Doctoral Dissertation
博士論文

Parameter Distribution Estimation for Physical Models by Bayesian Inference
(ベイズ推論による物理モデルに対するパラメータ分布推定)

A Dissertation Submitted for the Degree of Doctor of Philosophy

December 2020

令和２年 12 月博士 (理学) 申請

Department of Physics, Graduate School of Schience,
The University of Tokyo
東京大学大学院理学系研究科
物理学専攻

Shun Katakami
片上 舜

Abstract

In this thesis, I describe a theoretical study on parameter estimation of physical models using Bayesian inference. With the development of observation technology in recent years, enormous and diverse observation data have been obtained in various fields of natural science. The purpose of analyzing observational data in the natural sciences is to extract physical information in order to clarify the physical phenomenon of the observation target. To clarify more realistic phenomena, the observation data has become large-amount and high-dimensional, and it has become difficult to extract physical information. Experts in each field confirm the data and analyzed it using physical models to estimate the physical quantities, but the complexity of the observed data caused three problems. First, manual analysis could not make full use of the data. Second, the physical models became complicated and it became difficult to evaluate the models. Finally, it became difficult to evaluate the reliability of the estimated physical quantities. Therefore, We deal with these problems by using Bayesian inference. Bayesian inference has evolved in the field of information science. It infers the model parameters from the observed events as an inverse problem. In Bayesian inference, large-amount data can improve the estimation accuracy, and the format of the data does not matter by setting appropriate models. In addition, Bayesian free energy can be used to evaluate the suitability of complex models for data. Furthermore, since the estimation obtained by Bayesian inference is a distribution estimation, the physical quantities can be estimated with reliability by regarding the spread of the distribution as the reliability. In this thesis, we propose a method for estimating physical parameters for dispersion relation observation data obtained in scattering experiments. Furthermore, we show the solvability of the Markov random field model that corresponds to the diffusion equation when it does not have the boundary conditions, and propose a method for estimating the parameters analytically. Finally, since the computational cost of Bayesian inference is generally high, we propose a high-speed estimation method that reduces the computational cost.

Acknowledgements

This thesis summarized the work during Ph. D. course of Physics, Graduate School of Science, 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 mentors Dr. Shin MURATA and Dr. Sakamoto 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 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 studying and writing to me. I would like to thank Mr. Yuki Yoshida, Mr. Tomofumi Nakayama, Mr. Kensuke MUTO for their constant encouragement on the laboratory life. I would like to thank secretaries of OKADA lab. Ms. Ayako FUKUYA, Ms. Chiaki SAKAKIBARA 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

1.1 Data analysis and Bayesian inference

In this thesis, we describe a theoretical study on parameter estimation of physical models using Bayesian inference. With the development of observation technology in recent years, enormous and diverse observation data have been obtained in various fields of natural science. The purpose of analyzing observational data in the natural sciences is to extract physical information in order to clarify the physical phenomenon of the observation target. In neuroscience, neural activity is observed by calcium imaging[1, 2]. Calcium imaging can visualize the calcium ion concentration in cells and tissues by using a fluorescent substance that reacts with calcium ions. In geoscience, geoscientists use seismic tomography to study the subsurface of the earth[3, 4, 5]. Seismic tomography is a method of imaging the inside of the earth using the property that the velocity of seismic waves differs depending on the internal state. In order to know the fluid distribution and dynamics inside the crust, the diffusion of the fluid is analyzed using seismic waves. In condensed matter physics, various physical responses are observed in terms of elementary excitations. Dispersion relation of bosonic excitations has been observed by neutron scattering or X-ray scattering. The dispersion relations are also essential to understand sound propagation, specific heat, thermal conductivity, superconductivity, and so on. To clarify more realistic phenomena, the observation data has become large-amount and high-dimensional, and it has become difficult to extract physical information. Experts in each field confirm the data and analyzed it using physical models to estimate the physical quantities. Information could be obtained from intuitive knowledge for a small number of data and low-dimensional data, but three problems arose due to the complexity of the observation data. First, the manual analysis could not make full use of the data. When the amount of data is large, it is difficult to analyze by

considering all the data. When the data is high-dimensional, it is difficult to obtain knowledge unless the data is reduced to low dimensions. Second, the physical models became complicated and it became difficult to evaluate the models. With the development of observation technology, it became necessary to consider more complicated physical models. Therefore, it is necessary to evaluate the physical models itself. Finally, it became difficult to evaluate the reliability of the estimated physical quantities. As the model becomes more complicated, the estimated parameters correlate with each other because the model has multiple parameters. As a result, it is difficult to evaluate with the error bars.

To address these problems, we propose methods by using Bayesian inference for the physical parameters from data of inelastic scattering experiment and image data. Our goal is to estimate model parameters from the data, which is an inverse problem. To solve the inverse problem, we use Bayesian estimation. When there is a forward simulation model that generates data from parameters, the conditional probability distribution of the parameters can be computed given the data. This probability distribution is a solution to the inverse problem, and the estimated value of the parameters and the reliability of the estimation can be obtained from the probability distribution.

There have been many previous studies that solve inverse problems using Bayesian estimation. Nagata et al. proposed Bayesian spectral decomposition to analyze spectral data using Bayesian estimation [6]. In Bayesian spectral decomposition, the spectral shape parameters can be estimated: center position, width, and intensity of the basis function. The model parameters to be estimated in this study are deeper parameters because they determine the spectral shape parameters. Previous studies that have analyzed deep parameters include a study that deals with NMR spectral data [7] and one that deals with time-series spectral data [8]. Local optimization is often a problem when searching for spectral shape parameters or deeper parameter solutions. In a previous study by Nagata et al., the replica exchange Monte Carlo method (REMC) was used to get out of the local solution and approach the global optimal solution. We also used REMC in this study.

When dealing with natural science data, it is important to evaluate the reliability and performance of the estimation and to support the influence of various conditions on the estimation. Stochastic information processing based on the framework of Bayesian inference is a processing method suitable for theoretical analysis of estimation. Bayesian inference is a framework widely used in stochastic information processing, and prior knowledge can be incorporated into estimation by modeling prior knowledge as a probability distribution[9]. Especially in the case of natural science data, the method based on Bayesian estimation is effective because there is

usually prior knowledge about the data. In Bayesian inference, large-amount data improves the estimation accuracy, and the format of the data does not matter by establishing an appropriate probabilistic model. In addition, the Bayesian free energy calculated by Bayesian inference can be used to evaluate the suitability of the stochastic model for the data. Furthermore, the reliability of the estimation can be evaluated by interpreting the spread of the distribution as an error bar of the estimated values of the physical quantities by estimating the parameter distribution by Bayesian inference.

1.2 Dispersion relation

In condensed matter physics, dispersion relation spectrum data has been well analyzed for estimation of elastic constant. Since spectrum analysis usually involves human processes, it was not possible to analyze all data. The validity of the physical model is not evaluated statistically. In previous study[10], we assumed that the observation data generation model has two processes: a physical process that generates a dispersion relation and an observation process by superimposing noise and Lorentz functions. Then, they made it possible to estimate the distribution of the elastic constant by Bayesian inference, using the observed noise as Gaussian noise. The fluctuation of event numbers often follows the Poisson process. In this thesis, we improve the method for estimating the model parameter distributions by introducing the Poisson noise into the observation process.

1.3 Markov random field model

In this thesis, we analyze the Gaussian Markov random field (MRF) model as the model of image data. The MRF model is a probabilistic model often used in image processing, and is applied to many image processing methods such as image restoration and region division[11, 12, 13, 14, 15, 16, 17]. The MRF model was formulated as a stochastic model by Geman, Derin et al. based on the continuity of pixel values between adjacent pixels in an image[18, 19]. Geman, Derin et al. modeled image generation by two stochastic processes: an original image generation process with continuity between adjacent pixels and an observation process in which observation noise is added. The generation process and observation process of the original image are characterized by parameters called hyperparameters, respectively. The hyperparameter of the generation process indicates the smoothness of the image, and the hyperparameter of the observation process indicates the magnitude of the observation noise. In the MRF model, it is pointed out that the hyperparameters that represent the smoothness of the image correspond to the diffusion

coefficient[20]. Conventionally, the optimum value of hyperparameters has been estimated as a secondary factor for improving the performance of image processing[13, 21, 22, 23]. However, Nakanishi, Sakamoto et al. showed that the hyperparameters that represent the smoothness of the MRF model correspond to the diffusion coefficient and estimated the hyperparameters of the Gaussian MRF model, including the reliability of the data[20, 24]. They mean the spread of the distribution as an error bar for the estimated value of the diffusion coefficient. As a result, the hyperparameters are essential values as latent variables of images, and the reliability needs to be estimated. Previous research by Nakanishi, Sakamoto et al. has made it possible to extract information on important latent variables from image data and evaluate the reliability of estimation.

1.4 Bayesian inference

Since Bayesian inference is a framework equivalent to statistical mechanics, stochastic information processing based on the Bayesian inference framework has been the subject of statistical mechanics analysis. In particular, the MRF model was introduced in relation to the spin system in statistical mechanics[18, 19]. The property that the values between adjacent pixels are close to each other is the same as that of the ferromagnetic spin, and in fact, their formulation corresponds to the Ising model, which is a typical model of the ferromagnetic spin. In the analysis of the MRF model using statistical mechanics, Nishimori et al. have evaluated the performance of image restoration using the MRF model of binary images as a previous study[25]. There is a Gaussian MRF model as a model for grayscale images and color images, and an image restoration method has been proposed by Nishimori, Tanaka and others[22, 26, 27].

The information that humans need to give in Bayesian inference are prior distributions and probability models. Information can be estimated accurately by avoiding the introduction of arbitrary assumptions in the prior distribution and probability models to eliminate as much anthropogenic effects as possible when making the estimation[28, 29]. However, by incorporating a complex structure into the prior distribution and probability model, it becomes difficult to calculate the posterior distribution, and you have to rely on the numerical approximation method. Variational inference methods, belief propagation, Markov chain Monte Carlo methods, etc. have been proposed as approximate calculation methods[16, 30, 31, 32]. Since the variational inference method and belief propagation are approximation methods, there is no guarantee that the estimated posterior distribution will match the true posterior distribution. Since the Markov chain Monte Carlo method is a sampling method, it converges to a true posterior distribution by

sufficient sampling. However, since the Markov chain Monte Carlo method requires repeated numerical calculations, the amount of calculation is large and high-speed analysis cannot be performed. Therefore, it is necessary to reduce the calculation cost and improve the calculation efficiency.

1.5 Our Focus and Contribution

We focus on parameter estimation for physical models by using Bayesian inference. Our targets are event data obtained from inelastic scattering experiments and observed image data. We construct this thesis as follows (as shown in Fig. 1.1)

Bayesian Parameter Estimation for Dispersion Relation Observation Data with Poisson Process

In chapter 2, we present a practical analysis method for physical quantity estimation for dispersion relation observation data, and discuss the effect of the noise model on estimation. In the previous study[10], we assumed that Gaussian noise was added to the dispersion relation observation data, and proposed Bayesian estimation method for model parameters. However, Gaussian noise is usually not suitable for event data in physical measurement. Gaussian noise takes continuous values, but raw count data is integer data that does not take negative values. Furthermore, assuming that Gaussian noise with constant noise intensity is added over the entire data, the difference in noise intensity in the energy and momentum space cannot be considered. Therefore, we propose an estimation method that introduces Poisson noise that is more suitable for event data into the estimation model, and compare the proposed method with the method of previous study by numerical experiments. As a result, it was confirmed that the proposed method improved the accuracy of the estimation for the observation data with Poisson noise by about 100 times on the time scale compared to the method of the previous study. From this, it was clarified that whether the noise mechanism in the observed data and the noise model of the estimation model match or not has a great influence on the estimaton of model parameters.

Gaussian Markov Random Field Model without Boundary Conditions

In chapter 3, we propose the analytical method of hyperparameter estimation for Gaussian MRF model without boundary condition. In general, Bayesian posterior probabilities are difficult to evaluate analytically, but in previous studies[20, 24], periodic boundary conditions were intro-

duced into the generative model of images, and Bayesian posterior probabilities were calculated analytically. However, many natural images do not have periodicity, and when analyzing an image to obtain local hyperparameters of the image or for image segmentation, the image data is divided into small pieces and it is necessary to analyze small-sized image data that does not have periodicity. In that case, the method assuming periodic boundary conditions will result in incorrect inference. In order to make accurate inferences, it is necessary to rely on numerical calculations such as REMC, which have high calculation costs. To solve these problems, we show that Bayesian posterior probabilities of image data without periodic boundary conditions can be evaluated analytically, and propose a method for analytically estimating hyperparameters. We confirm that the proposed method works properly by numerical experience.

Bayesian Hyperparameter Estimation using Gaussian Process and Bayesian Optimization

In chapter 4, we propose a high-speed hyperparameter distribution estimation method for models that are difficult to calculate analytically. We apply Bayesian optimization and Gaussian processes used in the field of information science to efficient evaluation of Bayesian posterior probabilities. We use the Gaussian process to estimate the hyperparameter distribution from a small number of grid samplings. We search for the optimum value of hyperparameters by Bayesian optimization with a small number of samplings. Furthermore, we show a method for estimating the distribution using Bayesian optimization as a sampler for the Gaussian process. We evaluate the performance of the proposed methods by numerical experiments using the hyperparameter distribution of the Gaussia MRF model, which can be analyzed and calculated. The distance between the true distribution and the estimated distribution was evaluated using several metrics. As a result, it was clarified that Bayesian optimization and Gaussian process work well for Bayesian posterior probability evaluation and that the proposed method can efficiently estimate the hyperparameter distribution as a probability distribution from a small number of samplings.

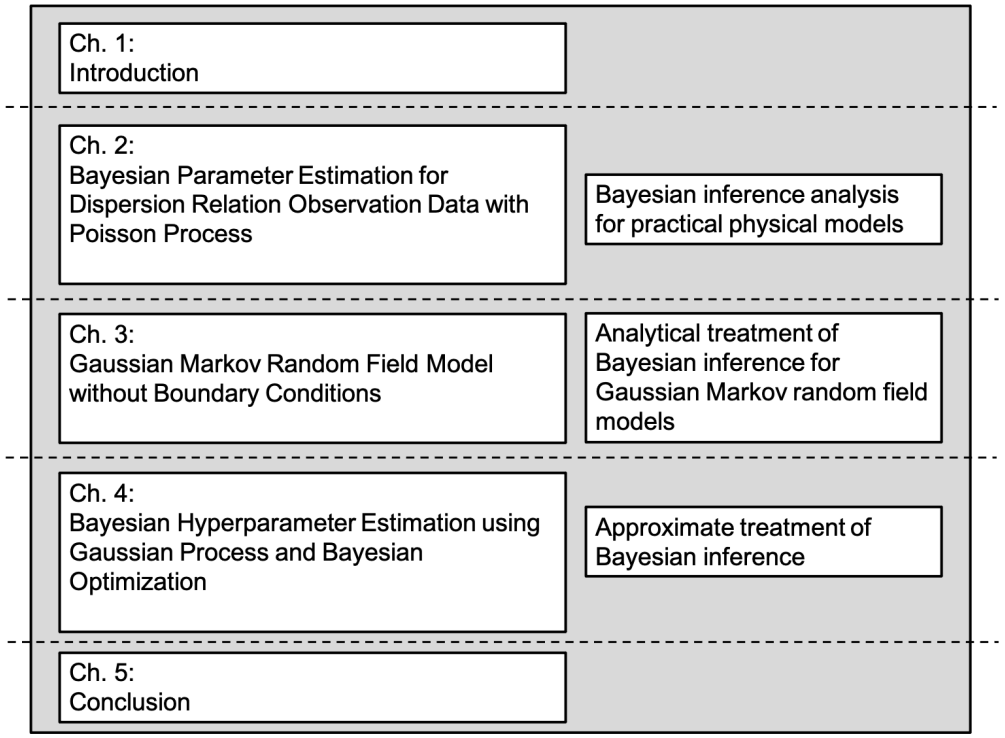


Fig. 1.1: Structure of this thesis

Chapter 2

Bayesian Parameter Estimation for Dispersion Relation Observation Data with Poisson Process

In condensed-matter physics, various physical responses are described in terms of elementary excitations. Each elementary excitation is characterized by its energy (frequency) and momentum (wave vector). The energy and momentum are connected by a dispersion relation, which is one of the more important characteristics of the elementary excitation. The dispersion relations are also essential to understand sound propagation, specific heat, thermal conductivity, superconductivity, and so on. Dispersion relations of bosonic excitations are observed by neutron scattering or X-ray scattering in general. The phonon dispersion relation was first observed in an inelastic neutron scattering measurement of germanium by Brockhouse and Iyengar[33, 34]. Since then, the inelastic neutron scattering technique has been widely used for studying phonons. In the early days of inelastic neutron scattering, a reactor-based neutron source was utilized. The obtained data were of a small size and could be analyzed in a simple way, irrelevant to the type of spectrometer. First, peak positions in the inelastic scattering spectrum at a certain momentum transfer $\boldsymbol{q}$ were estimated and regarded as the eigenfrequencies of phonons. Then, the obtained dispersion relation was fitted to a theoretical model with variable parameters. The validity of the model was evaluated rather intuitively. The establishment of high-power accelerator-based neutron sources like ISIS, SNS, J-PARC, and CSNS changed the situation. The combination of a high-power pulsed source and two-dimensional position-sensitive detectors remarkably improved the data acquisition efficiency. A time-of-flight neutron spectrometer designed for in-

elastic scattering measurements like MAPS [35], MERLIN [36], HYSPEC [37], ARCS [38], 4SEASONS [39], AMATERAS [40], and HRC [41] produces a large amount of event data in the four-dimensional (4D) frequency (ω)–momentum ($\boldsymbol{q}$) space that can be obtained within a couple of days. At present, the analysis of the event data are visualized as two-dimensional contour plot of one-dimensional spectrum for specified regions in the 4D space. Then, the physical model parameters are inferred to fit the estimated dispersion relations. However, the fitting is usually performed only for the visualized specified regions on experience. As far as the analysis is limited for the visualized data, it is impossible to perform the inference of the physical parameters by considering all the event data effectively.

To address these problems, we proposed a method to estimate the distribution of model parameters directly from high-dimensional event data. In the previous study[10], we assumed Gaussian noises in the observation process for the inference. The fluctuation of event numbers often follows the Poisson process. In this study, we improve the method for estimating the model parameter distributions by introducing the Poisson noise into the observation process.

We conduct numerical experiments using a simple model and artificial data. The dispersion relations are calculated as the vibration eigenmodes of the classical harmonic-oscillator lattice model of the body-centered cubic type. We assume that each eigenmode is expressed as a Lorentzian with identical peak height and width as a function of energy. Some sets of artificial histogramm data are generated from the Poisson process of the obtained spectral intensities. We show that the proposed method is superior to the method of the previous research for the data with a short observation time in terms of accuracy and reliability of Bayesian estimation.

2.1 Formulation of data generation process

In this study, we propose a method to estimate lattice parameters from the event data of historam, the analysis target is a body-centered cubic harmonic lattice model.

2.1.1 Dispersion relation

Let $\boldsymbol{u}(\boldsymbol{r}, t)$ be the displacement vector and the variable a be the lattice constant. In addition, the elastic constants for the first, second, and third nearest neighbors are set to α_1, α_2, and α_3. The eigen equation of motion is expressed as

$$-\omega^2 M\boldsymbol{u}(\boldsymbol{q}) = D\boldsymbol{u}(\boldsymbol{q}). \tag{2.1}$$

where, $u(\boldsymbol{q})$ is the Fourier transformation of the displacemment vector

$$\boldsymbol{u}(\boldsymbol{r},t) = \boldsymbol{u}(\boldsymbol{q})\mathrm{e}^{\mathrm{i}(\boldsymbol{q}\cdot\boldsymbol{r}-\omega t)}, \tag{2.2}$$

M is the mass of the atom and the coefficient matrix D is a q-dependent real symmetric matrix of 3×3.

The diagonal component d_{ii} and the off-diagonal component d_{ij} of D are

$$d_{ii} = \frac{8}{3}\alpha_1\left(\cos\left(\frac{aq_x}{2}\right)\cos\left(\frac{aq_y}{2}\right)\cos\left(\frac{aq_z}{2}\right) - 1\right) \tag{2.3}$$

$$+ 2\alpha_2\left(\cos(aq_i) - 1\right) + 2\alpha_3\left(\cos(aq_i)(\cos(aq_j) + \cos(aq_k)) - 2\right) \tag{2.4}$$

$$d_{ij} = -\frac{8}{3}\alpha_1 \sin\left(\frac{aq_i}{2}\right)\sin\left(\frac{aq_j}{2}\right)\cos\left(\frac{aq_k}{2}\right) - 2\alpha_3 \sin(aq_i)\sin(aq_j) \tag{2.5}$$

Here, $\lambda_1(\boldsymbol{q}), \lambda_2(\boldsymbol{q}), and \lambda_3(\boldsymbol{q})$, are three eigenvalues of the matrix D The dispersion relation can be described as

$$\omega_i = \sqrt{-\frac{\lambda_i(\boldsymbol{q})}{M}} \qquad (i = 1, 2, 3) \tag{2.6}$$

2.1.2 Observation process of data

The dispersion relation is observed as 4D histogram data including statistical noises. In this section, we introduce the observation process into the dispersion relation.

In general, every eigenmode has a life time. Thus the spectrum profile is represeted by a Lorentzian. The intesity for energy E_i and momentum $\boldsymbol{q}_j$ can be written as

$$I(E_i, \boldsymbol{q}_j; \boldsymbol{\alpha}) = \sum_{k=1}^{3} \eta_k(\boldsymbol{q}_j)\phi_k(E_i; \boldsymbol{q}_j, \boldsymbol{\alpha}, \gamma(\boldsymbol{q})), \tag{2.7}$$

$$\phi_k(E; \boldsymbol{q}, \boldsymbol{\alpha}, \gamma) \propto \frac{1}{\pi}\frac{\gamma^2}{(E - \hbar\omega_k(\boldsymbol{q}; \boldsymbol{\alpha}))^2 + \gamma^2}. \tag{2.8}$$

$\gamma_k(\boldsymbol{q})$ is the width of Lorentzian, $\eta_k(\boldsymbol{q})$ is the intensity of each mode basis, which is determined by the measurement system, the momentum q, temperature, and so on. γ_k is also dependent on the momentum $\boldsymbol{q}$ and temmperature, in general. $\Theta = \{\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\gamma}\}$ are considered as model parameters and $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_3\}$

In this study, for simplicity, we assume that γ_k and η_k are constants.

$$I(E_i, \boldsymbol{q}_j; \Theta) = \sum_{k=1}^{3} \eta_k\phi_k(E_i; \boldsymbol{q}_j, \boldsymbol{\alpha}, \gamma_k) \tag{2.9}$$

Assuming that the observation process follows Poisson distribution, and

The observed average of spectral intensities $\bar{y}_{ij}$ for the observation time T is represented as

$$\bar{y}_{ij} = (I(E_i, \boldsymbol{q}_j; \Theta) + B)T. \tag{2.10}$$

If the observation process follows Poisson distribution, the acquired spectral data $Y = \{y_{ij}\}$ can be

$$P(y_{ij}|I(E_i, \boldsymbol{q_j}; \Theta) = \frac{\bar{y}_{ij}^{y_{ij}} \exp(-\bar{y}_{ij})}{y_{ij}!}. \tag{2.11}$$

Here, we introduce uniform background B independent of energy or momentum.

2.1.3 Previous study: Observation process of data

In previous study[10], we assume that the observation process follows a Gauss distribution, the acquired spectral data $Y = \{y_{ij}\}$ can be

$$P(y_{ij}|I(E_i, \boldsymbol{q_j}; \Theta) = \mathcal{N}\left(I(E_i, \boldsymbol{q_j}; \Theta, \sigma^2_{\mathrm{REMC}}\right) \tag{2.12}$$

where σ_{REMC} is the lower bound noise intensisity for REMC. In this study, since it is generally impossible to know the noise intensity of data, noise estimation[42] is performed by giving the noise lower bound.

2.2 Analysis

2.2.1 Bayesian Inference

In this section, we show a method to estimate model parameters by Bayesian inference from the obtained observation data. The observation point is $X = \{x_{ij}|x_{ij} = (E_i, \boldsymbol{q}_j)\}$, and the observation value is $Y = \{y_{ij}\}$, We can write observation data $\mathcal{D} = \{X, Y\}$.

The number of data is $N = |D|$. Then, assuming that each observation follows the same independent distribution, the probability distribution of Y is

$$P(Y|X,\Theta) = \prod_{i,j} P(y_{ij}|I(E_i, \boldsymbol{q_j};\Theta) \tag{2.13}$$

$$= \prod_{i,j} \frac{\bar{y}_{ij}^{y_{ij}} \exp(-\bar{y}_{ij})}{y_{ij}!} \tag{2.14}$$

$$= \exp\left(-N\mathcal{E}(\Theta)\right) \tag{2.15}$$

$$\mathcal{E}(\Theta) \equiv \frac{1}{N} \sum_{ij} \left(\bar{y}_{ij} - y_{ij} \ln \bar{y}_{ij} + \ln y_{ij}!\right) \tag{2.16}$$

From Bayes' theorem, the posterior probability of the model parameter Θ is

$$P(\Theta|\mathcal{D}) = \frac{P(Y|X,\Theta)\varphi(\Theta)}{P(Y|X)} \tag{2.17}$$

$$= \frac{1}{P(Y|X)} \varphi(\Theta) \prod_{i,j} P(y_{ij}|I(E_i, \boldsymbol{q_j};\Theta) \tag{2.18}$$

$$= \frac{1}{Z(\mathcal{D})} \exp\left(-N\mathcal{E}\right) \varphi(\Theta), \tag{2.19}$$

where $\varphi(\Theta)$ is the prior distribution, $Z(\mathcal{D})$ is the marginal likelihood or normalization constant,

$$Z(\mathcal{D}) = \int \exp\left(-N\mathcal{E}(\Theta)\right) \varphi(\Theta) \mathrm{d}\Theta \tag{2.20}$$

We use the prior distributions $\varphi(\Theta) = \varphi_\alpha(\boldsymbol{\alpha})\varphi_\gamma(\gamma)\varphi_\eta(\eta)$. Each distribution is written as half-Cauchy distributions $\varphi_\alpha(\boldsymbol{\alpha}) = \prod_{k=1}^{3} C(\alpha_k;\iota_\alpha)$, $\varphi_\eta(\eta) = \prod_{k=1}^{3} C(\eta_k;\iota_\eta)$, and the inverse gamma distribution $\varphi_\gamma(\gamma) = \prod_{k=1}^{3} G(\gamma_k; a_\gamma, b_\gamma)$, where the half-Cauchy distribution is

$$C(\eta;\iota_\eta) = \begin{cases} \frac{1}{\pi}\frac{\iota_\eta}{\eta^2+\iota_\eta^2} & (\eta \geq 0), \\ 0 & (\eta < 0), \end{cases} \tag{2.21}$$

and the inverse gamma distribution is

$$G(\gamma; a_\gamma, b_\gamma) = \prod_i \frac{b_\gamma^{a_\gamma}}{\Gamma(a_\gamma)} \gamma_i^{-(a_\gamma+1)} e^{-b_\gamma/\gamma_i}. \tag{2.22}$$

2.2.2 Posterior Probability Calculation: Exchange Monte Carlo Method

When computing Bayes posterior probabilities, computing the partition function is unavoidable. The partition function is often computationally difficult to calculate, and the numerical cost is also enormous. In previous research[6, 10], the exchanged Markov chain Monte Carlo (REMC)

method [43] was utilized. The REMC method is a sampling method that reduces the amount of calculation by approximately calculating expected values. Here, we explain the exchange REMC method using the Bayesian framework.

For convenience, we introduce the inverse temperature β and define the function $z(\beta)$:

$$z(\beta) = \int \exp\left(-N\beta\mathcal{E}(\Theta)\right)\varphi(\Theta)\mathrm{d}\Theta \tag{2.23}$$

The partition function can be expressed as $Z(\theta|\mathcal{D}) = z(1)$. Therefore, L inverse temperatures $\boldsymbol{\beta} = \{\beta_1, \beta_2, \cdots, \beta_L | 0 = \beta_1 < \beta_2 < \cdots < \beta_L = 1\}$, then $z(0) = 1$

$$z(1) = \frac{z(\beta_L)}{z(\beta_{L-1})} \times \frac{z(\beta_{L-1})}{z(\beta_{L-2})} \times \cdots \times \frac{z(\beta_2)}{z(\beta_1)} = \prod_{l=1}^{L-1} \frac{z(\beta_{l+1})}{z(\beta_l)} \tag{2.24}$$

$$= \prod_{l=1}^{L-1} \frac{\int \exp\left(-N\beta_{l+1}\mathcal{E}(\Theta)\right)\varphi(\Theta)\mathrm{d}\Theta}{\int \exp\left(-N\beta_l\mathcal{E}(\Theta)\right)\varphi(\Theta)\mathrm{d}\Theta} \tag{2.25}$$

$$= \prod_{l=1}^{L-1} \frac{\int \exp\left(-N(\beta_{l+1}-\beta_l)\mathcal{E}(\Theta)\right) q(\Theta;\beta_l)\mathrm{d}\Theta}{\int q(\Theta;\beta_l)\mathrm{d}\theta} \tag{2.26}$$

$$= \prod_{l=1}^{L-1} \left\langle \exp\left(-N(\beta_{l+1}-\beta_l)\mathcal{E}(\Theta)\right)\right\rangle_{q(\times;\beta_l)} \tag{2.27}$$

$$q(\Theta;\beta) \propto \exp\left(-N\beta\mathcal{E}(\Theta)\right)\varphi(\Theta) \tag{2.28}$$

where, $\langle\cdot\rangle_{q(\Theta;\beta)}$ is the expected value for $q(\Theta;\beta)$.

From the formula (3.9), the marginal likelihood $Z(\mathcal{D})$ can be represented by the expected value that can be calculated by REMC method. In this case, the expected value for the probability density for different inverse temperatures is calculated, so it can be calculated in parallel.

2.3 Numerical experiment

In this section, we evaluate the performance of model parameter estimation method using Bayesian inference. We show numerical experiments comparing the performances of the proposed method and the previous method using the same synthetic data.

First, we generated the dispersion relation data of the body-centered cubic lattice model according to the Eq (2.11). The parameters used to generate the synthetic data are shown in Table 2.1. We choose points with high symmetry in the first Brillouin zone of momentum space, and sample the straight line connecting each point $\Gamma[0,0,0] \to H[0,0,2\pi/a] \to N[0,\pi/a,\pi/a] \to P[\pi/a,\pi/a,\pi/a] \to \Gamma[0,0,0]$. Examples of the synthetic data are shown in Fig 2.1. Figure 2.1

(a)-(d) is the spectrum data when the observation time is $T = 100, 10, 1, 0.1$. Model parameter estimation is obtained as a posterior probability of the spring constant sampled by the REMC method.

Table 2.1: Parameters of synthetic data.

N	50
$(\alpha_1, \alpha_2, \alpha_3)$[N/m]	(300, 200, 100)
(η_1, η_2, η_3)	(1, 1, 1)
$(\gamma_1, \gamma_2, \gamma_3)$[meV]	(0.0025, 0.0025, 0.0025)
M_{atom}[kg]	2×10^{-26}
B	0.8

* Note that the phonon energy is independent of the lattice constant a. The lattice constant is relevant only to the size of the first Brillouin zone.

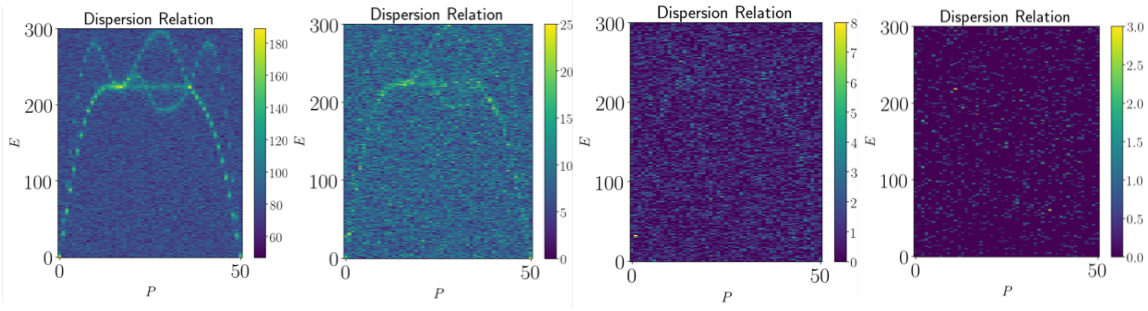
Table 2.2: Parameters of the priors and REMC conditions for parameter estimation.

L	48
M	120,000
Burn In Step	60,000
ι_α	150
ι_η	0.5
a_γ	4.001
b_γ[meV]	0.01001
ξ	1.4

Figure 2.2–2.5 show examples of spring constant estimations. The posterior probability distributions for the synthetic data with observation time $T = 100, 10, 1, 0.1$ (Fig. 2.1 (a)–(d)) are Fig. 2.2–2.5.

Since the spread of the distribution is evaluated as the reliability in the numerical experiments in this study, the estimated mean and the spread of the estimated distribution are shown statis-

tically. For simplicity, we consider the variance as the spread of the estimated distribution. In order to statistically evaluate the mean and variance obtained by each inference, Tables 2.3 and 2.4 show the average values and its variances obtained with 10 inferences and their standard deviations.



(a) $T = 100$ (b) $T = 10$ (c) $T = 1$ (d) $T = 0.1$

Fig. 2.1: Observation data of generated dispersion relation. (a)-(d) are spectrum data when $T = 100, 10, 1, 0.1$ respectively.

Table 2.3: Descriptive statistics values of posteriors by Poisson EMC.

Time	Posterior Mean			Posterior Std.		
	α_1 [N/m]	α_2 [N/m]	α_3 [N/m]	α_1 [N/m]	α_2 [N/m]	α_3 [N/m]
1000	300.00 ± 0.11	200.01 ± 0.09	100.03 ± 0.07	0.09 ± 0.00	0.11 ± 0.01	0.08 ± 0.01
100	299.93 ± 0.26	199.92 ± 0.40	100.07 ± 0.26	0.26 ± 0.01	0.33 ± 0.01	0.24 ± 0.02
10	300.33 ± 1.02	199.88 ± 1.05	99.69 ± 0.46	0.85 ± 0.09	1.10 ± 0.11	0.81 ± 0.08
1	301.72 ± 3.04	200.41 ± 4.46	98.65 ± 2.70	3.30 ± 1.39	3.67 ± 0.99	2.69 ± 0.96
0.1	1101.65 ± 924.92	1219.80 ± 2014.31	1056.90 ± 814.90	12382.66 ± 15656.31	9369.42 ± 19538.85	9769.06 ± 12901.07

Table 2.4: Descriptive statistics values of posteriors by Gaussian EMC.

Time	Posterior Mean			Posterior Std.		
	α_1 [N/m]	α_2 [N/m]	α_3 [N/m]	α_1 [N/m]	α_2 [N/m]	α_3 [N/m]
1000	300.00 ± 0.06	199.99 ± 0.06	100.01 ± 0.03	0.08 ± 0.00	0.10 ± 0.00	0.07 ± 0.00
100	299.88 ± 0.39	199.98 ± 0.63	100.09 ± 0.35	0.89 ± 0.42	1.24 ± 0.40	0.95 ± 0.32
10	2869.40 ± 4335.94	1288.29 ± 685.08	1037.25 ± 327.82	75030.23 ± 182473.8	18510.97 ± 19274.41	10078.06 ± 8481.60
1	1092.83 ± 793.79	1469.85 ± 1518.43	973.71 ± 344.36	28949.34 ± 54884.65	41824.27 ± 78722.12	18154.27 ± 26247.32
0.1	4154.65 ± 9067.73	1409.29 ± 1107.07	1061.90 ± 235.39	325395.19 ± 898216.3	45478.45 ± 74316.54	24386.52 ± 17130.80

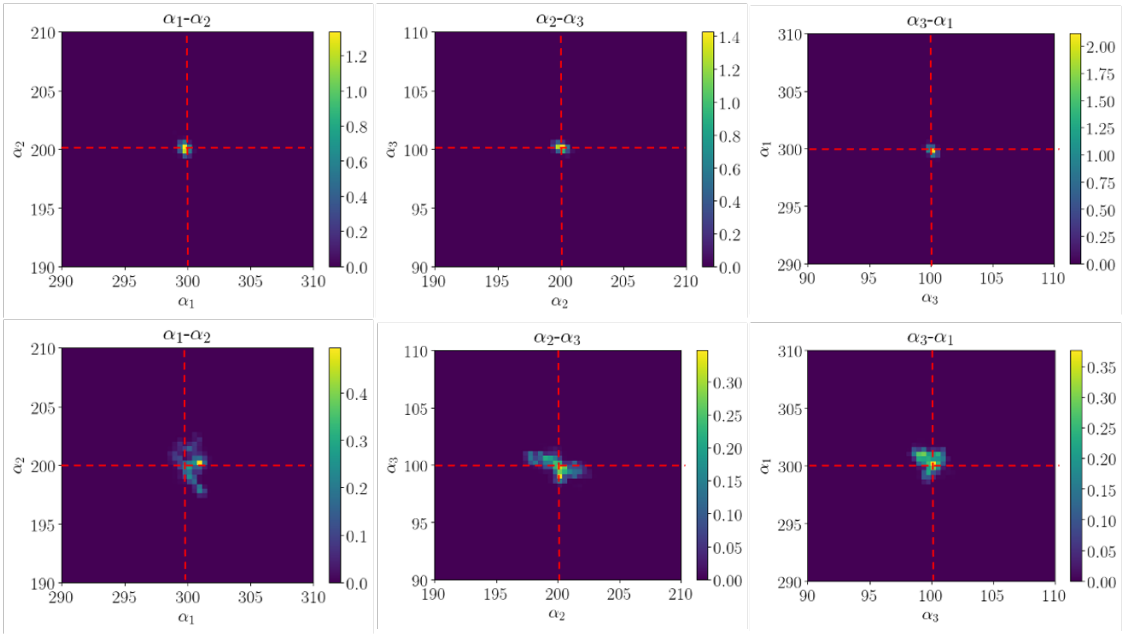


Fig. 2.2: Posterior distributions of $\boldsymbol{\alpha}$ of for the observation data generated by $T = 100$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\alpha}$.The parameters used for the REMC method for estimation of $\boldsymbol{\alpha}$ are shown in Table 2.1.

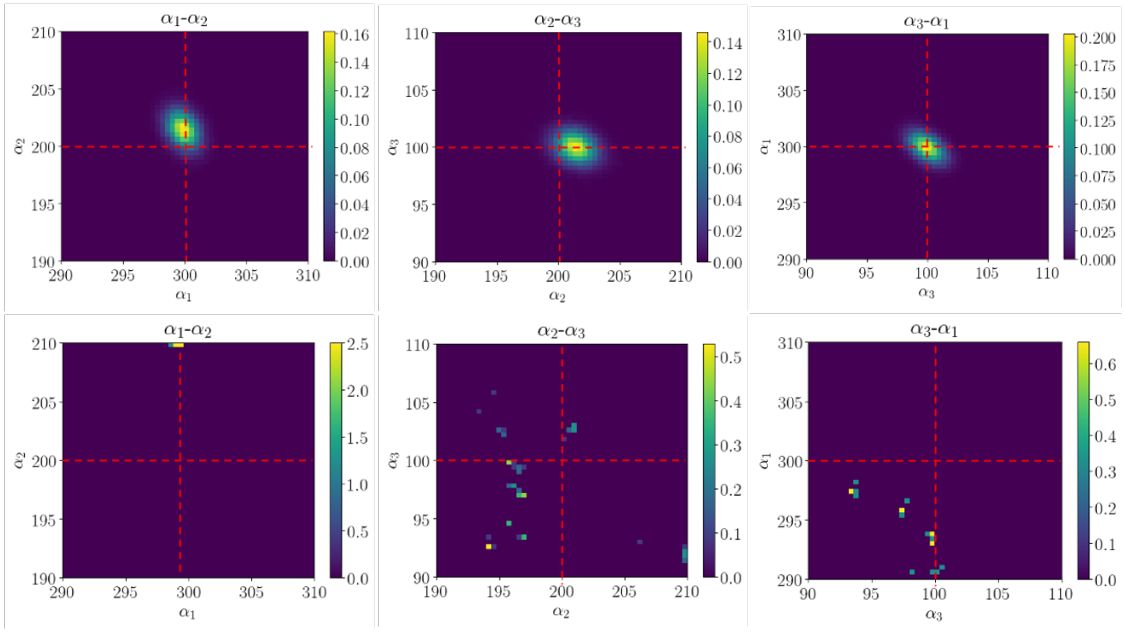


Fig. 2.3: Posterior distributions of $\boldsymbol{\alpha}$ of for the observation data generated by $T = 10$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\alpha}$.The parameters used for the REMC method for estimation of $\boldsymbol{\alpha}$ are shown in Table 2.1.

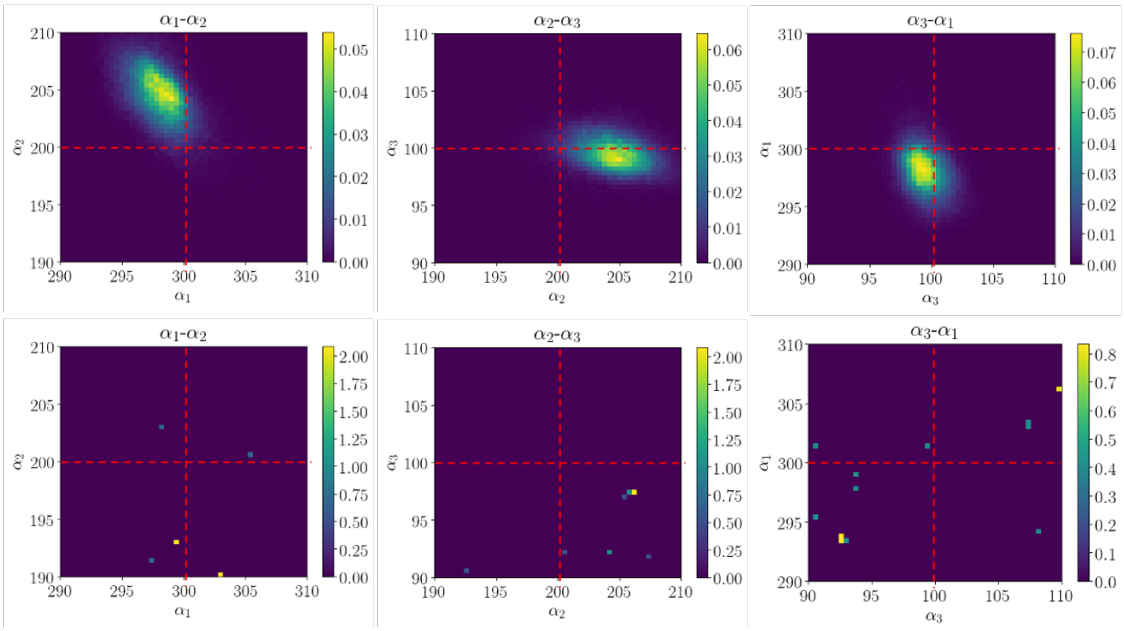


Fig. 2.4: Posterior distributions of $\boldsymbol{\alpha}$ of for the observation data generated by $T = 1$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\alpha}$.The parameters used for the REMC method for estimation of $\boldsymbol{\alpha}$ are shown in Table 2.1.

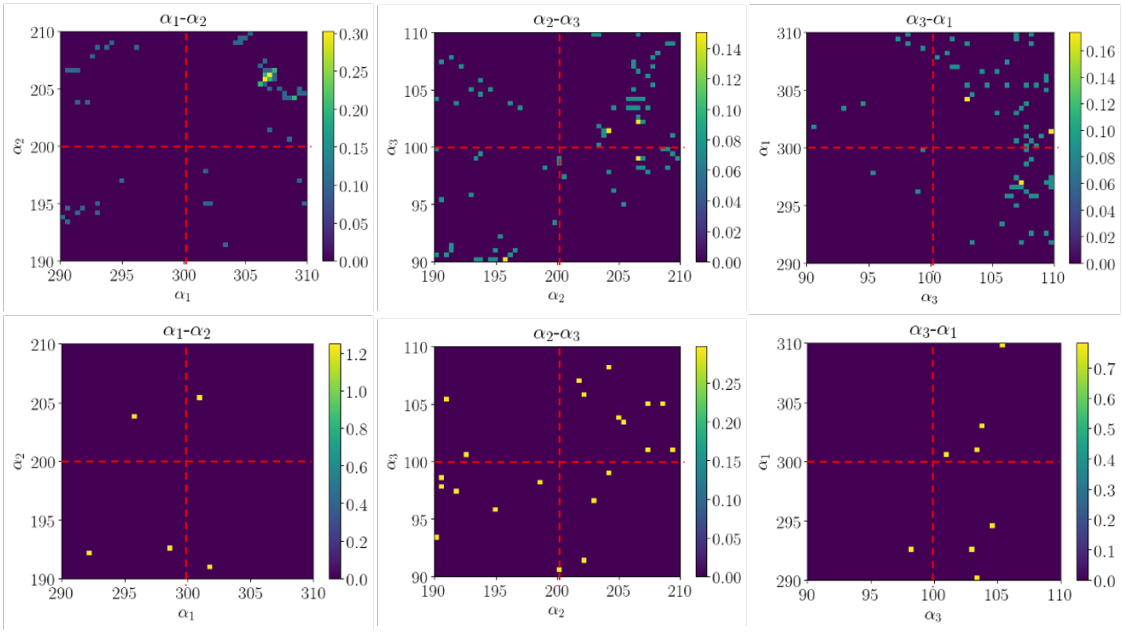


Fig. 2.5: Posterior distributions of $\boldsymbol{\alpha}$ of for the observation data generated by $T = 0.1$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\alpha}$.The parameters used for the REMC method for estimation of $\boldsymbol{\alpha}$ are shown in Table 2.1.

2.4 Discussion

In the numerical experiments in Sect. 2.3, we compared the proposed method with the method from the previous study and evaluated the performance of estimating physical model parameters from dispersion relation observation data. Since the observation data is supposed to be physical event measurement data, it was generated by the Poisson process.

From Figs. 2.2–2.5, it can be seen that the proposed method has superior accuracy compared to the method of the previous study. In addition, Tables 2.3 and 2.4 show that the proposed method improves the inference performance by about 100 times with respect to the observation time compared to the previous research.

This difference in inference performance is affected by whether the model and the likelihood function match. The likelihood function of the model of the proposed method has a Poisson distribution, and the model and the generation process of observation data match. However, in the method of the previous research, the likelihood function of the model has a Gaussian distribution, which does not match the generation process of the observation data. It was clarified that the estimation performance could be significantly reduced because the noise model of the estimation model does not match the noise mechanism of the observed data. Introducing an appropriate noise mechanism into the estimation model is necessary to improve the estimation performance.

In this study, we constructed a model for inelastic scattering experimental data of neutrons, and since the acquired data is stochastically measured count data, we introduced the Poisson noise model. Poisson noise can basically be used for the count data of stochastic events. In the Poisson distribution, the mean and variance are the same value, but in actual data, there are cases where the mean and variance of the distribution are not the same value even for count data. This is because there are noise factors that cannot be expressed by the Poisson distribution alone, such as the characteristics of measuring instruments or physical phenomena from other systems. In actual operation, if multiple noise models are possible, it is necessary to evaluate the noise model by model selection. In Bayesian inference, the appropriateness of the estimation model for the data can be measured by using the Bayesian free energy. First, the free energy is calculated for each model, and the model with the minimum free energy can be considered as the model most suitable for the data.

In our method, it is not necessary to estimate the noise level as in the case of the previous research method because our method does not include noise parameters in the observation model.

Therefore, the uncertainty due to noise estimation does not affect the parameter estimation, and stable analysis results can be obtained by our method.

In event measurement, it is desirable to keep the data as raw count data. In the fields of physical measurement and information measurement, measurement data is often processed and not retained as raw count data. Since our method is probabilistically modeled using a Poisson process on Bayesian inference and treats the data as raw count data, the accuracy is dramatically improved. The properties of count data may lead to improvement in estimation accuracy. Therefore, in event measurement, the data should be retained as raw count data.

2.5 Conclusion

In this study, we proposed a method for estimating the interaction parameters of the crystal lattice from the dispersion relation observation data, conducted numerical experiments using artificial data, and discussed performance evaluation. In the proposed method, the Poisson process, which is a physical observation process, was introduced into the inference model. Therefore, the proposed method has better inference accuracy than the conventional method, and it was possible to infer physical parameters with high accuracy from short-time measurement data.

As a future subject, we will perform more practical Bayesian modeling for actual data analysis. The shell model [44] can be handled as a more practical model for the calculation of the dispersion relation of lattice vibration. By extending our model, it is expected that it will be applied to more accurate model parameter estimation and optimization of experimental observation conditions.

Chapter 3

Gaussian Markov Random Field Model without Boundary Conditions

In this chapter, we deal with image data that does not have boundary conditions, and investigate the estimation method of latent variables of images and image restoration on Bayesian inference. In addition, we calculate the Bayesian free energy or Bayesian posterior probability to evaluate the reliability of the estimation.

In the natural sciences, various image data have come to be obtained with the development of observation technology. Image processing is an important technique for analyzing such image data, as images are obtained to examine the information to be observed. In neuroscience, calcium imaging is used to observe neural activity. In calcium imaging, fluorescent molecules that binds to calcium ions is introduced into the cells, and the concentration of calcium ions is detected by the change in fluorescence intensity. In earth science, seismic tomography is used to investigate the state inside the earth. In seismic tomography, the inside of the earth is imaged by utilizing the fact that the velocity of seismic waves depends on the state inside the earth. Seismic tomography makes it possible to analyze the diffusion of fluids inside the Earth and estimate the distribution and dynamics of fluids in the crust[3, 4, 5].

In this chapter, we study the Markov random field (MRF) model, which is a stochastic generative model of images. The MRF model is applied to image processing such as image restoration and image segmentation[19, 18, 11, 12, 13, 21, 14, 15, 16, 17, 22, 23]. Nishimori used the Gaussian MRF model to the monochrome image to restore the image using Fourier transform. The results and image restoration results by the Wiener filter are equivalent[45]. MRF models are stochastic generative models of images based on Bayesian inference. Therefore, stochastic observation noise can be naturally introduced into MRF models. The MRF model describes the

properties of image data using two hyperparameters, which represent the smoothness of the image and the magnitude of the observation noise. The MRF model has two hyperparameters for images. Its hyperparameters represent the intensity of the observed noise and the smoothness of the image. It is necessary to estimate the hyperparameters of the MRF model in order to improve the performance of image processing[13, 21, 22].

MRF models are known to correspond to physical models It was pointed out that the hyperparameter that can be interpreted as the smoothness of the image corresponds to the diffusion coefficient[20]. Therefore, when interpreting the smoothness hyperparameter as the diffusion coefficient, it is necessary to estimate not only the value but also the reliability. It was shown that the hyperparameter distribution can be calculated analytically for the MRF model with periodic boundary conditions[20, 24]; hence, hyperparameter distribution can be estimated analytically from observed images without using numerical calculations. Since many natural images do not satisfy the periodic boundary conditions, it is necessary to investigate a method for estimating the hyperparameter distribution of the MRF model that does not have the periodic boundary conditions. In this chapter, we propose an analytical hyperparameter distribution estimation method and image restoration method for MRF models without periodic boundary conditions. We also confirm by numerical experiments that the proposed method works properly compared to the conventional method with periodic boundary conditions.

3.1 Formulation

We consider images of a d-dimensional square lattice with N pixels using the MRF model. Let us denote the original image as $\boldsymbol{u} = \{u_1, u_2, \cdots, u_N\} \in \mathbb{R}^N$ and the observed image as $\boldsymbol{v} = \{v_1, v_2, \cdots, v_N\}$, where N is the number of pixels.

Each pixel value v_i is given by

$$v_i = u_i + n_i, \qquad (i = 1, 2, \cdots, N), \tag{3.1}$$

where n_i represents the observation noise. We assume that $\boldsymbol{n} = (n_1, n_2, \cdots, n_N) \in \mathbb{R}^N$ are independent and identically distributed random variables that follow a normal distribution whose mean and variance are respectively 0 and b [denoted as $\mathcal{N}(0, b)$]. Then the observed image $\boldsymbol{v}$ follows a multivariate normal distribution:

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

where the function Z_1 is a partition function and defined by $Z_1(b) = (2\pi b)^{N/2}$. The variable b is

a hyperparameter representing the magnitude of the observation noise.

By considering the smoothness of the original image, we assume that the original image $\boldsymbol{u}$ follows $p(\boldsymbol{u}|a)$ defined by

$$p(\boldsymbol{u}|a) = \frac{1}{Z_{\text{pri}}(a)} \exp\left[-\frac{1}{2a}\sum_{\langle i,j\rangle}(u_i - u_j)^2\right], \tag{3.3}$$

where $\langle i, j\rangle$ represents two adjacent sites and the function Z_{pri} is a partition function defined by

$$Z_{\text{pri}} = \int d\boldsymbol{u} \exp\left[-\frac{1}{2a}\sum_{\langle i,j\rangle}(u_i - u_j)^2\right]. \tag{3.4}$$

In this distribution, the closer the values of adjacent pixels are, the greater the probability becomes. Variable a is also a hyperparameter. This hyperparameter represents the smoothness of images.

3.2 Image Restoration and Hyperparameter Estimation

3.2.1 Image restoration

Methods of analytical image restoration are already known for images with periodic boundary conditions.[24, 45] Here, for images without boundary conditions, we describe analytical image restoration.

According to Bayes' theorem, the posterior distribution of the original image is written by

$$p(\boldsymbol{u}|\boldsymbol{v}, \hat{a}, \hat{b}) = \frac{p(\boldsymbol{v}|\boldsymbol{u}, \hat{b})p(\boldsymbol{u}|\hat{a})}{p(\boldsymbol{v}|\hat{a}, \hat{b})}, \tag{3.5}$$

where $\hat{a}$ and $\hat{b}$ are the estimated values of the hyperparameters a and b. The denominator is a normalized constant and is generally intractable, but in this case we can calculate it. From Eqs. (4.2), (4.3), and (3.5),

$$p(\boldsymbol{u}|\boldsymbol{v}, \hat{a}, \hat{b}) \propto \exp(-E), \tag{3.6}$$

where the exponent is defined as an energy function:

$$E = \frac{1}{2\hat{b}}\sum_{i=1}^{N}(v_i - u_i)^2 + \frac{1}{2\hat{a}}\sum_{\langle i,j\rangle}(u_i - u_j)^2. \tag{3.7}$$

We will restore images analytically by maximum a posteriori (MAP) estimation. Since MAP estimation is equivalent to minimizing the energy function, we will describe below the method to minimize the energy function.

For simplicity, we consider the energy function E of the one-dimensional MRF model. The energy function of the original image $\boldsymbol{u}$ with respect to the observed image $\boldsymbol{v}$ is expressed as

$$E(\boldsymbol{u}) = \frac{1}{2\hat{b}} \sum_{i=1}^{N} (v_i - u_i)^2 + \frac{1}{2\hat{a}} \sum_{i=1}^{N-1} (u_{i+1} - u_i)^2. \tag{3.8}$$

With the $N \times N$ matrix Λ,

$$\Lambda = \begin{pmatrix} 1 & -1 & 0 & & & \\ -1 & 2 & -1 & & \ddots & \\ 0 & -1 & 2 & -1 & & \\ & & & \ddots & & 0 \\ & \ddots & & -1 & 2 & -1 \\ & & & 0 & -1 & 1 \end{pmatrix}, \tag{3.9}$$

the energy function E is represented as

$$E(\boldsymbol{u}) = \frac{1}{2\hat{b}} |\boldsymbol{v} - \boldsymbol{u}|^2 + \frac{1}{2\hat{a}} \boldsymbol{u}^T \Lambda \boldsymbol{u}, \tag{3.10}$$

To analytically obtain the minimum value of this energy function, we will diagonalize Λ with a unitary matrix K whose elements are

$$K_{ij} = \begin{cases} \frac{1}{\sqrt{N}} & (j = 1) \\ \sqrt{\frac{2}{N}} \cos\left\{\left(i - \frac{1}{2}\right) \frac{(j-1)}{N} \pi\right\} & (j \neq 1) \end{cases}. \tag{3.11}$$

Therefore, the energy function E is rewritten as follows:

$$E(\tilde{\boldsymbol{u}}) = \frac{1}{2\hat{b}} |\tilde{\boldsymbol{v}} - \tilde{\boldsymbol{u}}|^2 + \frac{1}{2\hat{a}} \tilde{\boldsymbol{u}}^T \tilde{\Lambda} \tilde{\boldsymbol{u}}, \tag{3.12}$$

where $\tilde{\boldsymbol{v}} = K\boldsymbol{v}$, $\tilde{\boldsymbol{u}} = K\boldsymbol{u}$, and $\tilde{\Lambda} = K^T \Lambda K$ is an $N \times N$ diagonal matrix that has eigenvalues

$$\lambda_i = 4 \sin^2 \left(\frac{\pi}{2N} (i - 1) \right). \tag{3.13}$$

Therefore,

$$E(\tilde{\boldsymbol{u}}) = \sum_{i=1}^{N} \left[\frac{1}{2\hat{b}} (\tilde{v}_i - \tilde{u}_i)^2 + \frac{1}{2\hat{a}} \lambda_i \tilde{u}_i^2 \right] \tag{3.14}$$

$$= \sum_{i=1}^{N} \left(\frac{1}{2\hat{b}} + \frac{\lambda_i}{2\hat{a}} \right) \left\{ \tilde{u}_i - \left(1 + \frac{\hat{b}}{\hat{a}} \lambda_i \right)^{-1} \tilde{v}_i \right\}^2 + \frac{1}{2} \left(\hat{b} + \frac{\hat{a}}{\lambda_i} \right)^{-1} \tilde{v}_i^2. \tag{3.15}$$

From Eq. (3.15), the original image can be analytically estimated as

$$\tilde{u}_i = \left(1 + \frac{\hat{b}}{\hat{a}}\lambda_i\right)^{-1} \tilde{v}_i. \tag{3.16}$$

In the case of periodic boundary conditions, the above calculation is the same except for eigenvalues since the matrix Λ is different from Eq. (3.9) in that $\Lambda_{11} = \Lambda_{NN} = 2$ and $\Lambda_{1N} = \Lambda_{N1} = -1$. It is described in Sect. 3.5 that the eigenvalues are equal to Eq. (3.13) for $N \to \infty$.

Next, we will extend the one-dimensional MRF model to two dimensions. It can be extended to d-dimensional by the following procedure. Let us denote the original image as $u = \{u_{i,j}\}$ $(i, j = 1, 2, \cdots, N)$ and the observed image as $v = \{v_{i,j}\}$. Then, the energy function E is written as

$$E(u) = \frac{1}{2\hat{b}} \sum_{i,j=1}^{N} (v_{i,j} - u_{i,j})^2 + \frac{1}{2\hat{a}} \sum_{i,j=1}^{N-1} \left\{ (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 \right\}. \tag{3.17}$$

By using a matrix Λ'

$$\Lambda' = I \otimes \Lambda + \Lambda \otimes I, \tag{3.18}$$

where I is an $N \times N$ identity matrix, Λ is the matrix given by Eq. (3.9), and $\cdot \otimes \cdot$ is the Kronecker tensor product[46], the energy function E is rewritten by

$$E(\boldsymbol{u}) = \frac{1}{2\hat{b}} |\boldsymbol{v} - \boldsymbol{u}|^2 + \frac{1}{2\hat{a}} \boldsymbol{u}^T \Lambda' \boldsymbol{u}, \tag{3.19}$$

where $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \cdots, u_{iN})$, $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_N)^T$, $\boldsymbol{v}_i = (v_{i1}, v_{i2}, \cdots, v_{iN})$, and $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N)^T$. We will diagonalize Λ' to analytically obtain the minimum value of this energy function. This is accomplished by applying a unitary transformation

$$U = K \otimes K, \tag{3.20}$$

to this system.

$$\begin{aligned} U^T \Lambda' U &= U^T (K \otimes \Lambda K + \Lambda K \otimes K) && (3.21)\\ &= \left(K^T K \otimes K^T \Lambda K + K^T \Lambda K \otimes K^T K\right) && (3.22)\\ &= \left(I \otimes \tilde{\Lambda} + \tilde{\Lambda} \otimes I\right). && (3.23) \end{aligned}$$

Therefore, the eigenvalue of $\tilde{\Lambda}' = U^T \Lambda' U$ for $\tilde{u}_{i,j}$ is

$$\begin{aligned} \lambda_{i,j} &= \lambda_i + \lambda_j && (3.24)\\ &= 4 \sin^2 \left(\frac{\pi}{2N}(i-1)\right) + 4 \sin^2 \left(\frac{\pi}{2N}(j-1)\right). && (3.25) \end{aligned}$$

From this, the posterior distribution of the original image is obtained as

$$p(\tilde{\boldsymbol{u}}|\tilde{\boldsymbol{v}},\hat{a},\hat{b}) \propto \exp\left[\sum_{i,j=1}^{N}\left(\frac{1}{2\hat{b}}+\frac{\lambda_{i,j}}{2\hat{a}}\right)\left(\tilde{u}_{i,j}-\left(1+\frac{\hat{b}}{\hat{a}}\lambda_{i,j}\right)^{-1}\tilde{v}_{i,j}\right)^2\right], \tag{3.26}$$

where $\tilde{\boldsymbol{v}} = U\boldsymbol{v}$, and $\tilde{\boldsymbol{u}} = U\boldsymbol{u}$. From Eq. (3.26), in two dimensions, the original image can be analytically estimated as

$$\tilde{u}_{i,j} = \left(1+\frac{\hat{b}}{\hat{a}}\lambda_{i,j}\right)^{-1}\tilde{v}_{i,j}. \tag{3.27}$$

3.2.2 Hyperparameter estimation

Since the image restoration depends on the hyperparameters, those hyperparameters need to be estimated correctly. It is known that a hyperparameter corresponds to the diffusion coefficient as a physical quantity.[20] Therefore, to evaluate the confidence of estimation, the distribution of hyperparameters needs to be estimated. The distribution estimation of a model with periodic boundary conditions has already been calculated.[20, 24] Here, we will perform analytically estimate the distribution of the model without boundary conditions.

The marginal likelihood is represented by

$$p(\boldsymbol{v}|a,b) = \int d\boldsymbol{u}\; p(\boldsymbol{v}|\boldsymbol{u},b)p(\boldsymbol{u}|a) \tag{3.28}$$

$$\propto \int d\boldsymbol{u}\; \exp\left[-\left(\frac{1}{2\hat{b}}\sum_{i=1}^{N}(v_i-u_i)^2+\frac{1}{2\hat{a}}\sum_{\langle i,j\rangle}(u_i-u_j)^2\right)\right]. \tag{3.29}$$

This can be computed analytically by transformation with a unitary matrix and using Gaussian integration.

In the one-dimensional MRF model, by using the unitary matrix K, the marginal likelihood can be computed as

$$p(\tilde{\boldsymbol{v}}|a,b) \propto \int d\boldsymbol{u}\exp\left[-\sum_{i=1}^{N}\left(\frac{1}{2b}+\frac{\lambda_i}{2a}\right)\left\{\tilde{u}_i-\left(1+\frac{b}{a}\lambda_i\right)^{-1}\tilde{v}_i\right\}^2+\frac{1}{2}\left(\hat{b}+\frac{\hat{a}}{\lambda_i}\right)^{-1}\tilde{v}_i^2\right] \tag{3.30}$$

$$\propto \prod_i\left(b+\frac{a}{\lambda_i}\right)^{-1/2}\exp\left(-\frac{1}{2}\left(b+\frac{a}{\lambda_i}\right)^{-1}\tilde{v}_i^2\right). \tag{3.31}$$

Similarly, in the case of the two-dimensional MRF model, by using the unitary matrix U,

$$p(\tilde{\boldsymbol{v}}|a,b) \propto \int d\boldsymbol{u} \exp\left[-\sum_{i,j=1}^{N}\left(\frac{1}{2b}+\frac{\lambda_{i,j}}{2a}\right)\left\{\tilde{u}_{i,j}-\left(1+\frac{b}{a}\lambda_{i,j}\right)^{-1}\tilde{v}_{i,j}\right\}^2 + \frac{1}{2}\left(\hat{b}+\frac{\hat{a}}{\lambda_{i,j}}\right)^{-1}\tilde{v}_{i,j}^2\right] \tag{3.32}$$

$$\propto \prod_{i,j}\left(b+\frac{a}{\lambda_{i,j}}\right)^{-1/2}\exp\left(-\frac{1}{2}\left(b+\frac{a}{\lambda_{i,j}}\right)^{-1}\tilde{v}_{i,j}^2\right). \tag{3.33}$$

3.2.3 Expectation value of free energy

The marginal likelihood depends on probabilistically generated observation images. To conduct a theoretical evaluation, we will use an expectation value of free energy based on a statistical mechanics analogy.

Let us define the free energy F of the one-dimensional MRF model as follows:

$$F(a,b;\boldsymbol{v}) \equiv -\ln p(\boldsymbol{v}|a,b) \tag{3.34}$$

$$= -\frac{1}{2}\sum_{k}\left\{\ln\frac{1}{\frac{a}{\lambda_k}+b}-\frac{|\tilde{v}_k|^2}{\frac{a}{\lambda_k}+b}\right\}, \tag{3.35}$$

where λ_k is the eigenvalue of the coefficient matrix in the generative model. Note that the equals 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:

$$\langle F(a,b;\boldsymbol{v})\rangle_{\boldsymbol{v}|a_0,b_0} = \int d\boldsymbol{u}d\boldsymbol{v}F(a,b;\boldsymbol{v})p(\boldsymbol{v}|\boldsymbol{u},b_0)p(\boldsymbol{u}|a_0) \tag{3.36}$$

$$\equiv \left\langle\langle F(a,b;\boldsymbol{v})\rangle_{\boldsymbol{v}|\boldsymbol{u},b_0}\right\rangle_{\boldsymbol{u}|a_0}, \tag{3.37}$$

where $\langle\cdots\rangle_{\boldsymbol{v}|a_0,b_0}$ and $\langle\cdots\rangle_{\boldsymbol{u}|a_0}$ are the expectation values by probability distribution $p(\boldsymbol{v}|b_0)$ and $p(\boldsymbol{u}|a_0)$. a_0 and b_0 are true hyperparameters. The subscripts are omitted below. From Eq. (3.35), this is rewritten as

$$\langle F(a,b;\boldsymbol{v})\rangle = -\frac{1}{2}\sum_{k}\left\{\ln\frac{1}{\frac{a}{\lambda_k}+b}-\frac{\left\langle\left\langle|\tilde{v}_k|^2\right\rangle\right\rangle}{\frac{a}{\lambda_k}+b}\right\}. \tag{3.38}$$

Therefore, to obtain the expectation value of free energy, $\left\langle\left\langle|\tilde{v}_k|^2\right\rangle\right\rangle$ needs to be calculated. When the unitary transformations are applied to the model, the original image and the observed image are written as $\bar{u}_k \sim \mathcal{N}(0, \frac{a_0}{\bar{\lambda}_k})$ and $\bar{v}_k \sim \mathcal{N}(\bar{u}_k, b_0)$, respectively. $\bar{\lambda}_i$ is the eigenvalue of the

coefficient matrix in the estimation model. We will denote the unitary transformations that diagonalize the coefficient matrix of the generative model and estimation model as K_1 and K_2, respectively. Then, the observed image is represented as $\boldsymbol{v} = K_1^{-1}\bar{\boldsymbol{v}}$ and the observed image after unitary transformation of estimation model is represented as $\tilde{\boldsymbol{v}} = K_2\boldsymbol{v}$. Therefore,

$$\left\langle\left\langle|\tilde{v}_k^2|\right\rangle\right\rangle = \left\langle\left\langle\sum_{i,j} A_{ki}^* A_{kj}\bar{v}_i\bar{v}_j\right\rangle\right\rangle \tag{3.39}$$

$$= \sum_i |A_{ki}|^2 \left\langle\left\langle|\bar{v}_i^2|\right\rangle\right\rangle \tag{3.40}$$

$$= \sum_i |A_{ki}|^2 \left(\frac{a_0}{\bar{\lambda}_i} + b_0\right), \tag{3.41}$$

where $A = K_2K_1^{-1}$. Therefore, from Eqs. (3.38) and (3.41), the expectation of free energy can be obtained as

$$\langle F(a,b;\boldsymbol{v})\rangle = -\frac{1}{2}\sum_k \left\{\ln\frac{1}{\frac{a}{\lambda_k}+b} - \frac{1}{\frac{a}{\lambda_k}+b}\sum_i |A_{ki}|^2\left(\frac{a_0}{\bar{\lambda}_i}+b_0\right)\right\}. \tag{3.42}$$

3.2.4 Noiseless hyperparameter

Here, we derive hyperparameter from Eq. (4.3) in the case of the observed image without noise. For simplicity, we consider the one-dimensional MRF model. The extremization condition with respect to a is

$$\sum_i (u_i^* - u_{i+1}^*)^2 = \int d\boldsymbol{u}\sum_i (u_i - u_{i+1})^2 P(\boldsymbol{u}|\hat{a}), \tag{3.43}$$

where $\hat{a}$ is the estimated values of the hyperparameter a, and $\boldsymbol{u}^*$ is the observed original image. Gaussian integration of Eq. (3.43) yields

$$\hat{a} = \frac{1}{N}\sum_i (u_i^* - u_{i+1}^*)^2. \tag{3.44}$$

Regardless of the boundary conditions, this result is the same except for the range of sum. Without boundary conditions, $i = 1, \cdots, N-1$. On the other hand, with periodic boundary conditions, $u_1^* = u_{N+1}^*$ and $i = 1, \cdots, N$. Thus, in the case of the observed image without noise, boundary conditions affect the hyperparameter a.

3.2.5 Boundary conditions for a large system

Here, we indicate that the free energy of Eq. (3.42) for a large system becomes the same irrespective of boundary conditions. The eigenvalues with periodic boundary conditions can be

shown in ascending order as

$$\lambda_i = 4\sin^2\left(\frac{\pi}{N}\left\lceil\frac{i-1}{2}\right\rceil\right), \tag{3.45}$$

where $\lceil\cdots\rceil$ is the ceil function. For infinite limits of N,

$$\lambda_i = 4\sin^2\left(\frac{\pi}{N}\left\lceil\frac{i-1}{2}\right\rceil\right) \xrightarrow[N\to\infty]{} 4\sin^2\left(\frac{\pi}{2N}(i-1)\right). \tag{3.46}$$

Thus, eigenvalues with periodic boundary conditions is equal to those without boundary conditions for $N \to \infty$. Considering that the matrix A is an unitary matrix, this means that, regardless of the boundary conditions, the free energy is the same for a large system.

3.3 Numerical Experiment

An image is generated from a generative model with and without periodic boundary conditions formulated by Eqs. (4.2) and (4.3). For each image, hyperparameter estimation is performed by minimizing the free energy using the proposed method (which is an estimation model without boundary conditions) and the existing method (which is an estimation model with periodic boundary conditions). Furthermore, we confirm that the image is actually restored by using the estimated hyperparameter. We also compare the reliability of each method from the expectation value of free energy.

3.3.1 One-dimensional model

We consider the one-dimensional model where the number of pixels is $N = 256$. The true hyperparameter values were set to $(a, b) = (1.5, 1.5)$. Figure 3.1 (a) shows an example of an observed image generated by a generative model without boundary conditions. Figures 3.1 (b), and 3.1 (c) show the hyperparameter distributions for the image estimated by the proposed method and the existing method, respectively. The intersection of the lines indicates the value of the true hyperparameter. From Figs. 3.1 (b) and 3.1 (c), compared to the existing method, the proposed method shows that the distribution appears near the true hyperparameter for the image without boundary conditions.

Figure 3.1 (d) shows an example of an observed image generated by a generative model with periodic boundary conditions. Figures 3.1 (e) and 3.1 (f) show the hyperparameter distributions for the image estimated by the proposed method and the existing method, respectively. The intersection of the lines indicates the value of the true hyperparameter. From Figs. 3.1 (e) and 3.1 (f), for images with periodic boundary conditions, we can see that the hyperparameter

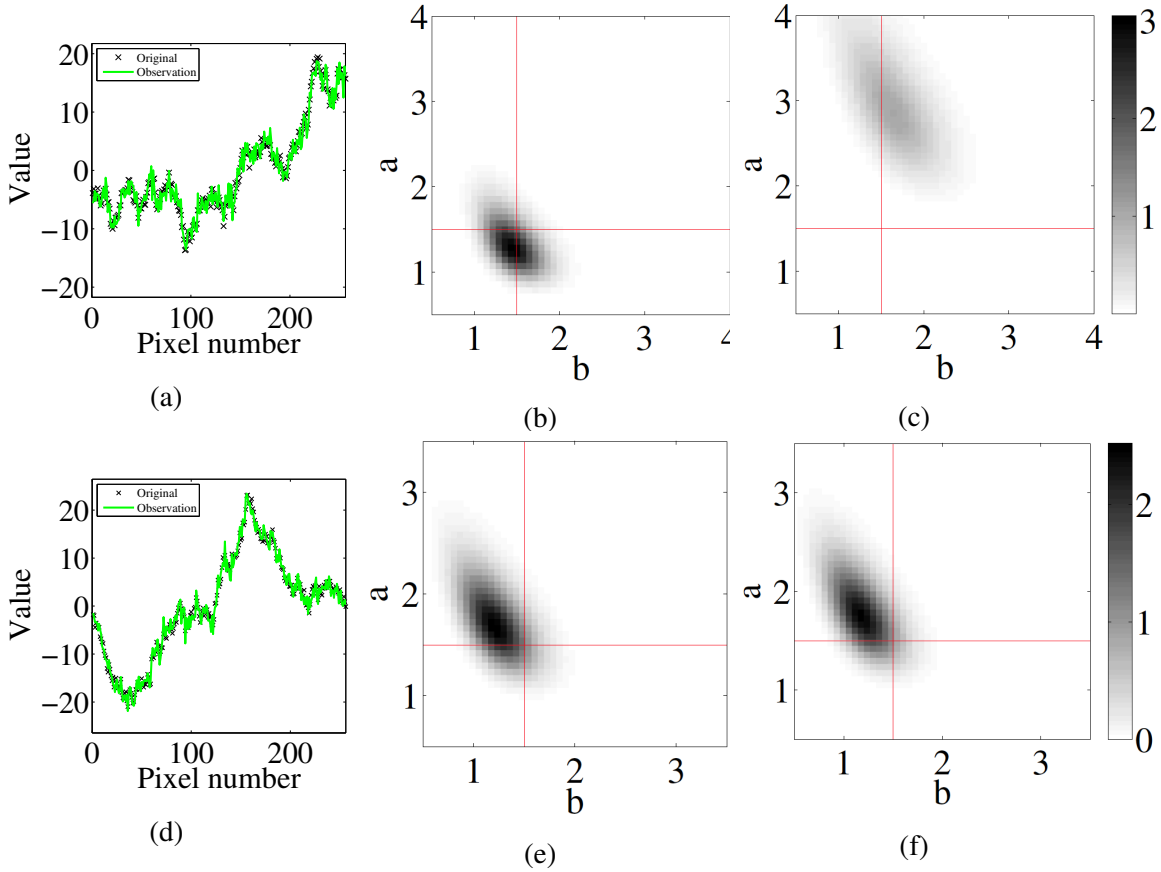


Fig. 3.1: Hyperparameter estimation ($N = 256$, $a = 1.5$, $b = 1.5$). Panels (a)–(c) generative model without boundary conditions. Panel (a) shows original and observed images. Panels (b) and (c) are hyperparameter distributions estimated by proposed method and existing method, respectively. Panels (d)–(f) generative model with periodic boundary conditions. Panel (d) shows original and observed images. Panels (e) and (f) are hyperparameter distributions estimated by proposed method and existing method, respectively.[47]

distributions of both the proposed method and the existing method are distributed near the true hyperparameters. From the above, it can be seen that the existing method can be used only when the image data is sufficiently periodic.

In addition, image restoration was performed using estimated hyperparameters by minimizing the free energy. Figures 3.2 (a) and 3.2 (b) show images restored by the proposed method and the existing method, respectively, in the case of the generative model without boundary conditions. Figure 3.2 (a) shows that the image can be restored. However, as shown in Fig. 3.2 (b) in the

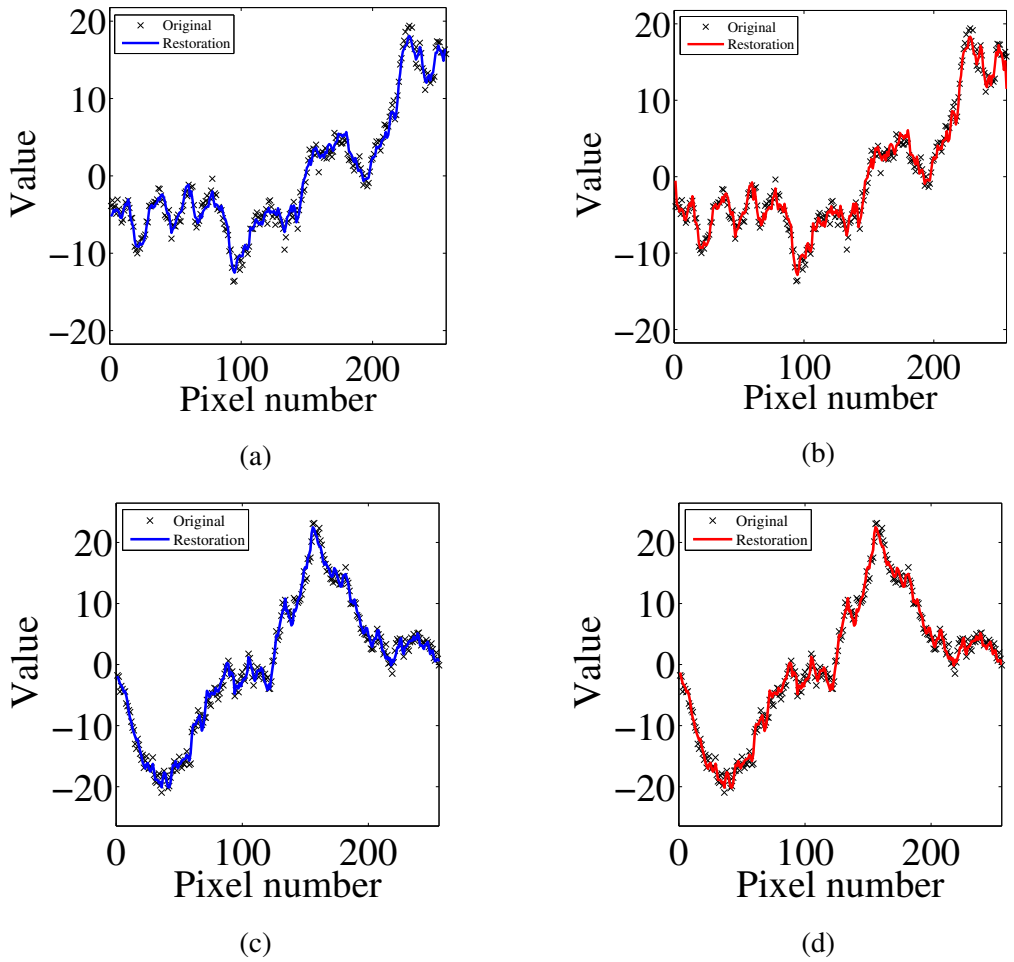


Fig. 3.2: Image restoration by estimated hyperparameters. In upper column, results for original images generated by generative model without boundary conditions. In lower column, results for original images generated by generative model with periodic boundary conditions. Panels (a) and (c) show restored images by proposed method. Panels (b) and (d) show restored images by existing method. [47]

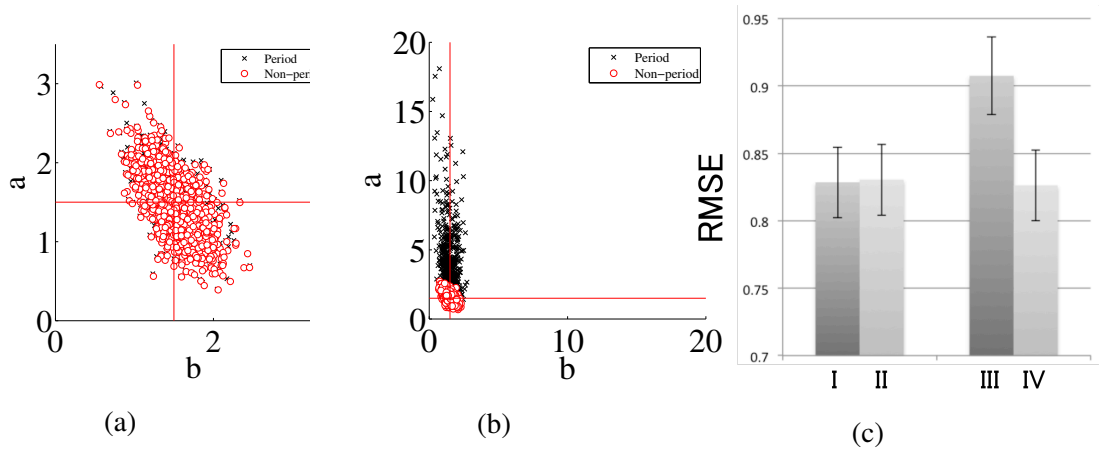


Fig. 3.3: Scatter diagram of hyperparameter estimate for 1000 images. Panels (a) and (b) are obtained by generative model with and without periodic boundary conditions, respectively. Panel (c) is RMSE between true images and restored images. Left and right bars show RMSE for original images with and without periodic boundary conditions. Bars I and III show RMSE for results restored by existing method. Bars II and IV show RMSE for results restored by proposed method. [47]

vicinity of the boundary, the restored image is separated from the original image by the effect of the assumption of the periodic boundary conditions.

Figures 3.2 (c) and 3.2 (d) show images restored by the proposed method and the existing method in the case of the generative model with periodic boundary conditions. As can be seen from Figs. 3.2 (c) and 3.2 (d), for images with periodic boundary conditions, each method can perform image restoration. From the results of hyperparameter estimation and image restoration, it was suggested that the proposed method is applicable regardless of the boundary conditions.

Furthermore, with $N = 256$, $a = 1.5$, and $b = 1.5$, 1000 images with no boundary conditions and 1000 images with periodic boundary conditions were generated. Hyperparameter estimation for those images was performed using each method, and the distributions of the estimated values was examined.

Figures 3.3 (a) and 3.3 (b) are scatter plots of hyperparameter estimation values by minimizing the free energies for images with and without periodic boundary conditions, respectively. The intersection of the lines indicates the value of the true hyperparameter. Figure 3.3 (c) shows the root mean square error (RMSE) between the true images and restored images. The left and right bars show RMSE for the original images with and without periodic boundary conditions. The bars I and III show RMSE for the results restored by the existing method, and the bars II and IV show RMSE for the results restored by the proposed method.

As can be seen from Fig. 3.3 (a), for images with periodic boundary conditions, hyperparameters are estimated in the vicinity of true hyperparameter values for each method. On the other hand, as shown in Fig. 3.3 (b), while the proposed method can estimate true hyperparameter values for images without boundary conditions, the existing method cannot estimate the true hyperparameter values. From comparing the bars I and II in Fig. 3.3 (c), we can see that RMSEs are about the same when the proposed and existing methods were applied to the images with periodic boundary conditions. On the other hand, the RMSE of the existing method for the image without boundary conditions was 10% worse than the RMSE of the proposed method for the image with boundary conditions. From this fact, the existing method degrades the performance for image restoration without boundary conditions, while the proposed method does not degrade the performance regardless of the boundary conditions.

Figures 3.4 (a) and 3.4 (b) are expectation values of free energy of the proposed method and the existing method, respectively, for the generative model without boundary conditions. It can be seen that for the generative models without boundary conditions, the existing methods cannot accurately estimate hyperparameters. This corresponds to the scatter diagram of the hyperparameter estimation of Fig. 3.3 (b). Figures 3.4 (c) and 3.4 (d) are expectation values of free

energy of the proposed method and the existing method, respectively, for the generative model with periodic boundary conditions. For generative models with periodic boundary conditions, hyperparameters can be estimated correctly for each method. This corresponds to the scatter diagram showing minimizing the free energy in Fig. 3.3 (a).

3.3.2 Two-dimensional model

We consider the two-dimensional model where the number of pixels is $N = 32 \times 32$. The true hyperparameter values were set to $(a, b) = (1.5, 1.5)$. Figure 3.5 (a) shows an example of an observed image generated by a generative model without boundary conditions. Figures 3.5 (b) and 3.5 (c) show the hyperparameter distributions for the image estimated by the proposed method and the existing method. The intersection of the lines indicates the value of the true hyperparameter. From Figs. 3.5 (b) and 3.5 (c), compared to the existing method, the proposed method shows that the hyperparameter estimation distribution appears near the true hyperparameter for the image without the boundary condition.

Figure 3.5 (d) shows an example of an observed image generated by a generative model with periodic boundary condition. Figures 3.5 (e) and 3.5 (f) show the distributions of the hyperparameters estimated by the proposed method and the existing method, respectively. The intersection of the lines indicates the value of the true hyperparameter. From Figs. 3.5 (e) and 3.5 (f), for images with periodic boundary conditions, we can see that the distributions of both the proposed method and the existing method are distributed near the true hyperparameters. From the above, it can be seen that the existing method can be used only when the image data is sufficiently periodic.

In addition, image restoration was performed using the estimated hyperparameters. Figures 3.6 (a) and 3.6 (b) are the original image and the observed image generated by a generative model without boundary conditions. Figures 3.6 (c) and 3.6 (d) show images restored by the proposed method and the existing method in the case of the generative model without boundary conditions. Figure 3.6 (c) shows that the image can be restored. However, as shown in Fig. 3.6 (d) in the vicinity of the boundary, the restored image is separated from the original image by the effect of the assumption of the periodic boundary condition.

Figures 3.6 (e) and 3.6 (f) are the original image and the observed image, respectively, generated by a generative model with periodic boundary conditions. Figures 3.6 (g) and 3.6 (h) show images restored by the proposed method and the existing method,respectively, in the case of the generative model with periodic boundary conditions. As can be seen from Figs. 3.6 (g) and 3.6 (h), for images with periodic boundary condition, each method can perform image restoration.

From the results of hyperparameter estimation and image restoration, it was suggested that the proposed method is applicable regardless of the boundary condition.

Furthermore, with $N = 32 \times 32$, $a = 1.5$, and $b = 1.5$, 1000 images with no boundary conditions and 1000 images with periodic boundary conditions were generated. Hyperparameter estimation for those images was performed using each method, and the distributions of the values was examined.

Figures 3.7 (a) and 3.7 (b) are scatter plots of hyperparameter estimation values by minimizing the free energy for images with and without periodic boundary conditions, respectively. The intersection of the lines indicates the value of the true hyperparameter. Figure 3.7 (c) shows the root mean square error (RMSE) between the true images and restored images. The left and right bars show RMSE for the original images with and without periodic boundary conditions. The bars I and III show RMSE for the results restored by the existing method, and the bars II and IV show RMSE for the results restored by the proposed method.

As can be seen from Fig. 3.7 (a), for images with periodic boundary conditions, hyperparameters are estimated in the vicinity of true hyperparameter values for each method. On the other hand, as shown in Fig. 3.7 (b), while the proposed method can estimate true hyperparameter values for images without boundary conditions, the existing method cannot estimate true hyperparameter values. From comparing the bars I and II in Fig. 3.7 (c), we can see that RMSEs are about the same when the proposed and existing methods were applied to the images with periodic boundary conditions. On the other hand, the RMSE of the existing method for the image without boundary conditions was 4% worse than the RMSE of the proposed method for the image with boundary conditions. From this fact, it can be seen that the image restoration by the existing method degrades the restore performance for images without boundary conditions, while the proposed method does not degrade the restore performance regardless of the boundary conditions.

Figures 3.8 (a) and 3.8 (b) are expectation values of free energy of the proposed method and the existing method, respectively, for the generative model without boundary conditions. It can be seen that for the generative models without boundary conditions, the existing methods cannot accurately estimate hyperparameters. This corresponds to the scatter diagram of the hyperparameter estimation of Fig. 3.8 (b). Figures 3.8 (c) and 3.8 (d) are expectation values of free energy of the proposed method and the existing method, respectively, for the generative model with the periodic boundary condition. For generative models with periodic boundary conditions, hyperparameters can be estimated correctly for each method. This corresponds to the scatter diagram showing the hyperparameter estimation in Fig. 3.7 (a).

3.4 Discussion

In Sec. 3.3, it was shown that the proposed method can analytically estimate hyperparameters with the reliability from one-dimensional and two-dimensional image data that do not have periodic boundary conditions. The proposed method is effective for hyperparameter estimation from small image data. Large-sized image data can be analyzed by the conventional method with periodic boundary conditions as described in Sec. 3.2.5. In the proposed method and the conventional method, the hyperparameters are constant in the observation space. Therefore, both the proposed method and the conventional method work for large-sized images that have uniform hyperparameters. However, in many actual observation images such as seismic tomography and calcium imaging, hyperparameters are not uniform and depend on the observation space. When the hyperparameters of an image are not globally uniform but locally uniform, it is necessary to divide the image into small areas and analyze small images. Since the divided image data does not always have periodicity, the proposed method works effectively and accurately on the image data. Moreover, since the distribution estimation can be interpreted as the reliability of the estimation, it can be expected that the estimated distribution of the proposed method can be treated as an index of whether the division of the image data is appropriate.

3.5 Conclusion

In this study, we proposed a method to estimate latent variables from image data without boundary conditions and analytically restore the image. The proposed method, which is an estimation model without boundary conditions, is applicable to wider kinds of images than the existing method. The numerical experiments and analytical solutions showed that the hyperparameter estimation by the existing method can be used only for images with periodic boundary conditions. For images without boundary conditions, the estimated hyperparameters are different from the true values. On the other hand, the proposed method, which is an estimation model independent of boundary conditions, showed that hyperparameters can be estimated regardless of the boundary conditions.

A future topic of discussion is the estimation accuracy of hyperparameters of each method depending on the size of image data. It is generally known that the effect of boundary conditions decreases as the size of data increases. For sufficiently large image data, it is expected that there will be no difference in the hyperparameter estimation by the proposed method and the existing method. It is important to investigate the dependency of hyperparameter estimation accuracy of

each method on the scale of image data.

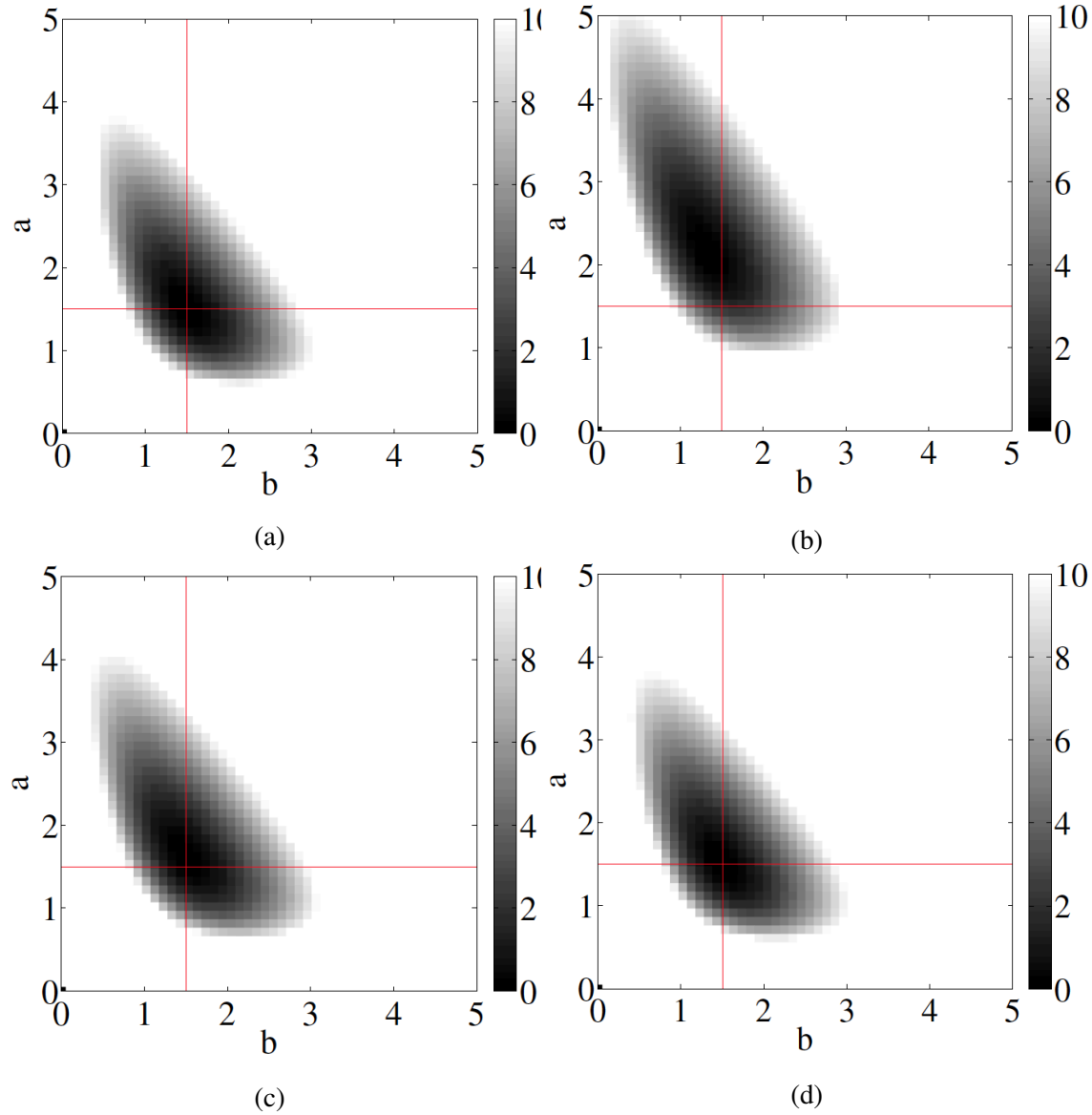


Fig. 3.4: Expectation value of free energy ($N = 256$, $a = 1.5$, $b = 1.5$). In upper column, results for original images generated by generative model without boundary conditions. In lower column, results for original images generated by generative model with periodic boundary conditions. Panels (a) and (c) show expectation values of free energy by proposed method. Panels (b) and (d) show expectation values of free energy by existing method. [47]

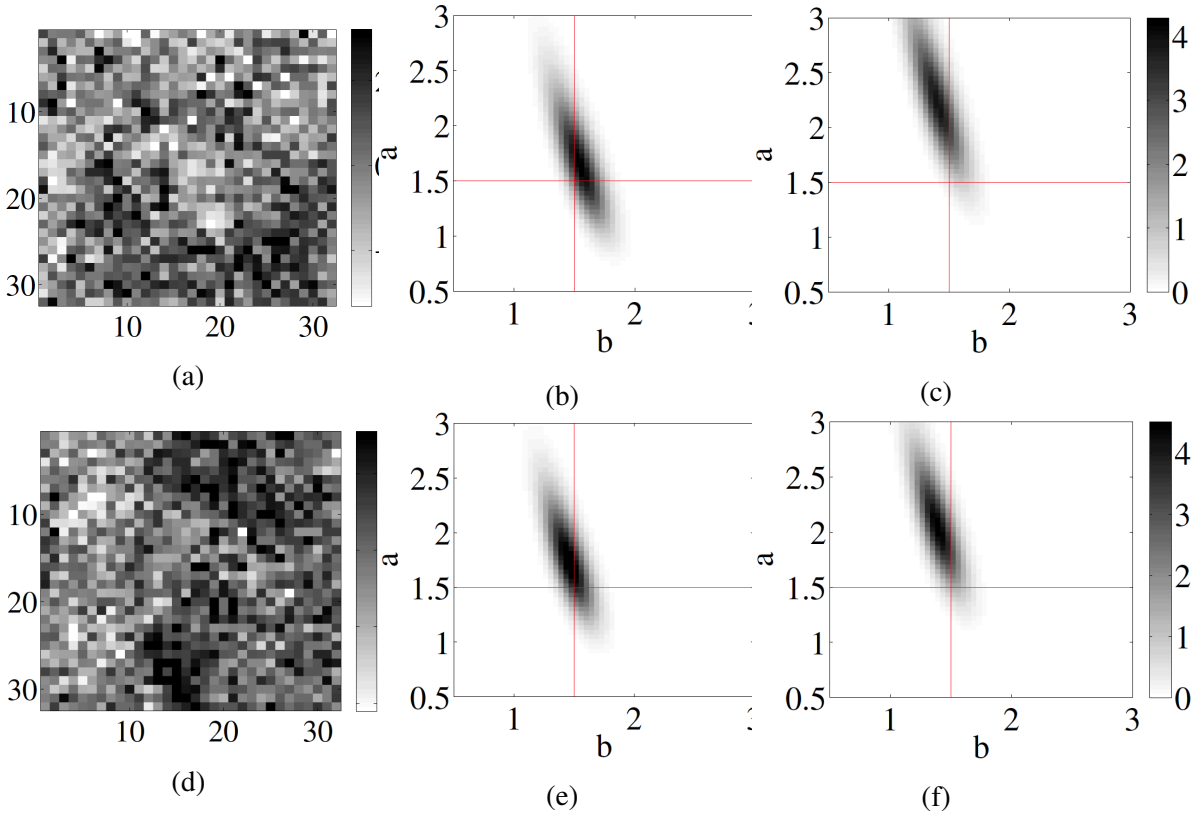


Fig. 3.5: Hyperparameter estimation ($N = 32 \times 32$, $a = 1.5$, $b = 1.5$). Panels (a)–(c) generative model without boundary conditions. Panel (a) shows observed images. Panels (b) and (c) are hyperparameter distributions estimated by proposed method and existing method, respectively. Panels (d)–(f) generative model with periodic boundary conditions. Panel (d) shows observed images. Panels (e) and (f) are hyperparameter distributions estimated by proposed method and existing method, respectively. [47]

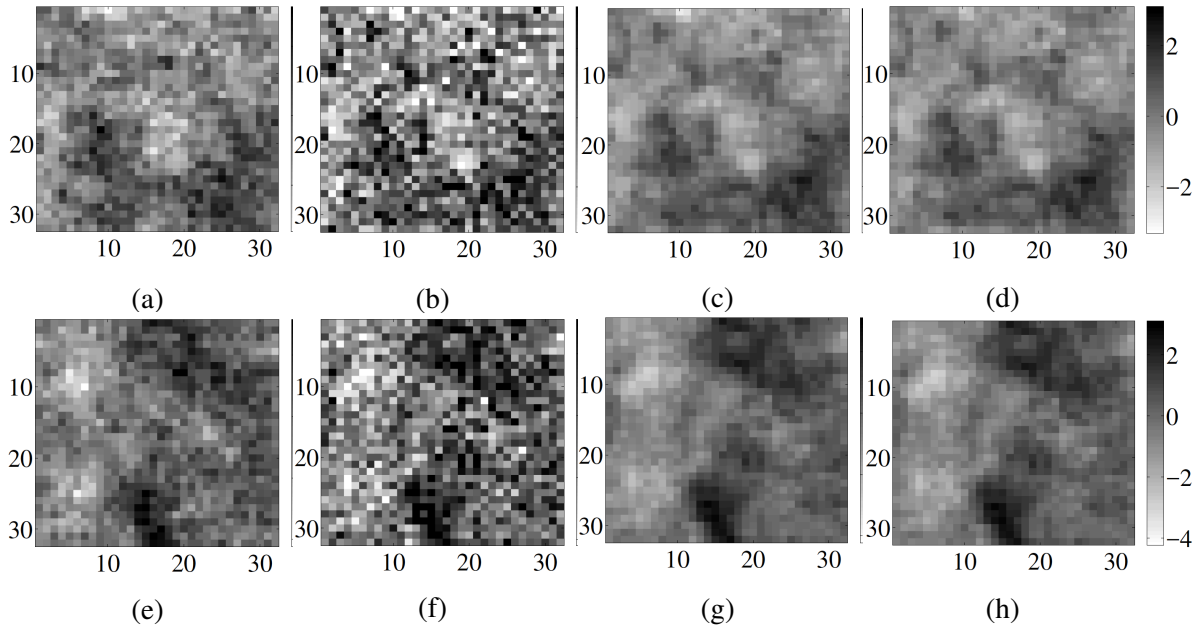


Fig. 3.6: Image restoration by estimated hyperparameters. In upper column, results for original images generated by generative model without boundary conditions. In lower column, results for original images generated by generative model with periodic boundary conditions. Panels (a) and (e) show original images. Panel (b) and (f) show observed images. Panels (c) and (g) show restored images by proposed method. Panels (d) and (h) show restored images by existing method. [47]

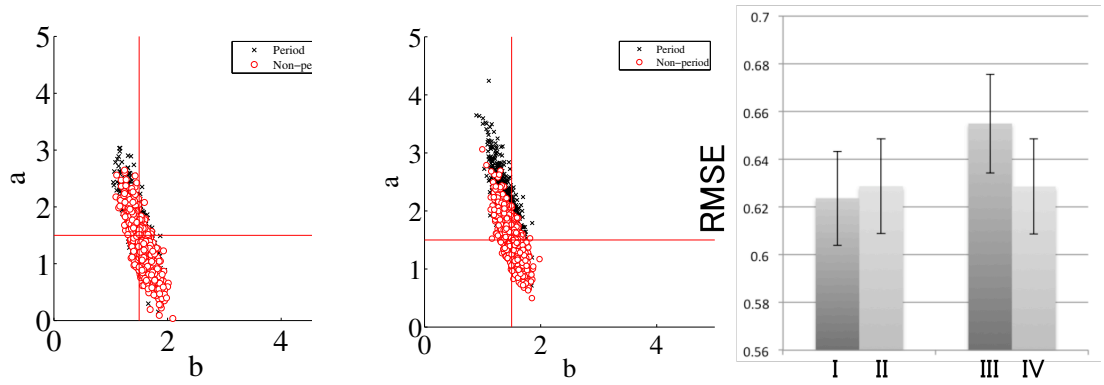


Fig. 3.7: Scatter diagram of hyperparameter estimate for 1000 images. Panels (a) and (b) are obtained by generative model with and without periodic boundary conditions, respectively. Panel (c) is RMSE between true images and restored images. Left and right bars show RMSE for original images with and without periodic boundary conditions. Bars I and III show RMSE for results restored by existing method. Bars II and IV show RMSE for results restored by proposed method. [47]

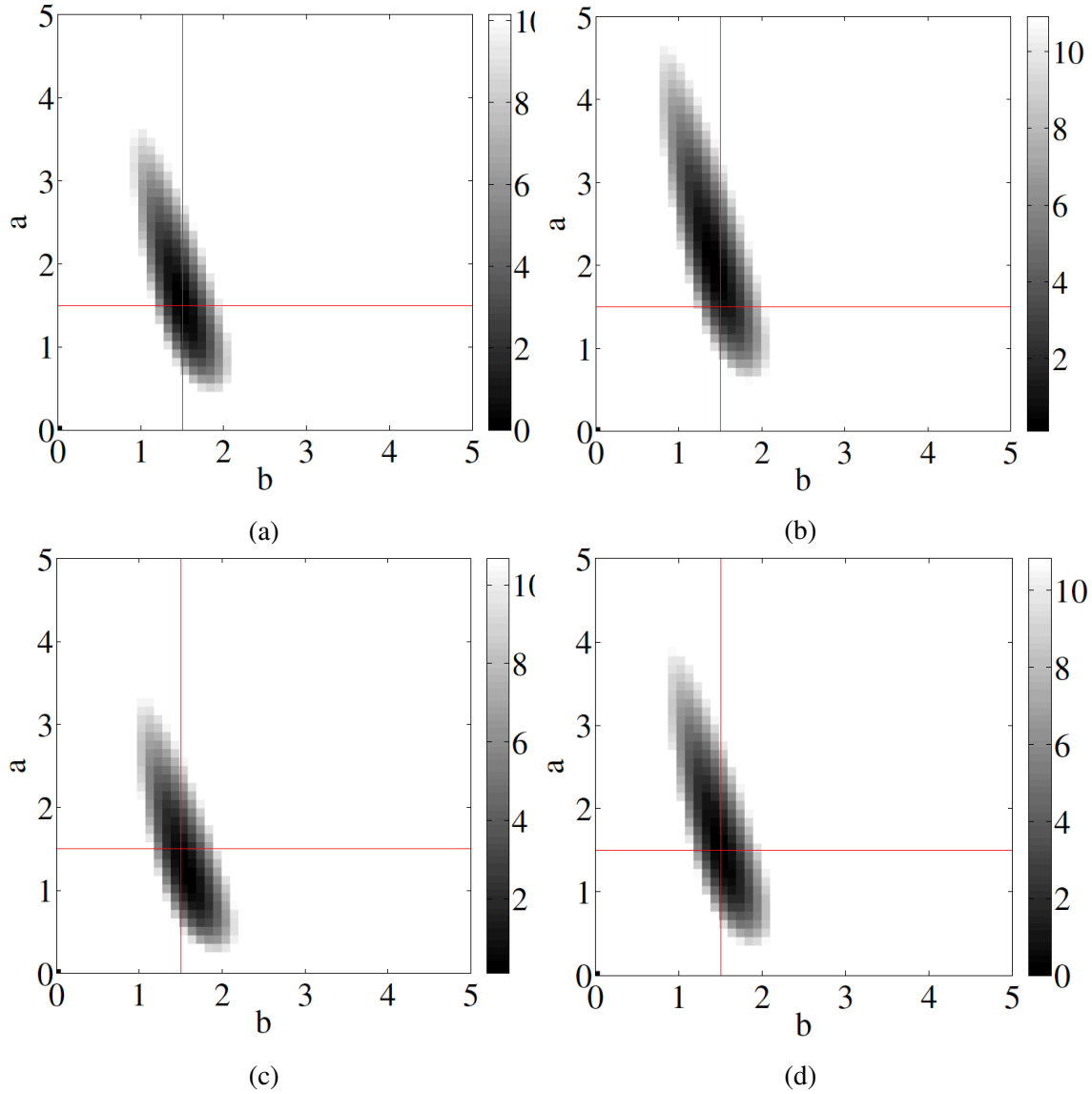


Fig. 3.8: Expectation value of free energy ($N = 32 \times 32$, $a = 1.5$, $b = 1.5$). In upper column, results for original images generated by generative model without boundary conditions. In lower column, results for original images generated by generative model with periodic boundary conditions. Panels (a) and (c) show expectation values of free energy by proposed method. Panels (b) and (d) show expectation values of free energy by existing method. [47]

Chapter 4

Bayesian Hyperparameter Estimation using Gaussian Process and Bayesian Optimization

With the development of observation technology, it has become possible to obtain information on observation targets from image data acquired in the natural sciences. The MRF model[19, 48, 49] used for image processing, which is also dealt with in Chapter 3, is often used for extracting physical information from images[50, 51, 4]. The MRF model corresponds to the lattice gas model[52] and has been analyzed and developed by statistical mechanics[53].

The MRF model is modeled on the hypothesis that adjacent pixels in images generally have similar values[19, 18, 22, 25], and a hyperparameters representing the smoothness of an image are introduced into the model. The MRF model corresponds to physical models such as the diffusion equation and the lattice gas model, and its smoothness hyperparameters correspond to the diffusion coefficient[20] and the coupling constant[52]. By estimating the hyperparameter of the MRF model from images, the physical features of the observation target can be inferred.

Many frameworks for estimating hyperparameters have been proposed for applications such as image restoration and domain segmentation. In information science, Bayesian optimization is known as an efficient method for optimizing model parameters[54]. In information statistical mechanics, the mean field approximation, the Bethe approximation, and the variational Bayes method are known as the approximate analysis methods to obtain hyperparameters[53]. These methods are point estimates, that is, they only estimate the valuesof the model's hyperparameters. The hyperparameters correspond to physical quantities such as diffusion coefficient and

coupling constant[20, 52]. In the estimation of physical quantities, not only the optimum value of the parameter but also the reliability of the estimation must be discussed. The distribution of hyperparameters can be interpreted as the error bar for parameters in the physical system, and the reliability of parameter estimation can be evaluated from the distribution. Therefore, the distribution of hyperparameters must be determined, but it is necessary to calculate the Bayesian posterior probability, which is also handled in Chapter 2 and Chapter 3, and the Bayesian posterior probability is generally difficult to calculate analytically. The hyperparameter distribution is often calculated numerically using the exchanged Markov chain Monte Carlo (REMC) method [43], but the method has a high calculation cost. To address these problems, we focus on algorithms to efficiently estimate hyperparameter distribution from a small number of samples.

We propose methods of applying Bayesian optimization and Gaussian process to estimate hyperparameter valuesand hyperparameter distributions. We estimate the hyperparameter distribution from a small number of samples using the Gaussian process. We estimate the hyperparameter values from a small number of samples using the Bayesian optimization. We also apply the Bayesian optimization as a sampler for the Gaussian process and estimate the distribution from a small number of samples. In numerical experiments, we use a Gaussian Markov random field (G–MRF) model[49, 11], which can analytically calculate a hyperparameter distribution[20, 24], and evaluate the efficiency of our method to estimate the values of hyperparameters and the hyperparameter distribution by comparing the true and estimated hyperparameter distributions.

In this chapter, our two main results are as follows. The first is that the distribution can be estimated with high accuracy from a small number of samples by using the Gaussian process. Bayesian optimization has the potential to be applied as a sampler for Gaussian processes to estimate the distribution of hyperparameters with high accuracy and low standard error from a smaller number of samples. The proposed method can estimate Bayesian posterior probability of the MRF model efficiently, and can be applied to not only the G–MRF model discussed here but also the general MRF model and other Bayesian frameworks.

4.1 Methods

This section explains the theories of the G–MRF model, which is used to test the effectiveness of the proposed methods; the Gaussian process[55, 56, 9], which is a technique to estimate the outputs of the objective function for unknown inputs; and Bayesian optimization[57, 58], which is an algorithm for searching for the optimal values of the objective function. In this work, the objective function is proportional to the Bayesian posterior probability distribution. This section

also explains the method of estimating the objective function from a small number of samples using Bayesian optimization as a sampler for the Gaussian process.

4.1.1 Bayesian estimation using G–MRF model

Here, we explain the theory of the 1-dimension G–MRF model used to test the proposed methods. We consider image data of a 1-dimensional square lattice with N pixels. Let us denote an original image as $\boldsymbol{u} = \{u_1, u_2, \cdots, u_N\} \in \mathbb{R}^N$ and an observed image as $\boldsymbol{v} = \{v_1, v_2, \cdots, v_N\} \in \mathbb{R}^N$. Each pixel value v_i is given by

$$v_i = u_i + n_i, \qquad (i = 1, 2, \cdots, N), \tag{4.1}$$

where n_i represents the observation noise. We assume that $\boldsymbol{n} = (n_1, n_2, \cdots, n_N) \in \mathbb{R}^N$ are independent and random variables, respectively, that follow a normal distribution whose mean and variance are 0 and b, respectively. Then, the observed image $\boldsymbol{v}$ follows a multivariate normal distribution:

$$p(\boldsymbol{v}|\boldsymbol{u}, b) = \frac{1}{Z_1(b)} \exp\left[-\frac{1}{2b}\sum_{i=1}^{N}(v_i - u_i)^2\right], \tag{4.2}$$

where the function Z_1 is a partition function and defined by $Z_1(b) = (2\pi b)^{N/2}$. The variable b is a hyperparameter representing the magnitude of the observation noise.

Considering the smoothness of the original image, we assume that the original image $\boldsymbol{u}$ follows $p(\boldsymbol{u}|a)$ defined by

$$p(\boldsymbol{u}|a) = \frac{1}{Z_{\text{pri}}(a)} \exp\left[-\frac{1}{2a}\sum_{\langle i,j\rangle}(u_i - u_j)^2\right], \tag{4.3}$$

where $\langle i, j\rangle$ represents two adjacent sites, and the function Z_{pri} is a partition function defined by

$$Z_{\text{pri}} = \int \mathrm{d}\boldsymbol{u} \exp\left[-\frac{1}{2a}\sum_{\langle i,j\rangle}(u_i - u_j)^2\right]. \tag{4.4}$$

In this distribution, the closer the values of adjacent pixels, the greater the probability of the original image. Variable a is also a hyperparameter and this hyperparameter represents the smoothness of the images.

According to Bayes' theorem, the posterior distribution of hyperparameters can be written as

$$p(a, b|\boldsymbol{v}) \propto p(\boldsymbol{v}|a, b)p(a, b) \tag{4.5}$$

$$\propto \int \mathrm{d}\boldsymbol{u}\ p(\boldsymbol{v}|\boldsymbol{u}, b)p(\boldsymbol{u}|a), \tag{4.6}$$

where the prior probability distribution of the hyperparameter is

$$p(a, b) = \text{const.}. \tag{4.7}$$

This posterior distribution means that there is no prior knowledge of hyperparameters. The posterior distribution of hyperparameters is represented by

$$p(a, b|\boldsymbol{v}) \propto \int \mathrm{d}\boldsymbol{u}\ p(\boldsymbol{v}|\boldsymbol{u}, b)p(\boldsymbol{u}|a) \tag{4.8}$$

$$\propto \int \mathrm{d}\boldsymbol{u}\ \exp\left[-\left(\frac{1}{2b}\sum_{i=1}^{N}(v_i - u_i)^2 + \frac{1}{2a}\sum_{\langle i,j\rangle}(u_i - u_j)^2\right)\right]. \tag{4.9}$$

This can be computed analytically using the Fourier transform and Gaussian integration. Finally, the hyperparameter distribution of the G–MRF model with a periodic boundary condition can be analytically derived in the following form[20, 24]:

$$F(a, b|\boldsymbol{v}) \equiv -\ln\left(p(a, b|\boldsymbol{v})\right) \tag{4.10}$$

$$= -\frac{1}{2}\sum_{k=1}^{N}\left[\ln\frac{1}{\frac{a}{\lambda_k} + b} - \frac{1}{\frac{a}{\lambda_k} + b}|\tilde{v}_k|^2\right] \tag{4.11}$$

$$\tilde{v}_k = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\exp\left(2\pi\mathrm{i}\frac{jk}{N}\right)v_j \qquad (k = 1, 2, \cdots, N), \tag{4.12}$$

where $\lambda_k = 2 - 2\cos(2\pi k/N)$ is an eigenvalue, i represents the imaginary unit, and $\tilde{}$ signifies the Fourier transform.

4.1.2 Gaussian process regression

Here, we explain the theory of Gaussian process regression[55, 56]. We consider M training data with D-dimensional input vector $\boldsymbol{x}_i$ and output value y_i, denoted by $\mathcal{D}_M = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{M}$. We consider the objective function f to be proportional to the Bayesian posterior distribution. We assume that the objective function f satisfies the following input-output relationship:

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i \qquad f(\boldsymbol{x}_i) = \boldsymbol{\phi}(\boldsymbol{x}_i)^T\boldsymbol{w}, \tag{4.13}$$

where the observation noise ϵ_i follows the normal distribution of the average 0 (the variance σ^2), the weight vector $\boldsymbol{w}$ follows the normal distribution of the average 0 and the variance covariance matrix Σ, and $\boldsymbol{\phi}$ is the function that maps the input vector $\boldsymbol{x}$ to the N-dimensional feature space. Given the input vector $\boldsymbol{x}$, the objective function f is a random variable that follows the normal distribution.

The mean of $f(x)$ is $E[f(\boldsymbol{x})]$, and the covariance $\mathrm{Cov}[f(\boldsymbol{x}), f(\boldsymbol{x}')]$ is

$$E[f(\boldsymbol{x})] = \boldsymbol{\phi}(\boldsymbol{x})^T E[\boldsymbol{w}] = 0 \tag{4.14}$$

$$\mathrm{Cov}[f(\boldsymbol{x})f(\boldsymbol{x}')] = \boldsymbol{\phi}(\boldsymbol{x})^T E[\boldsymbol{w}\boldsymbol{w}^T]\boldsymbol{\phi}(\boldsymbol{x}') \tag{4.15}$$

$$= \boldsymbol{\phi}(\boldsymbol{x})^T \Sigma \boldsymbol{\phi}(\boldsymbol{x}') \tag{4.16}$$

$$\equiv k(\boldsymbol{x}, \boldsymbol{x}'), \tag{4.17}$$

where $k(\boldsymbol{x}, \boldsymbol{x}')$ is a kernel function. Given a new input vector $\boldsymbol{x}_*$, the conditional probability distribution $p(f_*|\boldsymbol{x}_*, \mathcal{D}_M)$ of $f_* \equiv f(\boldsymbol{x}_*)$ for $\mathcal{D}_M$ follows the normal distribution. The probability distribution followed by the outputs of unknown inputs can be derived in the form of a normal distribution. To derive the conditional probability distribution $p(f_*|\boldsymbol{x}_*, \mathcal{D}_M)$, we calculate the posterior distribution of the output and weight vectors. The output is denoted as $\boldsymbol{y} = (y_1, y_2, \cdots, y_M)^T$. The posterior distribution is

$$p(\boldsymbol{y}|\Phi, \boldsymbol{w}) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \boldsymbol{\phi}(\boldsymbol{x}_i)^T \boldsymbol{w})^2}{2\sigma^2}\right) \tag{4.18}$$

$$= \mathcal{N}(\Phi^T \boldsymbol{w}, \sigma^2 I), \tag{4.19}$$

where $\Phi = (\boldsymbol{\phi}(\boldsymbol{x}_1), \boldsymbol{\phi}(\boldsymbol{x}_2), \cdots, \boldsymbol{\phi}(\boldsymbol{x}_M))$, and I is an N-dimensional unit matrix. From Equation (4.19) and Bayes' theorem, the posterior distribution of the weight vector $\boldsymbol{w}$ is

$$p(\boldsymbol{w}|\boldsymbol{y}, \Phi) = \frac{p(\boldsymbol{y}|\Phi, \boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{y}|\Phi)} \tag{4.20}$$

$$\propto p(\boldsymbol{y}|\Phi, \boldsymbol{w})p(\boldsymbol{w}) \tag{4.21}$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \Phi^T \boldsymbol{w})^T(\boldsymbol{y} - \Phi^T \boldsymbol{w})\right) \times \exp\left(-\frac{1}{2}\boldsymbol{w}^T \Sigma^{-1} \boldsymbol{w}\right) \tag{4.22}$$

$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{w} - \bar{\boldsymbol{w}})^T A^{-1}(\boldsymbol{w} - \bar{\boldsymbol{w}})\right) \tag{4.23}$$

$$\propto \mathcal{N}(\bar{\boldsymbol{w}}, A^{-1}), \tag{4.24}$$

where $A = \sigma^{-2}\Phi\Phi^T + \Sigma^{-1}$ and $\bar{\boldsymbol{w}} = \sigma^{-2}A^{-1}\Phi\boldsymbol{y}$. The conditional probability distribution $p(f_*|\boldsymbol{x}_*, \mathcal{D}_M)$ is derived as

$$p(f_*|\boldsymbol{x}_*, \mathcal{D}_M) = \int p(f_*|\boldsymbol{\phi}_*, \boldsymbol{w})p(\boldsymbol{w}|\boldsymbol{y}, \Phi)\mathrm{d}\boldsymbol{w} \tag{4.25}$$

$$= \mathcal{N}\left(\frac{1}{\sigma^2}\boldsymbol{\phi}_*^T A^{-1}\Phi\boldsymbol{y}, \boldsymbol{\phi}_*^T A^{-1}\boldsymbol{\phi}_*\right), \tag{4.26}$$

where $\boldsymbol{\phi}_* \equiv \boldsymbol{\phi}(\boldsymbol{x}_*)$. The average and variance of $f_*|\boldsymbol{x}_*, \mathcal{D}_M$ obtained in Equation (4.26) can be rewritten below using the kernel function defined in Equation (4.17):

$$E[f_*|\boldsymbol{x}_*, \mathcal{D}_M] = \boldsymbol{k}_*^T (K + \sigma^2 I)^{-1} \boldsymbol{y} \tag{4.27}$$

$$\mathrm{Cov}[f_*|\boldsymbol{x}_*, \mathcal{D}_M] = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^T (K + \sigma^2 I)^{-1} \boldsymbol{k}_* \tag{4.28}$$

where $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$,$\boldsymbol{k}_* = (k(\boldsymbol{x}_1, \boldsymbol{x}_*), \cdots, k(\boldsymbol{x}_M, \boldsymbol{x}_*))^T$. In other words, the estimate of the output of the objective function f for the unknown input vector $\boldsymbol{x}_*$ is $\boldsymbol{k}_*^T (K + \sigma^2 I)^{-1} \boldsymbol{y}$, and the reliability is $k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^T (K + \sigma^2 I)^{-1} \boldsymbol{k}_*$. By using these results, the outputs of the objective function for unknown inputs can be predicted. In this study, we assume a smooth posterior distribution and use a radial basis function (RBF)

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta \exp\left(-\frac{|\boldsymbol{x}_i - \boldsymbol{x}_j|^2}{2\beta}\right) \tag{4.29}$$

as a kernel function[9]. When an RBF is used as a kernel function, prediction with a Gaussian process can be conducted by specifying hyperparameters σ, θ, β. The values of hyperparameters have a largely affect the estimation accuracy, so the hyperparameters must be optimized. In our numerical experiments, we maximize the log-likelihood function

$$\ln p(\boldsymbol{y}|\sigma, \theta, \beta) = -\frac{1}{2} \ln |K + \sigma^2 I| - \frac{1}{2} \boldsymbol{y} (K + \sigma^2 I)^{-1} \boldsymbol{y} - \frac{N}{2} \ln(2\pi) \tag{4.30}$$

by Bayesian optimization, explained below, and optimize the hyperparameters[9].

4.1.3 Bayesian optimization

Here, we explain the theory of Bayesian optimization[57, 58]. We consider a situation wherein one is searching for the maximum value of the objective function $\mathcal{F}$ that is proportional to the Bayesian posterior distribution. Let us denote the space of the parameter being searched as $\mathbb{X} \subset \mathbb{R}^d (d \geq 1)$. The procedure of the search algorithm is as follows.

Depending on how the acquisition function is chosen, the type of Bayesian optimization will differ. Here, we use the expected improvement (EI) algorithm, which uses the expectation of the improvement function as the acquisition function[58]. This algorithm has been proven to converge to the optimal value[59]. The improvement function $\mathcal{I}$ is defined as follows:

$$\mathcal{I}(\boldsymbol{x}|\mathcal{D}_n) = \max\{0, \mathcal{F}(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+)\} \tag{4.31}$$

where $\boldsymbol{x}^+ = \arg\max_{\boldsymbol{x}_i \in \boldsymbol{x}_{1:n}} f(\boldsymbol{x}_i)$. The improvement function is positive when the prediction exceeds the maximum value so far; otherwise, it is 0. The acquisition function by the EI algorithm

Algorithm 1 Bayesian optimization

Step 1:

Select $\boldsymbol{x}_1 \in \mathbb{X}$ at random and set the training data $\mathcal{D}_1$.

Step 2:

Train the Gaussian process by the training data $\mathcal{D}_{n-1}$ to estimate the mean and covariance from Equations (4.27), (4.28).

Step 3:

Take $\boldsymbol{x} \in \mathbb{X}$ as $\boldsymbol{x}_n$ where the acquisition function $a(\boldsymbol{x}|\mathcal{D}_{n-1})$ defined by Equation (4.36) is maximized.

Step 4:

Add the new searched point $\boldsymbol{x}_n$ to the training data $\mathcal{D}_{n-1}$ and update the training data $\mathcal{D}_n = \{\mathcal{D}_{n-1}, (\boldsymbol{x}_n, y_n)\}$.

Step 5:

Steps 2–4 are repeated.

Step 6:

Finally, determine the maximum value of the objective function $\mathcal{F}$ that is proportional to the Bayesian posterior distribution from the training data $\mathcal{D}_n$.

is

$$a(\boldsymbol{x}|\mathcal{D}_n) = E[\mathcal{I}(\boldsymbol{x}|\mathcal{D}_n)]. \tag{4.32}$$

This can be advanced analytically to

$$\begin{aligned} E[\mathcal{I}(\boldsymbol{x}|\mathcal{D}_n)] &= \int_{\mathcal{I}=0}^{\mathcal{I}=\infty} \mathcal{I} \frac{1}{\sqrt{2\pi}\sigma(\boldsymbol{x})} \times \\ &\qquad \exp\left(-\frac{(\mu(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+) - \mathcal{I})^2}{2\sigma^2(\boldsymbol{x})}\right) \mathrm{d}\mathcal{I} \qquad (4.33) \\ &= \sigma(\boldsymbol{x}) \left[\frac{\mu(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+)}{\sigma(\boldsymbol{x})} C\left(\frac{\mu(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+)}{\sigma(\boldsymbol{x})}\right) \right. \\ &\qquad \left. + P\left(\frac{\mu(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+)}{\sigma(\boldsymbol{x})}\right) \right] \qquad (4.34) \end{aligned}$$

where

$$\mu(\boldsymbol{x}) = E[\mathcal{F}(\boldsymbol{x})|\boldsymbol{x}, \mathcal{D}_n], \qquad \sigma(\boldsymbol{x}) = \mathrm{Cov}[\mathcal{F}(\boldsymbol{x})|\boldsymbol{x}, \mathcal{D}_n]. \tag{4.35}$$

Therefore, the acquisition function can be calculated as

$$a(\boldsymbol{x}|\mathcal{D}_n) = \begin{cases} (\mu(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+))C(Z) + \sigma(\boldsymbol{x})P(Z) & (\sigma(\boldsymbol{x}) > 0) \\ 0 & (\sigma(\boldsymbol{x}) = 0) \end{cases}. \tag{4.36}$$

Here, $Z = (\mu(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{x}^+))/\sigma(\boldsymbol{x})$, $P(\cdot)$ is the probability density function, and $C(\cdot)$ is the cumulative distribution function of the standard normal distribution:

$$P(Z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{Z^2}{2}\right], \tag{4.37}$$

$$C(Z) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] \mathrm{d}t. \tag{4.38}$$

We use the EI algorithm in our numerical experiments. To maximize the acquisition function, thousands of points were first extracted from parameter space $\mathbb{X}$, and some of the best points to maximize the acquisition function were selected. In addition, a local search was performed around the selected points, and the best point was selected from among them.

4.1.4 Bayesian optimization for Gaussian process

Here, we explain a method for using Bayesian optimization as a sampler for the Gaussian process. Bayesian optimization takes samples from the search space $\mathbb{X}$ in the process of searching for the optimal value of the objective function $\mathcal{F}$. We estimate the shape of the objective function $\mathcal{F}$ that is proportional to the Bayesian posterior distribution using the points sampled with Bayesian optimization from the Gaussian process. The algorithm flow is almost the same as that of Bayesian optimization, but the goal is to estimate the objective function $\mathcal{F}$, not its maximum value.

4.2 Numerical Experiment

We tested the effectiveness of the methods for estimating the Bayesian posterior probability distribution and its maximum value using the Gaussian process and Bayesian optimization. Bayesian posterior probabilities are generally difficult to calculate analytically. Here, to evaluate the performance of the methods, we used the G–MRF model, where the Bayesian posterior probability can be calculated analytically. We generated observation image $\boldsymbol{v}$ with the number

Algorithm 2 Bayesian optimization for Gaussian process

Step 1:

Sample the training data $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ by using Bayesian optimization (Algorithm 1, steps 1–5).

Step 2:

Estimate the objective function $\mathcal{F}$ proportional to the Bayesian posterior distribution by using the Gaussian process (Equations (4.27), (4.28)) from $\mathcal{D}_n$.

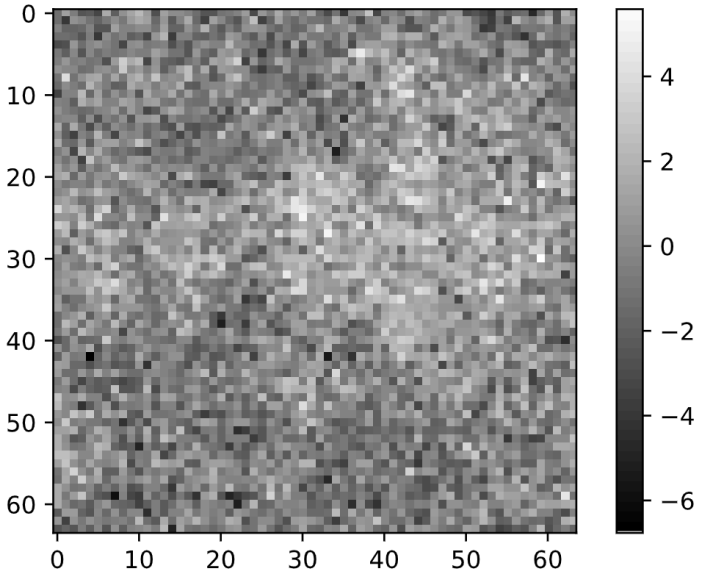


Fig. 4.1: Generated observation image with number of pixels $N = 64 \times 64$ and hyperparameters $a = b = 1.5$, according to Equations (4.2), (4.3).

of pixels $N = 64 \times 64$ (Fig. 4.1) and the hyperparameters $a = b = 1.5$, in accordance with Equations (4.2) and (4.3). Figure 4.1 is the generated observation image. Then, we calculated the hyperparameter distribution of the generated image numerically using Equation (4.11). We evaluate the performance of the methods using the hyperparameter distribution.

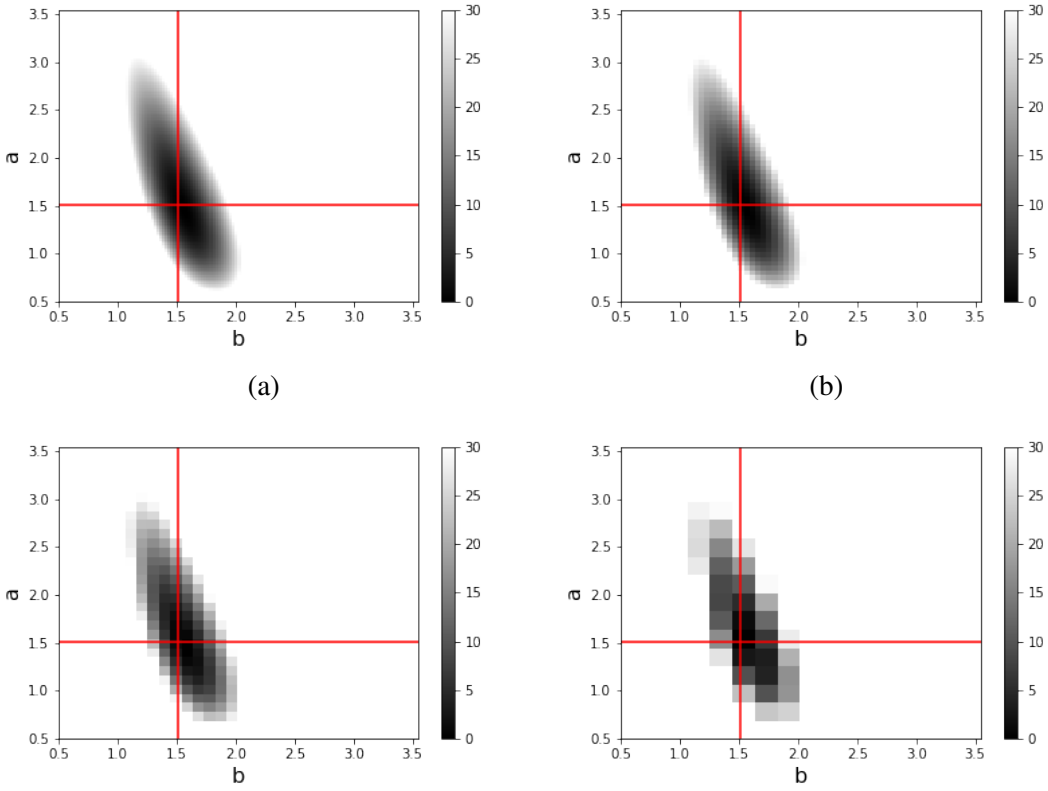
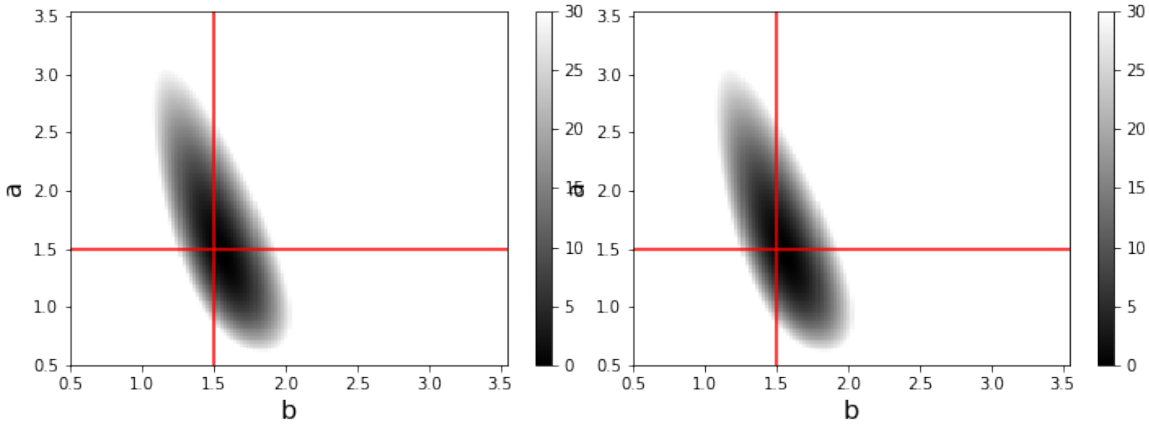


Fig. 4.2: Hyperparameter distribution with different number of grid points. We used numerical calculation to generate observation image $\boldsymbol{v}$ with pixel number $N = 64 \times 64$, $a = b = 1.5$, and output hyperparameter distribution. (a) Distribution with 128×128 number of grid points, (b) distribution with 64×64 number of grid points, (c) distribution with 32×32 number of grid points, and (d) distribution with 16×16 number of grid points.

4.2.1 Distribution interpolation using Gaussian process

The distribution and optimum values of hyperparameter estimation are often obtained by the grid search method. Here, we estimate the distribution from a small number of samples by the grid search method using the Gaussian process. We calculated the hyperparameter distribution for Fig. 4.1 in $0.5 \leq a, b \leq 3.5$ by the grid search method. Each distribution shown in Fig. 4.2 is a hyperparameter distribution with different number of grid points. Figure 4.2 (a)–(d) show the distribution with 128×128, 64×64, 32×32, and 16×16 number of grid points



(a) (b)

Fig. 4.3: Hyperparameter distribution with 128×128 number of grid points interpolated using hyperparameter distribution (Fig. 4.2 (d)) with 16×16 number of grid points as input. (a) Hyperparameter distribution interpolated by Gaussian process. (b) Hyperparameter distribution interpolated by spline interpolation.

respectively. We call the distribution with with 128×128 number of grid points (Fig. 4.2 (a)) the original distribution. We interpolated the distribution with 16×16 number of grid points (Fig. 4.2 (d)) by using the Gaussian process (Equations (4.27) and (4.28)) to estimate the true distribution. The distribution with 16×16 number of grid points (Fig. 4.2 (d)) was interpolated by the spline function[60] for performance comparison. Figure 4.3 is the result of interpolating the distribution with 16×16 number of grid points to the distribution with 128×128 number of grid points. Figure 4.3 (a) and 4.3 (b) are the interpolated distributions by Gaussian process and the spline function respectively.

Here, we evaluated the similarity between the distributions by the root-mean-square error (RMSE), Kullback–Leibler (KL) divergence, and Hellinger distance. In the evaluation by RMSE, we uses th estimated distribution of free energy (Equation (4.11)). On the other hand, since KL divergence and Hellinger distance can be calculated only for the probability distribution, we convert the free energy into the probability distribution for evaluation.

The each similarity between each distribution and the original distribution (Fig. 4.2 (a)) is shown in Table 4.1. From Tab. 4.1, the distribution estimated by the Gaussian process (Fig. 4.3 (a)) has better accuracy than the distribution with 64×64 number of grid points (Fig. 4.2-(b)). In particular, the distribution estimated by the Gaussian process was improved in the KL

Table 4.1: Similarity between distribution of number of grid points and hyperparameter distribution of 128×128 number of grid points (Fig. 4.2 (a)). Fig. 4.3 (a) is a distribution with 128×128 number of grid points estimated by Gaussian process using hyperparameter distribution (Fig. 4.2 (d)) with 16×16 number of grid points. Fig. 4.3 (b) is a distribution with 128×128 number of grid points estimated by Spline function using hyperparameter distribution (Fig. 4.2 (d)) with 16×16 number of grid points.

Grid points	64×64 (Fig. 4.2 (b))	32×32 (Fig. 4.2 (c))	16×16 (Fig. 4.2 (d))	GP (Fig. 4.3 (a))	Spline (Fig. 4.3 (b))
RMSE	11.7	31.9	75.6	0.08	0.63
KL	4.79×10^{-2}	2.62×10^{-1}	8.87×10^{-1}	8.38×10^{-6}	1.34×10^{-5}
Hellinger	2.41×10^{-2}	1.08×10^{-1}	3.97×10^{-1}	4.19×10^{-6}	6.71×10^{-6}

divergence and the Hellinger distance.

Gaussian process works well for interpolation of distributions. The peak of hyperparameter distribution is in the search area of the grid search. When the hyperparameter distribution is treated as free energy, the value of the hyperparameter distribution increases as the distance from the peak increases. Therefore, the result of RMSE which evaluated the hyperparameter distribution as free energy has a small improvement compared to other metrics. On the other hand, the evaluations by KL divergence and Hellinge distance evaluate use the free energy after converting it into the probability distribution. As a result, the error at points away from the center of the distribution is evaluated small KL divergence and Hellinge distance.

From these facts, it can be seen that the distribution estimation around the peak worked well for the target data. We found that the hyperparameter distribution can be estimated from small number of grid points by using the Gaussian process.

4.2.2 Hyperparameter estimation using Bayesian optimization

The optimum values of hyperparameter estimation are often obtained by the grid search method. Bayesian optimization also searches for the optimum value at high speed in consideration of the evaluation value of the sampling point. Here, we verify that Bayesian optimization (Algorithm 1) works effectively in the search for optimal values of hyperparameters using the hyperparameter distribution calculated analytically. We confirm the behavior of Bayesian optimization in comparison with simple random sampling. We calculated the hyperparameter distribution of the generated image (Fig. 4.1) numerically using Equation (4.11). We estimated the optimal value for the hyperparameter distribution using Bayesian optimization. We searched for the optimal

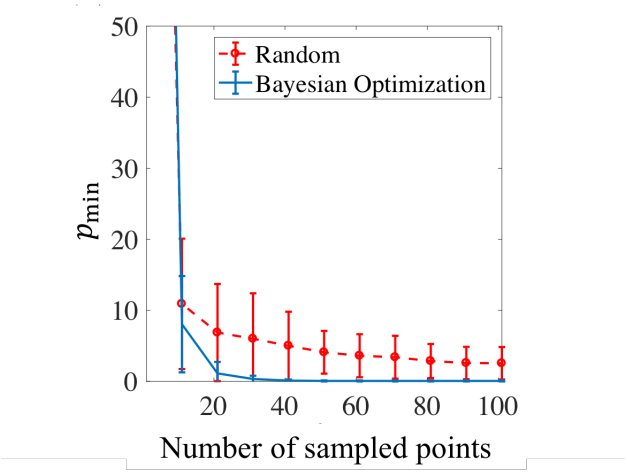


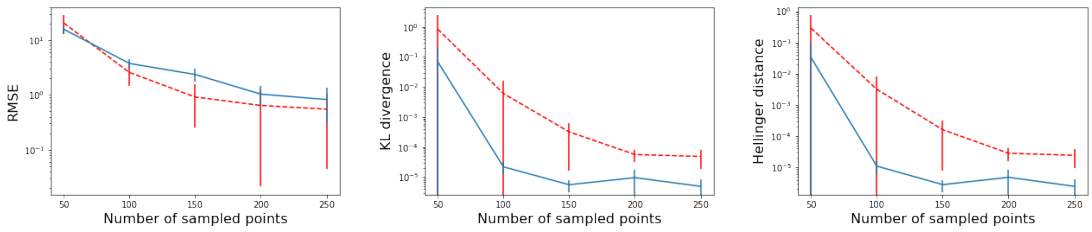
Fig. 4.4: Hyperparameter estimation by Bayesian optimization and random sampling. We generated observation image $\boldsymbol{v}$ with number of pixels $N = 64 \times 64$ and hyperparameters $a = b = 1.5$ and we calculated hyperparameter distribution of generated image (Fig. 4.1) numerically using Equation (4.11). We searched for the optimal value of hyperparameter distribution in $0.5 \leq a, b \leq 3.5$ using the EI algorithm with 100 sampled points. We denoted minimum value of hyperparameter distribution among sampled points as $p_{\min}$. Solid and dashed lines represent result Bayesian optimization (EI algorithm) and random sampling[61].

value of the hyperparameter distribution of the generated image (Fig. 4.1) in $0.5 \leq a, b \leq 3.5$ using the EI algorithm with 100 sampled points. Figure 4.4 shows the minimum value of the hyperparameter distribution among the sample points against the number of sampled points. The solid and dashed lines in Fig. 4.4 represent the results for Bayesian optimization and random sampling, respectively. We compared the Bayesian optimization results with the random

sampling results in terms of the number of sampled points required to converge to the true hyperparameters. Hyperparameter estimation values with 10 samples showed no difference between sampling by Bayesian optimization and random sampling. Hyperparameter estimation by Bayesian optimization converges to the true hyperparameters with 50 samples. However, hyperparameter estimation by random sampling does not converge with 100 samples. The standard deviation of Bayesian optimization is smaller than one of random sampling. From the result, Bayesian optimization works effectively for hyperparameter estimation and fluctuates little and highly reliable.

4.2.3 Hyperparameter distribution estimation by Bayesian optimization and Gaussian process

In the Bayesian optimization algorithm, sampling is performed from the search space. Here, we verify whether the hyperparameter distribution can be estimated with a small number of points sampled by Bayesian optimization using the Gaussian process. We also confirm the behavior of Bayesian optimization in comparison with simple random sampling. We searched for the optimal values of the hyperparameter distribution of the generated image (Fig. 4.1) in $0.5 \leq a, b \leq 3.5$ using the EI algorithm with 250 sampled points. We also randomly sampled 250 points



(a) (b) (c)

Fig. 4.5: (a) RMSE, (b) KL divergence, and (c) Hellinger distance between estimated distribution and original distribution (Fig. 4.2 (a)) against number ot sampled points for 10 trials. We searched for optimal values of hyperparameter distribution of generated image (Fig. 4.1) in $0.5 \leq a, b \leq 3.5$ using EI algorithm with 250 sampled points and also randomly sampled 250 points from same domain of hyperparameter distribution. We estimated hyperparameter distributions with the Gaussian process from sampled points by Bayesian optimization and random sampling. Solid and dashed lines show results for the estimated distributions using points sampled by Bayesian optimization and random sampling, respectively.

from the hyperparameter distribution of the generated image (Fig. 4.1) in $0.5 \leq a, b \leq 3.5$, and estimated the hyperparameter distributions with the Gaussian process from the sampled points by Bayesian optimization and random sampling. We evaluated the similarity between the estimated distributions by RMSE, KL divergence, and Hellinger distance. The statistical results of 10 trials are shown in Fig. 4.5. Figure 4.5 (a)–(c) show RMSE, KL divergence, and Hellinger distance between the estimated distribution and the original distribution (Fig. 4.2 (a)) against number ot sampled points respectively. The solid and dashed lines show the result for the estimated distributions using points sampled by Bayesian optimization and random sampling, respectively. Fig. 4.5, it can be seen that the hyperparameter distribution estimation fluctuated little and was reliable by Bayesian optimization. In particular, the improvement in KL divergence and Helliger distance is remarkable, indicating that the area around the peak of the distribution is improved. It can be seen that the estimated distribution by Bayesian optimization is improved around the peak of the distribution because the KL divergence and Helliger distances are significantly improved compared to the estimated distribution by random sampling.

Figure 4.6 shows the sampling points obtained by Bayesian optimization and random sampling in a trial. The upper column and lower column of Fig. 4.6 shows the sampling points by Bayesian optimization and random sampling respectively. Sampling points are indicated by blue

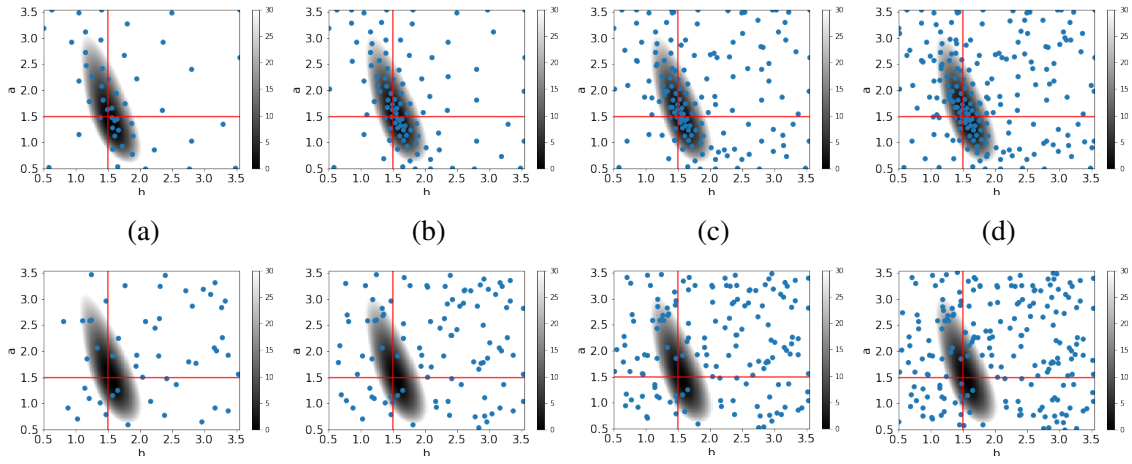


Fig. 4.6: Sampling points obtained by Bayesian optimization and random sampling in a trial. Sampling points are indicated by blue dots. In upper column, sampling points by Bayesian optimization. In lower column, sampling points by random sampling. Number of sampling points in each panel is 50, 100, 150, 200 in order from the left.

dots. The number of sampling points in each figure is 50, 100, 150, 200 in order from the left. From Fig. 4.6, it can be seen that Bayesian optimization focuses on the area around the peak of the distribution. Therefore, the estimated distribution with Bayesian optimization as a sampler has a smaller KL distance and Helliger distances than the distribution estimated with random sampling.

4.3 Discussion

In Sec. 4.2, we confirmed that Gaussian process and Bayesian optimization work effectively using the Bayesian posterior probabilities of the MRF model that can be calculated analytically. Since Bayesian optimization is an optimal value search method, the search is centered around the peak of the distribution as shown in Fig. 4.6. KL divergence largely depends on distribution shape around the peak of the distribution. Therefore, as you can see from Fig.4.6 (b) and Table 4.1, the estimation performances by Bayesian optimization in terms of KL divergence are high. On the other hand, since, in the evaluation by RMSE, we treated the hyperparameter distribution as free energy, the influence of the distribution shape around the distribution peak became small. As you can see from Fig.4.6 (a), the estimation performance by Bayesian optimization was low compared to the estimation performance by random sampling, which was extensively searched around the distribution.

The hyperparameter distribution used in the numerical experiment was a single peak and smooth distribution. Since Bayesian posterior probabilities are generally smooth, the proposed methods work well in many cases. However, it is known that Bayesian optimization is difficult to work when the search space for distribution parameters becomes high-dimensional or when the distribution becomes densely multi-peaked. It is necessary to investigate such a distribution as a future task.

4.4 Conclusion

We proposed a method to estimate hyperparameter distribution and its maximum value by using the Gaussian process and Bayesian optimization for an Markov random field (MRF) model and tested our method's effectiveness through numerical experiments. We have shown that the hyperparameter distribution of the MRF model can be estimated from a small number of sampling points. In other words, we showed that the reliability for the estimated value of the physical quantities can be estimated from a small number of data points. By using the Gaussian process, we found that the coarse-grained distribution can be interpolated and the resolution can be im-

proved more than fourfold. In addition, hyperparameter estimation using Bayesian optimization has faster convergence and higher accuracy from fewer samplings than random sampling. We also demonstrated that Bayesian optimization can be used not only for optimal-value searching but also as a sampler for the Gaussian process, so that hyperparameter distribution can be estimated from fewer samples. From these results, we proposed a method to estimate the hyperparameter distribution with a small amount of computation and demonstrated its effectiveness.

These methods are expected to be applicable to material science. The model parameters of materials are often estimated using Bayesian inference[62]. The posterior probability is estimated using Markov chain Monte Carlo methods that generally have a high calculation cost. The maximum a posteriori is estimated speedily using Bayesian optimization[63]. Not only the optimum value but also the reliability of the determined value needs to be estimated quickly. The proposed method can enable faster analysis of the posterior distribution. The forward calculations of physical system often have take a highly cost for computational resource. Our proposed method enable you to reduce the cost in the experiment design.

The applicability of this method could possibly be broadened by discussing its effectiveness for more complicated distributions. The hyperparameter distribution of the G–MRF model used in this work has a single peak and one optimal value. To discuss this method's applicability more generally, the effectiveness of the method for distributions with multiple local solutions needs to be investigated.

Chapter 5

Conclusion

In this thesis, we focused on on parameter estimation for physical models by using Bayesian inference. The target data are dispersion relation observation data and image data, and in each chapter, we show the practical approach, the analytical approach, and the approximate approach for the target data together with demonstrations.

In Chapter 2, we proposed a practical method for estimating the interaction parameters of the crystal lattice from the dispersion relation observation data. In the proposed method, Poisson noise was introduced as observed noise, and the interaction parameters were distribution estimated. We demonstrated the estimation of interaction parameters using artificial data for various measurement times. It was shown that parameters can be estimated from statistically accurate data, which was difficult to estimate until now. Distribution estimation made it possible to discuss estimation accuracy and it became clear that the estimation performance was improved compared to the previous method.

In Chapter 3, we proposed an analytical method for estimating hyperparameter of Gaussian MRF model from image data. We showed that Bayesian posterior probabilities can be calculated analytically for Gaussian MRF models with no boundary conditions, and this made it possible to analytically estimate the hyperparameter distribution of images with no boundary conditions, which had a high calculation cost and was difficult to calculate numerically. This made it possible to obtain parameters for more general image data at a lower calculation cost. With the proposed method, parameters can be obtained exactly from more general image data at a lower computational cost.

In Chapter 4, we proposed an approximate method for distribution estimation using Bayesian inferece. In general, the Bayesian posterior probability has a high calculation cost even when it is calculated numerically using sampling method such as REMC. Therefore, the optimum value and distribution of Bayesian posterior probabilities were obtained from a small number of

samplings by Gaussian process and Bayesian optimization. The performance of the proposed method was verified by numerical experiments using the GMRF model. As a result, it was confirmed that the true distribution shape can be obtained by interpolation from a small number of samplings and that the optimum value can be searched by a small number of sampling.

In conclusion, we propose the method to physical parameter estimation of physical models and evaluation of confidence of the data for the models. By utilizing the proposed method for measurement, we can improve designs of the measurement and processing procedure of the data.

A

Appendix

A.1 Modal analysis of monatomic body-centered cubic lattice model

Here, the monatomic cubic lattice model, which is the object of the numerical experiments in Chapter 2, is formulated as a classic spring model, and we analyze the model to obtain the stiffness matrix in the model system. First, an equation of motion is derived in consideration of the interaction of neighboring atoms in the model. Let the interatomic distance a. Eight first nearest neighbors exist for a given atom, and the eight direction vectors to them are

$$\boldsymbol{h}_1 = \frac{a}{2}\begin{pmatrix}1\\1\\1\end{pmatrix} \qquad \boldsymbol{h}_2 = \frac{a}{2}\begin{pmatrix}-1\\1\\1\end{pmatrix} \qquad \boldsymbol{h}_3 = \frac{a}{2}\begin{pmatrix}1\\-1\\1\end{pmatrix} \qquad \boldsymbol{h}_4 = \frac{a}{2}\begin{pmatrix}1\\1\\-1\end{pmatrix} \tag{A.1}$$

$$\boldsymbol{h}_i = -\boldsymbol{h}_{i-4} \qquad (5 \leq i \leq 8) \tag{A.2}$$

Six second nearest neighbors exist for the atom, and the six direction vectors to them are

$$\boldsymbol{n}_1 = a\begin{pmatrix}1\\0\\0\end{pmatrix} \qquad \boldsymbol{n}_2 = a\begin{pmatrix}0\\1\\0\end{pmatrix} \qquad \boldsymbol{n}_3 = a\begin{pmatrix}0\\0\\1\end{pmatrix} \tag{A.3}$$

$$\boldsymbol{n}_i = -\boldsymbol{n}_{i-3} \qquad (4 \leq i \leq 6) \tag{A.4}$$

12 third nearest neighbors exist for the atom, and the 12 direction vectors to them are

$$\boldsymbol{p}_1 = a\begin{pmatrix}1\\1\\0\end{pmatrix} \qquad \boldsymbol{p}_2 = a\begin{pmatrix}1\\-1\\0\end{pmatrix} \qquad \boldsymbol{p}_3 = a\begin{pmatrix}1\\0\\1\end{pmatrix} \tag{A.5}$$

$$\boldsymbol{p}_4 = a\begin{pmatrix}-1\\0\\1\end{pmatrix} \qquad \boldsymbol{p}_5 = a\begin{pmatrix}0\\1\\1\end{pmatrix} \qquad \boldsymbol{p}_6 = a\begin{pmatrix}0\\-1\\1\end{pmatrix} \tag{A.6}$$

$$\boldsymbol{p}_i = -\boldsymbol{p}_{i-6} \qquad (7 \le i \le 12) \tag{A.7}$$

Using these direction vectors, the equation of motion is written as

$$\begin{aligned} M\frac{\partial^2 \boldsymbol{u}}{\partial t^2}(\boldsymbol{r}_0) &= \alpha_1 \sum_{i=1}^{8} \left[(\boldsymbol{u}(\boldsymbol{r}_0+\boldsymbol{h}_i)-\boldsymbol{u}(\boldsymbol{r}_0))\cdot\hat{\boldsymbol{h}}_i\right]\hat{\boldsymbol{h}}_i \\ &+ \alpha_2 \sum_{i=1}^{6} \left[(\boldsymbol{u}(\boldsymbol{r}_0+\boldsymbol{n}_i)-\boldsymbol{u}(\boldsymbol{r}_0))\cdot\hat{\boldsymbol{n}}_i\right]\hat{\boldsymbol{n}}_i \\ &+ \alpha_3 \sum_{i=1}^{12} \left[(\boldsymbol{u}(\boldsymbol{r}_0+\boldsymbol{p}_i)-\boldsymbol{u}(\boldsymbol{r}_0))\cdot\hat{\boldsymbol{p}}_i\right]\hat{\boldsymbol{p}}_i \end{aligned} \tag{A.8}$$

$$\begin{aligned} &= \alpha_1 \sum_{i=1}^{4} \left[(\boldsymbol{u}(\boldsymbol{r}_0+\boldsymbol{h}_i)+\boldsymbol{u}(\boldsymbol{r}_0-\boldsymbol{h}_i)-2\boldsymbol{u}(\boldsymbol{r}_0))\cdot\hat{\boldsymbol{h}}_i\right]\hat{\boldsymbol{h}}_i \\ &+ \alpha_2 \sum_{i=1}^{3} \left[(\boldsymbol{u}(\boldsymbol{r}_0+\boldsymbol{n}_i)+\boldsymbol{u}(\boldsymbol{r}_0-\boldsymbol{n}_i)-2\boldsymbol{u}(\boldsymbol{r}_0))\cdot\hat{\boldsymbol{n}}_i\right]\hat{\boldsymbol{n}}_i \\ &+ \alpha_3 \sum_{i=1}^{6} \left[(\boldsymbol{u}(\boldsymbol{r}_0+\boldsymbol{p}_i)+\boldsymbol{u}(\boldsymbol{r}_0-\boldsymbol{p}_i)-2\boldsymbol{u}(\boldsymbol{r}_0))\cdot\hat{\boldsymbol{p}}_i\right]\hat{\boldsymbol{p}}_i \end{aligned} \tag{A.9}$$

where $\boldsymbol{u} \in \mathbb{R}^3$ is the displacement vector of the atom and $\boldsymbol{r}_0$ is the equilibrium position of the mass point. $\boldsymbol{h}_i$, $\boldsymbol{p}_i$, and $\boldsymbol{n}_i$ represent the relative position of neighboring atoms. $\hat{\cdot}$ means a unit vector. Introducing the angular frequency $\omega \in \mathbb{R}$ and the wave number $\boldsymbol{q} = (q_x, q_y, q_z) \in \mathbb{R}^3$, we have the displacement vector proportional to the wave component such

that $\boldsymbol{u} \propto \exp\left[i(\boldsymbol{q}\cdot\boldsymbol{r}_0 - \omega t)\right]$. Equation (A.9) is rewritten as follow

$$
\begin{aligned}
-\omega^2 M\boldsymbol{u} =& \alpha_1 \sum_{i=1}^{4} \left\{ \left(e^{i\boldsymbol{q}\cdot\boldsymbol{h}_i} + e^{-i\boldsymbol{q}\cdot\boldsymbol{h}_i} - 2 \right) \hat{\boldsymbol{h}}_i \cdot \hat{\boldsymbol{h}}_i^{\top} \right\} \boldsymbol{u} \\
&+ \alpha_2 \sum_{i=1}^{3} \left\{ \left(e^{i\boldsymbol{q}\cdot\boldsymbol{n}_i} + e^{-i\boldsymbol{q}\cdot\boldsymbol{n}_i} - 2 \right) \hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{n}}_i^{\top} \right\} \boldsymbol{u} \\
&+ \alpha_3 \sum_{i=1}^{6} \left\{ \left(e^{i\boldsymbol{q}\cdot\boldsymbol{p}_i} + e^{-i\boldsymbol{q}\cdot\boldsymbol{p}_i} - 2 \right) \hat{\boldsymbol{p}}_i \cdot \hat{\boldsymbol{p}}_i^{\top} \right\} \boldsymbol{u} \qquad \text{(A.10)} \\
=& \alpha_1 \sum_{i=1}^{4} \left\{ 2\left(\cos(\boldsymbol{q}\cdot\boldsymbol{h}_i) - 1\right) \hat{\boldsymbol{h}}_i \cdot \hat{\boldsymbol{h}}_i^{\top} \right\} \boldsymbol{u} \\
&+ \alpha_2 \sum_{i=1}^{3} \left\{ 2\left(\cos(\boldsymbol{q}\cdot\boldsymbol{n}_i) - 1\right) \hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{n}}_i^{\top} \right\} \boldsymbol{u} \\
&+ \alpha_3 \sum_{i=1}^{6} \left\{ 2\left(\cos(\boldsymbol{q}\cdot\boldsymbol{p}_i) - 1\right) \hat{\boldsymbol{p}}_i \cdot \hat{\boldsymbol{p}}_i^{\top} \right\} \boldsymbol{u} \qquad \text{(A.11)} \\
=& D\boldsymbol{u} \qquad \text{(A.12)}
\end{aligned}
$$

where D is the 3×3 stiffness matrix in the model system. From Eq. (A.11), we obtain the stiffness matrix $D = \{d_{ij}\}$ as

$$
\begin{aligned}
d_{ii} =& \frac{8}{3}\alpha_1 \left[\cos\left(\frac{aq_x}{2}\right) \cos\left(\frac{aq_y}{2}\right) \cos\left(\frac{aq_z}{2}\right) - 1 \right] \qquad \text{(A.13)} \\
&+ 2\alpha_2 \left[\cos(aq_i) - 1\right] + 2\alpha_3 \left\{ \cos(aq_i) \left[\cos(aq_j) + \cos(aq_k) \right] - 2 \right\} \qquad \text{(A.14)} \\
d_{ij} =& -\frac{8}{3}\alpha_1 \sin\left(\frac{aq_i}{2}\right) \sin\left(\frac{aq_j}{2}\right) \cos\left(\frac{aq_k}{2}\right) - 2\alpha_3 \sin(aq_i) \sin(aq_j) \qquad \text{(A.15)}
\end{aligned}
$$

where, the suffix $i, j, k \in \{x, y, z\}$ and $i \neq j \neq k$.

A.2 Estimation results of the spectrum parameters

Here, the estimation results of the other parameters of the proposed method in Chapter 2 are described below.

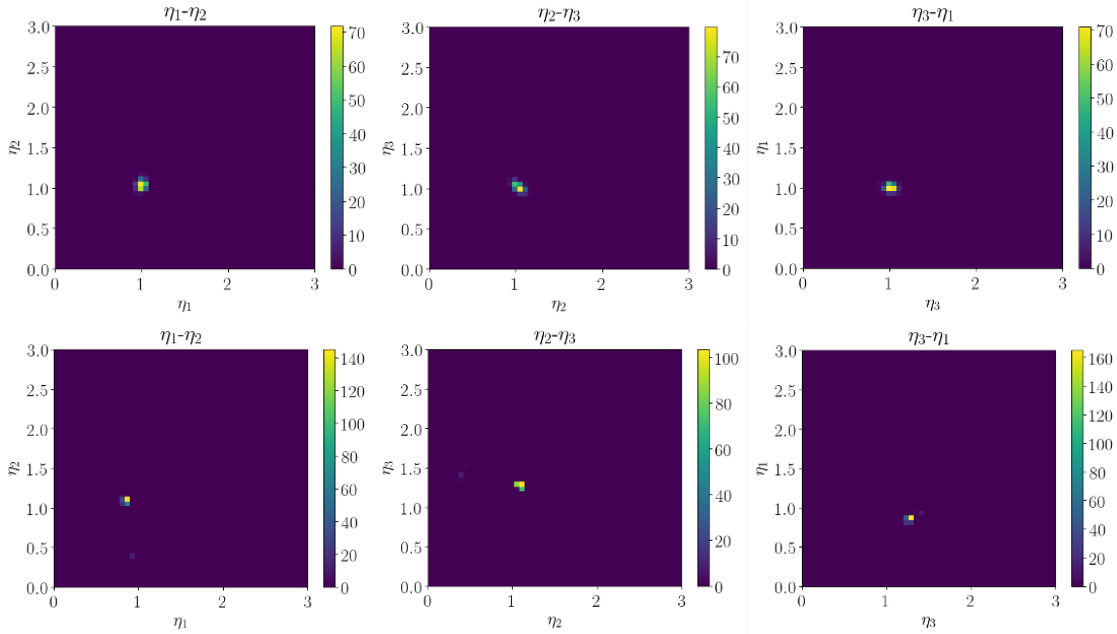


Fig. A.1: Posterior distributions of $\boldsymbol{\eta}$ of for the observation data generated by $T = 100$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\eta}$.The parameters used for the REMC method for estimation of $\boldsymbol{\eta}$ are shown in Table 2.1.

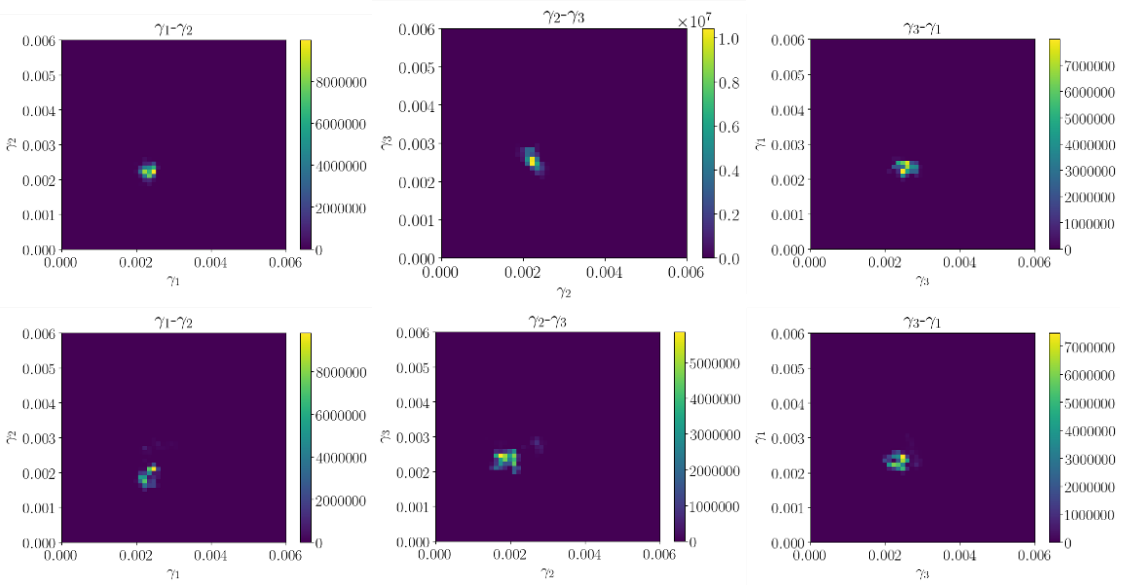


Fig. A.2: Posterior distributions of γ of for the observation data generated by $T = 100$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true γ.The parameters used for the REMC method for estimation of γ are shown in Table 2.1.

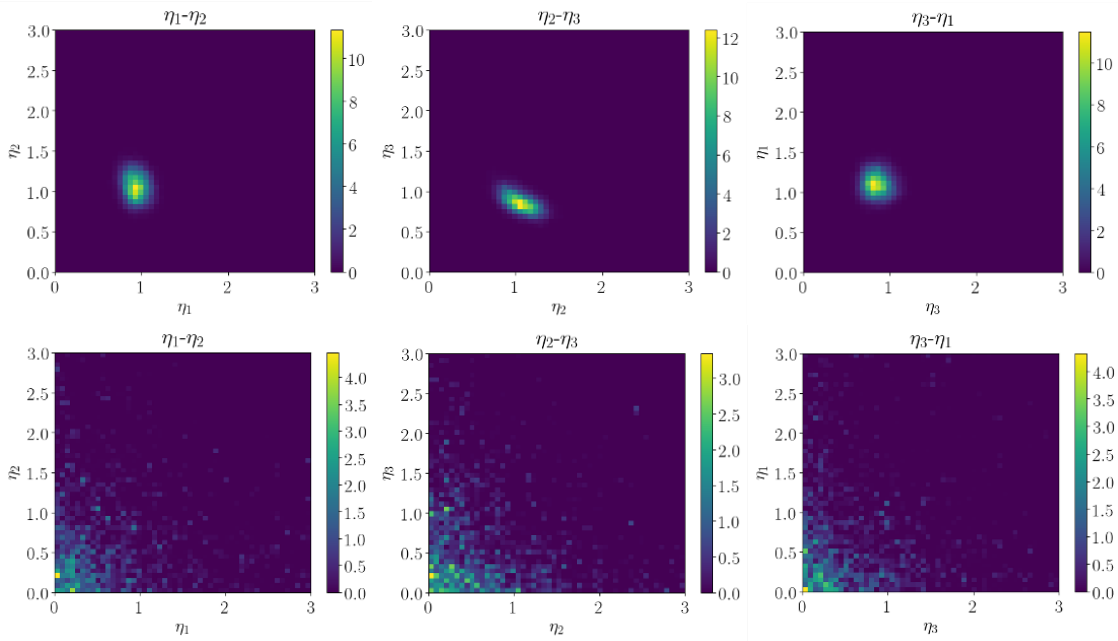


Fig. A.3: Posterior distributions of $\boldsymbol{\eta}$ of for the observation data generated by $T = 10$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\eta}$.The parameters used for the REMC method for estimation of $\boldsymbol{\eta}$ are shown in Table 2.1.

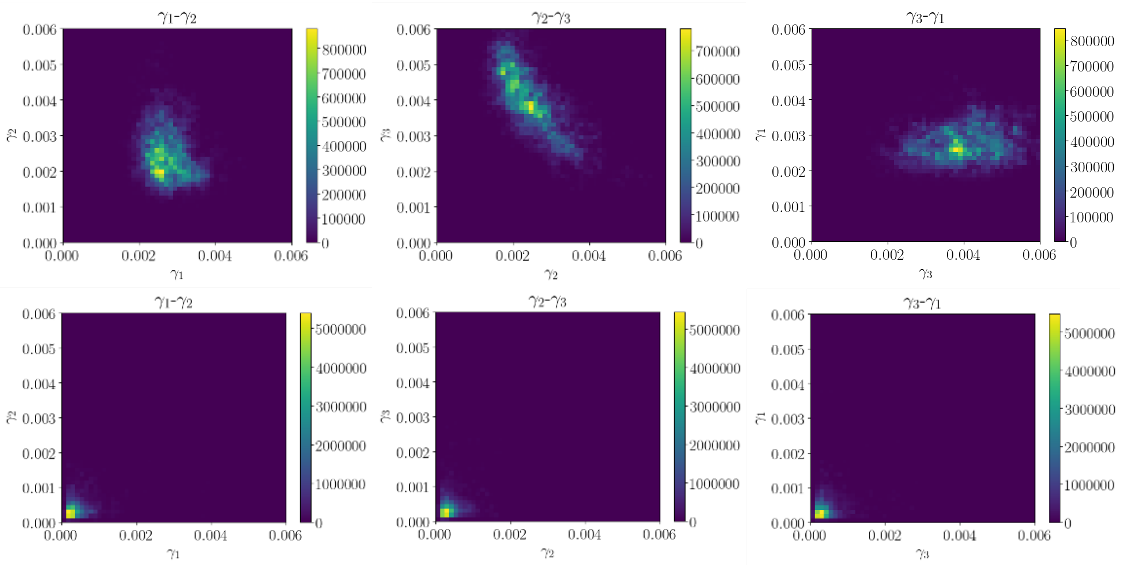


Fig. A.4: Posterior distributions of γ of for the observation data generated by $T = 10$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true γ.The parameters used for the REMC method for estimation of γ are shown in Table 2.1.

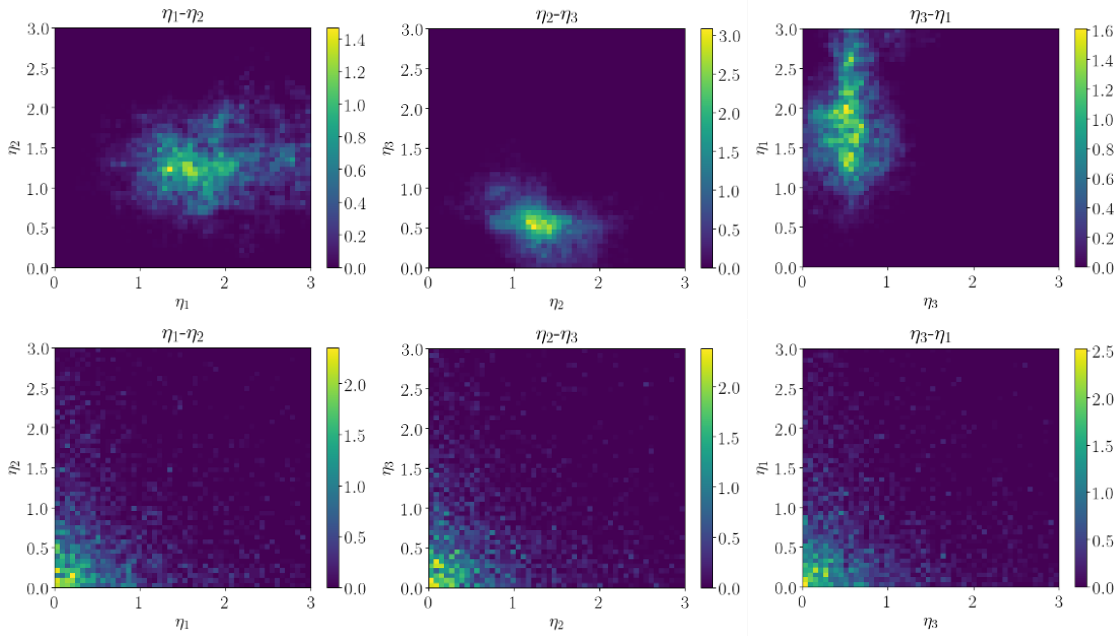


Fig. A.5: Posterior distributions of $\boldsymbol{\eta}$ of for the observation data generated by $T = 1$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\eta}$.The parameters used for the REMC method for estimation of $\boldsymbol{\eta}$ are shown in Table 2.1.

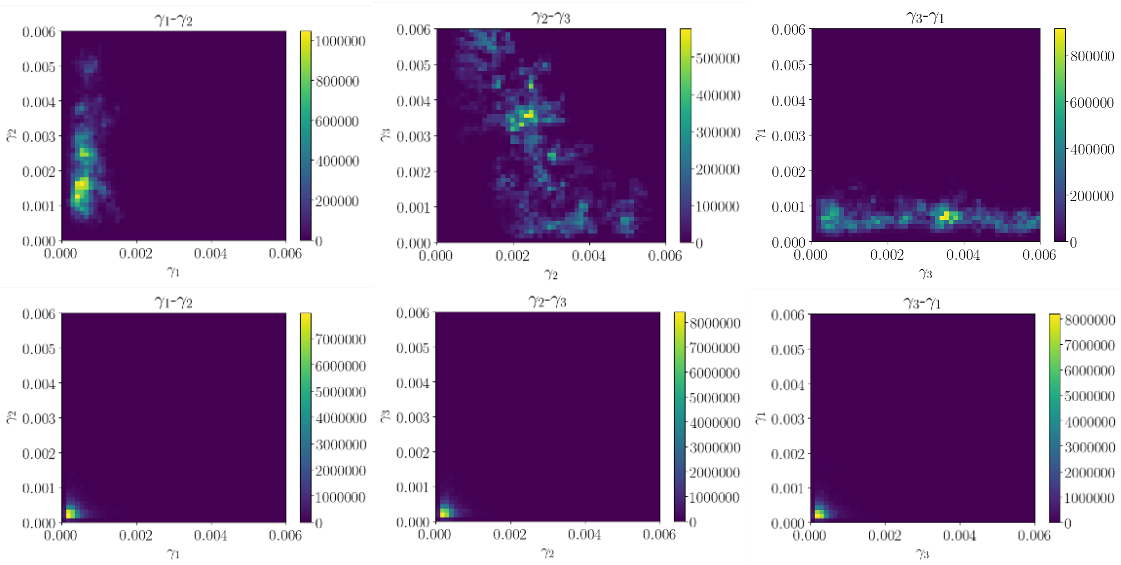


Fig. A.6: Posterior distributions of γ of for the observation data generated by $T = 1$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true γ.The parameters used for the REMC method for estimation of γ are shown in Table 2.1.

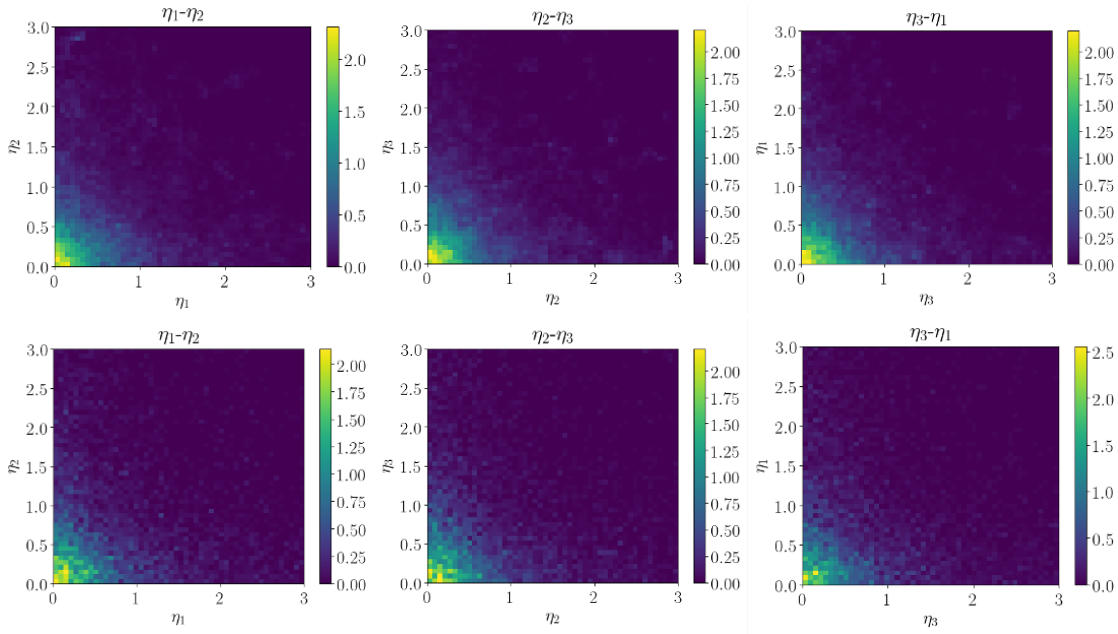


Fig. A.7: Posterior distributions of $\boldsymbol{\eta}$ of for the observation data generated by $T = 0.1$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true $\boldsymbol{\eta}$.The parameters used for the REMC method for estimation of $\boldsymbol{\eta}$ are shown in Table 2.1.

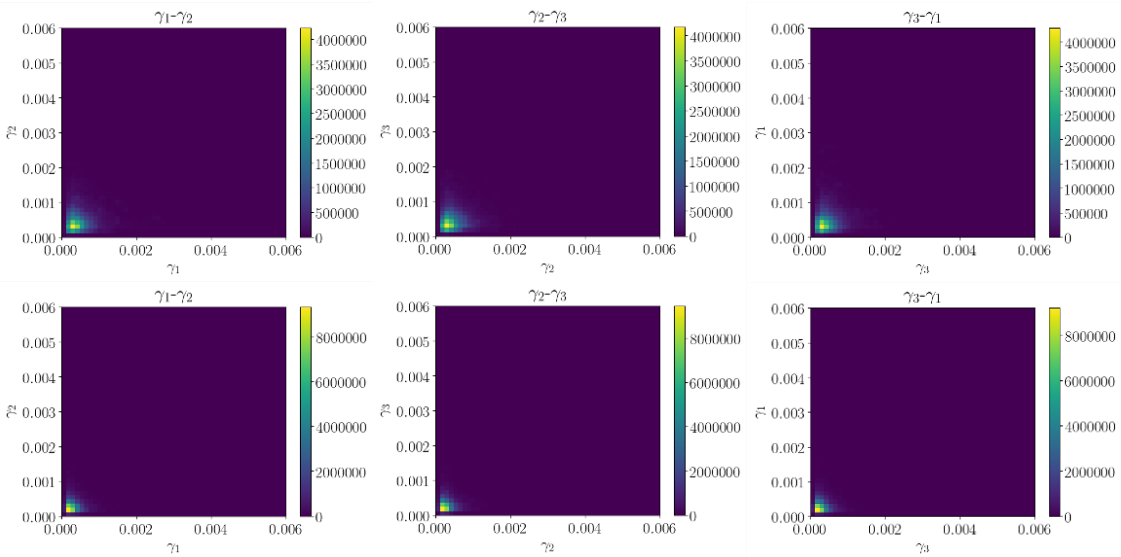


Fig. A.8: Posterior distributions of γ of for the observation data generated by $T = 0.1$. The results in the proposed method and previous method are shown on the upper and lower rows, respectively. The intersections of the red dot lines represent the true γ.The parameters used for the REMC method for estimation of γ are shown in Table 2.1.

Bibliography

[1] Christoph Stosiek, Olga Garaschuk, Knut Holthoff, and Arthur Konnerth. In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences*, Vol. 100, No. 12, pp. 7319–7324, 2003.

[2] Benjamin F Grewe, Dominik Langer, Hansjörg Kasper, Björn M Kampa, and Fritjof Helmchen. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature methods*, Vol. 7, No. 5, pp. 399–405, 2010.

[3] Yoshinori Ohno, Kenji Nagata, Tatsu Kuwatani, Hayaru Shouno, and Masato Okada. Deterministic algorithm for nonlinear markov random field model. *Journal of the Physical Society of Japan*, Vol. 81, No. 6, p. 064006, 2012.

[4] Tatsu Kuwatani, Kenji Nagata, Masato Okada, and Mitsuhiro Toriumi. Markov random field modeling for mapping geofluid distributions from seismic velocity structures. *Earth, Planets and Space*, Vol. 66, No. 1, p. 5, 2014.

[5] Tatsu Kuwatani, Kenji Nagata, Masato Okada, and Mitsuhiro Toriumi. Markov-random-field modeling for linear seismic tomography. *Physical Review E*, Vol. 90, No. 4, p. 042137, 2014.

[6] Kenji Nagata and Sumio 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.

[7] Takuma Kasai, Kenji Nagata, Masato Okada, and Takanori Kigawa. Nmr spectral analysis using prior knowledge. In *Journal of Physics: Conference Series*, Vol. 699, p. 012003. IOP Publishing, 2016.

[8] Shin Murata, Kenji Nagata, Makoto Uemura, and Masato 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.

[9] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[10] Hirotaka Sakamoto, Shun Katakami, Kensuke Muto, Kenji Nagata, Taka-hisa Arima, and Masato Okada. Bayesian parameter estimation using dispersion relation spectra. *Journal*

of the Physical Society of Japan, Vol. 89, No. 12, p. 124002, 2020.

[11] Julian E Besag and Patrick AP Moran. On the estimation and testing of spatial interaction in gaussian lattice processes. *Biometrika*, Vol. 62, No. 3, pp. 555–562, 1975.

[12] Sridhar Lakshmanan and Haluk 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.

[13] Guy 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.

[14] Jun 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.

[15] Jun Zhang, James W Modestino, and David 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.

[16] Seiji Miyoshi and Masato 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.

[17] Ryota Hasegawa, Masato Okada, and Seiji 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.

[18] Haluk Derin, Howard Elliott, Roberto Cristi, and Donald Geman. Bayes smoothing algorithms for segmentation of binary images modeled by markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 6, pp. 707–720, 1984.

[19] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 6, No. 6, pp. 721–741, 1984.

[20] Yoshinori Nakanishi-Ohno, Kenji Nagata, Hayaru Shouno, and Masato Okada. Distribution estimation of hyperparameters in markov random field models. *Journal of Physics A: Mathematical and Theoretical*, Vol. 47, No. 4, p. 045001, 2014.

[21] Jonathan Michael Pryce and AD Bruce. Statistical mechanics of image restoration. *Journal of Physics A: Mathematical and General*, Vol. 28, No. 3, p. 511, 1995.

[22] Kazuyuki Tanaka. Statistical-mechanical approach to image processing. *Journal of Physics A: Mathematical and General*, Vol. 35, No. 37, p. R81, 2002.

[23] Kazuyuki Tanaka and DM 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.

[24] Hirotaka Sakamoto, Yoshinori Nakanishi-Ohno, and Masato 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.

[25] Hidetoshi Nishimori and KY Michael Wong. Statistical mechanics of image restoration and error-correcting codes. *Physical Review E*, Vol. 60, No. 1, p. 132, 1999.

[26] 西森秀稔. 有限温度の画像修復 (多体問題としての情報処理-統計力学と情報科学の接点-, 研究会報告). 2000.

[27] Kazuyuki Tanaka and Tsuyoshi Horiguchi. Solvable markov random field model in color image restoration. *Physical Review E*, Vol. 65, No. 4, p. 046142, 2002.

[28] Hayaru Shouno and Masato Okada. Bayesian image restoration for medical images using radon transform. *Journal of the Physical Society of Japan*, Vol. 79, No. 7, p. 074004, 2010.

[29] Hayaru Shouno and Masato Okada. Poisson observed image restoration using a latent variational approximation with gaussian mrf. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)*, p. 197. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013.

[30] 田中和之. 確率モデルによる画像処理技術入門. 森北出版, 2006.

[31] 汪金吉, 田栗正章, 手塚集, 樺島祥介, 上田修功. 統計計算 i 確率計算の楽しい手法, 2003.

[32] 伊庭幸人, 種村正美, 大森裕浩, 和合肇, 佐藤整尚, 高橋明彦. 計算統計 II マルコフ連鎖モンテカルロ法とその周辺. 岩波書店, 2005.

[33] B. N. Brockhouse and P. K. Iyengar. Normal vibrations of germanium by neutron spectrometry. *Phys. Rev.*, Vol. 108, pp. 894–895, 1957.

[34] B. N. Brockhouse and P. K. Iyengar. Normal modes of germanium by neutron spectrometry. *Phys. Rev.*, Vol. 111, pp. 747–75, 1958.

[35] T.G. Perring, A.D. Taylor, R. Osborn, D.McK. Paul, A.T. Boothroyd, and G. Aeppli. Maps: A chopper spectrometer to measure high energy magnetic excitations in single crystals. *Proc. 12th Meeting of the Collaboration on Advanced Neutron Sources (ICANS-XII)*, pp. I–60–I–7, 1994.

[36] R.I. Bewley, R.S. Eccleston, K.A. McEwen, S.M. Hayden, M.T. Dove, S.M. Bennington, J.R. Treadgold, and R.L.S. Coleman. Merlin, a new high count rate spectrometer at isis. *Physica B*, Vol. 385–386, pp. 1029–103, 2006.

[37] B. Winn, U. Filges, V.O. Garlea, M. Graves-Brook, C. Jiang M. Hagen, M. Kenzelmann, L. Passell, S.M. Shapiro, X. Tong, and I. Zaliznyak. Recent progress on hyspec, and its polarization analysis capabilities. *EPJ Web of Conferences*, Vol. 83, p. 0301, 2015.

[38] D.L. Abernathy, M.B. Stone, M.J. Loguillo, M.S. Lucas, O. Delaire, X. Tang, J.Y.Y. Lin, and B. Fultz. Design and operation of the wide angular-range chopper spectrometer arcs at the spallation neutron source. *Rev. Sci. Instruments*, Vol. 83, p. 01511, 2012.

[39] Mitsutaka Nakamura, Ryoichi Kajimoto, Yasuhiro Inamura, Fumio Mizuno, Masaki Fujita, Tetsuya Yokoo, and Masatoshi 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.

[40] Kenji Nakajima, Seiko Ohira-Kawamura, Tatsuya Kikuchi, Mitsutaka Nakamura, Ryoichi Kajimoto, Yasuhiro Inamura, Nobuaki Takahashi, Kazuya Aizawa, Kentaro Suzuya, Kaoru 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.

[41] Shinichi Itoh, Tetsuya Yokoo, Setsuo Satoh, Shin-ichiro Yano, Daichi Kawana, Junichi Suzuki, and Taku 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.

[42] Satoru Tokuda, Kenji Nagata, and Masato 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.

[43] Koji Hukushima and Koji 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.

[44] 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.

[45] H Nishimori. Finite-temperature image restoration. *Bussei Kenkyu*, Vol. 73, No. 5, pp. 850–857, 2000.

[46] Alan J Laub. *Matrix analysis for scientists and engineers*, Vol. 91. Siam, 2005.

[47] Shun Katakami, Hirotaka Sakamoto, Shin Murata, and Masato Okada. Gaussian markov random field model without boundary conditions. *Journal of the Physical Society of Japan*, Vol. 86, No. 6, p. 064801, 2017.

[48] Paul Levy. A special problem of brownian motion, and a general theory of gaussian random functions. In *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*,

Vol. 2, pp. 133–175, 1956.

[49] John Woods. Two-dimensional discrete markovian fields. *IEEE Transactions on Information Theory*, Vol. 18, No. 2, pp. 232–240, 1972.

[50] Tatsu Kuwatani, Kenji Nagata, Masato Okada, and Mitsuhiro Toriumi. Precise estimation of pressure–temperature paths from zoned minerals using markov random field modeling: theory and synthetic inversion. *Contributions to Mineralogy and Petrology*, Vol. 163, No. 3, pp. 547–562, 2012.

[51] Jun Kitazono, Toshiaki Omori, Toru Aonishi, and Masato Okada. Estimating membrane resistance over dendrite using markov random field. *IPSJ Online Transactions*, Vol. 5, pp. 186–191, 2012.

[52] Martin Hassner and Jack Sklansky. The use of markov random fields as models of texture. *Computer Graphics and Image Processing*, Vol. 12, No. 4, pp. 357–370, 1980.

[53] Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*, Vol. 111. Clarendon Press, 2001.

[54] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pp. 2951–2959, 2012.

[55] Christopher KI Williams and Carl Edward Rasmussen. Gaussian processes for regression. In *Advances in neural information processing systems*, pp. 514–520, 1996.

[56] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, Vol. 1. MIT press Cambridge, 2006.

[57] Harold J Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, Vol. 86, No. 1, pp. 97–106, 1964.

[58] J. Mockus, V. Tiesis, and A. Zilinskas. *Towards global optimisation*, Vol. 2. North-Holland Amsterdam, 1978.

[59] Emmanuel Vazquez and Julien Bect. Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and inference*, Vol. 140, No. 11, pp. 3088–3095, 2010.

[60] Hiroshi Akima. A method of bivariate interpolation and smooth surface fitting based on local procedures. *Communications of the ACM*, Vol. 17, No. 1, pp. 18–20, 1974.

[61] Shun Katakami, Hirotaka Sakamoto, and Masato Okada. Bayesian hyperparameter estimation using gaussian process and bayesian optimization. *Journal of the Physical Society of Japan*, Vol. 88, No. 7, p. 074001, 2019.

[62] Ryo Tamura and Koji Hukushima. Method for estimating spin-spin interactions from magnetization curves. *Physical Review B*, Vol. 95, No. 6, p. 064407, 2017.

[63] Ryo Tamura and Koji Hukushima. Bayesian optimization for computationally extensive probability distributions. *PloS one*, Vol. 13, No. 3, p. e0193785, 2018.