Temperature-dependent performance of the erasure machine

Min Jae Jung and Beom Jun Kim*

Department of Physics, Sungkyunkwan University, Suwon 16419, Republic of Korea

Abstract

Interaction strengths in spin systems can be estimated from the maximum likelihood estimate (MLE) method based on information of spin configurations. The application of the standard MLE method has been known to face serious problems for large data set since it demands huge computation cost. We test the performance of the erasure machine, which has been recently proposed to overcome such drawbacks of MLE, for the Sherrington-Kirkpatrick (SK) model and two-dimensional Ising model. In this paper, we focus on how the performance of the erasure machine varies with the temperature and find that it exhibits the best performance not at criticality, but at somewhat higher temperature. We also attempt to explain such temperature-dependent performance of the erasure machine.

Keywords: Erasure Machine, Ising model, Regularization

*E-mail: beomjun@skku.edu

I. INTRODUCTION

Thanks to the advance of the technology for big data acquisition, it is becoming increasingly important to infer meaningful information within the data in a broad research area. Along this line of research, the inference of the interaction structure only from the observation of the spin configurations has been drawn much interest in statistical physics. In contrast to the original research framework of the Ising model, in which we have more interest in thermodynamic quantities of spin configurations for given interaction structure, the inverse Ising problem aims to infer the hidden structure of the spin interactions based on the observed information of spin configurations [1].

The inverse Ising problems can be applied in many areas to find the interaction structure and the strength in networks. The protein folding problem in biophysics, which has been computationally expensive, can be viewed as a physics problem of finding the low-energy structure of the 21-state Potts model composed of 20 different amino acids and the empty state [2,3]. For given sequence information of amino acids, the prediction of protein structure can be put in the framework of the inverse Ising problem. For neural networks, firing and resting states of a neuron can be represented as the Ising-like binary variable, and the firing information of neurons can be fed into the inverse Ising problem in which the interaction structure among neurons can be inferred [4, 5]. In the study of stock market, the inverse Ising problem is also applicable to find the underlying interaction structure, since the bullish and bearish states can be represented as the binary Ising spins [6,7].

Among various methods, the maximum likelihood estimate (MLE) [8–10], which has a broad and general applicability together with a mathematical rigor in predicting unknown parameters based on observed information, can be applied for the inverse Ising problem. For example, for a given sampled variables $\mathbf{x}$ obtained from any probability distribution $p(\mathbf{x}|\boldsymbol{\theta})$ with unknown parameters $\boldsymbol{\theta}$, we can infer $\boldsymbol{\theta}$ by maximizing a likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})$, i.e.,

$$\boldsymbol{\theta}^* = \text{argmax}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\mathbf{x}), \tag{1}$$

where the maximization is often achieved by the gradient ascent optimization. Despite of its theoretical clarity, the MLE has a critical drawback in practical applicability due to its infeasible computational cost for large systems.

Over the past several decades, other approaches have been developed for the inverse

Ising problem to overcome the limited applicability of MLE [11]. One of the most commonly used methods with much less computation cost is the the pseudolikelihood estimation method [12], in which the original likelihood function for N degrees of freedom in MLE is approximately substituted by N local pseudolikelihood functions. In Ref. [13], Aurell and Ekeberg have shown that the use of the ℓ_1 regularization can further improve the performance of pseudolikelihood estimation through the investigations on both the diluted Sherrington-Kirkpatrick (SK) model and the two-dimensional Ising model. The restricted Boltzmann machine (RBM) has also been applied by Cossu *et al.* at various temperatures to infer the coupling strengths in one- and two-dimensional Ising models [14]. Other approaches like the variational mean-field approach [15] and the screening method [**?**] have also been suggested.

Recently, a very efficient method called the erasure machine (EM) has been proposed [17]. In EM, reweighting of the distribution of the observed configuration has been shown to provide a reasonable approximation of the partition function, computation of which could be infeasible for large systems in the conventional MLE. Although the approximation scheme of the EM method can be interpreted as the high-temperature approximation in statistical physics, the effect of the temperature on the performance of the EM has not been investigated in detail, which composes our main research theme in the present paper.

In the present study, we use the Sherrington-Kirkpatrick (SK) model and the two-dimensional (2D) Ising model and investigate how the performance of the erasure machine (EM) changes with temperature. In our numerical investigations, we first generate interaction structures in the form of the weighted matrices for the SK model and the 2D Ising model as ground truths. The standard Metropolis Monte-Carlo simulations are then performed at a given temperature and the resulting spin configurations are fed into the EM formulation. The EM method produces the inferred interaction structures at various temperatures, which are to be compared with the actual underlying interaction structures.

The performance of the EM must be worse in two opposite limits: At very low temperatures, the spins are basically frozen regardless of the interaction structure and thus the EM method fails to detect interaction structure. On the other hand, in the high-temperature limit, the entropy effect prevails and the spins exhibit purely random fluctuation, again leading to the failure of the EM. Accordingly, it is very plausible that there must exist a temperature range in which the EM method can effectively reproduce the underlying

structure with minimum error between EM predicted interaction structure and the actual underlying structure. Without much consideration, one can expect that EM performance is maximized near the critical temperature of the the phase transition. We below test the expectation and find that it is not true and the EM performance is maximized at somewhat higher temperature than the critical temperature. We attempt to explain the origin of the discrepancy by the Monte-Carlo computation of thermodynamic quantities based on the EM inferred interaction structure.

II. SPIN MODELS AND MONTE-CARLO SIMULATIONS

To test the performance of the erasure machine, we use the Sherrington-Kirkpatrick (SK) model and the 2D Ising model. The Ising model on the 2D square lattice is described by the Hamiltonian

$$H = -J \sum_{\langle i,j \rangle} s_i s_j, \tag{2}$$

where the sum is over nearest-neighbor pairs without double counting in the square lattice with the periodic boundary condition, and s_i is the Ising spin variable with $s_i = \pm 1$. For the sake of simplicity we put $J = 1$, or the temperature of the system is measured in units of J/k_B with the Boltzmann constant k_B. The well-known ferromagnetic phase transition is detected by the conventional Ising order parameter defined by

$$m \equiv \left\langle \frac{1}{N} \sum_{i=1}^{N} s_i \right\rangle, \tag{3}$$

where $\langle \cdots \rangle$ is the sample average and N is the total number of spins. For the 2D square lattice, it is well-known that the critical temperature $T_c = 2/\ln(1+\sqrt{2}) \approx 2.269$ in our unit of the temperature.

For the 2D Ising model, we use the standard Metropolis Monte-Carlo (MC) simulation method. Starting from the random initial spin configuration at temperature $T = 5.0$ we thermalize for 10,000 MC steps to achieve equilibrium, and spin configurations are sampled and thermodynamic quantities are measured for next 10,000 MC steps. We then decrease the temperature by $\Delta T = 0.1$ and the final spin configuration at T is used as the initial spin configurations at $T - \Delta T$ for faster equilibration. The above procedure is repeated for 20 times and the average magnetization m shown in Fig. 1 is computed as a function of T.

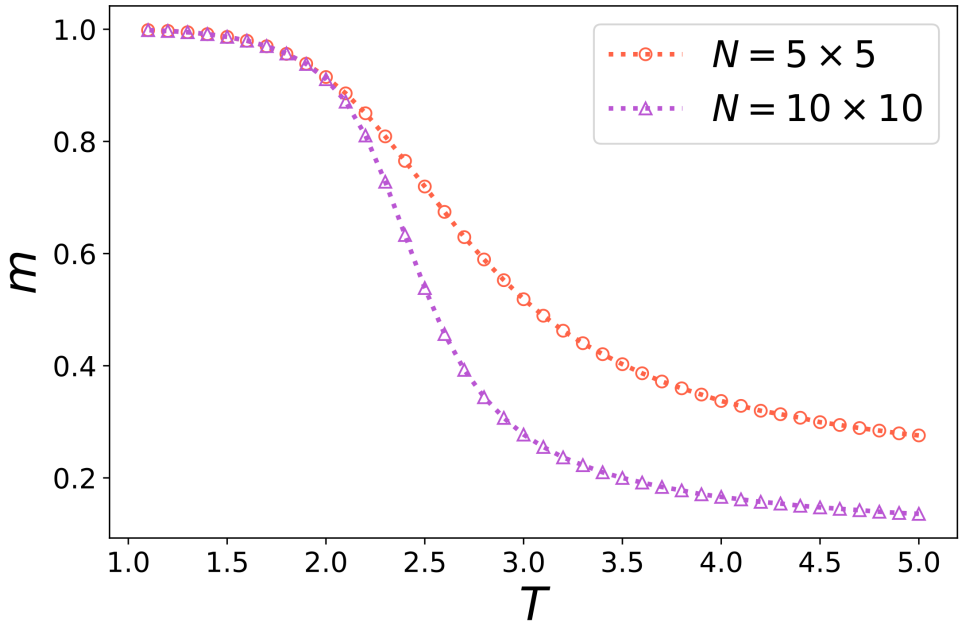

Fig. 1. The magnetization m of the 2D Ising model of the sizes $N = 5 \times 5$ and 10×10. The magnetization is shown to drop down rapidly near the true critical temperature $T_c \approx 2.269$ for both sizes.

Note that we do not aim to reproduce the well-known critical temperature and the critical exponents of the 2D Ising model, but we focus on producing spin configurations at various temperatures to be fed into the framework of the EM methodology to infer the underlying structure of the 2D square lattice.

The SK model has been suggested to explain the 'spin glass' phase in magnetic material with long-range interaction and its Hamiltonian in the absence of the external field is given by [18]

$$H = -\sum_{i<j} J_{ij} s_i s_j, \tag{4}$$

where $s_i = \pm 1$ $(i = 1, 2, \cdots, N)$ and the coupling strength J_{ij} between two spins i and j is produced from the Gaussian distribution $P(J_{ij}) \propto \exp\left[-(J_{ij} - J_0)^2/2J^2\right]$ with $J_0 = 0$ and $J = 1/\sqrt{N}$. The standard deviation J in the distribution function can measure the overall strength of the quenched random interaction J_{ij}, and the inverse square-root dependence on the number N of spins guarantees that the total energy of the SK model is an extensive quantity proportional to N. It has been well known that the SK model at $J_0 = 0$ has two phases: spin glass phase and the paramagnetic phase. The spin glass phase occurs below the

glass transition temperature T_g, which takes the value unity $T_g = 1$ for $J = 1/\sqrt{N}$. Below T_g the spin configurations are frozen in some orientation depending on their interaction structure. In other words, each spin has different orientation even when J_{ij} is fixed in the spin glass phase. Spin glass phase is different from the usual ferromagnetic phase since the conventional ferromagnetic order parameter is null.

The spin-glass transition can well be captured by the Edwards-Anderson (EA) order parameter [19], which utilizes the spin freezing in time. In detail, the spin configuration sampled at a particular time $\{s_i^{(1)}\}$ does not change over a long time later in the spin glass phase, and thus the spin configuration sampled at a different time $\{s_i^{(2)}\}$ will not be much different from $\{s_i^{(1)}\}$. The EA spin-glass order parameter is defined by

$$q \equiv \left[\left\langle \frac{1}{N} \sum_i s_i^{(1)} s_i^{(2)} \right\rangle \right], \tag{5}$$

where $\langle \cdots \rangle$ is the sample average for a given quenched disorder $\{J_{ij}\}$ and $[\cdots]$ is the average over different realizations of $\{J_{ij}\}$. We also measure the spin glass susceptibility in our Monte-Carlo simulation defined by

$$\chi_{SG} \equiv \left[\frac{1}{N} \sum_{i,j} (\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle)^2 \right], \tag{6}$$

which diverges at the spin-glass transition temperature [20].

For the SK model, we again use the Metropolis MC method with the similar annealing schedule for system sizes $N = 20, 40, 80, 160$, and 320: We start from random spin configurations at $T = 5.0$ and decrease temperature by $\Delta T = 0.1$ with 10,000 thermalization MC steps and 50,000 sampling MC steps. We repeat MC simulations for 50 different realizations of the quenched random interactions $\{J_{ij}\}$. The EA spin-glass order parameter q and the spin-glass susceptibility χ_{SG} are shown in Fig. 2. As N is increased, the transition temperature above which the spin-glass order parameter is close to null ($q \approx 0$) increases and the peak position of χ_{SG} also increases, in agreement of the known glass transition temperature $T_g = 1.0$ in thermodynamic limit [18].

III. ERASURE MACHINE

In order to initiate the computation machinery of the MLE formalism, we need to feed the observed information of spin configurations $D = \{\mathbf{s}^\mu\}$, where $\mu = 1, 2, \cdots, M$ with $M = |D|$,

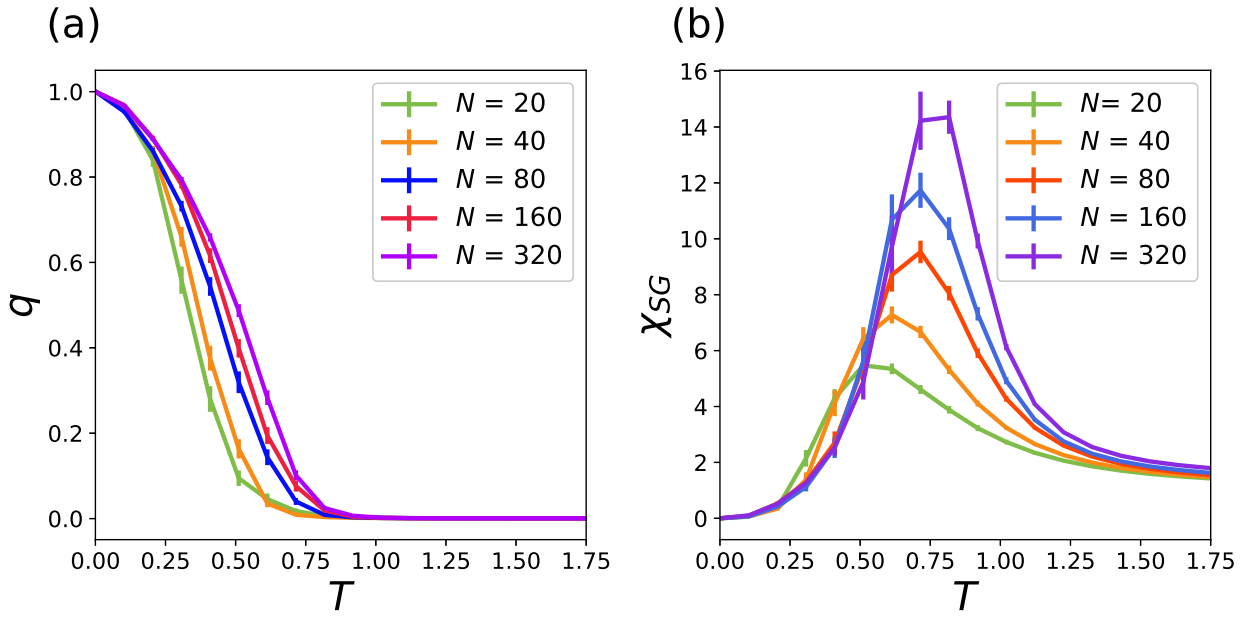

Fig. 2. The EA order parameters and spin glass susceptibilities for system size $N = 20$, 40, 80, 160, and 320. The spin glass susceptibility peak position represents the transition temperature for the SK model.

and $\mathbf{s}^\mu = (s_1^\mu, s_2^\mu, \cdots, s_N^\mu)$ with N being the total number of spins of the system. The likelihood function in Eq. (1) for our spin systems is written as $\mathcal{L}_D(\mathbf{J}) = p(D|\mathbf{J})$ with the true equilibrium Boltzmann distribution function $p(D|\mathbf{J})$ for the interaction matrix $\mathbf{J} \equiv \{J_{ij}\}$. The maximization of the likelihood function is commonly performed by the iterative gradient ascent method in practice [11,21] based on the observed distribution

$$f(\mathbf{s}) \equiv \frac{1}{M} \sum_\mu \delta_{\mathbf{s}^\mu, \mathbf{s}} \tag{7}$$

with the Kronecker δ.

We note that both Hamiltonians for the SK model and the 2D Ising model explained in Sec. II. can be written in the form

$$H_\mathbf{J}(\mathbf{s}) = -\sum_{i<j} J_{ij} s_i s_j, \tag{8}$$

although the interaction matrices have largely different forms. The model distribution of course takes the form of the equilibrium Boltzmann distribution $p(\mathbf{s}|\mathbf{J}) = e^{-\beta H_\mathbf{J}(\mathbf{s})}/Z$, where β is the inverse temperature and the canonical partition function $Z = \sum_\mathbf{s} e^{-\beta H_\mathbf{J}(\mathbf{s})}$ with the sum over all possible spin configurations. As the iterative maximization in the MLE

method proceeds, the equilibrium distribution $p(\mathbf{s}|\mathbf{J})$ is expected to approach the observed distribution $f(\mathbf{s})$.

For a practical convenience, the log-likelihood function is often used:

$$\ln \mathcal{L}_D(\mathbf{J}) = \ln p(D|\mathbf{J}) = \sum_{i<j} \sum_{\mu} J_{ij} s_i^{\mu} s_j^{\mu} - \sum_{\mathbf{s}} \ln Z, \tag{9}$$

which leads to

$$\ln \mathcal{L}_D(\mathbf{J}) = M \sum_{i<j} J_{ij} \langle s_i s_j \rangle_f - M \ln Z \tag{10}$$

with $\langle \cdots \rangle_f$ being the average over the observed distribution $f(\mathbf{s})$. The partial derivative of the log-likelihood function (divided by the overall constant M) to be used in the gradient ascent method is then written as

$$\frac{\partial \ln \mathcal{L}}{\partial J_{ij}} = \langle s_i s_j \rangle_f - \langle s_i s_j \rangle_p \,, \tag{11}$$

where $\langle \cdots \rangle_p$ is the average over the true model distribution $p(\mathbf{s}|\mathbf{J})$ for given $\mathbf{J}$ at the n-th step. The iterative maximization of the log-likelihood function is achieved by

$$J_{ij}^{n+1} = J_{ij}^{n} + \alpha \Big(\langle s_i s_j \rangle_f - \langle s_i s_j \rangle_p \Big) \tag{12}$$

for the n-th iteration with α being the tunable parameter of the learning rate. As the iterative maximization proceeds, the model parameters, which correspond to the interaction matrix $\mathbf{J}$ in this study, approach the ground truth and the model term $\langle s_i s_j \rangle_p$ becomes approximately equal to the data term $\langle s_i s_j \rangle_f$. In practice, the model term $\langle ... \rangle_p$ causes a severe computational problem since the average over all possible configurations whether they observed or not needs to be performed: For N spins, the computational complexity is $\mathcal{O}(2^N)$, and thus direct application of Eq. (12) is not possible even for system sizes not too large.

To overcome the impractical applicability of the original MLE, a very efficient methodology called the erasure machine (EM) [17] has been proposed, in which reweighting of the observed distribution and the model distribution, $f_\epsilon(\mathbf{s}) \propto f(\mathbf{s}) p^{\epsilon-1}(\mathbf{s})$ and $p_\epsilon(\mathbf{s}) \propto p(\mathbf{s}) p^{\epsilon-1}(\mathbf{s}) = p^{\epsilon}$, respectively, with the tunable reweighting parameter ϵ ($0 < \epsilon < 1$), plays an important role. The choice of the hyperparameter ϵ in EM has been shown to make only an insignificant difference in performance as long as sufficient number of samples is used [17]. It is interesting to note that the reweighting process resembles the high-temperature approximation

in statistical mechanics since p^ϵ puts the system at higher temperature from β to $\beta\epsilon$. The advantage of using the EM approach is that we can approximately compute the partition function for given ϵ:

$$\ln Z_\epsilon = N \ln 2 + \sum_{i<j} \ln \cosh (\epsilon J_{ij}) + O(\epsilon^3), \tag{13}$$

which yields

$$\langle s_i s_j \rangle_{p_\epsilon} = \frac{1}{\epsilon} \frac{\partial \ln Z_\epsilon}{\partial J_{ij}} \approx \epsilon J_{ij}. \tag{14}$$

Consequently, the iteration Eq. (12) in the MLE is given by

$$J_{ij}^{n+1} \approx J_{ij}^n + \alpha \Big(\langle s_i s_j \rangle_{f_\epsilon} - \epsilon J_{ij}^n \Big) \tag{15}$$

in the EM method. The computational cost for the model distribution term within the EM method is proportional to the size of the matrix size $N(N-1)$ of $\mathbf{J}$, which is a dramatic reduction from 2^N for the original MLE method as reported in Ref. [17].

IV. RESULTS

We are now ready to apply the EM method in Sec. III. to the SK model and the 2D Ising model in Sec. II. , with focus put on the temperature dependence of the performance of the EM. Only for a comparative convenience with Ref. [17], we first present our results for the SK model, which is followed by our results for the 2D Ising model. Differently from Ref. [17], we do not include the external field strength as a to-be-estimated parameter and the EM method is applied only to extract the most probable value of $\mathbf{J}$ based on the observed spin configurations from the MC simulations. Another difference from Ref. [17] is that we apply the ℓ_1 and ℓ_2 regularization methods with tunable parameters $\lambda_{1,2}$ in addition, as explained below in detail.

1. Sherrington-Kirkpatrick Model

We perform MC simulations of the SK model as described in Sec. II. to produce M observed spin configurations at each given temperature T for the fixed realization of the interaction matrix $\mathbf{J}^0$. Note that the superscript 0 for $\mathbf{J}^0$ is added for a notational convenience

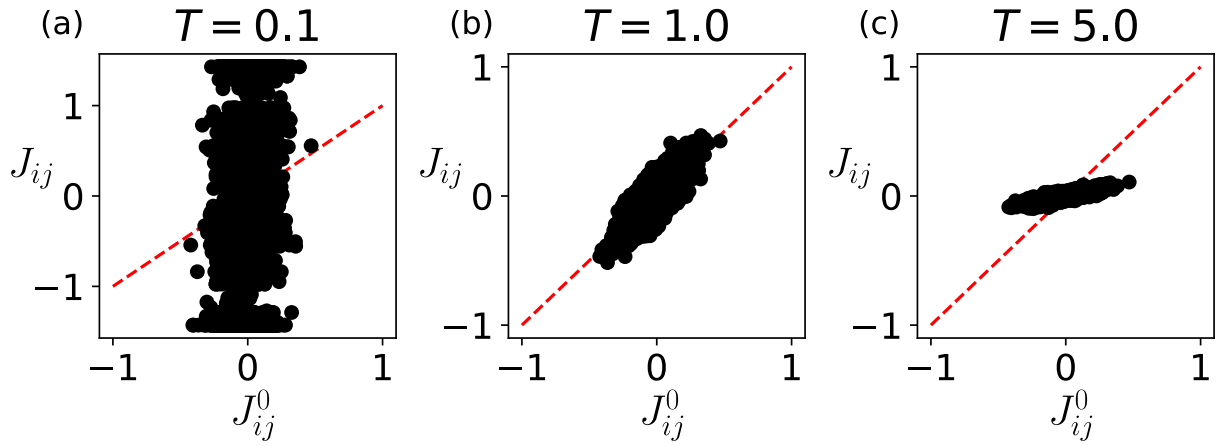

Fig. 3. Comparison between the original (J^0_{ij}) and the EM estimated (J_{ij}) interaction strengths for the SK model at temperatures $T =$ (a) 0.1, (b) 1.0, and (c) 5.0. We use the system size $N = 80$ without ℓ_2 regularization [$\lambda = 0$ in Eq. (20)] and the number of observed sample configurations is $M = 5000$. Note that the perfect performance of the EM is represented by $J_{ij} = J^0_{ij}$, denoted by the dashed red lines. It is shown clearly that the EM performance strongly depends on the temperature, and EM cannot recover correct interaction strengths at too low [(a) $T = 0.1$] and too high [(c) $T = 5.0$] temperatures.

to make distinction from the EM inferred interaction $\mathbf{J}$. As iterative maximization of the log-likelihood function proceeds through the use of Eq. (15), the difference between the ground truth $\{J^0_{ij}\}$ and the EM inference $\{J_{ij}\}$ is expected to decrease. In Fig. 3, we display J_{ij} versus J^0_{ij} at $T =$ (a) 0.1, (b) 1.0, and (c) 5.0, for $N = 80$ and $M = 5000$ after sufficient number (150) of iterations. We find that 150 iterations at the learning rate $\alpha = 0.1$ in Eq. (15) achieves stationarity of the EM inferred $\mathbf{J}$. We note that the EM performance becomes worse at very low and very high temperatures, and it appears that EM estimates the interaction strengths more accurately at an intermediate temperature around $T = 1.0$. The slope of scatter plot in Fig. 3(c) is less than unity, which indicates that the EM underestimates the strength of interaction at high temperatures.

In order to more quantitatively measure the performance of EM we compute the mean-square-error Δ defined by

$$\Delta \equiv \left[\frac{\sum_{ij} (J_{ij} - J^0_{ij})^2}{\sum_{ij} (J^0_{ij})^2} \right], \tag{16}$$

where $[\cdots]$ is the average over 50 different realizations of $\{J^0_{ij}\}$. Note that the better the EM

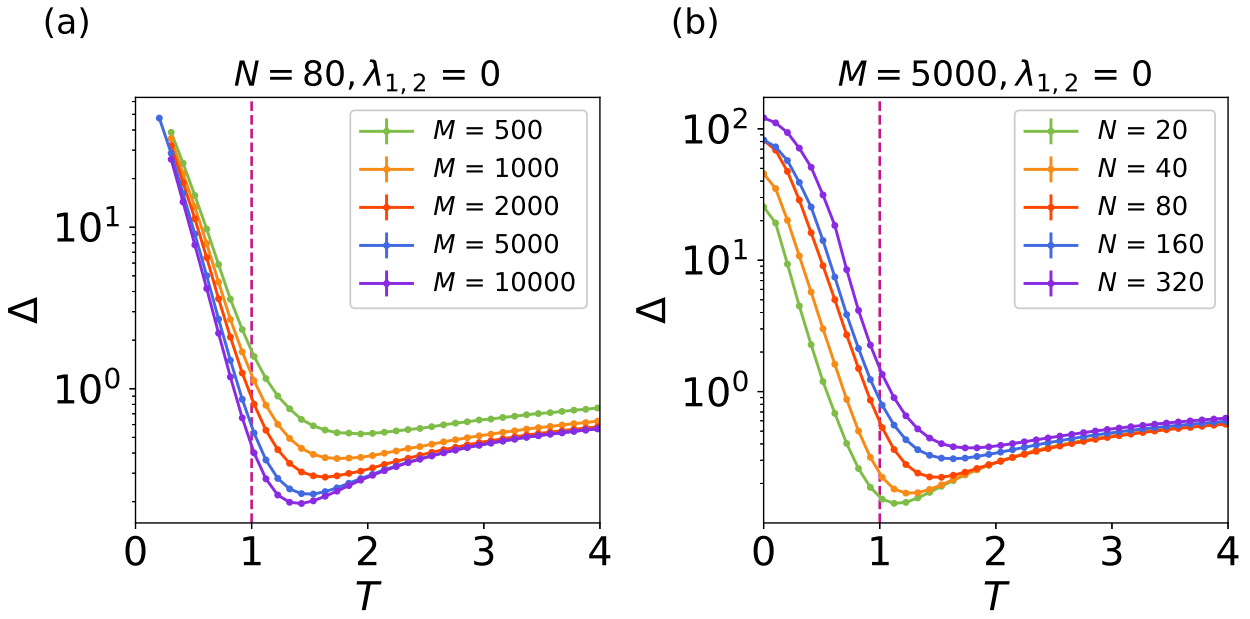

Fig. 4. Performance of the erasure machine for the SK model measured by the mean-square-error Δ in Eq. (16) versus temperature T for various (a) sample sizes M and (b) system sizes N without regularizations ($\lambda_{1,2} = 0$). All results are obtained from averages over 50 independent random realizations of $\mathbf{J}^0$ and we use 150 iterations at the learning rate $\alpha = 0.1$ in Eq. (16). (a) As more sample configurations are fed into the EM inference (i.e, as M is increased), Δ decreases and the minimum position of $\Delta(T)$ shifts to lower temperature. (b) However, the larger the size becomes the minimum of $\Delta(T)$ moves to higher temperature. For comparison, we also include the position of the glass transition temperature $T_g = 1.0$ as vertical lines in (a) and (b). The error bars, not included in the plots, are smaller than the symbol sizes.

performs, the smaller Δ becomes. We first investigate how the number M of observed sample configuration fed into EM affects the performance of EM. Figure 4(a) displays Δ versus T at $M = 500$, 1000, 2000, 5000, and 10000 for the system size $N = 80$ without regularization (see below for the regularization). As M is increased and thus more information is fed into for the EM inference, Δ becomes smaller at all temperatures. Interestingly, the minimum position of Δ appears to shift toward lower temperature as M is increased.

The observed spin configurations can be sampled during MC simulations in different intervals. We expect that the sampling interval may affect the EM performance especially at criticality where relaxation time often diverges in thermodynamic limit [22]. In order to guarantee the independence among sampled spin configurations, the sampling interval has

to be be chosen large enough in comparison to the relaxation time. This problem becomes very serious in the low-temperature glass phase in the SK model since each spin directions is almost frozen and does not change in very long time [20, 23]. We thus check how the sampling interval (Δt in units of MC time step) affects the EM performance by using two different sampling intervals $\Delta t = 1$ and $\Delta t = \tau$ with the relaxation time τ measured from the autocorrelation function of the EA order parameter q. Although not shown here, we find that the EM performance measured by Δ does not display much difference between sampling intervals $\Delta t = 1$ and $\Delta t = \tau$, as long as M is large enough. In short, we conclude that the difference between the minimum temperature of $\Delta(T)$ and the glass transition temperature T_g in Fig. 4(a) does not originate from the short sampling interval.

We next investigate how the EM performance changes with the system size N. As shown in Fig. 4(b) for $M = 5000$ without regularization, the minimum position of Δ is found to drift away toward higher temperature as the system size N is increased. This result deviates from our original expectation that the EM should best perform around the glass transition temperature $T_g = 1$ for large systems. It is still possible that the minimum of Δ occurs around the glass transition temperature if M is increased even further. However, it only indicates that the EM method becomes impractical for large systems.

The element J_{ij} of the interaction matrix $\mathbf{J}$ in the SK model is produced from the Gaussian distribution with null average. We believe that a proper regularization applied for the estimation of J_{ij} can significantly enhance the performance of the EM method for the SK model since it can penalize too large value of $|J_{ij}|$. In this respect, we write the log posterior in our Bayesian inference scheme as

$$\ln p(\mathbf{J}|D) = \ln \mathcal{L}_D(\mathbf{J}) + \lambda_1 \sum_{i<j} |J_{ij}| \tag{17}$$

for ℓ_1 regularization, and

$$\ln p(\mathbf{J}|D) = \ln \mathcal{L}_D(\mathbf{J}) + \frac{\lambda_2}{2} \sum_{i<j} J_{ij}^2, \tag{18}$$

for ℓ_2 regularization with λ_1 and λ_2 being the tunable parameter for each regularization process. Within the EM framework, it is straightforward to get

$$J_{ij}^{n+1} = J_{ij}^n + \alpha \left(\langle s_i s_j \rangle_{f_\epsilon} - \epsilon J_{ij}^n - \lambda_1 \frac{J_{ij}^n}{|J_{ij}^n|} \right) \tag{19}$$

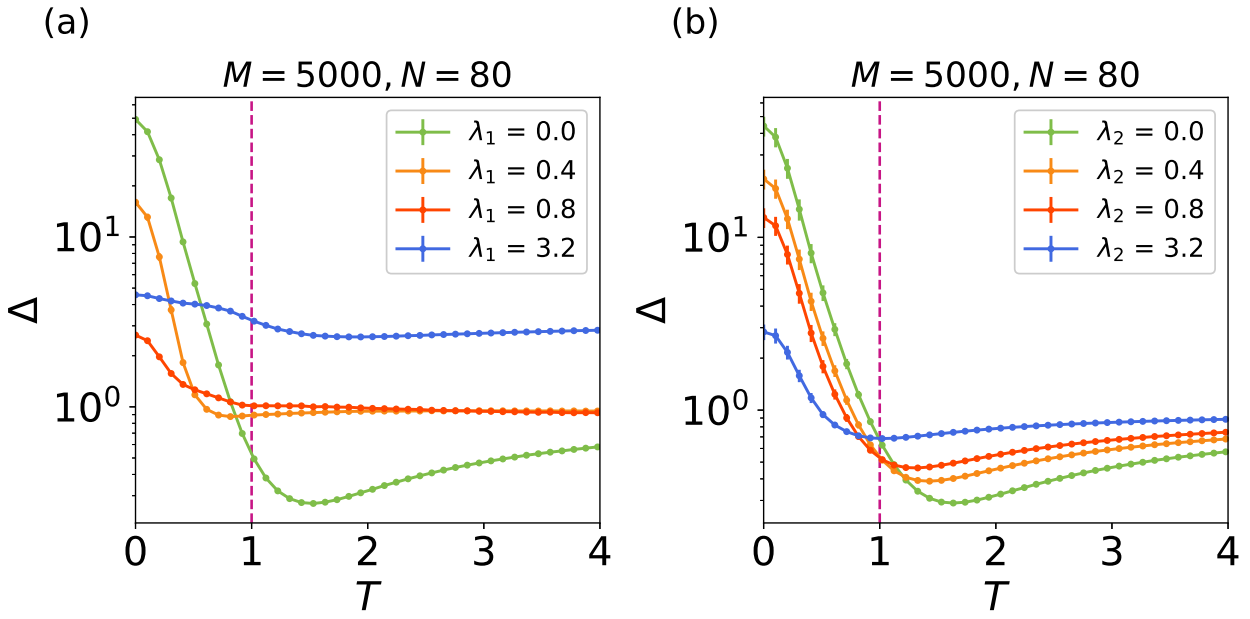

Fig. 5. The mean-square-error Δ versus temperature T for the SK model of the size $N = 80$ and the sample size $M = 5000$ at various values of (a) λ_1 and (b) λ_2 for ℓ_1 and ℓ_2 regularizations. (All other parameter values are the same as in Fig. 4). As the strength of the regularization is increased the minimum of $\Delta(T)$ shifts to lower temperature, but the minimum value tends to increase. The dashed vertical lines are at $T_g = 1.0$.

for ℓ_1 regularization, and

$$J_{ij}^{n+1} = J_{ij}^n + \alpha\Big(\langle s_i s_j\rangle_{f_\epsilon} - \epsilon J_{ij}^n - \lambda_2 J_{ij}^n\Big), \tag{20}$$

for ℓ_2 regularization, respectively, which substitute Eq. (15) without regularization.

We report our results on how the ℓ_1 and ℓ_2 regularizations affect the EM performance in Fig. 5 for (a) ℓ_1 and (b) ℓ_2 regularizations, respectively for $M = 5000$ and $N = 80$. The minimum position of $\Delta(T)$ shifts to lower temperature as $\lambda_{1,2}$ is increased, but the value of Δ at the minimum tends to increase. We thus conclude that both ℓ_1 and ℓ_2 regularizations applied to the EM inference for the SK model are not successful.

2. Two-dimensional Ising Model

We next investigate the temperature dependence of the EM performance for the Ising model on the two-dimensional square lattice. MC simulations of the 2D Ising model of the

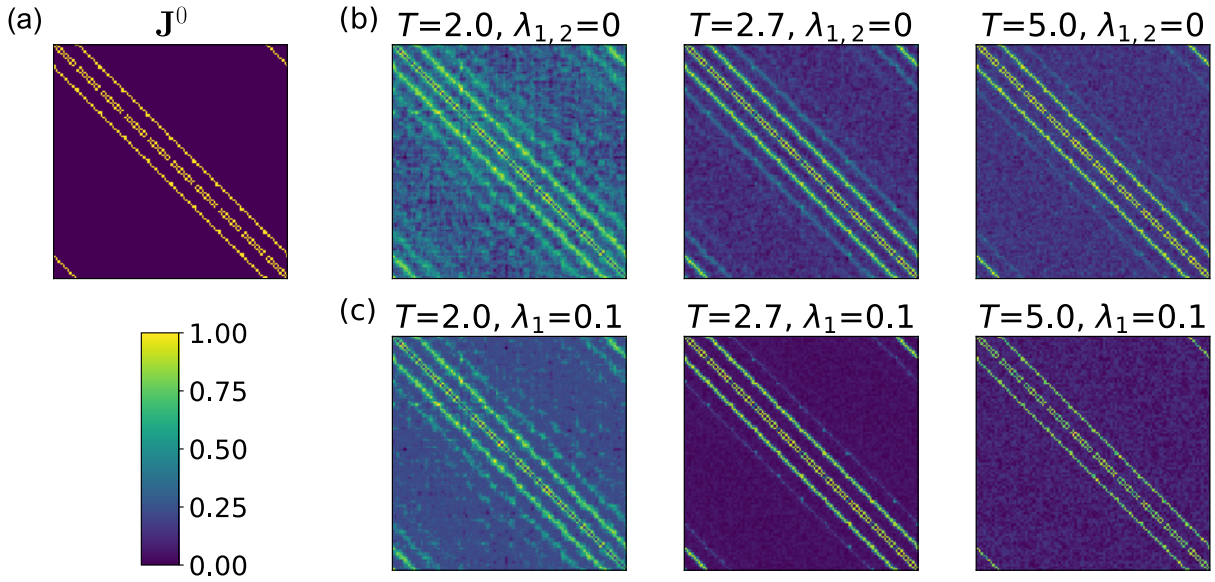

Fig. 6. The adjacency matrices of two-dimensional Ising model of the size 10×10 in the presence of the periodic boundary condition. (a) The ground truth interaction matrix $\mathbf{J}^0$. The EM inferred interaction matrix $\mathbf{J}$ without [(b)] and with [(c)] ℓ_1 regularization at temperatures $T = 2.0$, 2.7, and 5.0. We find that ℓ_1 regularization at $T = 2.7$ reproduces better the ground truth in comparison with other cases.

sizes 5×5 and 10×10 are performed as explained in Sec. II. to produce spin configurations. Differently from the SK model, $\mathbf{J}^0$ for the 2D Ising model contains only the nearest-neighbor interaction as shown in Fig. 6(a) for which the interaction strength J^0_{ij} is shown in density plot in the 2D plane of the lattice structure for the 10×10 lattice. The 2D Ising model is very well-known to exhibit the second-order phase transition between the low-temperature ferromagnetic phase and the high-temperature paramagnetic phase at the critical temperature $T_c \approx 2.27$ [24]. The EM inferred interaction matrices $\mathbf{J}$ are obtained at various temperatures with and without regularization, which are displayed in Fig. 6. For regularizations, we test both ℓ_1 and ℓ_2 regularizations (see above for both regularization methods), and find that the former performs better than the latter. We also test different regularization assuming two peaks at $J_{ij} = 0$ and 1, i.e., $\ln p(\mathbf{J}|D) = \ln \mathcal{L}_D(\mathbf{J}) + \frac{\lambda_2}{2} \sum_{i<j} [(-J^2_{ij}(J_{ij} - 1)^2]$, the performance of which is found not to be successful. As can be recognized in Fig. 6(c), the EM inference works better at $T = 2.7$, which is above $T_c \approx 2.27$, when we apply a suitable strength ($\lambda_1 = 0.1$) in the ℓ_1 regularization.

In Fig. 6, it is clearly seen that the EM inferred interaction matrix exhibits somewhat

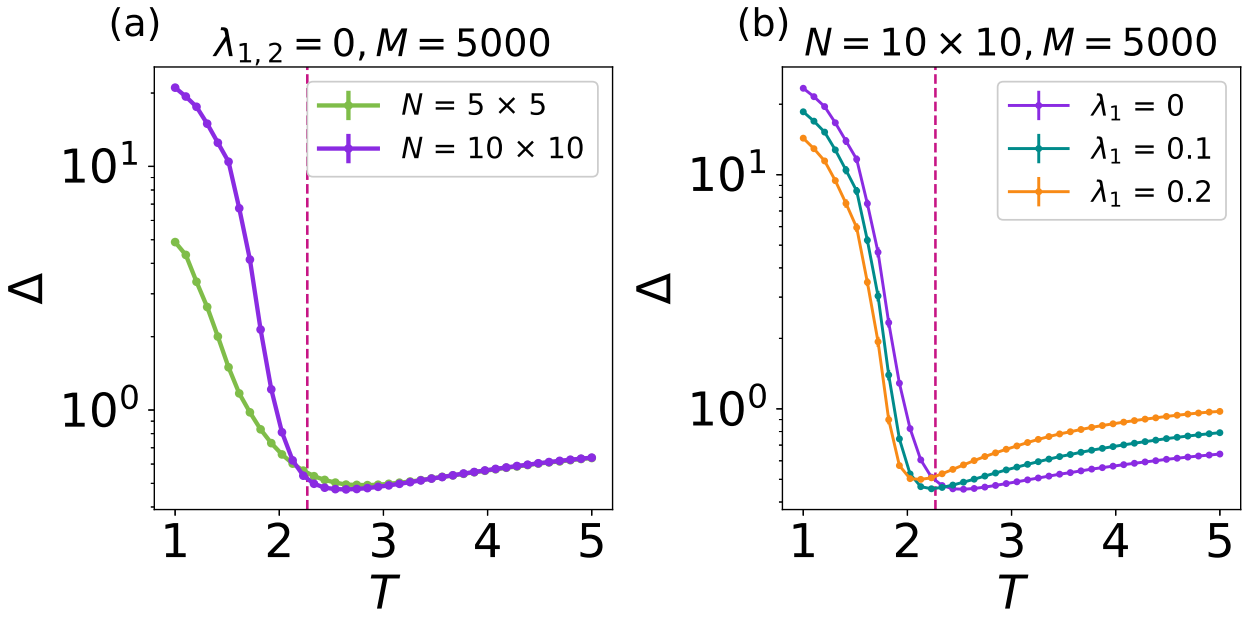

Fig. 7. Performance of the EM for two-dimensional Ising model is measured by the mean-square-error Δ in Eq. (16) as a function of the temperature T for (a) without regularization and (b) with ℓ_1 regularization. In (a), the minimum position of $\Delta(T)$ appears to approach $T_c \approx 2.27$ as the system size is increased. (b) For $N = 10 \times 10$, the use of $\lambda_1 = 0.1$ yields the minimum of $\Delta(T)$ approximately at $T_c \approx 2.27$. In (a) and (b) the critical temperature $T_c \approx 2.27$ for the 2D Ising model is represented by the vertical dashed lines.

blurred form at very low and very high temperatures, which suggests that even when $J^0_{ij} = 0$, the EM estimates nonzero value of J_{ij}. In order to reveal the temperature dependence of the EM performance we again use the mean-square-error Δ defined in Eq. (16) and report our results in Fig. 7. As the system size is increased the minimum of $\Delta(T)$ appears to approach the critical temperature $T_c \approx 2.27$ of the 2D Ising model as shown in Fig. 7(a), which is to be compared with different size dependence observed for the SK model in Fig. 4(b). We also find in Fig. 7(b) that ℓ_1 regularization with λ_1 for $N = 10 \times 10$ and $M = 5000$ results in minimum error around the critical temperature $T_c \approx 2.27$ of the 2D Ising model.

V. MONTE-CARLO RESULTS BASED ON INFERRED INTERACTION

We next study the difference between the original ($\mathbf{J}^0$) and the EM inferred ($\mathbf{J}$) interaction matrices through the direct comparison of the MC simulation results. In more detail, we

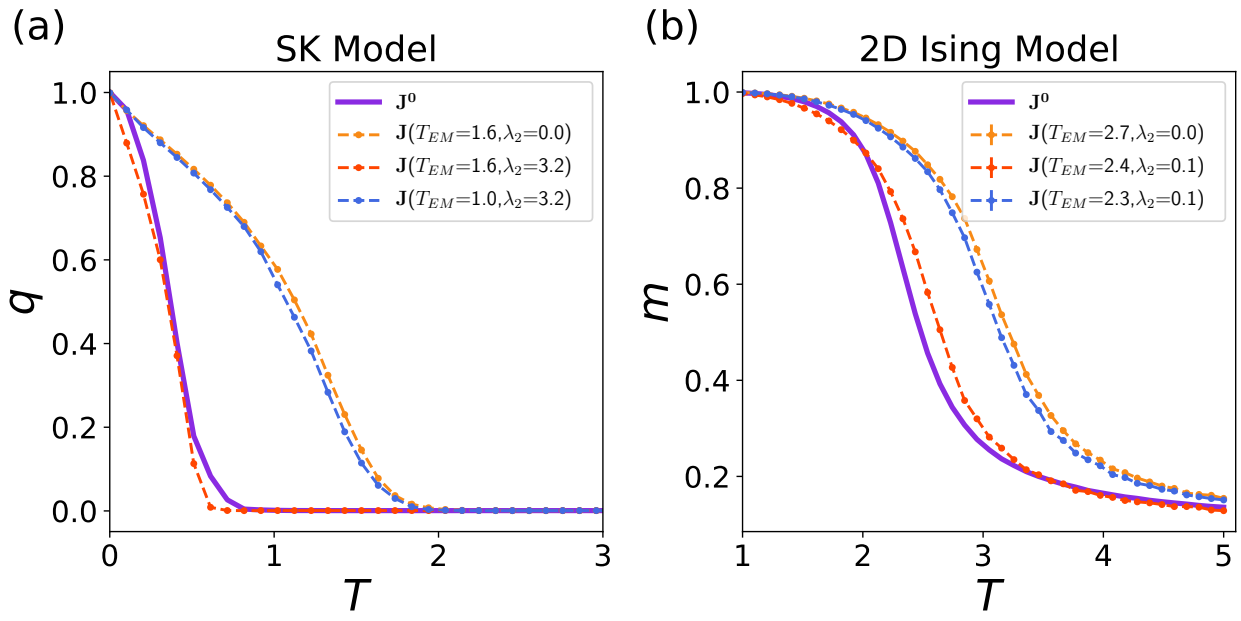

Fig. 8. Comparison of the Monte-Carlo results for the actual underlying interaction structure $\mathbf{J}^0$ and the EM inferred one $\mathbf{J}$ obtained at the temperature T_{EM}. (a) The EA order parameter q for the SK model of the size $N = 80$ as a function of the temperature T. We find that a proper regularization $\lambda_2 = 3.2$ helps to recover the MC result of the original SK model at $T_{EM} = 1.6$ (b) The ferromagnetic order parameter m for the 2D Ising model of the size $N = 10 \times 10$. We find that the ℓ_1 regularization at $\lambda_1 = 0.1$ exhibit relatively better performance at $T_{EM} = 2.4$.

use the EM inferred $\mathbf{J}(T_{EM})$ obtained at a given fixed temperature T_{EM} as an underlying interaction structure and perform MC simulations at various temperatures as described in Sec. II. . If the EM inference at T_{EM} has yielded the perfectly correct outcome, MC results based on $\mathbf{J}^0$ and $\mathbf{J}$ must be identical. We compute the order parameters, q for the SK model and m for the 2D Ising model, for given EM inferred $\mathbf{J}(T_{EM})$ with and without regularizations, and report our results in Fig. 8. We find in Sec. IV. that the performance of the EM is better for ℓ_2 regularization for the SK model, and for ℓ_1 regularization for the 2D Ising model, and thus include only results for corresponding regularizations in Fig. 8. . For the SK model in Fig. 8(a), we choose $T_{EM} = 1.6$ without regularization and $T_{EM} = 1.0$ and 1.6 for ℓ_2 regularization at $\lambda_2 = 3.2$. The best overlap with the MC results based on the actual underlying interaction matrix $\mathbf{J}^0$, as far as we can tell, occurs at $\ell_2 = 3.2$, but at somewhat higher temperature $T_{EM} = 1.6$ than the glass transition temperature $T_g = 1.0$. It is to be noted that the mean-square-error Δ at $\lambda_2 = 3.2$ shown in Fig. 5(b) does not

exhibit the minimum at $T_{EM} = 1.6$. For the 2D Ising model in Fig. 8(b), we display our comparisons of MC results at $T_{EM} = 2.7$ without regularization and at $T_{EM} = 2.3$ and 2.4 with ℓ_2 regularization. As also found for the SK model in Fig. 8(a) for the SK model, two MC results, based on $\mathbf{J}^0$ and $\mathbf{J}(T_{EM})$, display better at $T_{EM} \approx 2.4 > T_c = 2.27$ in the presence of a proper regularization at $\lambda_1 = 0.1$.

VI. SUMMARY AND DISCUSSION

In this paper, we have tested the temperature-dependent performance of the erasure machine (EM) [17] for the SK spin glass model and the 2D Ising model. Based on a simple reasoning, one can clearly expect that the EM fails to faithfully recover the original underlying interaction structure if the spin configurations are sampled at too low and too high temperatures: When the temperature is extremely low the spins are basically frozen for any given interaction structure, and when the temperature is extremely high, entropy effect prevails and thus spins only exhibit purely random fluctuation. We have started our investigation in the present work with an intuitive guess that the EM performs better near the criticality at which spin states must reflect long-rage correlation and thus can reveal the underlying genuine interaction structure. We regret to report that our results are close to only the half success. Through careful analyses of the mean-square-error and the Monte Carlo simulations based on the EM inferred interaction, we find that the EM performance is enhanced in the intermediate temperature region. However, the best EM performance is found not exactly at the critical temperatures, $T_g = 1.0$ for the SK model and $T_c \approx 2.27$, but at somewhat higher temperature. The reason can be that the EM framework basically uses the high-temperature approximation to make the method more efficient. We also find that if proper regularizations are introduced in the EM inference framework, the reliability of the obtained interaction structure is enhanced. Future research could extend the idea of erasure machine that approximate partition function by reweighting to other inverse problems where the cost is very large such as networks with hidden nodes [25], RBMs [14] and continuous variable problem [25].

ACKNOWLEDGMENTS

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) Grant No.2019R1A2C2089463.

REFERENCES

[1] R. H. Swendsen, Phys. Rev. Lett. **52**, 1165 (1984).

[2] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Proc. Natl. Acad. Sci. USA **106**, 67 (2009).

[3] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).

[4] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature **440**, 1007 (2006).

[5] T. Watanabe, S. Hirose, H. Wada, Y. Imai, T. Machida, I. Shirouzu, S. Konishi, Y. Miyashita, and N. Masuda, Nat. Commun. **4**, 1370 (2013).

[6] T. Bury, Physica A **392**, 1375 (2013).

[7] S. S. Borysov, Y. Roudi, and A. V. Balatsky, Eur. Phys. J. B **88**, 321 (2015).

[8] P. W. Holland, K. B. Laskey, and S. Leinhardt, Social Networks **5**, 109 (1983).

[9] D. Chase, IEEE Trans. Commun. **33**, 385 (1985).

[10] D. J. MacKay and D. J. Mac Kay, *Information theory, inference and learning algorithms* (Cambridge university press, 2003).

[11] H. C. Nguyen, R. Zecchina, and J. Berg, Adv. Phys. **66**, 197 (2017).

[12] P. Ravikumar, M. J. Wainwright, J. D. Lafferty, et al., Ann. Stat. **38**, 1287 (2010).

[13] E. Aurell and M. Ekeberg, Phys. Rev. Lett. **108**, 090201 (2012).

[14] G. Cossu, L. Del Debbio, T. Giani, A. Khamseh, and M. Wilson, Phys. Rev. B **100**, 064304 (2019).

[15] D. Wu, L. Wang, and P. Zhang, Phys. Rev. Lett. **122**, 080602 (2019).

[16] A. Y. Lokhov, M. Vuffray, S. Misra, and M. Chertkov, Sci. Adv. **4**, e1700791 (2018).

[17] J. Jo, D.-T. Hoang, and V. Periwal, Phys. Rev. E **101**, 032107 (2020).

[18] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).

[19] S. F. Edwards and P. W. Anderson, J. Phys. F: Met. Phys. **5**, 965 (1975).

[20] K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).

[21] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cogn. Sci. **9**, 147 (1985).

[22] M. E. J. Newman and G. T. Barkema, *Monte Carlo methods in statistical physics* (Clarendon Press, Oxford, 1999).

[23] K. Binder, Z. Phys. B: Condens. Matter **26**, 339 (1977).

[24] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Westview Press, 1992).

[25] C. Donner and M. Opper, Phys. Rev. E **96**, 062104 (2017)