Mean-field equations for higher-order quantum statistical models : an information geometric approach

N Yapage*

Department of Mathematics

University of Ruhuna,

Matara

Sri Lanka.

February 28, 2012

Abstract

This work is a simple extension of [1]. We apply the concepts of information geometry to study the mean-field approximation for a general class of quantum statistical models namely the higher-order quantum Boltzmann machines (QBMs). The states we consider are assumed to have at most third-order interactions with deterministic coupling coefficients. Such states, taken together, can be shown to form a quantum exponential family and thus can be viewed as a smooth manifold. In our work, we explicitly obtain naive mean-field equations for the third-order classical and quantum Boltzmann machines and demonstrate how some information geometrical concepts, particularly, exponential and mixture projections used to study the naive mean-field approximation in [1] can be extended to a more general case. Though our results do not differ much from those in [1], we emphasize the validity and the importance of information geometrical point of view for higher dimensional classical and quantum statistical models.

Keywords: mean-field theory, quantum statistical model, information geometry, quantum relative entropy, quantum exponential family

1 Introduction

The mean-field approximation uses a simple tractable family of density operators to calculate quantities related to a complex density operator including mutual interactions. Information geometry, on the other hand, studies intrinsic geometrical structure existing in the manifold of density operators [2]. Many authors have used mean-field approximation to classical statistical models like classical Boltzmann machines (CBMs) [3] and also have discussed the properties in the in the information geometrical point of view [4, 5]. In this work, we apply mean-field theory to the third-order CBMs and QBMs and derive the naive mean-field equations using information geometrical concepts.

2 Information geometry of mean-field approximation for third-order CBMs

Let us consider a network of n elements numbered as $1, 2, \ldots, n$. Let the value of each element $i \in \{1, 2, \ldots, n\}$ be $x_i \in \{-1, +1\}$. Then a state of the network can be represented

---

*nihal@maths.ruh.ac.lk

as $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \{-1, +1\}^n$. Each element $i \in \{1, 2, \ldots, n\}$ carries a threshold value $\theta_i \in \mathbb{R}$. The network also has a real-valued parameter w_{ij} for each pair of elements $\{i, j\}$, which is called the coupling coefficient between i and j. These parameters are assumed to satisfy conditions $w_{ij} = w_{ji}$, $w_{ii} = 0$. The other real-valued parameter is v_{ijk} and is symmetric on all pairs of indices. The equilibrium (stationary) distribution is given by the probability distributions of the form

$$p(\boldsymbol{x}, h, w, v) = \exp\Big\{\sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j + \sum_{i<j<k} v_{ijk} x_i x_j x_k - \psi(h, w, v)\Big\} \quad (1)$$

with

$$\psi(h, w, v) = \log \sum_{\boldsymbol{x}} \exp\Big\{\sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j + \sum_{i<j<k} v_{ijk} x_i x_j x_k\Big\}, \quad (2)$$

where $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{-1, +1\}^n$. Thus, noting that the correspondence $p_{h,w,v} \leftrightarrow (h, w, v)$ is one to one, we can, at least mathematically, identify each third-order CBM [6] with its equilibrium probability distribution.

Many good properties of such networks are consequences of the fact that the equilibrium distributions form an exponential family. Here, we discuss this important aspect of the CBM [7] briefly. Let $\mathcal{X}$ be a finite set or, more generally, a measurable space with an underlying measure $d\mu$. We denote the set of positive probability distributions (probability mass functions for a finite $\mathcal{X}$ and probability density functions for a general $(\mathcal{X}, d\mu)$) on $\mathcal{X}$ by $\mathcal{P} = \mathcal{P}(\mathcal{X})$. When a family of distributions, say

$$\mathcal{M} = \{p_\theta \mid \theta = (\theta^i); \ i = 1, \ldots, n\} \subset \mathcal{P}, \quad (3)$$

is represented in the form

$$p_\theta(x) = \exp\Big\{c(x) + \sum_{i=1}^n \theta^i f_i(x) - \psi(\theta)\Big\}, \quad x \in \mathcal{X}, \quad (4)$$

$\mathcal{M}$ is called an exponential family. Here, $\theta^i; \ i = 1, \ldots, n$ are $\mathbb{R}$-valued parameters, c and f_i are functions on $\mathcal{X}$ and $\psi(\theta)$ is a real-valued convex function. Further, we assume that the correspondence $\theta \mapsto p_\theta$ is one to one. These $\theta = (\theta^i)$ are called the natural coordinates of $\mathcal{M}$.

Now, for the exponential family $\mathcal{M}$, if we let

$$\eta_i(\theta) \stackrel{\text{def}}{=} \mathbb{E}_\theta[f_i] = \sum_{\boldsymbol{x}} p_\theta(\boldsymbol{x}) f_i(\boldsymbol{x})$$

then $\eta = (\eta_i)$ and $\theta = (\theta^i)$ are in one-to-one correspondence. That is, we can also use η instead of θ to specify an element of $\mathcal{M}$. These (η_i) are called the expectation coordinates of $\mathcal{M}$. The expectation coordinates are, in general, represented as

$$\eta_i = \partial_i \psi(\theta) \qquad \Big(\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \theta^i}\Big). \quad (5)$$

The set that consists of equilibrium probability distributions of third-order CBM (1) is one example of exponential family. In addition, threshold values, coupling coefficients (weights) and third-oder weights become the natural coordinates while $\mathbb{E}_\theta[x_i]$, $\mathbb{E}_\theta[x_i x_j]$ and $\mathbb{E}_\theta[x_i x_j x_k]$ become expectation coordinates. The notion of exponential family is very important in statistics and information geometry, and is also useful in studying properties of third-order CBMs with their mean-field approximations.

We now consider a hierarchy of exponential families. Let $\mathcal{P}_r$ be the set of rth-order CBMs. Then $\mathcal{P}_r$ also turns out to be an exponential family. Thus, we have a hierarchical structure of exponential families $\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_r$. In particular, $\mathcal{P}_1, \mathcal{P}_2$ and $\mathcal{P}_3$ can be represented by

$$
\begin{aligned}
\mathcal{P}_1 &= \{p_1(x_1)\cdots p_n(x_n)\} = \{\text{product distributions}\}, \\
\mathcal{P}_2 &= \{\text{equilibrium distributions of CBM}\}
\end{aligned}
\tag{6}
$$

and

$$
\mathcal{P}_3 = \{\text{equilibrium distributions of third-order CBM}\}
\tag{7}
$$

respectively.

In this subsection, we derive the naive mean-field equation for third-order CBM. When the system size is large, the partition function $\exp(\psi(h,w,v))$ is very difficult to calculate and thus explicit calculation of the expectations m_i is intractable. Therefore, due to that difficulty, we are led to obtain a good approximation of m_i for a given probability distribution $p_{h,w,v} \in \mathcal{P}_3$.

First, we consider the subspace $\mathcal{P}_1$ of $\mathcal{P}_3$. We parametrize each distribution in $\mathcal{P}_1$ by $\bar{h}$ and write as

$$
p_{\bar{h}}(\boldsymbol{x}) = \exp\left\{ \sum_i \bar{h}_i x_i - \psi(\bar{h}) \right\},
\tag{8}
$$

where

$$
\psi(\bar{h}) = \sum_i \log\left\{ \exp(\bar{h}_i) + \exp(-\bar{h}_i) \right\}.
$$

Then, $\mathcal{P}_1$ forms a submanifold of $\mathcal{P}_3$ specified by $w_{ij} = 0 = v_{ijk}$ and $\bar{h}_i$ as its coordinates. The expectations $\bar{m}_i := \mathbb{E}_{\bar{h}}[x_i]$ form another coordinate system of $\mathcal{P}_1$. For a given $p_{\bar{h}} \in \mathcal{P}_1$, it is easy to obtain $\bar{m}_i = \mathbb{E}_{\bar{h}}[x_i]$ from $\bar{h}_i$ because x_i's are independent. We can calculate $\bar{m}_i$ to be

$$
\bar{m}_i = \frac{\partial \psi(\bar{h})}{\partial \bar{h}_i} = \frac{\exp(\bar{h}_i) - \exp(-\bar{h}_i)}{\exp(\bar{h}_i) + \exp(-\bar{h}_i)} = \tanh(\bar{h}_i),
\tag{9}
$$

from which we obtain

$$
\bar{h}_i = \frac{1}{2} \log\left(\frac{1 + \bar{m}_i}{1 - \bar{m}_i} \right).
\tag{10}
$$

The simple idea behind the mean field approximation for a $p_{h,w,v} \in \mathcal{P}_3$ is to use quantities obtained in the form of expectation with respect to some relevant $p_{\bar{h}} \in \mathcal{P}_1$.

Now, we need a suitable criterion to measure the approximation of two probability distributions $q \in \mathcal{P}$ and $p_\theta \in \mathcal{M}$. For the present purpose, we adopt the Kullback-Leibler (KL) divergence (relative entropy)

$$
D(q \| p_\theta) \overset{\text{def}}{=} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p_\theta(\boldsymbol{x})}
\tag{11}
$$

. Given $p_{h,w,v} \in \mathcal{P}_3$, its e-(exponential) and m-(mixture) projections (see [2]) onto $\mathcal{P}_1$ are defined by

$$
\bar{p}^{(e)} = p_{\bar{h}^{(e)}} \overset{\text{def}}{=} \underset{p_{\bar{h}} \in \mathcal{P}_1}{\arg\min}\, D(p_{\bar{h}} \| p_\theta)
\tag{12}
$$

and

$$
\bar{p}^{(m)} = p_{\bar{h}^{(m)}} \overset{\text{def}}{=} \underset{p_{\bar{h}} \in \mathcal{P}_1}{\arg\min}\, D(p_\theta \| p_{\bar{h}})
\tag{13}
$$

respectively, where

$$
\bar{h}^{(e)} = \underset{\bar{h} = (\bar{h}_i)}{\arg\min}\, D(p_{\bar{h}} \| p_\theta)
\tag{14}
$$

3

and

$$\bar{h}^{(m)} = \arg\min_{\bar{h}=(\bar{h}_i)} D(p_\theta \| p_{\bar{h}}). \tag{15}$$

As necessary conditions, we have

$$\frac{\partial}{\partial \bar{h}_i} D(p_{\bar{h}} \| p_\theta) = 0 \tag{16}$$

and

$$\frac{\partial}{\partial \bar{h}_i} D(p_\theta \| p_{\bar{h}}) = 0, \tag{17}$$

which are weaker than (14) and (15). But sometimes (16) and (17) are chosen to be the definitions of e-, m- projections respectively for convenience. It can be shown that the m-projection $\bar{p}^{(m)}$ gives the true values of expectations, that is $m_i = \bar{m}_i$ or $\mathbb{E}_{(\theta)}[x_i] = \mathbb{E}_{\bar{h}}[x_i]$ for $p_{\bar{h}} = \bar{p}^{(m)}$. The e-projection $\bar{p}^{(e)}$ from $\mathcal{P}_3$ onto $\mathcal{P}_1$ gives the naive mean-field approximation for third-order CBM. Now we derive the naive mean-field equation for third-order CBM following [4]. Recall that the equilibrium distribution for third-order CBM is given by

$$p \stackrel{\text{def}}{=} p(\boldsymbol{x}, h, w, v) = \exp\Big\{ \sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j + \sum_{i<j<k} v_{ijk} x_i x_j x_k - \psi(h, w, v) \Big\} \tag{18}$$

with

$$\psi(h, w, v) = \log \sum_{\boldsymbol{x}} \exp\Big\{ \sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j + \sum_{i<j<k} v_{ijk} x_i x_j x_k \Big\}, \tag{19}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{-1, +1\}^n$. Now we define another function

$$\phi(p) = \phi(h, w, v) \stackrel{\text{def}}{=} \sum_{i<j<k} v_{ijk} \mathbb{E}_p[x_i x_j x_k] + \sum_{i<j} w_{ij} \mathbb{E}_p[x_i x_j] + \sum_i h_i \mathbb{E}_p[x_i] - \psi(p), \tag{20}$$

which coincides with the negative entropy

$$\phi(p) = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \log p(\boldsymbol{x}). \tag{21}$$

In particular, for a product distribution $p_{\bar{h}} \in \mathcal{P}_1$, using $\mathbb{E}_{\bar{h}}[x_i] = \bar{m}_i$, we have

$$\phi(p_{\bar{h}}) = \sum_i \left[\left(\frac{1 + \bar{m}_i}{2} \right) \log\left(\frac{1 + \bar{m}_i}{2} \right) + \left(\frac{1 - \bar{m}_i}{2} \right) \log\left(\frac{1 - \bar{m}_i}{2} \right) \right]. \tag{22}$$

The KL divergence between $p \in \mathcal{P}_3$ and $p_{\bar{h}} \in \mathcal{P}_1$ can be expressed in the following form:

$$
\begin{aligned}
D(p_{\bar{h}} \| p) &= \psi(p) + \phi(p_{\bar{h}}) - \sum_{i<j<k} v_{ijk} \mathbb{E}_p[x_i x_j x_k] - \sum_{i<j} w_{ij} \mathbb{E}_p[x_i x_j] - \sum_i h_i \mathbb{E}_{\bar{h}}[x_i] \\
&= \psi(p) + \phi(p_{\bar{h}}) - \sum_{i<j<k} v_{ijk} \bar{m}_i \bar{m}_j \bar{m}_k - \sum_{i<j} w_{ij} \bar{m}_i \bar{m}_j - \sum_i h_i \bar{m}_i. \\
&= \psi(p) + \frac{1}{2} \sum_i \left[(1 + \bar{m}_i) \log\left(\frac{1 + \bar{m}_i}{2} \right) + (1 - \bar{m}_i) \log\left(\frac{1 - \bar{m}_i}{2} \right) \right] \\
&\quad - \sum_{i<j<k} v_{ijk} \bar{m}_i \bar{m}_j \bar{m}_k - \sum_{i<j} w_{ij} \bar{m}_i \bar{m}_j - \sum_i h_i \bar{m}_i. \tag{23}
\end{aligned}
$$

4

Now consider the e-projection (16) from $p \in \mathcal{P}_3$ onto $p_{\bar{h}} \in \mathcal{P}_1$, i.e.

$$\frac{\partial}{\partial \bar{h}_i} D(p_{\bar{h}} \| p) = 0. \tag{24}$$

Noting that $\bar{h}$ and $\bar{m}$ are in one-to-one correspondence, we may consider instead

$$\frac{\partial}{\partial \bar{m}_i} D(p_{\bar{h}} \| p) = 0. \tag{25}$$

Since $\psi(p)$ does not depend on $\bar{m}_i$, we obtain from (23) that

$$\begin{aligned}
0 &= \frac{1}{2} \log\left(\frac{1 + \bar{m}_i}{1 - \bar{m}_i}\right) - \sum_{k \neq j \neq i} v_{ijk} \bar{m}_j \bar{m}_k - \sum_{j \neq i} w_{ij} \bar{m}_j - h_i \\
&= \bar{h}_i - \sum_{k \neq j \neq i} v_{ijk} \bar{m}_j \bar{m}_k - \sum_{j \neq i} w_{ij} \bar{m}_j - h_i,
\end{aligned} \tag{26}$$

where the second equality is from (10). Thus the naive mean-field equation is obtained from (9) and (26) as

$$\tanh^{-1}(\bar{m}_i) = \sum_{k \neq j \neq i} v_{ijk} \bar{m}_j \bar{m}_k + \sum_{j \neq i} w_{ij} \bar{m}_j + h_i \tag{27}$$

and this is usually written in the form

$$\bar{m}_i = \tanh\left(\sum_{k \neq j \neq i} v_{ijk} \bar{m}_j \bar{m}_k + \sum_{j \neq i} w_{ij} \bar{m}_j + h_i\right). \tag{28}$$

3 Definition of third-order quantum Boltzmann machines

Let us consider an n-element system of quantum spin-half particles. Each element is represented as a quantum spin with local Hilbert space $\mathbb{C}^2$, and the n-element system corresponds to $\mathcal{H} \equiv (\mathbb{C}^2)^{\otimes n} \simeq \mathbb{C}^{2^n}$. Let $\mathcal{S}$ be the set of strictly positive states on $\mathcal{H}$;

$$\mathcal{S} = \{\rho \mid \rho = \rho^* > 0 \text{ and } \operatorname{Tr} \rho = 1\}. \tag{29}$$

Here, each ρ is a $2^n \times 2^n$ matrix; $\rho = \rho^* > 0$ means that ρ is Hermitian and positive definite respectively; and $\operatorname{Tr} \rho = 1$ shows that the trace of the density matrix ρ is unity. Now an element of $\mathcal{S}$ is said to have at most rth-order interactions if it is written as

$$\rho_\theta = \exp\left\{\sum_{i,s} \theta_{is}^{(1)} \sigma_{is} + \sum_{i<j} \sum_{s,t} \theta_{ijst}^{(2)} \sigma_{is} \sigma_{jt} + \cdots + \sum_{i_1 < \cdots < i_r} \sum_{s_1 \ldots s_r} \theta_{i_1 \ldots i_r s_1 \ldots s_r}^{(r)} \sigma_{i_1 s_1} \cdots \sigma_{i_r s_r} - \psi(\theta)\right\}$$

$$= \exp\left\{\sum_{j=1}^r \sum_{i_1 < \cdots < i_j} \sum_{s_1 \ldots s_j} \theta_{i_1 \ldots i_j s_1 \ldots s_j}^{(j)} \sigma_{i_1 s_1} \cdots \sigma_{i_j s_j} - \psi(\theta)\right\} \tag{30}$$

with

$$\psi(\theta) = \log \operatorname{Tr} \exp\left\{\sum_{j=1}^r \sum_{i_1 < \cdots < i_j} \sum_{s_1 \ldots s_j} \theta_{i_1 \ldots i_j s_1 \ldots s_j}^{(j)} \sigma_{i_1 s_1} \cdots \sigma_{i_j s_j}\right\}, \tag{31}$$

where $\sigma_{is} = I^{\otimes(i-1)} \otimes \sigma_s \otimes I^{\otimes(n-i)}$, $\theta = (\theta_{i_1 \ldots i_j s_1 \ldots s_j}^{(j)})$. Here, I is the identity matrix on $\mathcal{H}$ and σ_s for $s \in \{1, 2, 3\}$ are the usual Pauli matrices given by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Letting $\mathcal{S}_r$ be the totality of states ρ_θ of the above form, we have the hierarchy $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \mathcal{S}_3 \subset \cdots \subset \mathcal{S}_n = \mathcal{S}$. Note that $\mathcal{S}_1$ is the set of product states $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$.

Corresponding to the classical case, an element of $\mathcal{S}_2$ is called a QBM (see [1]). The third-order quantum Boltzmann machines are given by the elements of $\mathcal{S}_3$ and those states can be explicitly written as

$$\rho_{h,w,v} = \exp\Big\{\sum_{i,s} h_{is}\sigma_{is} + \sum_{i<j}\sum_{s,t} w_{ijst}\sigma_{is}\sigma_{jt} + \sum_{i<j<k}\sum_{s,t,u} v_{ijkstu}\sigma_{is}\sigma_{jt}\sigma_{jt} - \psi(h,w,v)\Big\} \quad (32)$$

with

$$\psi(h,w,v) = \log \operatorname{Tr} \exp\Big\{\sum_{i,s} h_{is}\sigma_{is} + \sum_{i<j}\sum_{s,t} w_{ijst}\sigma_{is}\sigma_{jt} + \sum_{i<j<k}\sum_{s,t,u} v_{ijkstu}\sigma_{is}\sigma_{jt}\Big\}, \quad (33)$$

where $h = (h_{is}), w = (w_{ijst})$ and $v = (v_{ijkstu})$.

4 Some information geometrical concepts for quantum systems

We discuss in this section some information geometrical concepts for quantum systems [2]. Let us consider a manifold $\mathcal{S}$ of density operators and a submanifold $\mathcal{M}$ of $\mathcal{S}$. We define a quantum divergence function from $\rho \in \mathcal{S}$ to $\sigma \in \mathcal{S}$, which in this case turns out to be the quantum relative entropy and its reverse represented by

$$D^{(-1)}(\rho\|\sigma) \stackrel{\text{def}}{=} \operatorname{Tr}[\rho(\log \rho - \log \sigma)]; \quad D^{(+1)}(\rho\|\sigma) \stackrel{\text{def}}{=} \operatorname{Tr}[\sigma(\log \sigma - \log \rho)]. \quad (34)$$

The quantum relative entropy satisfies $D^{(\pm 1)}(\rho\|\sigma) \geq 0, \qquad D(\rho\|\sigma) = 0 \quad \text{iff} \quad \rho = \sigma$ but it is not symmetric.

Given $\rho \in \mathcal{S}$, the point $\tau^{(\pm 1)} \in \mathcal{M}$ is called the e, m-projection of ρ to $\mathcal{M}$, when function $D^{(\pm 1)}(\rho\|\tau)$, $\tau \in \mathcal{M}$ takes a critical value at $\tau^{(\pm 1)}$, that is

$$\frac{\partial}{\partial \xi} D^{(\pm 1)}(\rho\|\tau(\xi)) = 0 \quad (35)$$

at $\tau^{(\pm 1)}$ where ξ is a coordinate system of $\mathcal{M}$. the minimizer of $D^{(\pm 1)}(\rho\|\tau)$, $\tau \in \mathcal{M}$, is the ± 1-projection of ρ to $\mathcal{M}$.

Next we introduce a quantum version of exponential family (4) in the following. Suppose that a parametric family

$$\mathcal{M} = \{\rho_\theta \,|\, \theta = (\theta^i); \; i = 1, \dots, m\} \subset \mathcal{S} \quad (36)$$

is represented in the form

$$\rho_\theta = \exp\Big\{C + \sum_{i=1}^{m} \theta^i F_i - \psi(\theta)\Big\}, \quad (37)$$

where $F_i \, (i = 1, \dots, m), C$ are Hermitian operators and $\psi(\theta)$ is a real-valued function. We assume in addition that the operators $\{F_1, \dots, F_m, I\}$, where I is the identity operator, are linearly independent to ensure that the parametrization $\theta \mapsto \rho_\theta$ is one to one. Then $\mathcal{M}$ forms an m-dimensional smooth manifold with a coordinate system $\theta = (\theta^i)$. In this thesis, we call such an $\mathcal{M}$ a quantum exponential family or QEF for short, with natural coordinates $\theta = (\theta^i)$. Note also that for any $1 \leq k \leq n$ the set $\mathcal{S}_k$ of states (30) forms a QEF, including $\mathcal{S}_1$ of product states, $\mathcal{S}_2$ of QBMs and $\mathcal{S}_3$ of third-order QBMs.

If we let

$$\eta_i(\theta) \overset{\text{def}}{=} \text{Tr}[\rho_\theta F_i], \tag{38}$$

then $\eta = (\eta_i)$ and $\theta = (\theta^i)$ are in one-to-one correspondence. That is, we can also use η instead of θ to specify an element of $\mathcal{M}$. These (η_i) are called the expectation coordinates of $\mathcal{M}$.

In particular, the natural coordinates of $\mathcal{S}_3$ are given by $(h, w, v) = (h_{is}, w_{ijst}, v_{ijkstu})$ in (32), while the expectation coordinates are $(m, \mu, \iota) = (m_{is}, \mu_{ijst}, \iota_{ijkstu})$ defined by

$$m_{is} = \text{Tr}[\rho_{h,w,v}\,\sigma_{is}] \quad \text{and} \quad \mu_{ijst} = \text{Tr}[\rho_{h,w,v}\,\sigma_{is}\sigma_{jt}] \text{and} \quad \iota_{ijkstu} = \text{Tr}[\rho_{h,w,v}\,\sigma_{is}\sigma_{jt}\sigma_{ku}]. \tag{39}$$

On the other hand, the natural coordinates of $\mathcal{S}_1$ are $\bar{h} = (\bar{h}_{is})$ in (43), while the expectation coordinates are $\bar{m} = (\bar{m}_{is})$ defined by

$$\bar{m}_{is} = \text{Tr}[\tau_{\bar{h}}\sigma_{is}]. \tag{40}$$

In this case, the correspondence between the two coordinate systems can explicitly be represented as

$$\bar{m}_{is} = \frac{\partial \psi_i(\bar{h}_i)}{\partial \bar{h}_{is}} = \frac{\bar{h}_{is}}{||\bar{h}_i||}\tanh(||\bar{h}_i||) \tag{41}$$

or as

$$\bar{h}_{is} = \frac{\bar{m}_{is}}{||\bar{m}_i||}\tanh^{-1}(||\bar{m}_i||), \tag{42}$$

where $||\bar{m}_i|| \overset{\text{def}}{=} \sqrt{\sum_s (\bar{m}_{is})^2}$.

5 Information geometry of mean-field approximation for third-order QBMs

5.1 The submanifold of product states and its geometry

In this section, we briefly discuss the set $\mathcal{S}_1$. The elements of $\mathcal{S}_1$ are represented as τ_h by letting $w = 0$ and $v = 0$ in (32). In the sequel, we write them as

$$\tau_{\bar{h}} = \exp\left\{\sum_{i,s} \bar{h}_{is}\sigma_{is} - \psi(\bar{h})\right\} \tag{43}$$

by using new symbols τ and $\bar{h} = (\bar{h}_{is})$ when we wish to make it clear that we are treating $\mathcal{S}_1$ instead of $\mathcal{S}_3$. We have

$$\tau_{\bar{h}} = \bigotimes_{i=1}^{n} \exp\left\{\sum_s \bar{h}_{is}\sigma_s - \psi_i(\bar{h}_i)\right\}, \tag{44}$$

where $\bar{h}_i = (\bar{h}_{is})_s$ and

$$\begin{aligned}
\psi_i(\bar{h}_i) &= \log \text{Tr}\exp\left\{\sum_s h_{is}\sigma_s\right\} \\
&= \log\{\exp(||\bar{h}_i||) + \exp(-||\bar{h}_i||)\}
\end{aligned} \tag{45}$$

with $||\bar{h}_i|| \overset{\text{def}}{=} \sqrt{\sum_s (\bar{h}_{is})^2}$. Note that

$$\psi(\bar{h}) = \sum_i \psi_i(\bar{h}_i). \tag{46}$$

5.2 The exponential & mixture projections and mean-field approximation

In this section, we derive the naive mean-field equation for third-order QBMs explicitly from the viewpoint of information geometry. Suppose that we are interested in calculating the expectations $m_{is} = \text{Tr}[\rho_{h,w}X_{is}]$ from given $(h,w) = (h_{is}, w_{ijst})$. Since the direct calculation is intractable in general when the system size is large, we need to employ a computationally efficient approximation method. Mean-field approximation is a well-known technique for this purpose. The simple idea behind the mean-field approximation for a $\rho_{h,w,v} \in \mathcal{S}_3$ is to use quantities obtained in the form of expectation with respect to some relevant $\tau_{\bar{h}} \in \mathcal{S}_1$. T. Tanaka [4] has elucidated the essence of the naive mean-field approximation for classical spin models in terms of e-, m-projections. Our aim is to extend this idea to quantized spin models other than that considered in [1].

In the following arguments, we regard $\mathcal{S}_3$ as a QEF with the natural coordinates $(\theta^\alpha) = (h_{is}, w_{ijst}, v_{ijkstu})$ and the expectation coordinates $(\eta_\alpha) = (m_{is}, \mu_{ijst}, \iota_{ijkstu})$, where α is an index denoting $\alpha = (i,s)$ or $\alpha = (i,j,s,t)$ or $\alpha = (i,j,k,s,t,u)$. We follow a slightly different method to that of classical setting to obtain naive mean-field equation for third-order QBM.

Recall that the state for third-order QBM (32) is given by

$$\rho_{h,w,v} = \exp\Big\{\sum_{i,s} h_{is}\sigma_{is} + \sum_{i<j}\sum_{s,t} w_{ijst}\sigma_{is}\sigma_{jt} + \sum_{i<j<k}\sum_{s,t,u} v_{ijkstu}\sigma_{is}\sigma_{jt}\sigma_{jt} - \psi(h,w,v)\Big\} \quad (47)$$

with

$$\psi(h,w,v) = \log \text{Tr}\exp\Big\{\sum_{i,s} h_{is}\sigma_{is} + \sum_{i<j}\sum_{s,t} w_{ijst}\sigma_{is}\sigma_{jt} + \sum_{i<j<k}\sum_{s,t,u} v_{ijkstu}\sigma_{is}\sigma_{jt}\sigma_{kt}\Big\}, \quad (48)$$

where $h = (h_{is})$, $w = (w_{ijst})$ and $v = (v_{ijkstu})$.

Given $\rho_{h,w,v} \in \mathcal{S}_3$, its e-(+1) and m-(-1) projections (see [2]) onto $\mathcal{S}_1$ are defined by

$$\bar{\tau}^{(\pm 1)} = \tau_{\bar{h}^{(\pm 1)}} \stackrel{\text{def}}{=} \underset{\tau_{\bar{h}} \in \mathcal{S}_1}{\arg\min} D(\rho_{h,w,v}\|\tau_{\bar{h}}). \quad (49)$$

We denote by $\bar{m}_{is}^{(\pm 1)}[\rho_{h,w,v}]$ the expectation of σ_{is} with respect to $\bar{\tau}^{(\pm 1)}$, that is $\text{Tr}[\bar{\tau}^{(\pm 1)}\sigma_{is}]$. Then $\bar{m}_{is}^{(\pm 1)}[\rho_{h,w,v}]$ is given by

$$\frac{\partial}{\partial \bar{m}_i s} D^{(\pm 1)}(\rho_{h,w,v}\|\tau_{\bar{h}}) = 0. \quad (50)$$

From the information geometrical point of view, $\bar{\tau}^{(\pm 1)} \in \mathcal{S}_1$ is the ± 1-geodesic projection of $\rho_{h,w,v}$ to $\mathcal{S}_1$ in the sense that the ± 1-geodesic connecting $\rho_{h,w,v}$ and $\bar{\tau}$ is orthogonal to $\mathcal{S}_1$ at $\bar{\tau} = \bar{\tau}^{(\pm 1)}$. but we know that $\bar{\tau}^{(-1)}$ is m-projection of $\rho_{h,w,v}$ to $\mathcal{S}_1$. then we have $\bar{m}_{is}^{(-1)}\bar{m}_{is}[\rho_{h,w,v}]$ which is the quantity we want to obtain. This relation can be directly calculated by solving

$$\frac{\partial}{\partial \bar{m}_i s} D^{(-1)}(\rho_{h,w,v}\|\tau_{\bar{h}}) = 0, \quad (51)$$

because this is equivalent to

$$\frac{\partial}{\partial \bar{m}_i s} \text{Tr}[\rho_{h,w,v}(\log \rho_{h,w,v} - \log \bar{\tau})] \frac{\partial}{\partial \bar{m}_i s} \text{Tr}[\rho_{h,w,v}\log \bar{\tau}] = 0. \quad (52)$$

Hence $\bar{m}_{is}^{(-1)} = \text{Tr}[\rho_{h,w,v}\sigma_{is}]$ which is the quantity we have been searching for. But we cannot calculate $\text{Tr}[\rho_{h,w,v}\sigma_{is}]$ explicitly due to the difficulty in calculating $\psi(h,w,v)$ for $\rho_{h,w,v}$.

If we use the *e*-projection of $\rho \in \mathcal{S}_3$ to $\mathcal{S}_1$ instead of the *m*-projection, we have the naive mean-field approximation as in the classical case. to show this we calculate the *e*-projection (1-projection) of ρ to $\mathcal{S}_1$. Then we have

$$
D^{(1)}(\rho\|\tau) = D^{(-1)}(\tau\|\rho) = \mathrm{Tr}[\tau(\log\tau - \log\rho)] \tag{53}
$$

$$
= \mathrm{Tr}\Bigg[\tau\Bigg\{\Bigg(\sum_{i,s}\bar{h}_{is}\sigma_{is} - \psi(\bar{h})\Bigg)
$$

$$
-\Bigg(\sum_{i,s}h_{is}\sigma_{is} + \sum_{i<j}\sum_{s,t}w_{ijst}\sigma_{is}\sigma_{jt} + \sum_{i<j<k}\sum_{s,t,u}v_{ijkstu}\sigma_{is}\sigma_{jt}\sigma_{jt} - \psi(h,w,v)\Bigg)\Bigg\}\Bigg] \tag{54}
$$

$$
= \sum_{i,s}\bar{h}_{is}\bar{m}_{is} - \psi(\bar{h}) - \sum_{i,s}h_{is}\bar{m}_{is} + \sum_{i<j}\sum_{s,t}w_{ijst}\bar{m}_{is}\bar{m}_{jt}
$$

$$
+ \sum_{i<j<k}\sum_{s,t,u}v_{ijkstu}\bar{m}_{is}\bar{m}_{jt}\bar{m}_{ku} - \psi(h,w,v), \tag{55}
$$

where we define $\bar{m}_{is} = \mathrm{Tr}[\tau\sigma_{is}], \bar{m}_{is}\bar{m}_{jt} = \mathrm{Tr}[\tau\sigma_{is}\sigma_{jt}]$ and $\bar{m}_{is}\bar{m}_{jt}\bar{m}_{ku} = \mathrm{Tr}[\tau\sigma_{is}\sigma_{jt}\sigma_{ku}]$. Hence

$$
\frac{\partial}{\partial\bar{m}_i s}D^{(1)} = \frac{\partial}{\partial\bar{m}_i s}\Bigg[\sum_{i,s}\bar{h}_{is}\bar{m}_{is} - \psi(\bar{h})\Bigg] - h_{is} + \sum_{j\neq i}\sum_{s,t}w_{ijst}\bar{m}_{jt} + \sum_{k\neq j\neq k}\sum_{s,t,u}v_{ijkstu}\bar{m}_{jt}\bar{m}_{ku} \tag{56}
$$

$$
= \bar{h}_{is} - h_{is} + \sum_{j\neq i}\sum_{s,t}w_{ijst}\bar{m}_{jt} + \sum_{k\neq j\neq k}\sum_{s,t,u}v_{ijkstu}\bar{m}_{jt}\bar{m}_{ku} = 0. \tag{57}
$$

This gives

$$
\bar{h}_{is} = h_{is} + \sum_{j\neq i}\sum_{s,t}w_{ijst}\bar{m}_{jt} + \sum_{k\neq j\neq k}\sum_{s,t,u}v_{ijkstu}\bar{m}_{jt}\bar{m}_{ku} \tag{58}
$$

and $\bar{m}_{is}$ is given by

$$
\bar{m}_{is} = \frac{\partial\psi_i(\bar{h}_i)}{\partial\bar{h}_{is}} = \frac{\bar{h}_{is}}{||\bar{h}_i||}\tanh(||\bar{h}_i||). \tag{59}
$$

Both (58) and (59) together give the naive mean-field equations for third-order QBMs.

6   Concluding remarks

We have applied information geometry to the mean-field approximation for a general class of quantum statistical models. Here, we were able to derive only the naive mean-field equations. However, it is known that the naive mean-field approximation does not give a good approximation to the true value. Therefore, to improve the approximation we need to consider the higher order approximations and the information geometrical point of view is left open.

References

[1] N. Yapage, H. Nagaoka. An information geometrical approach to the mean-field approximation for quantum Ising spin models. *J. Phys. A: Math. Theor.* **41** (2008) 065005.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society and Oxford University Press, 2000.

[3] Manfred Opper and David Saad. *Advanced Mean Field Methods - Theory and Practice*, MIT Press, Cambridge, MA, 2001.

[4] T. Tanaka, Information Geometry of mean field Approximation. *Neural Computation*, **12** pp.1951–1968, 2000.

[5] S. Amari, K. Kurata and H. Nagaoka. Information Geometry of Boltzmann Machines. *IEEE Trans. on Neural Networks* Vol 3, No. 2, pp.260–271, 1992.

[6] T. J. Sejnowski. Higher-order Boltzmann Machines. *Conference Proceedings 151: Neural Networks for Computing*(Snowbird, Utah) 1986.

[7] H. Nagaoka and T. Kojima. Boltzmann Machine as a Statistical Model. *Bulletin of The Computational Statistics of Japan* Vol.1, pp.61–81, 1995 (in Japanese).