and this indicates that averaging preprocessing should not be done to maintain the precision of MAP estimation of hyper-parameters.

The performance of hyper-parameter estimation can also be evaluated by the performance of image restoration. The previous subsection explained how the performance of image restoration was evaluated when the true values of hyper-parameters were used. However, the true values are unknown beforehand in practical situations and they should be adequately determined by using hyper-parameter estimation. If hyper-parameter estimation provides the true values, the performance of image restoration is also maximized [23]. The results of 1000 trials of image restoration are shown in Figs. 5.9 and 5.10. Fig. 5.9 shows boxplots of MSE. Fig. 5.10 shows the mean value of MSE and its standard error to evaluate the typical performance of image restoration. We can see that MSE is smaller as there are more averaged images. This indicates that hyper-parameter estimation performs well in terms of image restoration when raw image data are used as they are as much as possible.

It is important to evaluate the confidence of estimated values when the estimation target is a physical quantity such as a diffusion coefficient, which is related to the hyper-parameter a. Most conventional estimation methods such as MAP estimation are point estimations, which only provide estimated values, and it is impossible to evaluate the confidence of estimated values unless data acquisition and parameter estimation are repeated many times. In contrast, distribution estimation can simultaneously provide an

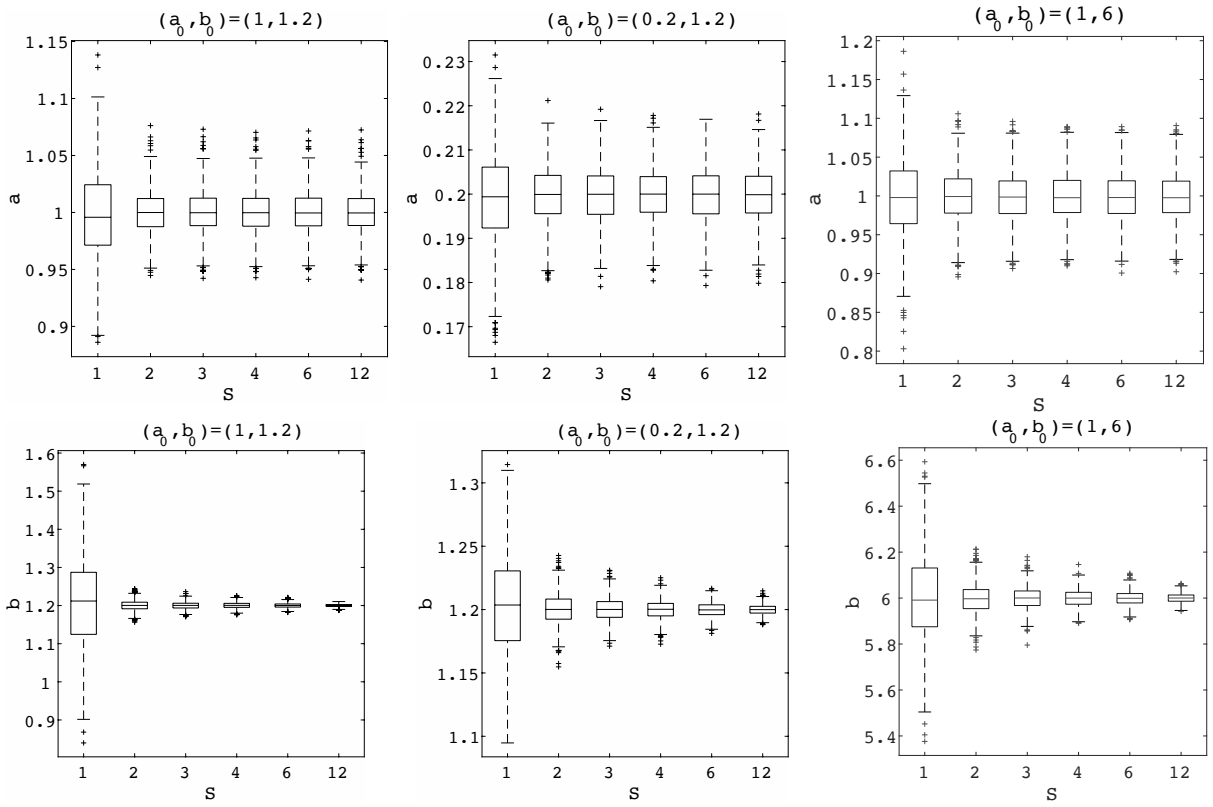

Figure 5.8: Results of 1000 trials of hyper-parameter estimation. The MAP value of hyper-parameters is shown in the form of a boxplot. See caption to Fig. 5.6 for details. The images are two-dimensional $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, and $(1, 6)$ from left to right.

estimated value and its confidence [16, 15]. Figures 5.11–5.13 present the results obtained from estimating hyper-parameter distributions. We dealt with a two-dimensional case where the number of pixels was $N = 128^2$, the number of raw images was $T = 12$, and the true values of the hyper-parameters were given by $(a_0, b_0) = (1, 1.2)$, $(0.2, 1.2)$, and $(1, 6)$. The number of averaged images was set to $S = 1, 2$, or 12 to evaluate how averaging influenced the posterior distribution. We can see that the posterior distribution has a sharper peak close to the true values as the number of averaged images is smaller. This indicates that hyper-parameter estimation can be achieved with higher precision without averaging.

We examined posterior distributions in terms of the mutual information of the two hyper-parameters a and b to evaluate the influence of averaging on distribution estimation in detail. Mutual information is useful in measuring the dependence between two random variables and it is defined as

$$I(a; b) = \int dadb P(a, b) \ln \frac{P(a, b)}{P(a)P(b)}. \tag{5.33}$$

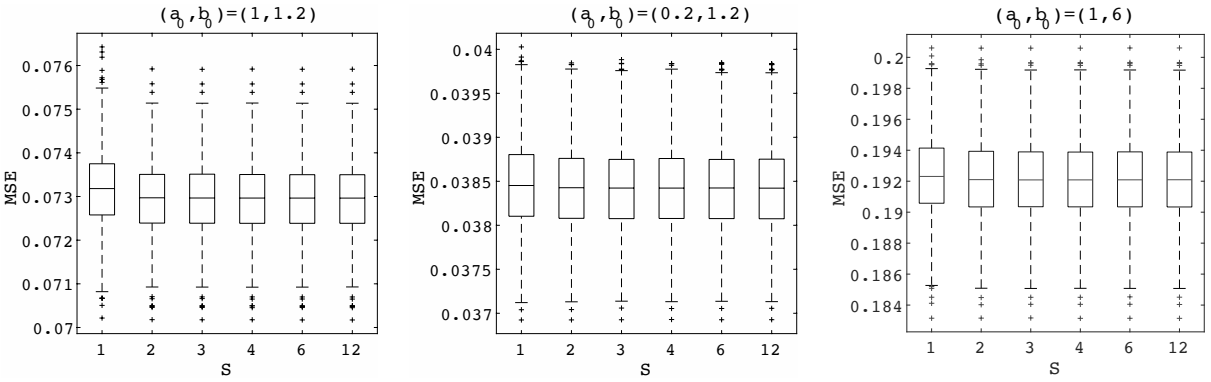

Figure 5.9: Results of 1000 trials of image restoration using estimated hyper-parameters. The MSE of restored images with respect to the original images is shown in the form of a boxplot. See caption to Fig. 5.6 for details. The images are two-dimensional $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, and $(1, 6)$ from left to right.

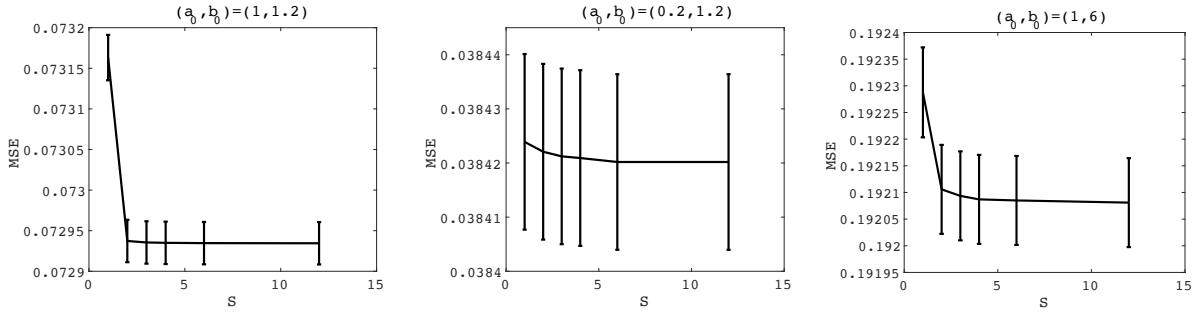

Figure 5.10: Results of 1000 trials of image restoration using estimated hyper-parameters. The MSE of restored images with respect to original images is plotted against S. The solid line and its error bars represent the mean value and its standard error, respectively. The images are two-dimensional $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2), (0.2, 1.2)$, and $(1, 6)$ from left to right.

Since the definition is based on the Kullback–Leibler divergence, the value of mutual information is always non-negative, and it is equal to zero if and only if the two variables are independent of each other. Figure 5.14 plots the results of mutual information. Each point and its error bar respectively represent the mean value and its standard error in 1000 trials. We can see that mutual information decreases to zero with increasing number of averaged images. When random variables to be estimated are regarded as independent of each other, approximation methods such as variational Bayesian methods could work well. This possibility will be discussed later.

It is also important to analyze the typical performance of hyper-parameter estimation. Let us define a free-energy function by

$$F_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S) \;=\; -\ln P_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S), \tag{5.34}$$

and its expected value with respect to the marginalized likelihood function can be obtained analytically[15]. According to Eq. (5.29), the expected value is given by

$$\left\langle F_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S)\right\rangle_{\{\boldsymbol{m}^s\}_{s=1}^S|a_0,b_0} = \left\langle \left\langle F_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S)\right\rangle_{\{\boldsymbol{m}^s\}_{s=1}^S|\boldsymbol{u},b_0}\right\rangle_{\boldsymbol{u}|a_0}, \tag{5.35}$$

where the outer and inner expected values on the right-hand side are respectively calculated with respect to Eqs. (5.4) and (5.9). Equation (5.31) is substituted to obtain

$$\begin{aligned}\left\langle F_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S)\right\rangle_{\{\boldsymbol{m}^s\}_{s=1}^S|a_0,b_0} \;\simeq\; & \sum_{k=1}^N \left\{ \frac{1}{2}\frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}}\left\langle\left\langle|\tilde{\mu}_k|^2\right\rangle_m\right\rangle_u - \frac{1}{2}\ln\frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}}\right. \\ & \left. +(S-1)\left(\frac{T/S}{2b}\left\langle\left\langle\tilde{\eta}_k\right\rangle_m\right\rangle_u - \frac{1}{2}\ln\frac{1}{b}\right)\right\},\end{aligned} \tag{5.36}$$

where the subscripts regarding expected values have been abbreviated on the right-hand side. The expected values $\left\langle\left\langle|\tilde{\mu}_k|^2\right\rangle_m\right\rangle_u$ and $\left\langle\left\langle\tilde{\eta}_k\right\rangle_m\right\rangle_u$ should be calculated. Because $\tilde{\mu}_k$ is the mean value of $\tilde{v}_k^t$, each of which follows $\mathcal{N}(\tilde{u}_k, b_0)$, and the kth Fourier component of an original image $\tilde{u}_k$ follows $\mathcal{N}(0, a_0/\lambda_k)$, we obtain $\left\langle\left\langle|\tilde{\mu}_k|^2\right\rangle_m\right\rangle_u = \frac{a_0}{\lambda_k}+\frac{b_0}{T}$. The other expected value, $\tilde{\eta}_k$, is the variance of $\tilde{v}_k^t$. Since the random variable $\tilde{\eta}_k' = \frac{T/S}{b_0}(S-1)\tilde{\eta}_k$ follows a chi-squared distribution with $S-1$ degrees of freedom[62], we obtain $\left\langle\left\langle\tilde{\eta}_k\right\rangle_m\right\rangle_u = b_0$. As a result, the expected value of free energy is given by

$$\begin{aligned}\left\langle F_{T\to S}(a,b|\{\boldsymbol{m}^s\}_{s=1}^S)\right\rangle_{\{\boldsymbol{m}^s\}_{s=1}^S|a_0,b_0} \;\simeq\; & \sum_{k=1}^N \left\{ \frac{1}{2}\frac{\frac{a_0}{\lambda_k}+\frac{b_0}{T}}{\frac{a}{\lambda_k}+\frac{b}{T}} - \frac{1}{2}\ln\frac{1}{\frac{a}{\lambda_k}+\frac{b}{T}}\right. \\ & \left. +(S-1)\left(\frac{1}{2}\frac{b_0}{b} - \frac{1}{2}\ln\frac{1}{b}\right)\right\},\end{aligned} \tag{5.37}$$

and its partial derivatives are given by

$$\frac{\partial\langle F_{T\to S}\rangle}{\partial a} = \sum_{k=1}^N \frac{1}{2\lambda_k}\frac{\frac{a-a_0}{\lambda_k}+\frac{b-b_0}{T}}{\left(\frac{a}{\lambda_k}+\frac{b}{T}\right)^2} \tag{5.38}$$

and

$$\frac{\partial\langle F_{T\to S}\rangle}{\partial b} = \sum_{k=1}^N \left\{ \frac{1}{2T}\frac{\frac{a-a_0}{\lambda_k}+\frac{b-b_0}{T}}{\left(\frac{a}{\lambda_k}+\frac{b}{T}\right)^2} + (S-1)\frac{b-b_0}{2b^2}\right\}. \tag{5.39}$$

These derivatives are made to equal zero to find whether the true values of the hyper-parameters minimize the typical free energy. Since the minimizer of the free energy corresponds to the MAP-estimated values, we have ensured that the MAP estimation is unbiased regardless of whether averaging is performed.

Figure 5.15 presents the expected value of free energy calculated under the same conditions as those in Figs. 5.11–5.13. We can see that the free-energy function has a sharper valley as the number of averaged images increases. This property of free energy can be deduced from its formula in Eq. (5.37). The last half of Eq. (5.37) depends on S and it reflects the influence of averaging. Since the second coefficient of its Taylor expansion at $b = b_0$ is proportional to $S - 1$, free energy turns out to have a sharper valley as there are more averaged images. The confidence of data determined from the peak sharpness of the posterior distribution decreases monotonically as there are fewer averaged images.

The term independent of S outside the braces on the right of Eq. (5.31) accounts for the correlation between a and b. This independence is indicated from the second partial derivatives of the expected value of free energy where $(a, b) = (a_0, b_0)$, which are

$$\left.\frac{\partial^2 \langle F_{T\to S}\rangle}{\partial a^2}\right|_{a=a_0,b=b_0} = \sum_{k=1}^{N} \frac{1}{2\lambda_k^2} \frac{1}{\left(\frac{a_0}{\lambda_k} + \frac{b_0}{T}\right)^2}, \tag{5.40}$$

and

$$\left.\frac{\partial^2 \langle F_{T\to S}\rangle}{\partial b^2}\right|_{a=a_0,b=b_0} = \sum_{k=1}^{N} \left\{ \frac{1}{2T^2} \frac{1}{\left(\frac{a_0}{\lambda_k} + \frac{b_0}{T}\right)^2} + (S-1)\frac{1}{2b_0^2} \right\}. \tag{5.41}$$

We can see that only the second derivative of b is influenced by averaging. When S is large, the second term of Eq. (5.41) is dominant and the free energy has a sharper valley as S increases. Hence, the value of b is estimated precisely and independently of the estimated value of a when there are many items of averaged images.

5.3 Discussion

Our analysis of the Bayesian posterior distribution and our numerical experiments revealed that the performance of image restoration with predetermined hyper-parameters is independent of averaging. The hyper-parameter estimation is unbiased, but the confidence decreases as there are fewer items of averaged image data. Our analysis of the estimated hyper-parameter distributions indicated that averaging results in the correlation and dependence of the estimated hyper-parameters. The performance of image restoration using estimated hyper-parameters degrades as there are fewer averaged images. As a result, we deduced that averaging is not problematic as long as image analysis does

not involve any hyper-parameter estimation, but averaging has an adverse effect on the performance of image analysis when it requires hyper-parameter estimation.

The performance of preprocessing has often been discussed with the data processing inequality, which states that the mutual information between data and what is to be estimated never increases by averaging, viz.

$$I(a, b, \boldsymbol{u}; \{\boldsymbol{v}^t\}_{t=1}^T) \geq I(a, b, \boldsymbol{u}; \{\boldsymbol{m}^s\}_{s=1}^S), \tag{5.42}$$

since the random variables consist of the Markov chain $\{a, b, \mathbf{u}\} \rightarrow \{\mathbf{v}^t\}_{t=1}^T \rightarrow \{\mathbf{m}^s\}_{s=1}^S$[70]. Our results are considered to be consistent with the data processing inequality in the sense that averaging preprocessing does not improve the performance of image restoration or hyper-parameter estimation. Moreover, our detailed analysis indicated that averaging has no influence on image restoration with predetermined hyper-parameters and that it does not cause any bias but deteriorates the confidence in the results of image analysis that requires hyper-parameter estimation.

The computational cost is often too high when many raw image data are directly used for image analysis. It is important to discuss approximate methods based on the posterior distributions examined in this study. For example, it is well known that variational Bayesian methods are effective in approximately calculating posterior distributions when the random variables are independent[14]. As shown in Fig. 5.14, the mutual information of the hyper-parameters a and b regarding our posterior distribution decreases toward zero as there are more items of averaged image data, and because the mutual information of two independent variables is zero by definition, the hyper-parameters can be regarded as relatively independent if averaging preprocessing is not performed. In addition, we analyzed this independence by calculating the second partial derivatives of free energy, as given by Eqs. (5.40) and (5.41). We confirmed that the free-energy distribution becomes independent as using more averaged images. Consequently, the variational Bayesian method could be effective when the number of items of averaged image data S is kept at a large value.

Although we studied image analysis tasks based on Gaussian MRF models, it is also important to evaluate the influence of averaging preprocessing on other tasks. A related study regarding the least absolute shrinkage and selection operator (LASSO) within the context of compressed sensing has proved that averaging should not be performed to maximize performance[71]. Compressed sensing is a framework for signal reconstruction from a few items of data by utilizing the sparseness of signals[72, 73]. LASSO is formulated as an ℓ_1-norm regularized least-squares method, viz.

$$\min_{\boldsymbol{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \right\}, \tag{5.43}$$

where $\mathbf{y}$ is a data vector, $\mathbf{A}$ is a measurement matrix, and $\mathbf{x}$ is a signal vector[74]. $\|\cdot\|_2$ denotes the ℓ_2-norm defined as $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$. $\|\cdot\|_1$ denotes the ℓ_1-norm defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$. LASSO can be interpreted in the framework of Bayesian

inference and is therefore worth comparing with MRF models. The least-squares term in LASSO reflects Gaussian noise in a similar way to that in Eq. (5.2). The ℓ_1-norm term represents prior knowledge of signal sparseness because the spatial continuity and diffusive properties of image data are described by MRF models, as can be seen from Eq. (5.4). The regularization coefficient λ can be regarded as a hyper-parameter of LASSO, which controls the degree of sparseness. We expected that the performance would deteriorate in LASSO due to averaging caused by hyper-parameter estimation, similarly to for our settings in this work, and that when a predetermined hyper-parameter was used, the performance of LASSO would be uneffected by averaging preprocessing. Therefore, it will be important in future work to find the influence of averaging on LASSO from the viewpoint of hyper-parameter distribution estimation.

We dealt with a Gaussian model, in which noise components are assumed to follow a Gaussian distribution and the prior distribution of original images is described by a Gaussian MRF. It is important to study different measurement models to discuss many other real environments. For example, if the exposure time for obtaining image data is extremely short, the quantum effects of counting photons might become dominant. In such a case, we have to consider Poissonian noise, not Gaussian noise, as a generative model for image data in this study. A previous study proposed a method of image restoration with respect to image data blurred by Poissonian noise[75]. It will be interesting to examine the relation between the influence of averaging on image analysis and hyper-parameter estimation also in the case of Poissonian generative models. In another real situation, dynamical systems are often studied using video data, viz., a time series of image data. In such a case, prior distributions should be discussed further considering not only spatial patterns based on MRF models but also spatio-temporal patterns described by partial differential equations. In addition, it is necessary to balance another averaging-related trade-off between spatial and temporal resolutions in acquiring video data. A promising future work will be to evaluate the influence of averaging preprocessing on image analysis based on realistic models.

5.4 Summary of this Section

We investigated the influence of averaging preprocessing on image analysis in this study, such as that in image restoration and hyper-parameter estimation, with an MRF model. We obtained two main results by analyzing Bayesian posterior distributions. First, the performance of image restoration with true hyper-parameters is independent of averaging. Second, averaging adversely affects the performance of hyper-parameter estimation. Averaging does not produce bias in hyper-parameter estimation but degrades the confidence and the performance of image restoration with the estimated hyper-parameters. As a result, averaging is not problematic as long as we use predetermined hyper-parameters for image restoration, but averaging is not appropriate for image restoration with hyper-parameter estimation.

Hyper-parameter estimation without image averaging requires a method of analysis for large amounts of image data. The analysis of free energy suggests that the variational Bayesian method is an effective approximation approach to hyper-parameter estimation, which takes the independence between hyper-parameters into consideration. It will be important to investigate the cases of images on various scales and movie data, such as when systems are dynamical and noise follows Poisson distributions, in future developments.

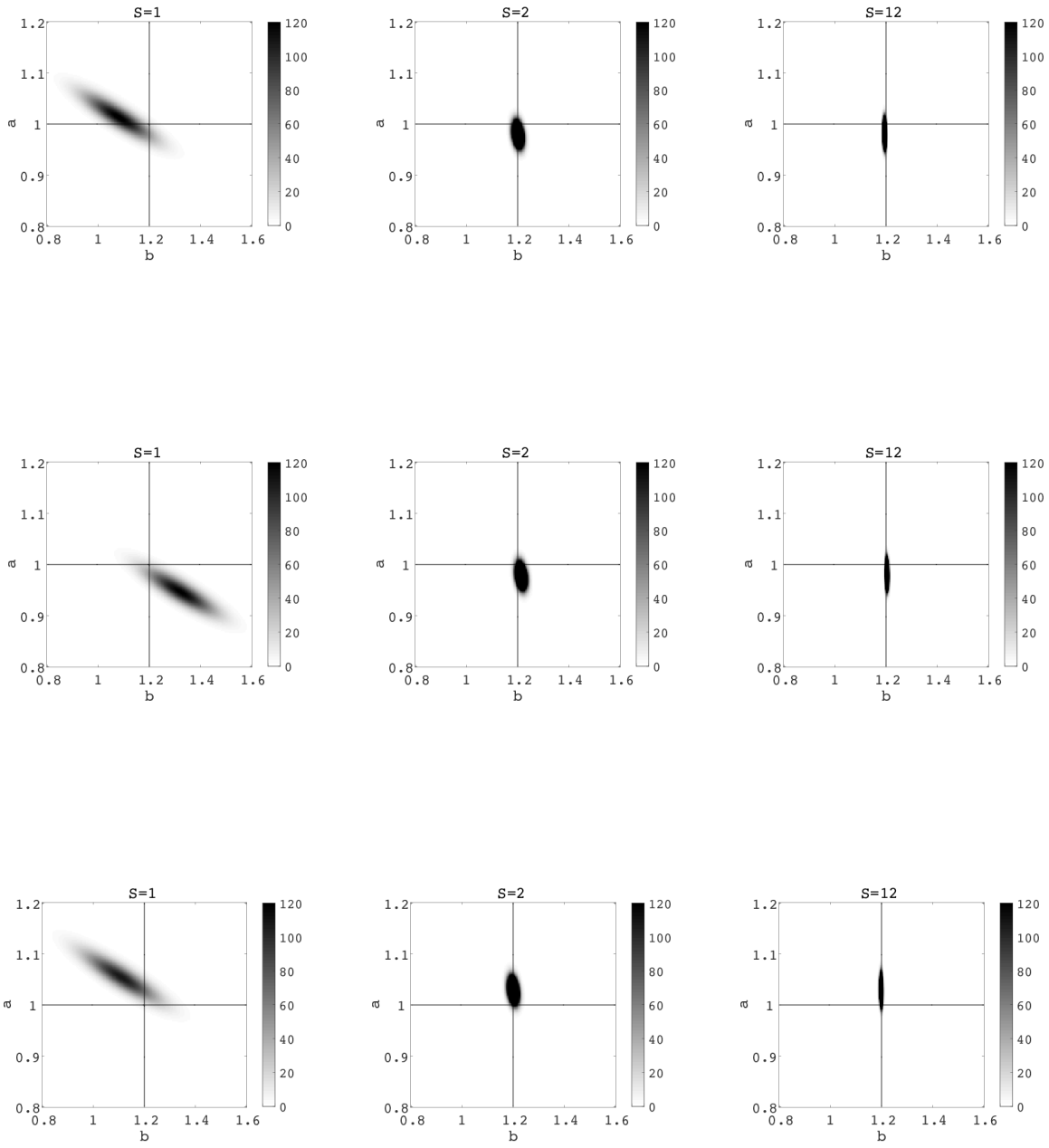

Figure 5.11: Examples of posterior distribution of hyper-parameters $P_S(a, b|\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S)$. Where, $N = 128^2$ pixels, $T = 12$, $(a_0, b_0) = (1, 1.2)$. The posteriors in the each columns are the results where the numbers of averaged images are $S = 1, 2$, and 12, respectively.

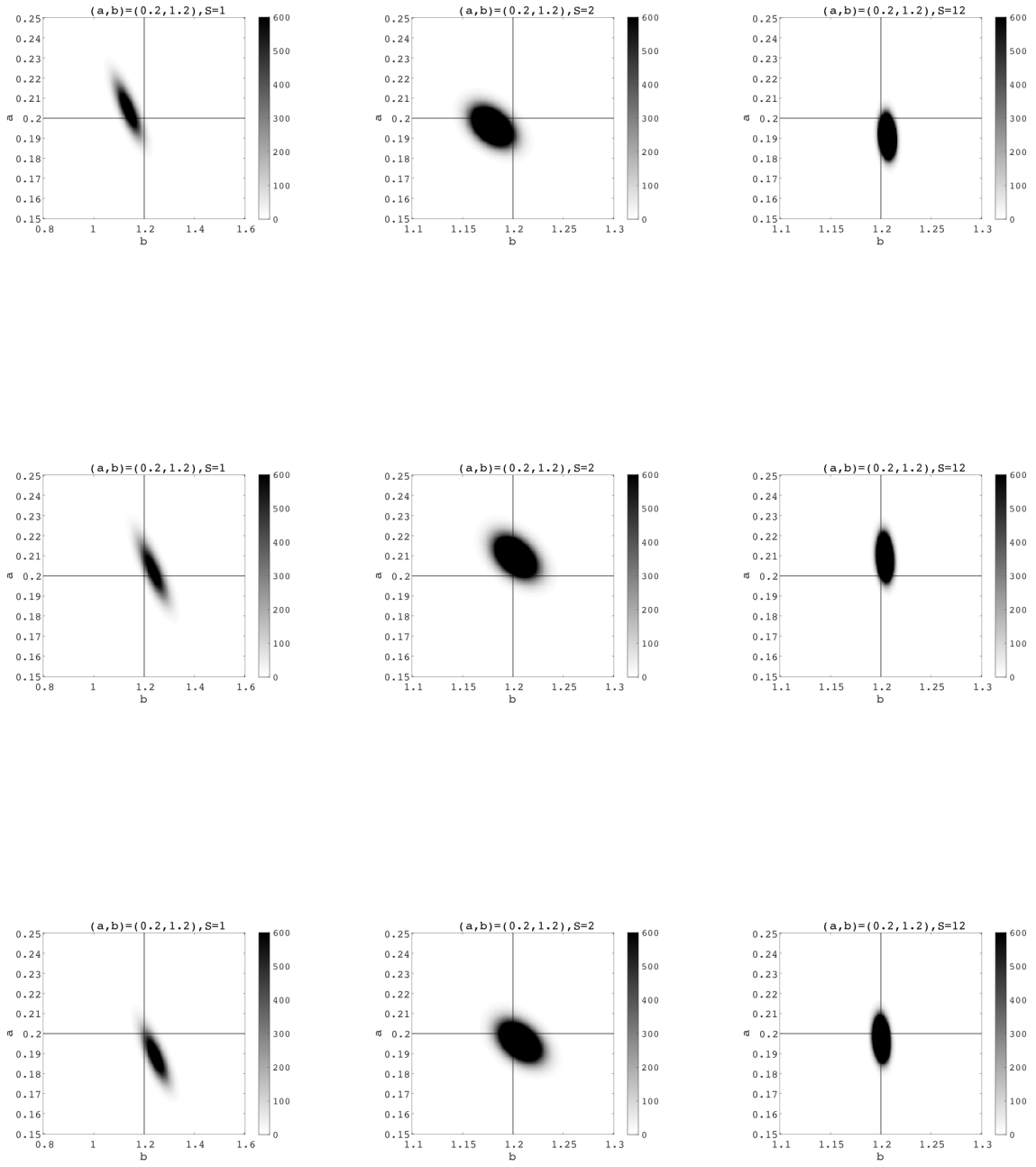

Figure 5.12: Examples of posterior distribution of hyper-parameters $P_S(a, b|\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S)$. This figure is the same as Fig. 5.11 except that the true hyper-parameters are $(a_0, b_0) = (0.2, 1.2)$.

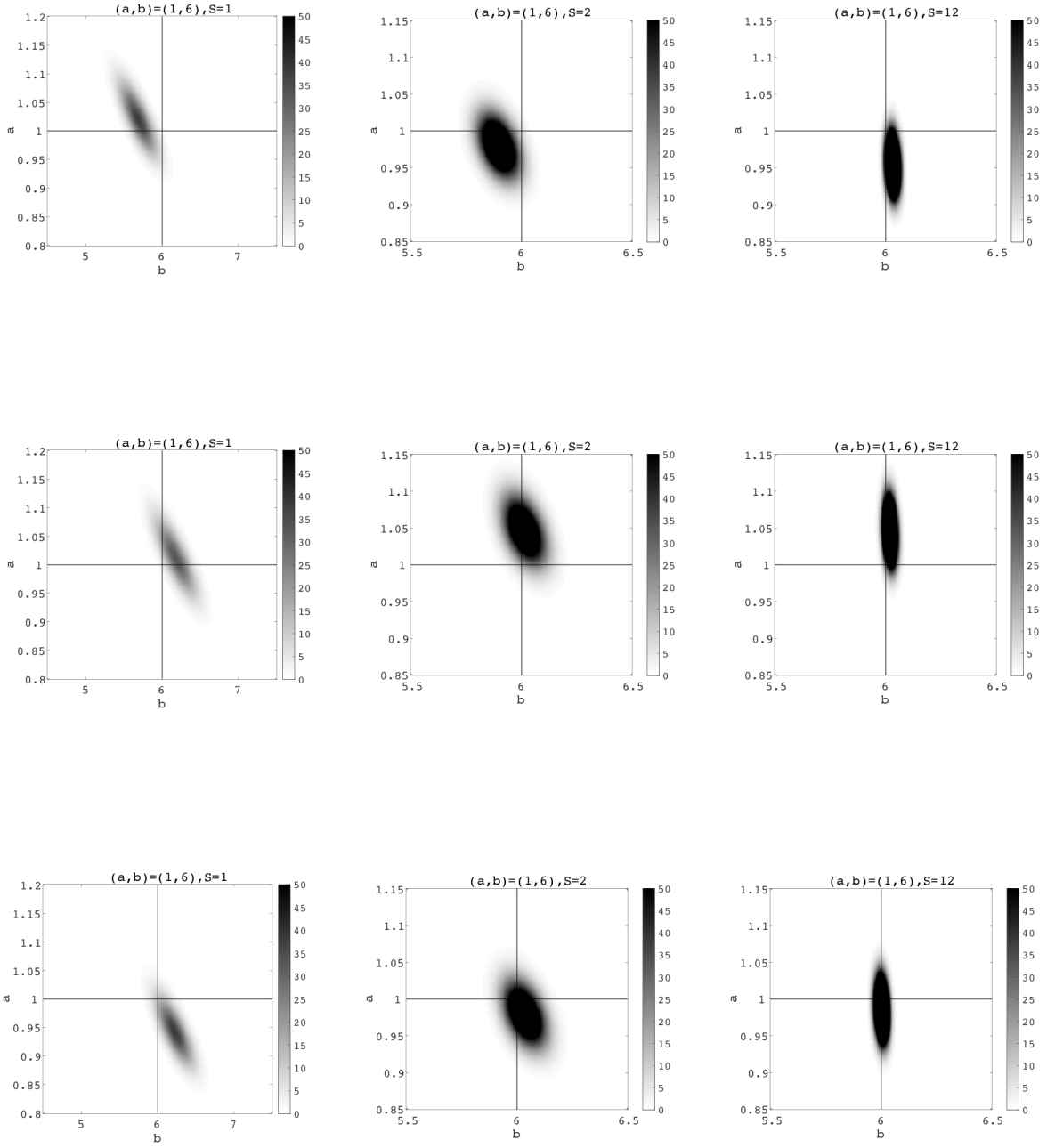

Figure 5.13: Examples of posterior distribution of hyper-parameters $P_S(a, b|\{\tilde{\boldsymbol{m}}^s\}_{s=1}^S)$. This figure is the same as Fig. 5.11 except that the true hyper-parameters are $(a_0, b_0) = (1, 6)$.

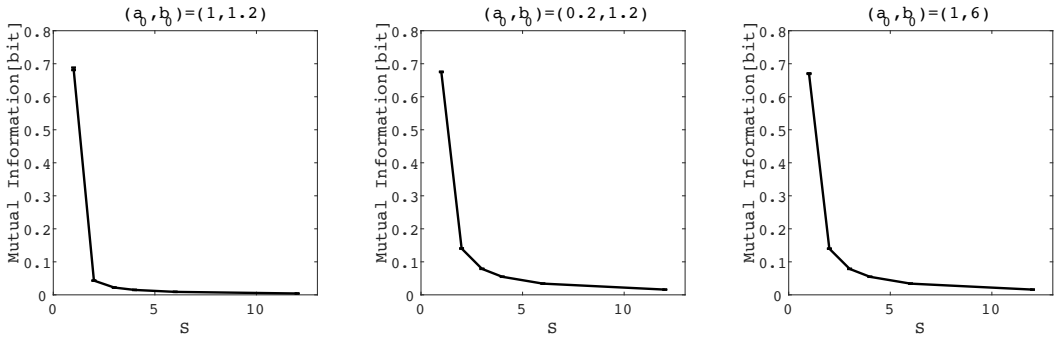

Figure 5.14: Results of 1000 trials of hyper-parameter estimation. Mutual information $I(a;b)$ is plotted against S. The solid line and its error bars represent the mean value and its standard error, respectively. The images are two-dimensional $N = 128^2$ pixels. The number of images is $T = 12$. The true hyper-parameters are $(a_0, b_0) = (1, 1.2)$, $(0.2, 1.2)$, and $(1, 6)$ from left to right.

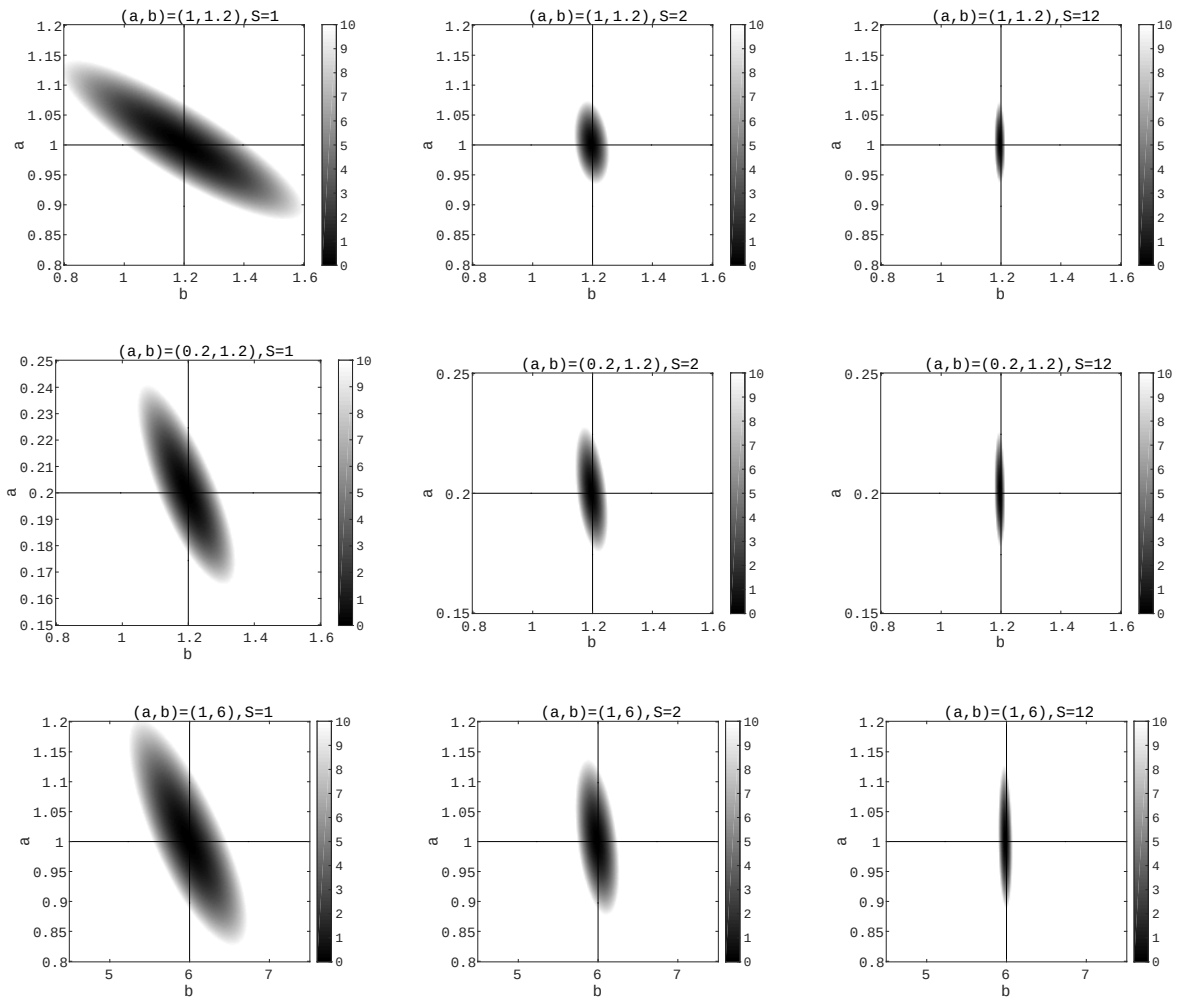

Figure 5.15: Configurational average of free energy $\left\langle F(a,b|\{\tilde{\boldsymbol{m}}_i^s\}_{s=1}^S)\right\rangle_{\{\tilde{\boldsymbol{m}}_i^s\}_{s=1}^S}$. The images are two-dimensional $N = 128^2$ pixels. The number of images is $T = 12$. The top, middle, and bottom rows correspond to the true hyper-parameters $(a_0, b_0) = (1, 1.2)$, $(0.2, 1.2)$, and $(1, 6)$, respectively. The left, middle, and right columns correspond to $S = 1$, 2, and 12, respectively.

Chapter 6

Bayesian Analysis of Crystal Lattice System Using Dispersion Relation Spectra

第 6 章は雑誌掲載が予定される内容を含むため，インターネット公表できません.

Chapter 7

Conclusion

In this thesis, we focused on the image data quantification and the design of measurement by utilizing the quantified image. We defined the image data quantification as the latent parameter estimation of the lattice models by using image data. Two observed methods obtain the image data; one is the mapping to real-space, and the other one is to Fourier-space. The thesis showed the approach and demonstration for Bayesian measurement modeling, for the latent parameter estimation of the images and design of the measurement and preprocessing.

We treated the Gaussian Markov random field (GMRF) models as the model of the image mapped to the real-space as the lattice system in Chapter 3, 4, and 5. In Chapter 3, we showed the exact analysis of the GMRF based on the Bayesian framework. We calculated the expectation values of the estimated model parameter distribution where the true data model is given. We demonstrate to evaluate the theoretical confidence of the data following the GMRF model. In the analysis, we present the limitation of confidence on the estimation for the model parameters even if we know the true model form. The proposed method to analyze the expectation values was utilized in Chapter 5.

In Chapter 4, we considered the two Bayesian model for the original image; one of the models was to generate data, and the other one was to estimate. In this chapter, we focused on evaluating the influence of sampling which is essential on the measurement of image data. The cause of why that we use both models is to represent the difference of the treatment between the generating the image and estimating from the image, from the point of view of sampling on the measurement. By using those models, we demonstrated that the sampling and estimation when the sampling deteriorates the image data. We showed the results to estimate the posterior distribution by using those models and evaluated the effect of down-sampling for the bias and spread of the posteriors. We proposed the method to correction of the bias by down-sampling of given image data. The shape changing with respect to down-sampling indicated the effect of the sampling step. Besides, the discussion in this chapter, we suggested that the deformation of the graphical model cause of this bias. This suggestion shows that the model evaluation is necessary for the model parameter estimation.

In Chapter 5, we demonstrate the method to evaluate the image processing step in the image analysis by using the Bayesian inference. In this Chapter, we focused on the trade-off of the image averaging. Through the latent parameters distribution estimation and image restoration tasks, the image averaging sometimes give the adverse effects on the performance of the tasks. By utilizing this evaluation of the influence of averaging, we can propose the better way of preprocessing image data for the image modeled by GMRFs. We show the way of design of the measurement process based on the Bayesian framework, due to the analysis of the influence of averaging which is the basic preprocessing for image data.

The work in Chapter 6 study the estimation of the latent parameter which is the interaction parameter of the crystal lattice by using the image mapped to Fourier-space. The image mapped to Fourier-space is obtained the spectral data. We use the Bayesian framework to estimate the latent parameters where the spectra by using the dispersion relation are treated as input data.

In Chapter 6, we proposed the methods to estimate the model parameters and evaluate its confidence which denote the interaction between the atom of the crystal lattice by using dispersion relation spectra that mirror the lattice vibration. The proposed methods show the results to estimate latent parameter by using the image of the dispersion relation. We can treat the data which has not been utilized for latent parameter estimation by using proposed method. Furthermore, we compared the method to estimate directly and the method based on the conventional way via the spectral deconvolution as the preprocessing of the image data. We show that the direct method has the advantages and better performance than the indirect method. Besides, we demonstrates the method to compare the better sampling method from Fourier space by using Bayesian analysis of dispersion relations. This method contributes to the design of the measurement process by using lattice observation from Fourier space such as neutron diffraction.

In conclusion, we propose the method to latent parameter estimation of the image data and evaluation of confidence of the data for the lattice systems. By utilizing the proposed modeling method for measurement, we can improve designs of the measurement and processing procedure of the image data.

Bibliography

[1] J. Simmons, P. Chuang, M. Comer, J. Spowart, M. Uchic, and M. De Graef, "Application and further development of advanced image processing algorithms for automated analysis of serial section image data," Modelling and Simulation in Materials Science and Engineering, vol.17, no.2, p.025002, 2008.

[2] M. Russina and F. Mezei, "First implementation of repetition rate multiplication in neutron spectroscopy," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol.604, no.3, pp.624–631, 2009.

[3] G. Binnig and H. Rohrer, "Scanning tunneling microscopy," Surface science, vol.126, no.1-3, pp.236–244, 1983.

[4] T. Hashimoto, ekkususen・hikari・chūseishisanran no genri to ōyō (Principles and Applications of X-ray, Light and Neutron Scattering), Koudansha, 2017.

[5] W.L.B. M.A., R.J. M.A., and C. Bosanquet, "Xxix. the intensity of reflexion of x-rays by rock-salt," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol.41, no.243, pp.309–337, 1921.

[6] C.M. Bishop, Pattern recognition and machine learning, Springer, 2006.

[7] K. Nagata and S. Watanabe, "Exchange monte carlo sampling from bayesian posterior for singular learning machines," IEEE Transactions on Neural Networks, vol.19, no.7, pp.1253–1266, 2008.

[8] K. Nagata, S. Sugita, and M. Okada, "Bayesian spectral deconvolution with the exchange monte carlo method," Neural Networks, vol.28, pp.82–89, 2012.

[9] S. Watanabe, Mathematical theory of Bayesian statistics, CRC Press, 2018.

[10] C. Kittel, P. McEuen, and P. McEuen, Introduction to solid state physics, Wiley New York, 1996.

[11] A. Woods, W. Cochran, and B. Brockhouse, "Lattice dynamics of alkali halide crystals," Physical Review, vol.119, no.3, p.980, 1960.

[12] H. Bilz and W. Kress, "Phonon dispersion relations and phonon models," in Phonon Dispersion Relations in Insulators, pp.6–16, Springer, 1979.

[13] V. Minkiewicz, G. Shirane, and R. Nathans, "Phonon dispersion relation for iron," Physical Review, vol.162, no.3, p.528, 1967.

[14] H. Attias, "A variational baysian framework for graphical models," Advances in neural information processing systems, pp.209–215, 2000.

[15] H. Sakamoto, Y. Nakanishi-Ohno, and M. Okada, "Theory of distribution estimation of hyperparameters in markov random field models," Journal of the Physical Society of Japan, vol.85, no.6, p.063801, 2016.

[16] Y. Nakanishi-Ohno, K. Nagata, H. Shouno, and M. Okada, "Distribution estimation of hyperparameters in markov random field models," Journal of Physics A: Mathematical and Theoretical, vol.47, no.4, p.045001, 2014.

[17] H. Sakamoto, Y. Nakanishi-Ohno, and M. Okada, "Influence of averaging preprocessing on image analysis with a markov random field model," Journal of the Physical Society of Japan, vol.87, no.2, p.024802, 2018.

[18] P. LAivY, "A special problem of brownian motion, and a general theory of gaussian random functions," Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, pp.133–175, 1972.

[19] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," Journal of the Royal Statistical Society. Series B (Methodological), pp.192–236, 1974.

[20] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," IEEE Transactions on pattern analysis and machine intelligence, no.6, pp.721–741, 1984.

[21] H. Derin, H. Elliott, R. Cristi, and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by markov random fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, no.6, pp.707–720, 1984.

[22] K. Tanaka and D. Titterington, "Statistical trajectory of an approximate em algorithm for probabilistic image processing," Journal of Physics A: Mathematical and Theoretical, vol.40, no.37, p.11285, 2007.

[23] H. Nishimori, Statistical physics of spin glasses and information processing: an introduction, Clarendon Press, 2001.

[24] N. Sourlas, "Spin-glass models as error-correcting codes," Nature, vol.339, no.6227, p.693, 1989.

[25] J. Besag, "On the statistical analysis of dirty pictures," Journal of the Royal Statistical Society. Series B (Methodological), pp.259–302, 1986.

[26] H. Nishimori, "Finite-temperature image restoration," Bussei Kenkyu, vol.73, no.5, pp.850–857, 2000.

[27] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, no.8, pp.799–813, 1989.

[28] J. Zhang, J.W. Modestino, and D.A. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," IEEE transactions on image processing, vol.3, no.4, pp.404–420, 1994.

[29] R. Hasegawa, M. Okada, and S. Miyoshi, "Image segmentation using region-based latent variables and belief propagation," Journal of the Physical Society of Japan, vol.80, no.9, p.093802, 2011.

[30] H. Nishimori and K.M. Wong, "Statistical mechanics of image restoration and error-correcting codes," Physical Review E, vol.60, no.1, p.132, 1999.

[31] J. Simmons, C. Przybyla, S. Bricker, D.W. Kim, and M. Comer, "Physics of mrf regularization for segmentation of materials microstructure images," Image Processing (ICIP), 2014 IEEE International Conference on, pp.4882–4886, IEEE, 2014.

[32] T. Kennett, W. Prestwich, and A. Robertson, "Bayesian deconvolution i: Convergent properties," Nuclear Instruments and Methods, vol.151, no.1-2, pp.285–292, 1978.

[33] D.V. Rubtsov and J.L. Griffin, "Time-domain bayesian detection and estimation of noisy damped sinusoidal signals applied to nmr spectroscopy," Journal of Magnetic Resonance, vol.188, no.2, pp.367–379, 2007.

[34] S.G. Razul, W. Fitzgerald, and C. Andrieu, "Bayesian model selection and parameter estimation of nuclear emission spectra using rjmcmc," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol.497, no.2-3, pp.492–510, 2003.

[35] P.J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," Biometrika, vol.82, no.4, pp.711–732, 1995.

[36] M.W. Cann, R. Nicholls, P. Roney, A. Blanchard, and F. Findlay, "Spectral line shapes for carbon dioxide in the 4.3-μm band," Applied optics, vol.24, no.9, pp.1374–1384, 1985.

[37] D. Zajfman, Z. Vager, R. Naaman, R. Mitchell, E. Kanter, T. Graber, and A. Belkacem, "The structures of c2h+ and c2h+ 2 as measured by coulomb explosion imaging," The Journal of chemical physics, vol.94, no.10, pp.6377–6387, 1991.

[38] M. Fujimoto, S.i. Aoshima, M. Hosoda, and Y. Tsuchiya, "Analysis of instantaneous profiles of intense femtosecond optical pulses propagating in helium gas measured by using femtosecond time-resolved optical polarigraphy," Physical Review A, vol.64, no.3, p.033813, 2001.

[39] K. Iwamitsu, S. Aihara, M. Okada, and I. Akai, "Bayesian analysis of an excitonic absorption spectrum in a cu2o thin film sandwiched by paired mgo plates," Journal of the Physical Society of Japan, vol.85, no.9, p.094716, 2016.

[40] S. Murata, K. Nagata, M. Uemura, and M. Okada, "Extraction of latent dynamical structure from time-series spectral data," Journal of the Physical Society of Japan, vol.85, no.10, p.104003, 2016.

[41] T. Kasai, K. Nagata, M. Okada, and T. Kigawa, "Nmr spectral analysis using prior knowledge," Journal of Physics: Conference Series, p.012003, IOP Publishing, 2016.

[42] N. Metropolis and S. Ulam, "The monte carlo method," Journal of the American statistical association, vol.44, no.247, pp.335–341, 1949.

[43] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," The journal of chemical physics, vol.21, no.6, pp.1087–1092, 1953.

[44] W.K. Hastings, "Monte carlo sampling methods using markov chains and their applications," 1970.

[45] A.E. Gelfand and A.F. Smith, "Sampling-based approaches to calculating marginal densities," Journal of the American statistical association, vol.85, no.410, pp.398–409, 1990.

[46] G.O. Roberts and A.F. Smith, "Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms," Stochastic processes and their applications, vol.49, no.2, pp.207–216, 1994.

[47] K. Hukushima and K. Nemoto, "Exchange monte carlo method and application to spin glass simulations," Journal of the Physical Society of Japan, vol.65, no.6, pp.1604–1608, 1996.

[48] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," Journal of statistical physics, vol.34, no.5-6, pp.975–986, 1984.

[49] Y. Ogata, "A monte carlo method for an objective bayesian procedure," Annals of the Institute of statistical Mathematics, vol.42, no.3, pp.403–433, 1990.

[50] M. Jerrum and A. Sinclair, "Polynomial-time approximation algorithms for the ising model," SIAM Journal on computing, vol.22, no.5, pp.1087–1116, 1993.

[51] S. Tokuda, K. Nagata, and M. Okada, "Simultaneous estimation of noise variance and number of peaks in bayesian spectral deconvolution," Journal of the Physical Society of Japan, vol.86, no.2, p.024001, 2016.

[52] J. Zhang, "The mean field theory in em procedures for markov random fields," IEEE Transactions on signal processing, vol.40, no.10, pp.2570–2583, 1992.

[53] S. Miyoshi and M. Okada, "Image restoration and segmentation using region-based latent variables: Bayesian inference based on variational method," Journal of the Physical Society of Japan, vol.80, no.1, p.014802, 2010.

[54] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.37, no.12, pp.2024–2036, 1989.

[55] J.M. Pryce and A. Bruce, "Statistical mechanics of image restoration," Journal of Physics A: Mathematical and General, vol.28, no.3, p.511, 1995.

[56] K. Tanaka, "Statistical-mechanical approach to image processing," Journal of Physics A: Mathematical and General, vol.35, no.37, p.R81, 2002.

[57] K. Tanaka, Introduction to Image Processing Techniques by Probabilistic Models, Morikita Shuppan, 2006.

[58] Y. Ohno, K. Nagata, T. Kuwatani, H. Shouno, and M. Okada, "Deterministic algorithm for nonlinear markov random field model," Journal of the Physical Society of Japan, vol.81, no.6, p.064006, 2012.

[59] T. Kuwatani, K. Nagata, M. Okada, and M. Toriumi, "Markov random field modeling for mapping geofluid distributions from seismic velocity structures," Earth, Planets and Space, vol.66, no.1, p.5, 2014.

[60] T. Kuwatani, K. Nagata, M. Okada, and M. Toriumi, "Markov-random-field modeling for linear seismic tomography," Physical Review E, vol.90, no.4, p.042137, 2014.

[61] Y. Igarashi, K. Nagata, T. Kuwatani, T. Omori, Y. Nakanishi-Ohno, and M. Okada, "Three levels of data-driven science," Journal of Physics: Conference Series, p.012001, IOP Publishing, 2016.

[62] B.C. Gupta and I. Guttman, Statistics and probability with applications for engineers and scientists, John Wiley & Sons, 2014.

[63] M. Yasuda, S. Kataoka, and K. Tanaka, "Statistical analysis of loopy belief propagation in random fields," Physical Review E, vol.92, no.4, p.042120, 2015.

[64] S. Kataoka, M. Yasuda, and K. Tanaka, "Statistical analysis of gaussian image inpainting problems," Journal of the Physical Society of Japan, vol.81, no.2, p.025001, 2012.

[65] S.C. Park, M.K. Park, and M.G. Kang, "Super-resolution image reconstruction: a technical overview," IEEE signal processing magazine, vol.20, no.3, pp.21–36, 2003.

[66] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp.417–424, ACM Press/Addison-Wesley Publishing Co., 2000.

[67] C.K. Williams and C.E. Rasmussen, "Gaussian processes for regression," Advances in neural information processing systems, pp.514–520, 1996.

[68] J.M. Bernardo and A.F. Smith, "Bayesian theory," 2001.

[69] T. Kazutaka and H. Nishimori, sōten-i ' rinkaigenshō to kurikomigun, Maruzen Shuppan, 2017.

[70] T.M. Cover and J.A. Thomas, "Elements of information theory 2nd edition," Willey-Interscience: NJ, 2006.

[71] K. Sano, R. Matsushita, and T. Tanaka, "To average or not to average: trade-off in compressed sensing with noisy measurements," Information Theory (ISIT), 2014 IEEE International Symposium on, pp.1316–1320, IEEE, 2014.

[72] D.L. Donoho, "Compressed sensing," IEEE Transactions on information theory, vol.52, no.4, pp.1289–1306, 2006.

[73] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," IEEE signal processing magazine, vol.25, no.2, pp.21–30, 2008.

[74] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), pp.267–288, 1996.

[75] H. Shouno, "Bayesian image restoration for poisson corrupted image using a latent variational method with gaussian mrf," Information and Media Technologies, vol.10, no.2, pp.363–372, 2015.

[76] M. Nakamura, R. Kajimoto, Y. Inamura, F. Mizuno, M. Fujita, T. Yokoo, and M. Arai, "First demonstration of novel method for inelastic neutron scattering measurement utilizing multiple incident energies," Journal of the Physical Society of Japan, vol.78, no.9, pp.093002–093002, 2009.

[77] K. Nakajima, S. Ohira-Kawamura, T. Kikuchi, M. Nakamura, R. Kajimoto, Y. Inamura, N. Takahashi, K. Aizawa, K. Suzuya, K. Shibata, *et al.*, "Amateras: a cold-neutron disk chopper spectrometer," Journal of the Physical Society of Japan, vol.80, no.Suppl. B, p.SB028, 2011.

[78] S. Itoh, T. Yokoo, S. Satoh, S.i. Yano, D. Kawana, J. Suzuki, and T.J. Sato, "High resolution chopper spectrometer (hrc) at j-parc," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol.631, no.1, pp.90–97, 2011.

[79] "The first brilluoin zone of a body centered cubic lattice(nov. 8, 2018, 11:35 jst). in graz university of technology phy.f20 molecular and solid state physics." http://lampx.tugraz.at/ hadley/ss1/bzones/bcc.php.

[80] T. Satoru, Effects of Measurement Noise on Bayesian Spectral Deconvolution: Degenerate or Not, Ph.D. thesis, University of Tokyo, 2017.

[81] R.A. Cowley, W. Cochran, B.N. Brockhouse, and A.D.B. Woods, "Lattice dynamics of alkali halide crystals. iii. theoretical," Phys. Rev., vol.131, pp.1030–1039, Aug 1963.