Hereditary structure in Hamiltonians: Information geometry of Ising spin chains

Akio Fujiwara* and Shigeru Shuto
Department of Mathematics, Osaka University
Toyonaka, Osaka 560-0043, Japan

Abstract

This paper deals with the problem of approximating the canonical state of an Ising spin chain by a hierarchical series of independent and identically distributed cluster states. Based on information geometry, it is shown that the structure of the effective Hamiltonian for each cluster state is inherited from the total Hamiltonian. This fact partly justifies the methodology of mean field theories. The issue of a phase transition is further analyzed from the point of view of statistical hypothesis testing.

Keywords:
cluster approximation, hierarchical model, information geometry, mathematical statistics, mean field theory, phase transition

PACS:
02.40.-k (Geometry, differential geometry, and topology)
02.50.-r (Probability theory, stochastic processes, and statistics)
75.10.Pq (Spin chain models)
89.70.Cf (Entropy and other measures of information)

1 Introduction

In statistical mechanics, one often invokes approximations in calculating the partition function of the system. A cluster approximation method, such as the Weiss molecular field approximation [1], the Bragg-Williams approximation [2], the Bethe approximation [3, 4], and the coherent anomaly method [5, 6], can be viewed as a method of approximating the total system by a family of independent and identically distributed clusters. The structure of each cluster is specified by an "effective" Hamiltonian which is defined by imitating the structure of the total Hamiltonian for all but the boundary particles of the cluster. The parameters for the boundary particles are determined afterwards, through the requirement of self-consistency. The quality of an approximation is usually assessed by its power of describing the critical behavior of the system[1].

From a purely mathematical point of view, however, there is no *a priori* reason for restricting ourselves to effective Hamiltonians that imitate the total Hamiltonian. Put another way, one might be able to obtain a better approximation by taking account of longer range and/or many body interactions which do not exist in the total Hamiltonian. To the best of the authors' knowledge, such a question has not been investigated systematically. The purpose of this paper is to point out that imitating the structure of the total Hamiltonian is actually valid for systems that have

*fujiwara@math.sci.osaka-u.ac.jp

[1]Related issues are also argued in connection with statistical learning theory; see, for example, [7, 8].

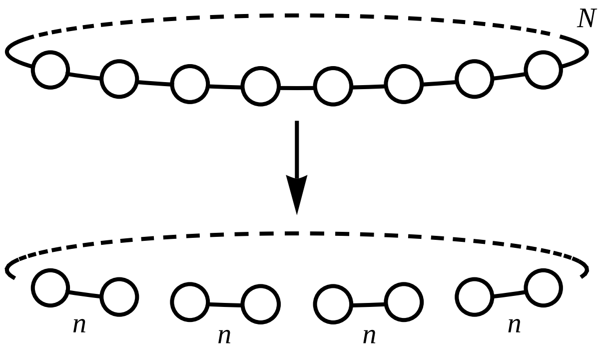

Figure 1: The N-spin Ising chain is decomposed into m clusters ($m \geq 2$), each comprising n consecutive spins. The clusters are assumed to be independent and identically distributed.

a high degree of symmetry. This fact partly justifies the methodology of mean field theories. We demonstrate this by proving a hereditary property of Hamiltonian for the one-dimensional Ising spin model, the simplest model for which the partition function is exactly calculable and has a thermodynamic limit. We also mention how the problem of a phase transition is treated from the point of view of statistical hypothesis testing.

This paper is organized as follows. The hereditary property of Hamiltonian for Ising spin chains is stated in Section 2, and is proved in Section 3. The issue of a phase transition is investigated in Section 4 from the point of view of mathematical statistics. Throughout the paper, we assume some basic knowledge of information geometry [9].

2 Main result

Let us consider an N-spin Ising ring specified by the Hamiltonian

$$H := H(\{S_\xi\}) := -h\sum_{\xi=1}^{N} S_\xi - J\sum_{\xi=1}^{N} S_\xi S_{\xi+1}, \qquad (J > 0),$$

where each random variable S_ξ ($1 \leq \xi \leq N$) takes values in $\{-1, +1\}$, and the index $\xi + 1$ is understood modulo N. We denote the corresponding canonical distribution as

$$q := q(\{S_\xi\}) := \exp\left[-\beta H - \tilde{\psi}(h, J)\right], \tag{1}$$

where β is the inverse temperature, and $\tilde{\psi}(h, J)$ the logarithm of the partition function. Geometrically, the distribution q is regarded as a point on the $(2^N - 1)$-dimensional manifold $\mathcal{P}_N$ of all probability distributions of N spins.

Let us partition those N spins into m clusters ($m \geq 2$), each comprising n consecutive spins, so that $N = m \times n$. The ith spin $S_i^{(\lambda)}$ in the λth cluster ($1 \leq i \leq n$ and $1 \leq \lambda \leq m$) corresponds to the ξth spin S_ξ in the original chain, where

$$\xi = (\lambda - 1)n + i.$$

We address the problem of approximating the N-spin probability distribution q by means of a certain n-spin probability distribution, regarding m clusters as independent and identically distributed (i.i.d.), see Figure 1.

Any n-spin probability distribution is represented in the form of an exponential family [9]

$$\exp\left[\sum_{\ell=1}^{n}\sum_{1\le i_1<i_2<\cdots<i_\ell\le n}\theta^{\langle i_1 i_2\cdots i_\ell\rangle}S_{i_1}S_{i_2}\cdots S_{i_\ell}-\psi(\theta)\right], \tag{2}$$

where $\theta := (\theta^{\langle i_1 i_2\cdots i_\ell\rangle})$ is a set of real parameters, and $\psi(\theta)$ is the normalization. By using an abridged notation

$$S_{\langle i_1 i_2\cdots i_\ell\rangle} := S_{i_1}S_{i_2}\cdots S_{i_\ell},$$

the distribution (2) is rewritten as

$$\exp\left[\theta^a S_a - \psi(\theta)\right],$$

where the index a runs over the set

$$I=\bigcup_{\ell=1}^{n} I_\ell, \qquad I_\ell=\{\langle i_1 i_2\cdots i_\ell\rangle\ ;\ 1\le i_1<i_2<\cdots<i_\ell\le n\}$$

of multi-indices, and Einstein's summation convention is used. Note that $\{S_a\}_{a\in I}$ and a nonzero constant function together form a set of linearly independent functions on $\{-1,+1\}^n$. As a consequence, the totality of probability distributions

$$p_\theta := p_\theta(\{S_\xi\}) := \prod_{\lambda=1}^{m}\exp\left[\theta^a S_a^{(\lambda)}-\psi(\theta)\right] \tag{3}$$

of m i.i.d. clusters can be regarded as a (2^n-1)-dimensional submanifold $\mathcal{M}_n$ of $\mathcal{P}_N$ having the coordinate system $\theta=(\theta^a)_{a\in I}$. We shall call each p_θ a *cluster state*.

In order to formulate the problem of approximating the canonical distribution q by a cluster state $p_\theta\in\mathcal{M}_n$, let us consider the relative entropy (also called the Kullback-Leibler divergence):

$$D(p_\theta\|q) := E_{p_\theta}\left[\log\frac{p_\theta}{q}\right],$$

where $E_p[\,\cdot\,]$ denotes the expectation with respect to p. By a simple algebra, we get

$$D(p_\theta\|q)=\beta\left[F(p_\theta)-F_q\right], \tag{4}$$

where

$$F(p) := E_p[H]-\frac{1}{\beta}S(p)$$

is the "free energy" of p with $S(p):=E_p[-\log p]$ the Shannon entropy, and

$$F_q := -\frac{1}{\beta}\tilde{\psi}(h,J)$$

the true Helmholtz free energy of the Ising spin chain. Since $F_q=\min_{p\in\mathcal{P}_N}F(p)=F(q)$, (4) is rewritten as

$$D(p_\theta\|q)=\beta\left[F(p_\theta)-F(q)\right]. \tag{5}$$

We see from (5) that minimizing the free energy $F(p_\theta)$ with respect to $p_\theta\in\mathcal{M}_n$ is equivalent to minimizing $D(p_\theta\|q)$, which in turn amounts to finding a ∇^e-projection[2] from q onto the submanifold $\mathcal{M}_n$, as illustrated in Figure 2. We regard the point $p_{\hat\theta}\in\mathcal{M}_n$ that minimizes $D(p_\theta\|q)$ as approximating the canonical distribution $q\in\mathcal{P}_N$. Since $\mathcal{M}_n$ is not a ∇^m-autoparallel submanifold (in fact it is ∇^e-autoparallel), the approximating point is not always unique [9, Section 3.4].

[2] ∇^e and ∇^m stand for the exponential connection and the mixture connection [9, Section 2.3].

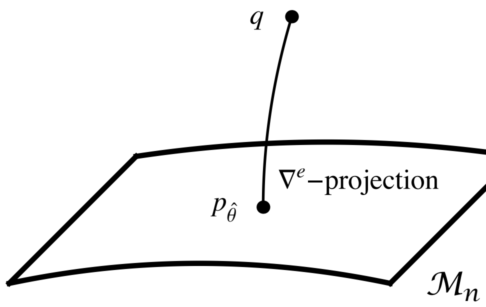

Figure 2: The canonical distribution q is approximated by the point $p_{\hat{\theta}}$ on $\mathcal{M}_n$ at which the function $p_\theta \mapsto D(p_\theta \| q)$ takes the minimum. This is a ∇^e-projection from q onto $\mathcal{M}_n$ [9, Theorem 3.10].

When $n = 1$, this approximation is nothing but the Bragg-Williams approximation, or equivalently, the Weiss molecular field approximation. The non-uniqueness of the ∇^e-projection is sometimes interpreted as the existence of broken symmetry states with spontaneous magnetization. When $n = 3$, on the other hand, the present approximation is different from the Bethe approximation, because the ∇^e-projection $p_{\hat{\theta}}$ onto $\mathcal{M}_3$ does not satisfy the self-consistency $E_{p_{\hat{\theta}}}[S_1^{(\lambda)}] = E_{p_{\hat{\theta}}}[S_2^{(\lambda)}]$.

The present geometrical approach has the following advantage [10]: it allows us to study, in a unified manner, the series of approximations associated with the natural hierarchy[3]

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots \subset \mathcal{M}_n \subset \cdots .$$

The main result of this paper is the following

Theorem 1. *Let $n \geq 2$ and let $\hat{\theta}$ be the minimizer of the function $\theta \mapsto D(p_\theta \| q)$ on $\mathcal{M}_n$. Then*

$$\hat{\theta}^{\langle 1 \rangle} = \hat{\theta}^{\langle n \rangle}$$

$$\hat{\theta}^{\langle 2 \rangle} = \hat{\theta}^{\langle 3 \rangle} = \cdots = \hat{\theta}^{\langle n-1 \rangle} = \beta h$$

$$\hat{\theta}^{\langle 12 \rangle} = \hat{\theta}^{\langle 23 \rangle} = \cdots = \hat{\theta}^{\langle n-1,n \rangle} = \beta J$$

and the rest are all zero. The only remaining free parameter $\hat{\theta}^{\langle 1 \rangle}$ is determined by the equation:

$$\hat{\theta}^{\langle 1 \rangle} = \beta \left(h + J E_{p_{\hat{\theta}}}[S_1] \right). \tag{6}$$

Theorem 1 implies that the parameters of Hamiltonian are inherited for all but the boundary parameters $\theta^{\langle 1 \rangle}$ and $\theta^{\langle n \rangle}$, and the parameter $\theta^{\langle 1 \rangle}(= \theta^{\langle n \rangle})$ is determined by (6).

Remark 2. *The case $n = 1$ is exceptional because there is no bond in a cluster. In this case (see Remark 4), the parameter $\hat{\theta}^{\langle 1 \rangle}$ is determined by the standard mean field equation*

$$\hat{\theta}^{\langle 1 \rangle} = \beta \left(h + 2J E_{p_{\hat{\theta}}}[S_1] \right).$$

Unless otherwise stated, we always assume that $n \geq 2$ in what follows.

[3] To be precise, the series $\{\mathcal{M}_n\}_n$ form a directed system [11], in that $\mathcal{M}_k \subset \mathcal{M}_\ell$ if ℓ is a multiple of k.

3 Proof of Theorem 1

For the sake of notational simplicity, the inverse temperature β shall be incorporated into the parameters h and J, so that symbols h and J in this section actually stand for βh and βJ. Observe that

$$\log p_\theta = \sum_{\lambda=1}^{m} \ell^{(\lambda)},$$

where

$$\ell^{(\lambda)} := \theta^a S_a^{(\lambda)} - \psi(\theta)$$

are i.i.d. random variables. Due to normalization, we have

$$E_{p_\theta}[\partial_a \ell^{(\lambda)}] = 0, \qquad (1 \le \forall \lambda \le m) \tag{7}$$

where $\partial_a = \partial/\partial\theta^a$. The Fisher metric $g = (g_{ab})$ of p_θ is then given by

$$g_{ab} := E_{p_\theta}\left[(\partial_a \log p_\theta)(\partial_b \log p_\theta)\right] = \sum_{\lambda=1}^{m} E_{p_\theta}\left[\left(\partial_a \ell^{(\lambda)}\right)\left(\partial_b \ell^{(\lambda)}\right)\right] = m E_{p_\theta}\left[\left(\partial_a \ell^{(1)}\right)\left(\partial_b \ell^{(1)}\right)\right].$$

Note that the dual coordinate system $\eta = (\eta_a)$ defined by

$$\eta_a := \partial_a \psi(\theta) = E_{p_\theta}[S_a^{(\lambda)}]$$

enjoys the following identity

$$\partial_a \ell^{(\lambda)} = S_a^{(\lambda)} - \eta_a. \tag{8}$$

Since

$$\partial_a D(p_\theta \| q) = \partial_a \sum_{\{S_\xi\}} p_\theta \left(\log p_\theta - \log q\right) = \sum_{\{S_\xi\}} p_\theta (\partial_a \log p_\theta)\left(\log p_\theta - \log q\right),$$

the function $\theta \mapsto D(p_\theta \| q)$ takes the minimum only if the orthogonality condition

$$E_{p_\theta}\left[(\log p_\theta - \log q)(\partial_a \log p_\theta)\right] = 0, \qquad (\forall a \in I) \tag{9}$$

is satisfied (cf. [9, Theorem 3.10]). In order to evaluate (9), let us introduce auxiliary parameters $x = (x^a)_{a\in I}$ and $y = (y^a)_{a \in I}$ as follows.

$$x^a = \begin{cases} \theta^{\langle i \rangle} - h, & a = \langle i \rangle \\ \theta^{\langle ij \rangle} - J, & a = \langle ij \rangle \text{ and } j = i+1 \\ \theta^{\langle ij \rangle}, & a = \langle ij \rangle \text{ and } j > i+1 \\ \theta^{\langle i_1 i_2 \cdots i_\ell \rangle}, & a = \langle i_1 i_2 \cdots i_\ell \rangle \text{ and } \ell \ge 3 \end{cases}$$

$$y^a = \begin{cases} \eta_{\langle n \rangle}, & a = \langle 1 \rangle \\ \eta_{\langle 1 \rangle}, & a = \langle n \rangle \\ 0, & \text{otherwise} \end{cases}$$

Lemma 3. *The condition (9) is equivalent to*

$$x^a = J y^a, \qquad (\forall a \in I).$$

Proof. The canonical state (1) is rewritten as

$$\begin{aligned} \log q &= h \sum_{\xi=1}^{N} S_\xi + J \sum_{\xi=1}^{N} S_\xi S_{\xi+1} - \tilde{\psi}(h, J) \\ &= \sum_{\lambda=1}^{m} \left\{ h \sum_{i=1}^{n} S_i^{(\lambda)} + J \sum_{i=1}^{n-1} S_i^{(\lambda)} S_{i+1}^{(\lambda)} + J S_n^{(\lambda)} S_1^{(\lambda+1)} \right\} - \tilde{\psi}(h, J), \end{aligned}$$

where the superscript $\lambda+1$ is understood modulo m. Consequently, we see from (3) that

$$\begin{aligned}\log p_\theta - \log q &= \sum_{\lambda=1}^{m}\left\{\sum_{i=1}^{n}\left(\theta^{\langle i\rangle}-h\right)S^{(\lambda)}_{\langle i\rangle}+\sum_{i=1}^{n-1}\left(\theta^{\langle i,i+1\rangle}-J\right)S^{(\lambda)}_{\langle i,i+1\rangle}-JS^{(\lambda)}_{\langle n\rangle}S^{(\lambda+1)}_{\langle 1\rangle}\right.\\ &\qquad\left.+\sum_{j>i+1}\theta^{\langle ij\rangle}S^{(\lambda)}_{\langle ij\rangle}+\sum_{|b|\geq 3}\theta^{b}S^{(\lambda)}_{b}\right\}-m\psi(\theta)+\tilde\psi(h,J)\\ &= \sum_{\lambda=1}^{m}\left\{L^{(\lambda)}-JS^{(\lambda)}_{\langle n\rangle}S^{(\lambda+1)}_{\langle 1\rangle}\right\}-m\psi(\theta)+\tilde\psi(h,J),\end{aligned}$$

where

$$L^{(\lambda)}:=\sum_{i=1}^{n}\left(\theta^{\langle i\rangle}-h\right)S^{(\lambda)}_{\langle i\rangle}+\sum_{i=1}^{n-1}\left(\theta^{\langle i,i+1\rangle}-J\right)S^{(\lambda)}_{\langle i,i+1\rangle}+\sum_{j>i+1}\theta^{\langle ij\rangle}S^{(\lambda)}_{\langle ij\rangle}+\sum_{|b|\geq 3}\theta^{b}S^{(\lambda)}_{b}.$$

The condition (9) is therefore rewritten as

$$E_{p_\theta}\left[\left(\sum_{\lambda=1}^{m}L^{(\lambda)}\right)\left(\sum_{\mu=1}^{m}\partial_a\ell^{(\mu)}\right)\right]=E_{p_\theta}\left[\left(\sum_{\lambda=1}^{m}JS^{(\lambda)}_{\langle n\rangle}S^{(\lambda+1)}_{\langle 1\rangle}\right)\left(\sum_{\mu=1}^{m}\partial_a\ell^{(\mu)}\right)\right].$$

Here we used the identity (7), as well as the fact that $\psi(\theta)$ and $\tilde\psi(h,J)$ are not random variables. Since random variables that belong to different clusters are independent, this is also equivalent to

$$\sum_{\lambda=1}^{m}E_{p_\theta}\left[L^{(\lambda)}\partial_a\ell^{(\lambda)}\right]=J\sum_{\lambda=1}^{m}E_{p_\theta}\left[S^{(\lambda)}_{\langle n\rangle}S^{(\lambda+1)}_{\langle 1\rangle}\left(\partial_a\ell^{(\lambda)}+\partial_a\ell^{(\lambda+1)}\right)\right]. \tag{10}$$

Let us first evaluate the right-hand side (RHS) of (10). By using the relation (8),

$$\begin{aligned}\text{(RHS)} &= J\sum_{\lambda=1}^{m}E_{p_\theta}\left[\left(\partial_{\langle n\rangle}\ell^{(\lambda)}+\eta_{\langle n\rangle}\right)\left(\partial_{\langle 1\rangle}\ell^{(\lambda+1)}+\eta_{\langle 1\rangle}\right)\left(\partial_a\ell^{(\lambda)}+\partial_a\ell^{(\lambda+1)}\right)\right]\\ &= J\sum_{\lambda=1}^{m}\left\{E_{p_\theta}\left[\left(\partial_{\langle n\rangle}\ell^{(\lambda)}+\eta_{\langle n\rangle}\right)\left(\partial_a\ell^{(\lambda)}\right)\right]E_{p_\theta}\left[\partial_{\langle 1\rangle}\ell^{(\lambda+1)}+\eta_{\langle 1\rangle}\right]\right.\\ &\qquad\left.+E_{p_\theta}\left[\partial_{\langle n\rangle}\ell^{(\lambda)}+\eta_{\langle n\rangle}\right]E_{p_\theta}\left[\left(\partial_{\langle 1\rangle}\ell^{(\lambda+1)}+\eta_{\langle 1\rangle}\right)\left(\partial_a\ell^{(\lambda+1)}\right)\right]\right\}\\ &= J\left(g_{a\langle n\rangle}\eta_{\langle 1\rangle}+g_{a\langle 1\rangle}\eta_{\langle n\rangle}\right)\\ &= Jg_{ab}\,y^b\end{aligned} \tag{11}$$

In a quite similar way, we can calculate the left-hand side (LHS) of (10) as follows.

$$\begin{aligned}\text{(LHS)} &= \sum_{i=1}^{n}g_{a\langle i\rangle}\left(\theta^{\langle i\rangle}-h\right)+\sum_{i=1}^{n-1}g_{a\langle i,i+1\rangle}\left(\theta^{\langle i,i+1\rangle}-J\right)+\sum_{j>i+1}g_{a\langle ij\rangle}\theta^{\langle ij\rangle}+\sum_{|b|\geq 3}g_{ab}\theta^b\\ &= g_{ab}\,x^b.\end{aligned}$$

In summary, (9) is equivalent to

$$g_{ab}\,x^b=Jg_{ab}\,y^b,\qquad(\forall a\in I).$$

Since the metric g is positive definite, this is also equivalent to

$$x^a=Jy^a,\qquad(\forall a\in I),$$

and the claim is verified. □

Lemma 3 implies that the condition (9) leads to

$$\hat{\theta}^{\langle 1\rangle} = h + J\hat{\eta}_{\langle n\rangle}$$
$$\hat{\theta}^{\langle n\rangle} = h + J\hat{\eta}_{\langle 1\rangle}$$
$$\hat{\theta}^{\langle 2\rangle} = \hat{\theta}^{\langle 3\rangle} = \cdots = \hat{\theta}^{\langle n-1\rangle} = h$$
$$\hat{\theta}^{\langle 12\rangle} = \hat{\theta}^{\langle 23\rangle} = \cdots = \hat{\theta}^{\langle n-1,n\rangle} = J$$

and the other parameters are all zero.

Remark 4. *The assumption $n \geq 2$ is used at the last equality of (11). When $n = 1$, the last line of (11) becomes $2Jg_{a\langle 1\rangle}\eta_{\langle 1\rangle}$. This leads to $\hat{\theta}^{\langle 1\rangle} = h + 2J\hat{\eta}_{\langle 1\rangle}$ as pointed out in Remark 2.*

Remark 5. *The argument presented in the proof of Lemma 3 can be generalized to obtain an analogous hereditary property of a Hamiltonian with longer range and/or many body interactions as long as it has periodicity n.*

To complete the proof of Theorem 1, we need the following

Lemma 6. *When $J > 0$, the simultaneous equations*

$$\begin{cases} \theta^{\langle 1\rangle} = h + J\eta_{\langle n\rangle} \\ \theta^{\langle n\rangle} = h + J\eta_{\langle 1\rangle} \end{cases}$$

with other values of parameters θ specified as above, lead to a unique symmetric solution

$$\theta^{\langle 1\rangle} = \theta^{\langle n\rangle}.$$

Proof. Observe that

$$\theta^{\langle 1\rangle} - \theta^{\langle n\rangle} = -J\left(\eta_{\langle 1\rangle} - \eta_{\langle n\rangle}\right) = -J\left(\frac{\partial}{\partial\theta^{\langle 1\rangle}} - \frac{\partial}{\partial\theta^{\langle n\rangle}}\right)\psi(\theta). \tag{12}$$

Let us transform the coordinate system $(\theta^{\langle 1\rangle}, \theta^{\langle n\rangle})$ into $(X, Y) := (\theta^{\langle 1\rangle} - \theta^{\langle n\rangle}, \theta^{\langle 1\rangle} + \theta^{\langle n\rangle})$. Then (12) is equivalent to

$$X = -2J\,\frac{\partial\psi}{\partial X}. \tag{13}$$

Since $\psi(\theta)$ is strictly convex in θ, it follows that $\partial^2\psi/\partial X^2 > 0$, so that the function $X \mapsto \partial\psi/\partial X$ is monotone increasing. Moreover, since the function $\psi(\theta)$ is symmetric in $\theta^{\langle 1\rangle}$ and $\theta^{\langle n\rangle}$, we see that

$$\left.\frac{\partial\psi}{\partial X}\right|_{X=0} = 0.$$

As a consequence, the equation (13) has a unique solution $X = 0$ for each Y. □

4 Discussions

As mentioned in Section 2, the foot of ∇^e-projection is not in general unique. One might associate this non-uniqueness with broken symmetry states as in the Weiss molecular field approximation. Since the exact result for the Ising spin chain does not show a phase transition [12], it is expected that this non-uniqueness would disappear as the cluster size n tends to infinity. Let us investigate this issue in more detail.

In order to determine the free parameter $\hat{\theta}^{\langle 1\rangle}$ through (6), we need to evaluate the expectation $E_{p_{\hat{\theta}}}[S_1]$. To concentrate on the issue of a phase transition, we let $h = 0$.

Lemma 7. *For $n \geq 2$,*

$$E_{p_{\hat{\theta}}}[S_1] = \frac{\sinh 2\hat{\theta}^{\langle 1 \rangle}}{\cosh 2\hat{\theta}^{\langle 1 \rangle} + a_n}, \tag{14}$$

where

$$a_n := 1 - \frac{2}{1 + \tanh^{1-n} \beta J}.$$

Proof. Let us evaluate the partition function

$$Z_n(\theta^{\langle 1 \rangle}, \theta^{\langle n \rangle}) := \sum_{\{S_i\}_{i=1}^n} \exp \left[\theta^{\langle 1 \rangle} S_1 + \theta^{\langle n \rangle} S_n + J \sum_{i=1}^{n-1} S_i S_{i+1} \right]. \tag{15}$$

Here J stands for βJ. For $n = 2$ and 3, it is easy to verify (14) by a direct calculation of (15). For $n \geq 4$, the following recursion formula holds.

$$\begin{aligned} Z_n(\theta^{\langle 1 \rangle}, \theta^{\langle n \rangle}) &= e^{\theta^{\langle 1 \rangle} + \theta^{\langle n \rangle}} Z_{n-2}(J, J) + e^{\theta^{\langle 1 \rangle} - \theta^{\langle n \rangle}} Z_{n-2}(J, -J) \\ &\quad + e^{-\theta^{\langle 1 \rangle} + \theta^{\langle n \rangle}} Z_{n-2}(-J, J) + e^{-\theta^{\langle 1 \rangle} - \theta^{\langle n \rangle}} Z_{n-2}(-J, -J) \\ &= 2 \cosh(\theta^{\langle 1 \rangle} + \theta^{\langle n \rangle}) Z_{n-2}(J, J) + 2 \cosh(\theta^{\langle 1 \rangle} - \theta^{\langle n \rangle}) Z_{n-2}(J, -J). \end{aligned} \tag{16}$$

Here the identities $Z_n(J, J) = Z_n(-J, -J)$ and $Z_n(J, -J) = Z_n(-J, J)$ are used. Letting

$$\mathbf{Z}_n := \begin{bmatrix} Z_n(J, J) \\ Z_n(J, -J) \end{bmatrix},$$

we have from (16) that

$$\mathbf{Z}_n = A \, \mathbf{Z}_{n-2}, \tag{17}$$

where

$$A := 2 \begin{bmatrix} \cosh 2J & 1 \\ 1 & \cosh 2J \end{bmatrix}.$$

The recursion (17) is easily solved by using

$$\mathbf{Z}_2 = \begin{bmatrix} e^{3J} + 3e^{-J} \\ e^{-3J} + 3e^{J} \end{bmatrix}, \qquad \mathbf{Z}_3 = \begin{bmatrix} e^{4J} + e^{-4J} + 6 \\ 4(e^{2J} + e^{-2J}) \end{bmatrix}.$$

In particular,

$$\frac{Z_n(J, -J)}{Z_n(J, J)} = 1 - \frac{2}{1 + \tanh^{-(n+1)} J} = a_{n+2}.$$

As a consequence, it follows from (16) that

$$\begin{aligned} E_{p_\theta}[S_1] &= \frac{1}{Z_n(\theta^{\langle 1 \rangle}, \theta^{\langle n \rangle})} \frac{\partial Z_n(\theta^{\langle 1 \rangle}, \theta^{\langle n \rangle})}{\partial \theta^{\langle 1 \rangle}} \bigg|_{\theta^{\langle n \rangle} = \theta^{\langle 1 \rangle}} \\ &= \frac{2 \sinh 2\theta^{\langle 1 \rangle} Z_{n-2}(J, J)}{2 \cosh 2\theta^{\langle 1 \rangle} Z_{n-2}(J, J) + 2 Z_{n-2}(J, -J)} \\ &= \frac{\sinh 2\theta^{\langle 1 \rangle}}{\cosh 2\theta^{\langle 1 \rangle} + a_n}. \end{aligned}$$

This completes the proof. □

When $h = 0$, the equation (6) that determines $\theta^{\langle 1 \rangle}$ becomes

$$\theta^{\langle 1 \rangle} = \beta J E_{p_\theta}[S_1]. \tag{18}$$

This equation always has a solution $\theta^{\langle 1\rangle}=0$. Let us investigate if there are other solutions. Since $a_n \nearrow 1$ as $n\to\infty$, the slope of the function $\theta^{\langle 1\rangle}\mapsto E_{p_\theta}[S_1]$ at $\theta^{\langle 1\rangle}=0$ exhibits

$$\frac{\partial}{\partial\theta^{\langle 1\rangle}}E_{p_\theta}[S_1]\bigg|_{\theta^{\langle 1\rangle}=0}=\frac{2}{1+a_n}\searrow 1\qquad (n\to\infty).$$

As a consequence, $\theta^{\langle 1\rangle}=0$ is the only solution of (18) from some natural number n onwards if and only if $\beta J<1$. When $\beta J\ge 1$, on the other hand, the equation (18) has two further solutions $\theta^{\langle 1\rangle}=\pm\theta_n$ $(0<\theta_n<\beta J)$ no matter how large n is.

Let $p_\pm^{(N,n)}$ denote the probability distributions on $\mathcal{M}_n$ that correspond to the solutions $\theta^{\langle 1\rangle}=\pm\theta_n$. The above observation implies that, if the temperature is low enough to satisfy $\beta J\ge 1$, there are two states $p_\pm^{(N,n)}$ on $\mathcal{M}_n$ $(\subset\mathcal{P}_N)$ that approximate the exact canonical state $q\in\mathcal{P}_N$ no matter how large N and n are. Apparently, this contradicts the fact that the Ising spin chain does not have a phase transition. We show that this paradox is resolved through a mathematical statistical investigation. For some technical reasons, we assume in what follows that N is even.

Let us consider the hypothesis testing problem [13, Chapter 4]:

$$p_+^{(N,n)}\quad\text{against}\quad p_-^{(N,n)}.$$

Let μ_N be the probability that we misjudge $p_-^{(N,n)}$ to be true when $p_+^{(N,n)}$ is actually true. Conversely, let λ_N be the probability that we misjudge $p_+^{(N,n)}$ to be true when $p_-^{(N,n)}$ is actually true. The probabilities μ_N and λ_N are called the *error probabilities of the first* and *the second kind*.

Due to symmetry, we restrict ourselves to the constraint that $\mu_N\to 0$ as $N\to\infty$. Under this constraint, λ_N is usually written in the form

$$\lambda_N\simeq e^{-NR}.$$

In statistics, it is customary to seek the largest such $R\ge 0$, which is sometimes referred to as the *supremum achievable error probability exponent*. According to [13, Theorem 4.1.1], it is given by the spectrum inf-divergence rate:

$$\underline{D}(p_+^{(N,n)}\|p_-^{(N,n)}):=\text{p-}\liminf_{N\to\infty}\frac{1}{N}\log\frac{p_+^{(N,n)}}{p_-^{(N,n)}},\tag{19}$$

where "p- lim inf" (the limit inferior in probability) for a sequence $\{X_N\}_N$ of real-valued random variables is defined by

$$\text{p-}\liminf_{N\to\infty}X_N:=\sup\left\{\alpha\,;\ \lim_{N\to\infty}P(X_N<\alpha)=0\right\}.$$

Let us evaluate the right-hand side of (19). Since $Z_n(\theta_n,\theta_n)=Z_n(-\theta_n,-\theta_n)$, the likelihood ratio is evaluated as

$$\left|\log\frac{p_+^{(N,n)}}{p_-^{(N,n)}}\right|=2m\theta_n|S_1+S_n|<4m\beta J,$$

or

$$\frac{1}{N}\left|\log\frac{p_+^{(N,n)}}{p_-^{(N,n)}}\right|<\frac{4\beta J}{n}.$$

Since this inequality holds for all $N\in\mathbb{N}$ and all factor n of N $(n<N)$, we have, by letting $n=N/2$, that

$$\lim_{N\to\infty}\frac{1}{N}\left|\log\frac{p_+^{(N,N/2)}}{p_-^{(N,N/2)}}\right|=0.$$

As a consequence, the error exponent (19) with $n = N/2$ is zero, which implies that there is no testing strategy for making the error probability λ_N decrease exponentially fast. More stringently, due to [13, Theorem 4.3.1], this hypothesis testing satisfies the strong converse property, in that $\mu_N \to 1$ as $N \to \infty$ under the constraint that $\lambda_N \le e^{-NR}$ for some (in fact, any) $R > 0$.

In summary, the probability measures $p_+^{(N,N/2)}$ and $p_-^{(N,N/2)}$ are, in effect, statistically indistinguishable for sufficiently large N. This fact could be paraphrased by saying that the states $p_+^{(N,N/2)}$ and $p_-^{(N,N/2)}$ are macroscopically identified.

5 Concluding remarks

We have studied the problem of approximating the canonical state q of an Ising spin chain by an i.i.d. cluster state p_θ that minimizes the relative entropy $D(p_\theta \| q)$, or equivalently, that minimizes the free energy $F(p_\theta)$. Theorem 1, the main result of this paper, states that the effective Hamiltonian corresponding to each cluster inherits the same structure as the original Hamiltonian. It is important to notice that equation (6) that determines the boundary term is different from the conventional self-consistency equation. This fact implies in particular that our method provides a tool for improved mean field calculations.

In the proof of Theorem 1, an information geometrical technique played an essential role. This strongly suggests its potential applicability to other problems in statistical physics such as the issue of a phase transition in two or higher dimensional Ising models, the renormalization group techniques, and the linear response theory. We address these issues in subsequent work.

References

[1] P. Weiss, J. de Phys., **6**, 661-690 (1907).

[2] W. L. Bragg and E. J. Williams, Proc. R. Soc. London, Ser. A **145**, 699-730 (1934); A **151**, 540-566 (1935); A **152**, 231-252 (1935).

[3] H. A. Bethe, Proc. R. Soc. London, Ser. A **150**, 552-575 (1935).

[4] R. Peierls, Proc. Camb. Phil. Soc., **32**, 477-481 (1936).

[5] M. Suzuki, J. Phys. Soc. Japan, **55**, 4205-4230 (1986).

[6] M. Katori and M. Suzuki, J. Phys. Soc. Japan, **56**, 3113-3125 (1987).

[7] T. Tanaka, Neural Computation, **12**, 1951-1968 (2000).

[8] S. Ikeda, T. Tanaka, and S.-I. Amari, Neural Computation, **16**, 1779-1810 (2004).

[9] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry* (AMS and Oxford, Providence, RI, 2000).

[10] S.-I. Amari, IEEE Trans. Inform. Theory, **47**, 1701-1711 (2001).

[11] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, I: Functional Analysis* (Academic, 1980).

[12] M. Plischke and B. Bergersen, *Equilibrium Statistical Physics*, 3rd edition (World Scientific, 2006).

[13] T. S. Han, *Information-Spectrum Methods in Information Theory* (Springer, Berlin, 2003).