Method for estimating spin-spin interactions from magnetization curves

Ryo Tamura[1,2,*] and Koji Hukushima[2,3,†]

[1]*Computational Materials Science Unit, National Institute for Materials Science,*
1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
[2]*Center for Materials Research by Information Integration, National Institute for Materials Science,*
1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan
[3]*Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo,*
Komaba, Meguro, Tokyo 153-8902, Japan

We develop a method to estimate the spin-spin interactions in the Hamiltonian from the observed magnetization curve by machine learning based on Bayesian inference. In our method, plausible spin-spin interactions are determined by maximizing the posterior distribution, which is the conditional probability of the spin-spin interactions in the Hamiltonian for a given magnetization curve with observation noise. The conditional probability is obtained with the Markov chain Monte Carlo simulations combined with an exchange Monte Carlo method. The efficiency of our method is tested using synthetic magnetization curve data, and the results show that spin-spin interactions are estimated with a high accuracy. In particular, the relevant terms of the spin-spin interactions are successfully selected from the redundant interaction candidates by the l_1 regularization in the prior distribution.

I. INTRODUCTION

The importance of data-driven techniques using machine learning [1,2] is recognized in both academic and industrial fields. Machine learning is generally a statistical tool used to extract the inherent structure from a finite set of observed data. In condensed-matter physics, machine learning techniques have been used for, for example, interpolation of density-functional theory (DFT) [3–6] and dynamical-mean field theory [7] calculations and for model selection of strongly correlated systems [8] and the Ginzburg-Landau equation [9]. In this paper, we apply Bayesian inference from the field of machine learning [10–12] to estimate the spin-spin interactions in the Hamiltonian given an observed magnetization curve as an input (Fig. 1).

Magnetization is the average of the magnetic moments normalized by the amplitude of the magnetic moment, which is induced in a magnetic material by a magnetic field and is a fundamental physical quantity in both experiments and theories of magnetism. The magnetic-field dependence of the magnetization is called the magnetization curve [13,14] and has been measured in many magnetic materials using devices such as a superconducting quantum interference device (SQUID) in the laboratory. In general, the magnetization curve for a fixed temperature increases monotonically as a nonlinear function of the magnetic field. Depending on the magnetic material, various characteristic features, such as plateaus [15–21], ramps [22–24], and sharp increases [25], may be observed in the magnetization curve. Consequently, the magnetization curve is an important quantity for understanding the magnetic properties of individual magnetic materials.

To understand the essence of the properties in magnetic materials, an effective Hamiltonian is often derived in materials science. Many methods have been proposed to determine Hamiltonians, and they can be divided roughly into two groups: one in which the model parameter values in the Hamiltonian are calculated by *ab initio* electronic structure calculations when basic data of the magnetic materials are inputted [26–29] and one in which the model parameter values are evaluated by fitting the physical quantities observed in a target magnetic material [30–34].

The aim of this paper is to establish a Hamiltonian estimation method which can be classified as the latter group where the model parameters in the Hamiltonian of a target material are predicted from experimental data (Fig. 1). In particular, we propose a method which selects relevant terms for the model parameters from redundant model parameter candidates by machine learning based on the l_1 regularization. The framework of this method can be adopted for any measurement data, but herein we focus on the case where an observed magnetization curve is used as the input data to estimate the spin-spin interactions.

The rest of the paper is organized as follows. In Sec. II, we introduce a method to estimate the model parameters in the Hamiltonian from the observed magnetization curve. We first present the forward modeling for an observed magnetization
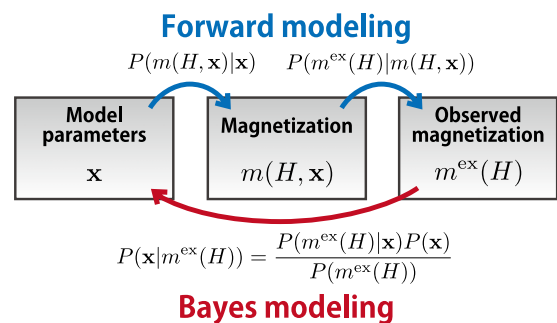

FIG. 1. Schematic research flows for forward modeling and Bayes modeling. $P(A)$ is the probability distribution of A, and $P(A|B)$ is the conditional probability of A given B.

*tamura.ryo@nims.go.jp
†hukusima@phys.c.u-tokyo.ac.jp

curve. By using Bayesian inference, we construct the posterior distribution, that is, the conditional probability of the model parameters in the Hamiltonian given an observed magnetization curve (the Bayes modeling). In this case, the plausible model parameters maximize the posterior distribution. Next, we explain the simulation methods, which employ the Markov chain Monte Carlo (MCMC) method and exchange Monte Carlo method, to analyze the posterior distribution. In the last part of Sec. II, we describe cross validation to avoid overfitting. Cross validation makes it possible to calculate prediction errors and allows a hyperparameter in the prior distribution to be determined. To validate our estimation method, we determine the spin-spin interactions from a synthetic magnetization curve in Sec. III. We prepare an input observed magnetization curve using a theoretical model with fixed spin-spin interactions to consider the case where the solution of the estimation problem is known in advance. We show that the spin-spin interactions can be accurately estimated from the observed magnetization curve using our method. Section IV presents the discussion and summary.

II. METHOD FOR ESTIMATING MODEL PARAMETERS

This section presents our estimation method for model parameters in the Hamiltonian determined from the observed magnetization curve. Using Bayesian inference, the conditional probability of the spin-spin interactions is obtained for a given observed magnetization curve. Furthermore, the simulation methods used to analyze the conditional probability and the cross validation to avoid overfitting are introduced.

A. Bayesian inference for the magnetization curve

1. Forward modeling

Let us assume a class of the Hamiltonian $\mathcal{H}(\{\mathbf{s}_i\}; \mathbf{x})$ for a magnetic system with a set of model parameters denoted as $\mathbf{x} = (x_1, \ldots, x_K)$, where K is the number of model parameters. Here, $\mathbf{s}_i$ is the spin variable at site i with the absolute value $|\mathbf{s}|$, and $\{\mathbf{s}_i\}$ denotes a set of spin configurations. In statistical physics, the magnetization under a magnetic field H at a temperature T for a given Hamiltonian $\mathcal{H}(\{\mathbf{s}_i\}; \mathbf{x})$ is calculated as

$$m(H, \mathbf{x}) = \left| \frac{1}{N|\mathbf{s}|} \sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H, \mathbf{x}} \right|, \quad (1)$$

$$\langle \mathbf{s}_i \rangle_{H, \mathbf{x}} = \frac{\mathrm{Tr}\, \mathbf{s}_i \exp\left[-\left(\mathcal{H}(\{\mathbf{s}_i\}; \mathbf{x}) - H \sum_{i=1}^{N} s_i^z\right)/T\right]}{\mathrm{Tr} \exp\left[-\left(\mathcal{H}(\{\mathbf{s}_i\}; \mathbf{x}) - H \sum_{i=1}^{N} s_i^z\right)/T\right]}, \quad (2)$$

where s_i^z is the z component of $\mathbf{s}_i$ and N is the number of spins. Throughout the paper, the g factor, the Bohr magneton, and the Boltzmann constant are set to unity. In the forward modeling, the conditional probability of the observed magnetization in the given model parameters is examined. Because the magnetization $m(H, \mathbf{x})$ is uniquely obtained in the thermodynamic limit by statistical physics, the conditional probability of the magnetization $m(H, \mathbf{x})$ given $\mathbf{x}$ is expressed as

$$P(m(H, \mathbf{x})|\mathbf{x}) = \delta\left(m(H, \mathbf{x}) - \left| \frac{1}{N|\mathbf{s}|} \sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H, \mathbf{x}} \right|\right), \quad (3)$$

where $\delta(\cdots)$ is the Dirac delta function and $P(A|B)$ expresses the conditional probability of A given B.

The measurements in an experiment always have some uncertainty. Taking this uncertainty into account as the observation noise ϵ, the observed magnetization in experiments $m^{\mathrm{ex}}(H)$ is expressed as

$$m^{\mathrm{ex}}(H) = m(H, \mathbf{x}) + \epsilon. \quad (4)$$

At this point, we assume that the observation noise follows a Gaussian distribution with a mean of zero and a standard deviation of σ:

$$P(\epsilon) \propto \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right). \quad (5)$$

The conditional probability of $m^{\mathrm{ex}}(H)$ for a given $m(H, \mathbf{x})$ is then

$$P(m^{\mathrm{ex}}(H)|m(H, \mathbf{x})) \propto \exp\left\{-\frac{1}{2\sigma^2}[m^{\mathrm{ex}}(H) - m(H, \mathbf{x})]^2\right\}. \quad (6)$$

Thus, in the forward modeling, the conditional probability of the observed magnetization $m^{\mathrm{ex}}(H)$ given the model parameters $\mathbf{x}$ is written as

$$P(m^{\mathrm{ex}}(H)|\mathbf{x})$$
$$\propto \int dm(H, \mathbf{x}) P(m^{\mathrm{ex}}(H)|m(H, \mathbf{x})) P(m(H, \mathbf{x})|\mathbf{x}) \quad (7)$$
$$\propto \exp\left[-\frac{1}{2\sigma^2}\left(m^{\mathrm{ex}}(H) - \left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle \mathbf{s}_i \rangle_{H, \mathbf{x}}\right|\right)^2\right]. \quad (8)$$

By using this formula, the probabilistic prediction of the magnetization observed in the experiments is obtained as the conditional probability for the given model parameters. This can be compared with the experimental data. Figure 1 shows the flow of the forward modeling.

2. Bayes modeling

The Bayes modeling provides a general framework to estimate the model parameters $\mathbf{x}$ from the observed magnetizations. In this work, we consider the case where the magnetization under various magnetic fields is observed experimentally; that is, the observed magnetization curve is expressed as $\{m^{\mathrm{ex}}(H_l)\}_{l \in D}$ for a series of magnetic fields $\{H_l\}_{l \in D}$, where the data are indexed by $D = (1, \ldots, L)$ and L is the number of observed magnetizations. By using Bayes's theorem, the posterior distribution, which is the conditional probability of the model parameters $\mathbf{x}$ given the observed magnetization curve $\{m^{\mathrm{ex}}(H_l)\}_{l \in D}$, is expressed as

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l \in D}) = \frac{P(\{m^{\mathrm{ex}}(H_l)\}_{l \in D}|\mathbf{x})P(\mathbf{x})}{P(\{m^{\mathrm{ex}}(H_l)\}_{l \in D})}, \quad (9)$$

where $P(\mathbf{x})$ is the prior distribution of $\mathbf{x}$. On the right-hand side, the denominator is independent of the model parameters. In contrast, the prior distribution $P(\mathbf{x})$ explicitly contributes to the estimation in the posterior distribution. For example, in the case without *a priori* knowledge of the model parameters, a uniform distribution of $P(\mathbf{x})$ may be used. In this work, we assume that the number of important model parameters

is small, indicating that the relevant parameters are sparse in the model Hamiltonian. In this situation, the prior distribution $P(\mathbf{x})$ based on the l_1 regularization is frequently represented as

$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_{k=1}^{K} |x_k|\right), \tag{10}$$

where λ is called the hyperparameter given before the analysis. The inference scheme using the prior distribution is called the least absolute shrinkage and selection operator (LASSO) [35], and the value of λ controls the strength of the regularization.

In addition, we assume that $m^{\mathrm{ex}}(H_l)$ is independently obtained from different magnetic fields. That is, the previous measurement in a magnetization process does not affect the present measurement, which gives

$$P(\{m^{\mathrm{ex}}(H_l)\}_{l\in D}|\mathbf{x}) = \prod_{l=1}^{L} P(m^{\mathrm{ex}}(H_l)|\mathbf{x}). \tag{11}$$

Combined with the prior distribution of Eq. (10), the posterior distribution of $\mathbf{x}$ given $\{m^{\mathrm{ex}}(H_l)\}_{l\in D}$ based on Bayes's theorem is expressed as

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D})$$

$$\propto \prod_{l=1}^{L} P(m^{\mathrm{ex}}(H_l)|\mathbf{x}) P(\mathbf{x}) \tag{12}$$

$$\propto \exp\left[-\frac{1}{2\sigma^2} \sum_{l=1}^{L} \left(m^{\mathrm{ex}}(H_l) - \left|\frac{1}{N|\mathbf{s}|} \sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H_l,\mathbf{x}}\right|\right)^2 \right.$$
$$\left. - \lambda \sum_{k=1}^{K} |x_k|\right]. \tag{13}$$

From the view point of the maximum *a posteriori* (MAP) estimation, the plausible model parameters $\mathbf{x}^*$ are obtained by the maximizer of Eq. (13). In other words, we search the model parameters so that Eq. (13) is maximized in the Bayes modeling, where σ, λ, and K are the previously fixed hyperparameters. Although the posterior distribution is constructed using the magnetization curve as the input data, it should be emphasized that the framework of our estimation method can be adopted for any measurement data.

B. Simulation method

In this study, we use the MCMC method and the exchange Monte Carlo method to analyze the posterior distribution $P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D})$. This combination significantly contributes to finding the global maximum of the posterior distribution in systems where many local maxima exist. When the energy function $E(\mathbf{x}|\sigma,\lambda,K)$ is defined as

$$E(\mathbf{x}|\sigma,\lambda,K) := \frac{1}{2\sigma^2} \sum_{l=1}^{L} \left(m^{\mathrm{ex}}(H_l) - \left|\frac{1}{N|\mathbf{s}|} \sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H_l,\mathbf{x}}\right|\right)^2$$
$$+ \lambda \sum_{k=1}^{K} |x_k|, \tag{14}$$

the posterior distribution can be written as

$$P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D}) \propto \exp\left[-\frac{1}{T_{\mathrm{R}}} E(\mathbf{x}|\sigma,\lambda,K)\right], \tag{15}$$

where $T_{\mathrm{R}} = 1$ in the case of Eq. (13). This distribution formally has the same form as the Boltzmann distribution by regarding $E(\mathbf{x}|\sigma,\lambda,K)$ and T_{R} as the energy of the system and the virtual temperature, respectively. Thus, a statistical-mechanical approach is promising for sampling from the distribution [36]. It should be noted that the dynamical variables in this scheme are the model parameters in the Hamiltonian, which is unusual in the study of magnetism as well as in statistical physics.

In our MCMC simulations, we employ the Metropolis-type transition probability from $\mathbf{x}$ to $\mathbf{x}'$:

$$w(\mathbf{x}'|\mathbf{x}) = \min\left\{1, \exp\left[-\frac{1}{T_{\mathrm{R}}} \Delta E(\mathbf{x}',\mathbf{x})\right]\right\}, \tag{16}$$

$$\Delta E(\mathbf{x}',\mathbf{x}) := E(\mathbf{x}'|\sigma,\lambda,K) - E(\mathbf{x}|\sigma,\lambda,K). \tag{17}$$

By sampling, we investigate the properties of $P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D})$ and search for the model parameters $\mathbf{x}^*$ that maximize $P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D})$.

In addition, we use the exchange Monte Carlo method [37] to enhance the sampling efficiency and to find the global maximum of $P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D})$ in a system where many local maxima exist. In the method, several MCMC simulations using Eq. (16) with different virtual temperatures are performed in parallel, and the model parameters $\mathbf{x}$ are exchanged between the different MCMC simulations with an exchange probability expressed as

$$w_{\mathrm{ex}}(\mathbf{x}^j|\mathbf{x}^i) = \min\left\{1, \exp\left[\left(\frac{1}{T_{\mathrm{R}}^j} - \frac{1}{T_{\mathrm{R}}^i}\right)\Delta E(\mathbf{x}^j,\mathbf{x}^i)\right]\right\}, \tag{18}$$

$$\Delta E(\mathbf{x}^j,\mathbf{x}^i) := E(\mathbf{x}^j|\sigma,\lambda,K) - E(\mathbf{x}^i|\sigma,\lambda,K), \tag{19}$$

where $\mathbf{x}^i$ and T_{R}^i indicate the model parameters and the virtual temperature in the ith MCMC simulation, respectively. By using this technique, we can find the global maximum of $P(\mathbf{x}|\{m^{\mathrm{ex}}(H_l)\}_{l\in D})$ and obtain the plausible model parameters $\mathbf{x}^*$ with a high probability. Without a loss in generality, the units of the parameter λ and the virtual temperature T_{R} can be taken as σ^2. Thus, we set $\sigma = 1$ throughout the paper.

C. Cross validation

In general, the observed magnetization curve is well fitted when the number of model parameters is sufficiently large. Often, the estimated model in machine learning cannot properly predict unknown data not used in the fitting, even when the observed data are well fitted by the estimated model. This is known as the overfitting problem.

To determine the hyperparameter λ while preventing overfitting, we perform cross validation in which the hyperparameter is chosen to minimize the prediction error. Consider the case where the value of K is fixed. In cross validation, the data set D is divided into S groups as $D_s = ((s-1)L/S + 1, (s-1)L/S + 2, \ldots, sL/S)$, with $s = 1, \ldots, S$. One of the S groups is regarded as the test data, while the remaining $S-1$
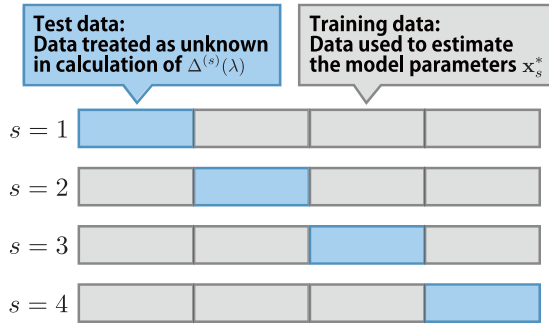
FIG. 2. Cross validation for $S = 4$. We evaluate $\Delta^{(s)}(\lambda)$ in each case. The average of $\Delta^{(s)}(\lambda)$ over the four cases is used to estimate λ^*.

groups are used as the training data. Figure 2 shows an example for $S = 4$. The number of pieces of test data and that of the training data are given by $L_{\text{te}} := N/S$ and $L_{\text{tr}} := N(S - 1)/S$, respectively. For each data subset $G_s = D \setminus D_s$, the plausible model parameters $\mathbf{x}_s^*$ are determined by the MAP estimation according to λ. The mean-square deviation between the test data and the estimated magnetization curve when using $\mathbf{x}_s^*$ for various λ is evaluated as

$$\Delta^{(s)}(\lambda) := \frac{1}{L_{\text{te}}} \sum_{l_{\text{te}} \in D_s} \left(m^{\text{ex}}(H_{l_{\text{te}}}) - \left| \frac{1}{N|\mathbf{s}|} \sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H_{l_{\text{te}}}, \mathbf{x}_s^*} \right| \right)^2.$$
(20)

In the cross validation, $\Delta^{(s)}(\lambda)$ represents the prediction error when the test data D_s are treated as unknown data. Using the average of $\Delta^{(s)}(\lambda)$ over different data sets, we evaluate the optimal value of λ^*, where the averaged prediction error takes a minimum as a function of λ. Furthermore, the plausible model parameters $\mathbf{x}^*$ are obtained such that $\Delta^{(s)}(\lambda)$ is the smallest among S selections for the test data for λ^*. It should be noted that λ^* depends on the input magnetization curve by performing cross validation.

III. DEMONSTRATION OF ESTIMATION OF SPIN-SPIN INTERACTIONS

In this section, we confirm that the estimation method proposed in Sec. II can determine the spin-spin interactions in the Hamiltonian. The inputted observed magnetization curve was a zero-temperature magnetization curve calculated from a theoretical model. The results confirm that our estimation method is highly accurate and the relevant terms of the spin-spin interactions are successfully selected from the redundant interaction candidates by the l_1 regularization.

A. Model and observed magnetization curve

We examine the efficiency of our proposed method by synthetic data for the magnetization curve. To generate the input data, we used the classical Heisenberg model with bilinear and biquadratic interactions, which has the following Hamiltonian:

$$\mathcal{H} = \sum_{i,j} J_{ij} [\mathbf{s}_i \cdot \mathbf{s}_j - b_{ij}(\mathbf{s}_i \cdot \mathbf{s}_j)^2] - H \sum_i s_i^z. \quad (21)$$
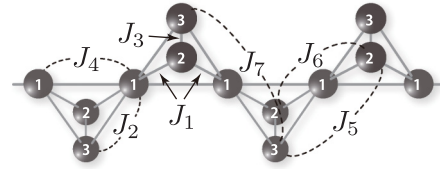

FIG. 3. Lattice structure and types of spin-spin interactions in the model Hamiltonian used to generate the synthetic magnetization curve. Numbers indicate the types of lattice points.

Here, $b_{ij} = b J_{ij}$, and $\mathbf{s}_i = (s_i^x, s_i^y, s_i^z)$ denotes the three-component vector spin with a length $|\mathbf{s}| = 1/2$. The value of b represents the amplitude of the biquadratic interactions. Depending on the lattice and sets of spin-spin interactions, this model exhibits fruitful magnetization curves [38–40]. Here, the inputted observed magnetization curve is given by the model on a tetrahedral chain at a zero temperature. The lattice structure is shown in Fig. 3. For simplicity, we consider seven types of spin-spin interactions J_k ($k = 1, \ldots, 7$). Thus, the model parameter can be expressed by $\mathbf{x} = (J_1, J_2, J_3, J_4, J_5, J_6, J_7, b)$ with the dimension $K = 8$.

Figure 4(a) shows the zero-temperature magnetization process $\{m(H_l, \mathbf{x}')\}_{l \in D}$ with the number of data points $L = 160$, which is obtained by $\mathbf{x}' = (1, 4, 5, 6, 0, 0, 0, 0.1)$. The magnetization curve exhibits a huge 2/3 magnetization plateau. Furthermore, we assume that the observed magnetization curve $\{m^{\text{ex}}(H_l)\}_{l \in D}$ can be represented by adding Gaussian noise with a mean of zero and a standard deviation of $\sigma = 0.004$. Figure 4(b) shows the synthesized magnetization curve $\{m^{\text{ex}}(H_l)\}_{l \in D}$.

For the magnetization curve shown in Fig. 4(b), we estimate model parameters $\mathbf{x} = (J_1, J_2, J_3, J_4, J_5, J_6, J_7, b)$ using our proposed method. For the cross validation, we randomly divided the observed magnetization curve into four sets of data points ($S = 4$). Each set of training data contains 120 data points ($L_{\text{te}} = 120$) in the magnetization curve.

B. Estimation of spin-spin interactions

This section explains the details of the numerical simulations to evaluate the posterior distribution $P(\mathbf{x} | \{m^{\text{ex}}(H_l)\}_{l \in G_s})$,
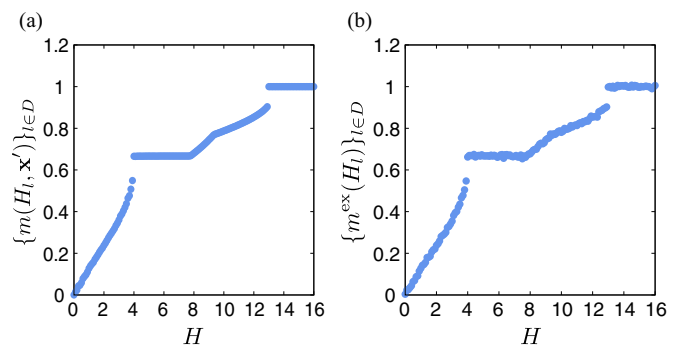

FIG. 4. (a) Magnetization curve $\{m(H_l, \mathbf{x}')\}_{l \in D}$ for $L = 160$ at the zero temperature in the theoretical model defined by Eq. (21) with model parameter $\mathbf{x}' = (1, 4, 5, 6, 0, 0, 0, 0.1)$. (b) Observed magnetization curve $\{m^{\text{ex}}(H_l)\}_{l \in D}$ represented by $\{m(H_l, \mathbf{x}')\}_{l \in D}$ with added Gaussian noise of the mean zero and standard deviation of 0.004.

TABLE I. Number of each type of interaction J_k per tetrahedron, distance between spins connected by each interaction J_k within the nearest-neighbor lattice spacing, spin-spin interactions used in the input magnetization curve, and spin-spin interactions estimated using our estimation method.

	J_1	J_2	J_3	J_4	J_5	J_6	J_7	b
Number of interactions n_k	2	2	1	1	2	1	1	
Distance between spins r_k	1	1	1	1	$\sqrt{3}$	2	2	
Input spin-spin interactions $\mathbf{x}'$	1.000	4.000	5.000	6.000	0.000	0.000	0.000	0.100
Estimated spin-spin interactions $\mathbf{x}^*$	1.074	3.850	5.012	6.051	0.011	-0.051	0.002	0.102

with $s = 1, \ldots, 4$. Our numerical method consists of double-loop calculations, which include a calculation of the magnetization curve for a given set of model parameters and the sampling of the model parameters. We used the Hamiltonian defined by Eq. (21) with periodic boundary conditions where the total number of spins was six as a theoretical model to obtain $\langle \mathbf{s}_i \rangle_{H_l, \mathbf{x}}$. The zero-temperature magnetization curve was obtained by the steepest descent method. This is the inner-loop calculation, which was performed in a few seconds. Monte Carlo simulations were employed to sample the model parameters $\mathbf{x}$ in the outer loop. Typically, the number of the Monte Carlo steps to update the model parameters was 10^4. Furthermore, we prepared 20 replicas with virtual temperatures T_R between 0.001 and 10 for the exchange Monte Carlo method.

We considered three types of prior distributions for the model parameters $\mathbf{x}$ based on the l_1 regularization:

$$\text{Type I}: \ P(\mathbf{x}) \propto \exp\left[-\lambda\left(\sum_{k=1}^{7} |J_k| + |b|\right)\right], \quad (22)$$

$$\text{Type II}: \ P(\mathbf{x}) \propto \exp\left[-\lambda\left(\sum_{k=1}^{7} |n_k J_k| + |b|\right)\right], \quad (23)$$

$$\text{Type III}: \ P(\mathbf{x}) \propto \exp\left[-\lambda\left(\sum_{k=1}^{7} |n_k r_k J_k| + |b|\right)\right], \quad (24)$$

where n_k is the number of each type of interaction J_k per tetrahedron and r_k is the distance between the spins connected by interaction J_k. Table I summarizes the values used in this demonstration. Type I is the general form of the l_1 regularization. Type II considers the number of interactions per tetrahedron. In the theoretical model, the number of interactions in the lattice depends on J_k. Thus, we constructed the l_1 regularization from the viewpoint of the sum of the absolute values of the spin-spin interactions (i.e., the number of interactions per tetrahedron was introduced in the prior distribution). Type III considers both the number and the distance of the interactions. In general, the amplitude of the spin-spin interactions decreases as the distance between spins increases. Therefore, we assumed that the effect of the long-range interactions on the physical properties was smaller than that of the short-range interactions. To add a penalty to the long-range interactions, we introduced the distance between spins in the prior distribution in type III.

Figure 5 shows the λ dependence of the estimated spin-spin interactions that maximize $P(\mathbf{x}|\{m^{\text{ex}}(H_l)\}_{l \in G_s})$, with $s = 1, \ldots, 4$. The results are the averages over the spin-spin interactions estimated by the four sets of training data $\mathbf{x}_s^*$, with $s = 1, \ldots, 4$, and the error bars calculated from the standard deviation. Large error bars indicate that the estimated spin-spin interactions differ across the four sets of training data. The dashed lines in Fig. 5 show the values of the spin-spin interactions used in the inputted magnetization curve, that is, the solution of the estimation problem. Between $\lambda = 0.002$ and 0.008 for type III, which is shown in the shaded area in Fig. 5(c), the estimated spin-spin interactions are roughly the same as the values used in the inputted magnetization curve. However, these figures do not conclusively show which estimated spin-spin interactions are plausible for the inputted magnetization curve.

Using cross validation, we calculated the prediction error $\Delta_{\text{av}}(\lambda)$, which is the average of $\Delta^{(s)}(\lambda)$, with $s = 1, \ldots, 4$, defined in Eq. (20). Figure 6 shows the λ dependence of $\Delta_{\text{av}}(\lambda)$ for each type of prior distribution. The minimum value of $\Delta_{\text{av}}(\lambda)$ occurs at $\lambda = 0.004$ in type III, which produces the highest prediction accuracy. We conclude that the estimated spin-spin interactions for $\lambda = 0.004$ in type III are the most likely spin-spin interactions for the inputted magnetization curve. Table I, which shows the estimated spin-spin interactions for $\lambda = 0.004$ in type III, where $\Delta^{(s)}(\lambda)$ is the smallest among $s = 1, \ldots, 4$, confirms that the estimated spin-spin interactions are almost the same as the values used in the inputted magnetization curve. This demonstrates that our proposed estimation method can estimate the spin-spin interactions with high accuracy from the observed magnetization curve. Furthermore, the estimated values of J_5, J_6, and J_7 are very close to zero, indicating that the important spin-spin interactions can be selected among the many candidate interactions in the Hamiltonian. In addition, Fig. 6 shows that overfitting occurs in the case of $\lambda = 0$ when using nonzero values for the additional interaction terms. Consequently, overfitting can be avoided by introducing a regularization term based on the l_1 regularization. Finally, Fig. 7 compares the estimated magnetization curve for $\lambda = 0.004$ in type III with the inputted magnetization curve.

The model parameters in the Hamiltonian can be estimated with high accuracy from the observed magnetization curve using the estimation method introduced in Sec. II. In addition, we confirm that the model parameters are well estimated for certain input magnetization curves with the Hamiltonian defined in Eq. (21).

IV. DISCUSSION AND SUMMARY

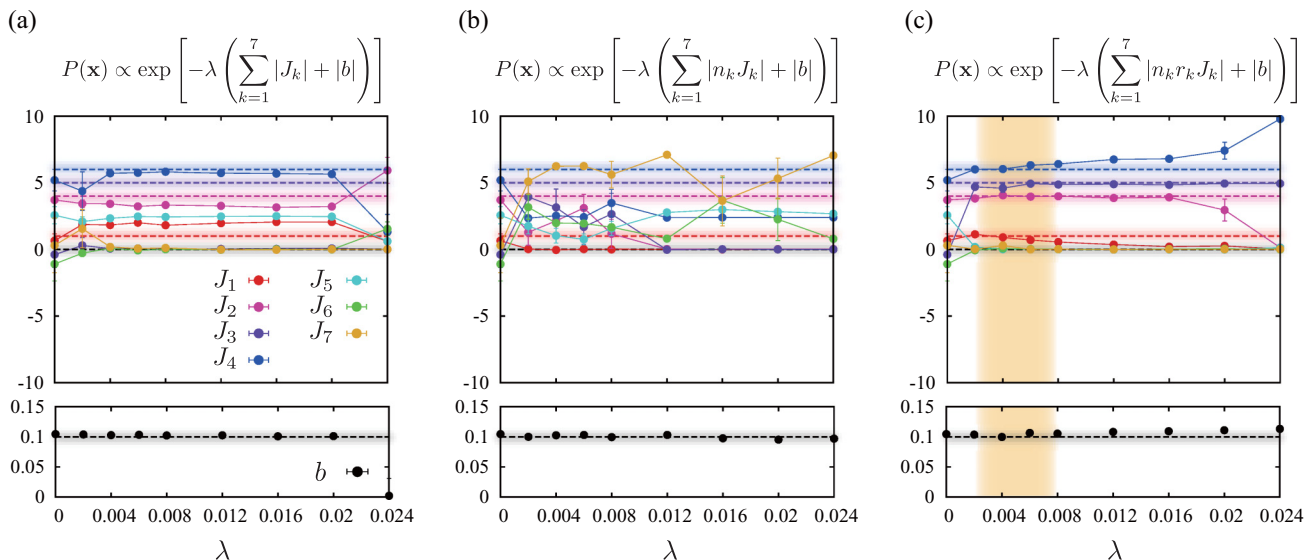We developed a method for estimating the model parameters in the Hamiltonian from the observed magnetization

FIG. 5. The λ dependence of estimated spin-spin interactions J_k and b using the prior distributions of (a) type I, (b) type II, and (c) type III, respectively. The results are the averages over the four sets of training data. Dashed lines indicate the input spin-spin interactions J_k and b. The shaded area in (c) indicates that the estimated spin-spin interactions are roughly the same as the values used in the inputted magnetization curve.

curve. Bayesian inference is initially used to determine the posterior distribution (i.e., the conditional probability of model parameters in the Hamiltonian for a given observed magnetization curve). We adopted the l_1 regularization as the prior distribution of the model parameters. In this scheme, the plausible model parameters are determined by maximizing the posterior distribution. To analyze the posterior distribution, we adopted the Markov chain Monte Carlo method and the exchange Monte Carlo method because this combination significantly contributes to finding the global maximum of a posterior distribution where many local maxima exist. Cross validation was introduced to avoid overfitting and to determine the hyperparameter in the prior distribution. Note that the framework of our estimation method can be used not only for magnetization curves but also for any measured data as the input data.
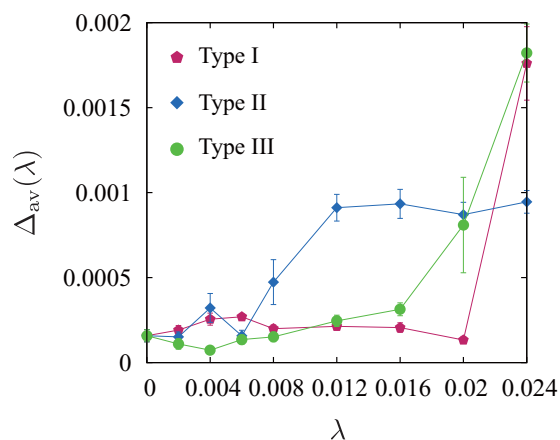
The effectiveness of our method is demonstrated by estimating the spin-spin interactions in the Hamiltonian from the observed magnetization curve. We used a magnetization curve synthesized by adding Gaussian noise prepared by a theoretical model with fixed spin-spin interactions. Specifically, we employed a classical Heisenberg model with bilinear and biquadratic interactions on a tetrahedral chain to consider the case where the solution of the estimation problem is known in advance. Furthermore, we introduced three types of prior distributions of the spin-spin interactions based on the l_1 regularization. These prior distributions include the hyperparameter λ, which controls the strength of the regularization. The spin-spin interactions are sampled from the posterior distribution for a given λ by numerical simulations based on the Monte Carlo method. We also determined the

FIG. 6. The λ dependence of the prediction error $\Delta_{av}(\lambda)$ obtained by cross validation for each type of prior distribution. $\Delta_{av}(\lambda)$ takes a minimum for $\lambda = 0.004$ in type III.
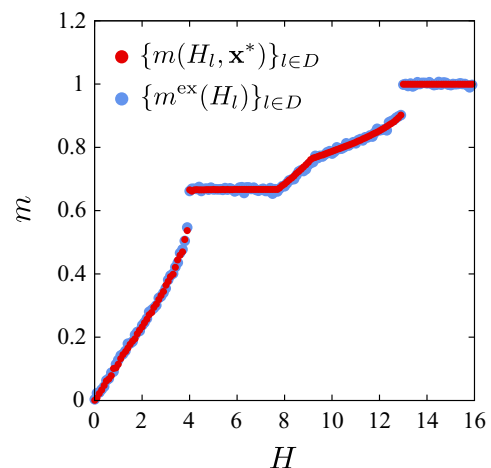

FIG. 7. Estimated magnetization curve $\{m(H_l, \mathbf{x}^*)\}_{l \in D}$ for $L = 160$ when $\lambda = 0.004$ in type III and inputted observed magnetization curve $\{m^{ex}(H_l)\}_{l \in D}$. Table I shows the estimated spin-spin interactions $\mathbf{x}^*$.

value of λ and the type of the prior distribution by minimizing the prediction error by cross validation. Spin-spin interactions consistent with the input data can be estimated. It is noted that our estimation method successfully selects the relevant spin-spin interactions from the many redundant interactions in the Hamiltonian. The prior distribution taking into account both the number and the distance of the interactions plays a relevant role in this study, confirming that our method can estimate the spin-spin interactions from an observed magnetization curve with high accuracy.

In general, it is difficult to obtain physical quantities measured by indirect measurements such as magnetic specific heat and magnetic entropy. However, our estimation method should predict these quantities once the Hamiltonian for a target magnetic material is obtained. Furthermore, we can directly obtain the predicted spin snapshots as well as predict the magnetic structure and structure factor. Thus, the information from the estimated Hamiltonian should provide preliminary information before using large facilities such as those for neutron diffraction measurements.

Our final objective is to assist with measurements by providing additional estimation information. However, this research shows only the case of inputting a synthetic magnetization curve obtained from a classical theoretical model, including bilinear and biquadratic interactions. To achieve our broader objectives, we must address two issues. First, we need to establish an estimation method that can be used for many types of magnetic materials described by classical or quantum Hamiltonians with many model parameters such as various types of spin-spin interactions, spin-lattice interactions, and spin anisotropy. In this case, a prior distribution, which estimates the correct model parameters, will depend on the types of target model parameters and the target material. Thus, it is important to develop a method that provides a useful prior distribution. Second, we need to improve our proposed

estimation method to produce a more robust Hamiltonian that can simultaneously explain various physical properties of a target material. To achieve this, the input data in the estimation method should include multiple data sets such as some magnetization curves at different temperatures and the magnetic field dependences of the magnetic specific heat. These are the future directions of our research and will be discussed elsewhere [41].

Finally, we emphasize that our proposed estimation method is a kind of physics-based machine learning technique because it employs a Hamiltonian based on physical laws. Our proposed method should form the basis for new applications of machine learning in physics and interdisciplinary physics. In particular, in materials informatics [42,43], which is an interdisciplinary field between materials science based on the physical law and informatics, our proposed estimation method should be a useful tool to establish important concepts.

[1] T. M. Mitchell, *Machine Learning* (McGraw-Hill, New York, 1997).

[2] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).

[3] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Phys. Rev. Lett. **108**, 253002 (2012).

[4] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, Phys. Rev. B **88**, 054104 (2013).

[5] F. Zhou, W. Nielson, Y. Xia, and V. Ozolins, Phys. Rev. Lett. **113**, 185501 (2014).

[6] Z. Li, J. R. Kermode, and A. DeVita, Phys. Rev. Lett. **114**, 096405 (2015).

[7] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).

[8] H. Takenaka, K. Nagata, T. Mizokawa, and M. Okada, J. Phys. Soc. Jpn. **83**, 124706 (2014).

[9] S. Taverniers, T. S. Haut, K. Barros, F. J. Alexander, and T. Lookman, Phys. Rev. E **92**, 053301 (2015).

[10] U. Toussaint, Rev. Mod. Phys. **83**, 943 (2011).

[11] G. D. Agostini, Rep. Prog. Phys. **66**, 1383 (2003).

[12] V. Dose, Rep. Prog. Phys. **66**, 1421 (2003).

[13] T. Nagamiya, K. Nagata, and Y. Kitano, Prog. Theor. Phys. **27**, 1253 (1962).

[14] J. Kanamori, Prog. Theor. Phys. **35**, 16 (1966).

[15] K. Hida, J. Phys. Soc. Jpn. **63**, 2359 (1994).

[16] M. Oshikawa, M. Yamanaka, and I. Affleck, Phys. Rev. Lett. **78**, 1984 (1997).

[17] K. Totsuka, Phys. Rev. B **57**, 3454 (1998).

[18] H. Kitazawa, H. Suzuki, H. Abe, J. Tang, and G. Kido, Phys. B (Amsterdam, Neth.) **259–261**, 890 (1999).

[19] K. Onizuka, H. Kageyama, Y. Narumi, K. Kindo, Y. Ueda, and T. Goto, J. Phys. Soc. Jpn. **69**, 1016 (2000).

[20] H. Kikuchi, Y. Fujii, M. Chiba, S. Mitsudo, T. Idehara, T. Tonegawa, K. Okamoto, T. Sakai, T. Kuwai, and H. Ohta, Phys. Rev. Lett. **94**, 227201 (2005).

[21] R. Ishii, S. Tanaka, K. Onuma, Y. Nambu, M. Tokunaga, T. Sakakibara, N. Kawashima, Y. Maeno, C. Broholm, D. P. Gautreaux, J. Y. Chan, and S. Nakatsuji, Europhys. Lett. **94**, 17001 (2011).

[22] V. Bindilatti, N. F. Oliveira, Jr., Y. Shapira, G. H. McCabe, M. T. Liu, S. Isber, S. Charar, M. Averous, E. J. McNiff, Jr., and Z. Golacki, Phys. Rev. B **53**, 5472 (1996).

[23] X. Gratens, V. Bindilatti, N. F. Oliveira, Jr., Y. Shapira, S. Foner, Z. Golacki, and T. E. Haas, Phys. Rev. B **69**, 125209 (2004).

[24] H. Nakano and T. Sakai, J. Phys. Soc. Jpn. **79**, 053707 (2010).

[25] K. Katsumata, H. Hori, T. Takeuchi, M. Date, A. Yamagishi, and J. P. Renard, Phys. Rev. Lett. **63**, 86 (1989).

[26] I. I. Mazin, Phys. Rev. B **76**, 140406(R) (2007).

[27] K. Nakamura, Y. Yoshimoto, R. Arita, S. Tsuneyuki, and M. Imada, Phys. Rev. B **77**, 195126 (2008).

[28] J. M. Pruneda, J. Íñiguez, E. Canadell, H. Kageyama, and M. Takano, Phys. Rev. B **78**, 115101 (2008).

[29] M. Hirayama, T. Miyake, and M. Imada, Phys. Rev. B **87**, 195144 (2013).

[30] K. Takubo, T. Mizokawa, J.-Y. Son, T. Nambu, S. Nakatsuji, and Y. Maeno, Phys. Rev. Lett. **99**, 037203 (2007).

[31] R. Tamura, N. Kawashima, T. Yamamoto, C. Tassel, and H. Kageyama, Phys. Rev. B **84**, 214408 (2011).

[32] M. Matsumoto, H. Kuroe, T. Sekine, and M. Hase, J. Phys. Soc. Jpn. **81**, 024711 (2012).

[33] R. A. DiStasio, Jr., É. Marcotte, R. Car, F. H. Stillinger, and S. Torquato, Phys. Rev. B **88**, 134104 (2013).

[34] É. Marcotte, R. A. DiStasio, Jr., F. H. Stillinger, and S. Torquato, Phys. Rev. B **88**, 184432 (2013).

[35] R. Tibshirani, J. R. Statist. Soc., Ser. B **58**, 267 (1996).

[36] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2009).

[37] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[38] K. Penc, N. Shannon, and H. Shiba, Phys. Rev. Lett. **93**, 197203 (2004).

[39] Y. Motome, K. Penc, and N. Shannon, J. Magn. Magn. Mater. **300**, 57 (2006).

[40] N. Shannon, H. Ueda, Y. Motome, K. Penc, H. Shiba, and H. Takagi, J. Phys. Conf. Ser. **51**, 31 (2006).

[41] R. Tamura and K. Hukushima (unpublished).

[42] K. Rajan, Mater. Today **8**, 38 (2005).

[43] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 011002 (2013).