A Quantum Extension of Boltzmann Machine: An Information Geometrical Approach

Nihal Yapage[1] *　　Hiroshi Nagaoka[1] †

[1] *Graduate School of Information Systems, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan.*

Abstract. We extend the Classical Boltzmann Machine (CBM) to the quantum setting, which we call the Quantum Boltzmann Machine (QBM), in the viewpoint of Gibbs sampler, exponential family and information geometry. We also introduce a restricted class of the QBM called the Strongly Separable Quantum Boltzmann Machine (SSQBM). The information geometrical structure of the SSQBM is shown to be equivalent to that of the CBM. Moreover, the idea of Gibbs sampler is applied to the SSQBM to yield a state renewal rule, for which the SSQBM is obtained as the equilibrium state.

Keywords: Boltzmann machine, exponential family, Gibbs sampler, quantum Boltzmann machine, information geometry (IG), quantum information theory (QIT)

1 The classical Boltzmann machine

A CBM [1] is a neural net of n elements $1, 2, \cdots, n$ where the value of each element $i \in \{1, 2, \cdots, n\}$ is $x_i \in \{0, 1\}$. Then a state of the CBM is given by $\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in \{0,1\}^n$. The CBM has parameters $h_i \in \mathbb{R}$, the threshold for each element i and $w_{ij} \in \mathbb{R}$ for each pair $\{i, j\}$, the weight between i and j. The weights satisfy $w_{ij} = w_{ji}$ and $w_{ii} = 0$. When the CBM is in the state $\boldsymbol{x}$, the input to the element i is $I_i(\hat{\boldsymbol{x}}) = \sum_{j=1}^{n} w_{ij} x_j + h_i$, where $\hat{\boldsymbol{x}} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$. The state renewal rule $\boldsymbol{x} \to \boldsymbol{x}'$ of the CBM is

$$\mathrm{Prob}\{x'_i := 1\} = 1/\left[1 + \exp(-I_i(\hat{\boldsymbol{x}})/T)\right], \tag{1}$$

where $T \in \mathbb{R}$ is temperature. This state update of the CBM is sequential. The equilibrium distribution is

$$P(\boldsymbol{x}) = \frac{1}{\mathcal{Z}} \exp\Big[\Big(\sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j\Big)/T\Big], \tag{2}$$

where $\mathcal{Z}$ is a normalization and, in the sequel, we assume $T = 1$. Thus, we can identify each CBM with (2).

The (2) form an exponential (e-) family [2]. Let $\mathcal{X}$ be an arbitrary finite set. In general, when a family of distributions $\mathcal{M} = \{P_\theta | \theta = [\theta^\alpha]_{\alpha \in A} \in \mathbb{R}^A\}$ on $\mathcal{X}$ is as

$$P_\theta(\boldsymbol{x}) = \exp\Big[c(\boldsymbol{x}) + \sum_{\alpha \in A} \theta^\alpha f_\alpha(\boldsymbol{x}) - \psi(\theta)\Big], \tag{3}$$

$\mathcal{M}$ is called an *e- family*. Then $\theta = [\theta^\alpha]$ are called the *natural coordinates* of $\mathcal{M}$. If we let $\eta_\alpha(\theta) \stackrel{\mathrm{def}}{=} E_\theta[f_\alpha]$ then $\eta = [\eta_\alpha]$ and $\theta = [\theta^\alpha]$ are in one-to-one. These $[\eta_\alpha]$ are called the *expectation coordinates* of $\mathcal{M}$.

Let $\mathcal{P}$ be the set of distributions P on the finite set $\{0,1\}^n$ satisfying $P(\boldsymbol{x}) > 0$. Note that $\mathcal{P}$ itself is an e-family. For $k \in \{1, \cdots, n\}$, let $\mathcal{P}_k$ be the set of distributions

$$p(\boldsymbol{x}) = \frac{1}{\mathcal{Z}} \exp\Big[\sum_i \theta_i^{(1)} x_i + \sum_{i<j} \theta_{ij}^{(2)} x_i x_j + \cdots + \sum_{i_1 < \cdots < i_k} \theta_{i_1 \cdots i_k}^{(k)} x_{i_1} \cdots x_{i_k}\Big]. \tag{4}$$

Then $\mathcal{P}_k$ is also an e- family. Thus, we have a hierarchical structure of e- families $\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_n = \mathcal{P}$.

From the viewpoint of IG [3], each $\mathcal{P}_k$ is dually flat with respect to the e- and mixture (m-) connections together with the Fisher metric and, for $\ell < k$, $\mathcal{P}_\ell$ is an autoparallel submanifold of $\mathcal{P}_k$ with respect to the e-connection.

Given an $\mathcal{M} \subset \mathcal{P}$ of the form (3) and an arbitrary distribution $Q \in \mathcal{P}$ outside $\mathcal{M}$, consider the approximation of Q by an element $P_\theta \in \mathcal{M}$. We take the Kullback divergence $D(Q \parallel P_\theta) \stackrel{\mathrm{def}}{=} \sum_{\boldsymbol{x}} Q(\boldsymbol{x}) \log(Q(\boldsymbol{x}) - P_\theta(\boldsymbol{x}))$ as a criterion of approximation. Our interest is to find θ which minimizes $D(Q||P_\theta)$. Then, $\theta^* = \arg\min_\theta D(Q||P_\theta)$ iff $\eta_\alpha(\theta^*) = E_Q[f_\alpha], \forall\alpha$. An algorithm for computing this is the *gradient method* in which a positive-definite symmetric matrix $[\gamma^{\alpha\beta}(\theta)] \in \mathbb{R}^{A \times A}$ is specified for each $\theta \in \mathbb{R}^A$, and a small constant $\varepsilon > 0$ is given. Then, starting from an arbitrary initial value, this process recurrently updates θ for sufficiently many times according to $\theta^\alpha := \theta^\alpha + \Delta\theta^\alpha$ where $\Delta\theta^\alpha := -\varepsilon \sum_\beta \gamma^{\alpha\beta}(\theta)\{E_Q[f_\beta] - \eta_\beta(\theta)\}$ until P_θ converges to P_{θ^*}.

2 The quantum Boltzmann machine

An n-element quantum system corresponds to $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$. Let $\mathcal{S}$ be the set of faithful states on $\mathcal{H}$. Analogous to (4), we introduce a set $\mathcal{S}_k \subset \mathcal{S}$ of states

$$\rho = \frac{1}{\mathcal{Z}} \exp\Big[\sum_{i,s} \theta_{is}^{(1)} \pi_{is} + \sum_{i<j} \sum_{s,t} \theta_{ijst}^{(2)} \pi_{is} \pi_{jt} + \cdots + \sum_{i_1 < \cdots < i_k} \sum_{s_1 \cdots s_k} \theta_{i_1 \cdots i_k s_1 \cdots s_k}^{(k)} \pi_{i_1 s_1} \cdots \pi_{i_k s_k}\Big], \tag{5}$$

where $\pi_{is} = I^{\otimes(i-1)} \otimes \pi_s \otimes I^{\otimes(n-i)}$. Here, $\pi_s = \frac{1}{2}(I + \sigma_s)$, I, the identity and σ_s for $s = x, y, z$, the Pauli matrices. Note that π_s is a projection corresponding to $x_i \in \{0, 1\}$, whereas the set $\mathcal{S}_k$ is unchanged even if we replace π_s with σ_s in (5). We have the hierarchy $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_n = \mathcal{S}$. Now, corresponding to (2), we define the elements of $\mathcal{S}_2$ to be the QBMs. Letting $h_{is} = \theta_{is}^{(1)}$ and $w_{ijst} = \theta_{ijst}^{(2)}$, a QBM can be represented by

$$\rho = \frac{1}{\mathcal{Z}} \exp\Big[\sum_{i,s} h_{is} \pi_{is} + \sum_{i<j} \sum_{s,t} w_{ijst} \pi_{is} \pi_{jt}\Big]. \tag{6}$$

*nihal@hn.is.uec.ac.jp
†nagaoka@is.uec.ac.jp

Let $\mathcal{M} = \{\rho_\theta\}$ be a parametric family of faithful states

$$\rho_\theta = \exp\Big[C + \sum_{\alpha \in A} \theta^\alpha F_\alpha - \psi(\theta)\Big], \tag{7}$$

where F_α, C are Hermitian and $\psi(\theta)$ is a $\mathbb{R}$-valued function. Here, though (7) is only one among the possible several definitions, we call such an $\mathcal{M}$ a *quantum e- family* (QEF) and $\theta = [\theta^\alpha]$ its natural coordinates. The expectation coordinates $\eta = [\eta_\alpha]$ of $\mathcal{M}$ are $\eta_\alpha(\theta) = \mathrm{Tr}[\rho_\theta F_\alpha]$. It is easy to see that $\mathcal{S}_k$ is a QEF. An IG characterization of (7) is given in the following theorem with respect to the IG structure of $\mathcal{S}$ [3].

Theorem 1 *$\mathcal{S}_k$ is an autoparallel submanifold of $\mathcal{S}$ with respect to the* e-*connection and is dually flat with respect to $(g, \nabla^{(e)}, \nabla^{(m)})$ which is the IG structure induced from the quantum relative entropy $D(\rho\|\tau) \overset{\text{def}}{=} \mathrm{Tr}(\rho\log\rho - \rho\log\tau)$. In particular, g is the BKM (Bogoliubov-Kubo-Mori) Fisher information.*

The approximation problem for QEF corresponding to the classical case is described in the next theorem.

Theorem 2 *Given $\tau \in \mathcal{S}$ and a QEF (7), consider $\min_\theta D(\tau\|\rho_\theta)$ where D is the quantum relative entropy. Then, $\theta^* = \arg\min_\theta D(\tau\|\rho_\theta)$ iff $\eta_\alpha(\theta^*) = \mathrm{Tr}[\tau F_\alpha], \forall\alpha$. The gradient algorithm for computing θ^* is $\theta^\alpha := \theta^\alpha + \Delta\theta^\alpha$ where $\Delta\theta^\alpha := -\varepsilon\sum_\beta \gamma^{\alpha\beta}(\theta)\{\mathrm{Tr}[\tau F_\beta] - \eta_\beta(\theta)\}$.*

3 The strongly separable QBM

Separability is well-known in QIT. A state $\rho \in \mathcal{H}$ is called separable if there exists finite sets $\mathcal{X}_1, \cdots, \mathcal{X}_n$, a distribution P on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and $\{\tau^{(i)}_{x_i} \mid x_i \in \mathcal{X}_i\} \subset \mathcal{S}(\mathbb{C}^2), i = 1, \cdots, n$, such that

$$\rho = \sum_{\boldsymbol{x}} P(\boldsymbol{x})\tau_{\boldsymbol{x}}, \tag{8}$$

where $\tau_{\boldsymbol{x}} = \tau^{(1)}_{x_1} \otimes \cdots \otimes \tau^{(n)}_{x_n}$.

Definition 3 *A separable state (8) is strongly separable (SS) if $[\tau^{(i)}_{x_i}, \tau^{(i)}_{x'_i}] = 0 \quad \forall i, x_i, x'_i \in \mathcal{X}_i$.*

This is equivalent to the existence of $\mathcal{U} = \{\boldsymbol{u}_i\}_{i=1}^n$, where each $\boldsymbol{u}_i$ is a unit vector in $\mathbb{R}^3$, and a distribution P on $\{0,1\}^n$ such that

$$\rho = \sum_{x_1,\cdots,x_n} P(x_1, \cdots, x_n)\pi^{\boldsymbol{u}_1}_{x_1} \otimes \cdots \otimes \pi^{\boldsymbol{u}_n}_{x_n}, \tag{9}$$

where $\pi^{\boldsymbol{u}}_1 = \frac{1}{2}(I + \sum_{s=x,y,z} u_s\sigma_s)$ for $\boldsymbol{u} = (u_s)$ and $\pi^{\boldsymbol{u}}_0 = I - \pi^{\boldsymbol{u}}_1$. Note that $\pi^{\boldsymbol{u}} = (\pi^{\boldsymbol{u}}_1, \pi^{\boldsymbol{u}}_0)$ represents the Stern-Gerlach measurement of direction $\boldsymbol{u}$. We call $\mathcal{U} = \{\boldsymbol{u}_i\}$ a *frame* of a SS state ρ. Next theorem gives the necessary and sufficient conditions for a state (5) to be SS.

Theorem 4 *A state ρ represented in the form (5) in terms of the parameters $[\theta^{(j)}_{i_1,\cdots,i_j s_1,\cdots,s_j}]$ is SS with frame $\mathcal{U} = \{\boldsymbol{u}_i\}$ iff $\exists[\theta^{(j)\prime}_{i_1,\cdots,i_j}]$ such that $\forall j, \forall i_1 < \cdots < \forall i_j, \forall s_1, \cdots, \forall s_j,\ \theta^{(j)}_{i_1,\cdots,i_j s_1,\cdots,s_j} = \theta^{(j)\prime}_{i_1,\cdots,i_j} u_{i_1 s_1}, \cdots, u_{i_j s_j}$, where $\boldsymbol{u}_i = (u_{is})_{s=x,y,z}$.*

Let us now define $\mathcal{S}'(\mathcal{U})$ as the set of SS states $\rho \in \mathcal{S}$ with frame $\mathcal{U}$ and $\mathcal{S}'_k(\mathcal{U}) := \mathcal{S}'(\mathcal{U}) \cap \mathcal{S}_k$. Then, we have natural diffeomorphisms $\mathcal{S}'(\mathcal{U}) \simeq \mathcal{P}$ and $\mathcal{S}'_k(\mathcal{U}) \simeq \mathcal{P}_k$. Next theorem describes the IG structure of $\mathcal{S}'_k(\mathcal{U})$.

Theorem 5 *For an arbitrary frame $\mathcal{U}$, $\mathcal{S}'_k(\mathcal{U})$ is a QEF and therefore is autoparallel in $\mathcal{S}$ with respect to the e-connection. The induced IG structure $(g, \nabla^{(e)}, \nabla^{(m)})$ on $\mathcal{S}'_k(\mathcal{U})$ is dually flat and is equivalent to that of $\mathcal{P}_k$.*

Now, we define the elements of the $\mathcal{S}'_2(\mathcal{U})$ to be the SSQBMs. First, we give a corollary of Theorem 4.

Corollary 6 *A QBM $\rho_\theta, \theta = [h_{is}, w_{ijst}]$, (6) is SS with frame $\mathcal{U} = \{\boldsymbol{u}_i\}$ iff $\exists\theta' = [h'_i, w'_{ij}]$ such that $\forall i < j, \forall s, t$,*

$$h_{is} = h'_i u_{is}, \quad w_{ijst} = w'_{ij} u_{is} u_{jt}, \tag{10}$$

where $\boldsymbol{u}_i = (u_{is})_{s=x,y,z}$. In particular, if $\boldsymbol{h}_i \neq 0, \forall i$, the necessary and sufficient condition for a QBM ρ_θ to be SS (with some frame) is that $\forall i, j, \quad W_{ij} \propto \boldsymbol{h}_i\boldsymbol{h}_j^\top$, where $\boldsymbol{h}_i = [h_{ix} \quad h_{iy} \quad h_{iz}]^\top$, $W_{ij} = (w_{ijst})_{s,t=x,y,z}$ and $\top$ denotes the transpose.

We apply the idea of *Gibbs sampler* [4] to construct a state renewal process for the SSQBMs, which is given in the next theorem.

Theorem 7 *When the target state of a SSQBM is given by (6) and (10), the state renewal is carried out by the following procedure, starting from an arbitrary initial state $\rho \in \mathcal{S}$ and data $\boldsymbol{x} \in \{0,1\}^n$.*

(i) Choose i randomly.

(ii) Using data $\boldsymbol{x}$, renew the state of the i-th element to $(I + \sigma_{\boldsymbol{v}})/2$, where $\sigma_{\boldsymbol{v}} = \sum_{s=x,y,z} v_s\sigma_s$ and $\boldsymbol{v} = \tanh\Big((\sum_j w'_{ij}x_j + h'_i)/2\Big)\boldsymbol{u}_i$.

(iii) Perform the measurement $\pi^{\boldsymbol{u}_i}$ to the i-th element and update x_i in $\boldsymbol{x}$ by the measurement outcome.

Finally, we revisit the approximation problem of Theorem 2 for SSQBMs in the following theorem.

Theorem 8 *Given $\tau \in \mathcal{S}$, consider the problem of approximating τ by a SSQBM in $\mathcal{S}'_2(\mathcal{U}) = \{\rho_\theta\}$ where $\mathcal{U} = \{\boldsymbol{u}_i\}$ is an arbitrarily fixed frame. The approximation process to find $\theta^* = \arg\min_\theta D(\tau\|\rho_\theta)$ is decomposed into two parts: $\tau' = \arg\min_{\sigma\in\mathcal{S}'(\mathcal{U})} D(\tau\|\sigma)$ and $\theta^* = \arg\min_\theta D(\tau'\|\rho_\theta)$. The second part turns out equivalent to the approximation problem for the CBM by $\mathcal{S}'_2(\mathcal{U}) \simeq \mathcal{P}_2$.*

References

[1] D.H. Ackley, G.E. Hinton, T.J. Sejnowski, *Cognitive Science*, Vol.9, pp. 147-169, 1985.

[2] H. Nagaoka and T. Kojima, *Bulletin of The Computational Statistics of Japan*, Vol.1, pp. 61–81, 1995 (in Japanese).

[3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS / Oxford University Press, 2000.

[4] S. Geman, D. Geman, *IEEE Trans. on Pattern Analysis & Machine Intelligence*, PAMI-6, 6, pp. 721-741, 1984.