Information Geometrical Study of Quantum Boltzmann Machines

Nihal Yapage

The University of Electro-Communications

June 2008

Information Geometrical Study of Quantum Boltzmann Machines

Nihal Yapage

A thesis submitted to the

Graduate School of Information Systems
The University of Electro-Communications (UEC)
Tokyo, Japan

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

June 2008

Information Geometrical Study of Quantum Boltzmann Machines

Aproved by the supervisory Committee:

Chairperson	:	Prof. Hiroshi Nagaoka (UEC)
Member	:	Prof. Masahiro Sowa (UEC)
Member	:	Prof. Tsutomu Kawabata (UEC)
Member	:	Prof. Yutaka Sakaguchi (UEC)
Member	:	Prof. Akio Fujiwara (Osaka University)

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Hiroshi Nagaoka, for his excellent guidance while working on this work. His extraordinary teaching method was very attractive. The knowledge I retrieved from the frequent stimulating discussions with him is enormous and undoubtedly led to this thesis. He taught me not only the subject matter but the logical way of writing a technical paper also. His inspiring suggestions enhanced the presentation of this thesis. On the other hand, my interest to work on such a highly interdisciplinary area of research was almost fulfilled working with him. I am indebted to him for spending his time on me despite the heavy work load.

Secondly, I would like to thank Prof. Masahiro Sowa, Prof. Tsutomu Kawabata, Prof. Yutaka Sakaguchi and Prof. Akio Fujiwara for serving on my thesis committee and for their valuable suggestions. My thanks also go to all members of the Nagaoka-Ogawa (formerly Han-Nagaoka) laboratory for their friendship. Since my arrival here in 2002, the support I received from them (including the former members) in different ways is gratefully acknowledged. Special thanks are due to Prof. T. S. Han, Dr. T. Kojima, Dr. T. Ogawa and Dr. M. Iwamoto for participating in my seminars and for their encouraging comments. Specially, I thank Dr. Ogawa from the bottom of my heart for his great support in my difficulties.

Thirdly, but most importantly, I must offer my gratitude to my family and others. I dedicate this thesis to my loving parents for educating me amongst many a hardship. I am mostly indebted to my loving wife Roshinie and daughter Shanudi for their love, support, encouragement and endurance, although they were far away from me, during the past few years. Specially, their support was very important to maintain the physical and mental stability. It is difficult to find words to express my gratefulness to my wife for her helping hand extended to me, particularly for taking all the burden on her shoulders, during the period of my PhD studies. Also I extend my deepest thanks to all those who supported me throughout my academic career.

Finally, I would like to thank the Ministry of Education, Culture, Sports, Science and Technology (Japan) for financial support under the Japanese Government (Monbukagakusho) Scholarship program.

Nihal Yapage
Graduate School of Information Systems
The University of Electro-Communications
June 2008

量子ボルツマンマシンに関する情報幾何学的研究

ヤーパーゲー ニハール

概要

この論文では、確率的ニューラルネットワークの一種として知られているボルツマンマシン（classical Boltzmann machine, 略して CBM）の量子力学系への拡張を考える。具体的には、CBM の平衡確率分布の拡張とみなせるような量子状態（密度作用素）を考え、それを量子ボルツマンマシン (quantum Boltzmann machine, 略して QBM）と呼ぶ。CBM と同様に、QBM も二種類の実数値パラメータを持つ。すなわち、しきい値（物理的には外場を表す）h と結合定数 w である。これらのパラメータは、スピングラスの場合のような確率変数ではなく、任意ではあるが確定した（deterministic な）値を持つものとする。CBM は、単にニューラルネットワークの一種と言うにとどまらず、統計学的および物理学的な観点からさまざまな重要な性質を持つ。 これらに対応する QBM の性質を明らかにすることがこの研究の目的である。

CBM の平衡分布の全体は指数型分布族を成す。指数型分布族は統計学および情報幾何学においてきわめて重要な概念であり、その量子版として量子指数型分布族というものを考えると、QBM の全体はその一例になる。また、QBM の全体は滑らかな多様体とみなせ、そこに CBM の場合と同様の情報幾何学的構造を導入することができる。これらの数学的枠組みにもとづいて、まず、与えられた任意の状態を QBM で近似する問題を考察した。近似の規範として相対エントロピーを用いるとき、CBM の場合と同様の性質および計算アルゴリズムを導いた。ただし、CBM においては近似プロセスを学習問題（パラメータ推定）に適用することができたが、QBM ではそのような直接的応用は見いだせなかった。また、学習と同様に、CBM の持っていた確率的状態変化則に相当するダイナミクスも、QBM の場合にはその対応物を見いだすことができなかった。ただし、強可分 QBM （strongly separable QBM, 略して SSQBM）という特殊なクラスでは、ギブス・サンプラーの考え方にもとづいて CBM と同様の状態変化則を導入することができ、幾何学的にも SSQBM の成す多様体と CBM の成す多様体は同等な構造を持つことを示した。

続いて、QBM に関する平均場近似を情報幾何学的観点から考察した。これらは、T. Tanaka（田中利幸）によって示されていた CBM に関する結果の量子系への拡張と位置づけられる。まず、QBM の成す多様体上の情報幾何構造にもとづいて e-射影と m-射影という二種類の射影概念を定義し、それらを用いてナイーブ平均場近似の幾何学的意味を明らかにするとともに、その条件を与える平均場方程式を陽に与えた。さらに、統計力学における Plefka 展開のアイデアによるナイーブ平均場近似の精密化に関して考察を行った。特に、この精密化の本質を情報幾何学的観点から明らかにするとともに、Plefka 展開の係数が計量や接続などの幾何学的量によって表されることを示した。

Information Geometrical
Study of
Quantum Boltzmann Machines

Nihal Yapage

Abstract

In this thesis, we consider quantum extension of the well-known stochastic neural network model called (classical) Boltzmann machine (CBM) from an information geometrical point of view. The new model is called *quantum Boltzmann machine (QBM)*. We investigate some properties and geometrical aspects of QBM analogous to those of CBM. Furthermore, we study the mean-field approximation for such a model from the information geometrical point of view. This is in some sense motivated by the application of mean-field approximation for probabilistic inference in the graphical models in the classical probability theory. Although the problem tackled in the present thesis is somewhat deviated from the major field of quantum information theory, we have elucidated the relationships among several well-known fields such as statistical physics, differential geometry, information theory and statistics using the concepts of the new emerging subject of *quantum information geometry*.

We first define QBMs which can be considered as a general class of quantum Ising spin models. The states we consider are assumed to have at most second-order interactions with arbitrary but deterministic coupling coefficients. We call such a state a QBM for the reason that it can be regarded as a quantum extension of the equilibrium distribution of CBM. The totality of QBMs is then shown to form a quantum exponential family and thus can be considered as a smooth manifold having similar geometrical structures to those of CBMs. The information geometrical structure of the manifold of QBMs is discussed and the problem of approximating a given quantum state (density operator) by a QBM is also treated.

We also define a restricted class of QBMs called the *strongly separable QBMs (SSQBMs)*. We consider the dynamics of SSQBMs and propose a new state renewal rule based on that of CBM. The geometrical structure of the totality of SSQBMs is shown to be equivalent to that of the totality of CBMs. Approximation process for SSQBMs is also studied. Finally, we briefly discuss the parameter estimation of a SSQBM.

Next, we study the *mean-field approximation* for QBMs from an information geometrical point of view. We elaborate on the significance and usefulness of information

geometrical concepts, in particular the e-(exponential) and m-(mixture) projections, in studying the naive mean-field approximation for QBMs and derive the naive mean-field equation explicitly. We also discuss the higher-order corrections to the naive mean-field approximation based on the idea of *Plefka expansion* in statistical physics. We elucidate the geometrical essence of the corrections and provide the expansion coefficients with expressions in terms of information geometrical quantities. Here, one may note this work as the information geometrical interpretation of [Ple06] and as the quantum extension of [Tan00].

Contents

List of Figures

List of Symbols and Abbreviations

CBM	:	classical Boltzmann machine
QBM	:	quantum Boltzmann machine
SSQBM	:	strongly separable quantum Boltzmann machine
QEF	:	quantum exponential family
PVM	:	projection-valued measure
$\mathbb{R}$	:	set of real numbers
$\mathbb{C}$	:	set of complex numbers
ρ, σ, τ	:	density matrices (operators)
$\mathcal{S} = \mathcal{S}(\mathcal{H})$	:	totality of strictly positive density operators (faithful states) on a Hilbert space $\mathcal{H}$
X_1, X_2, X_3	:	Pauli spin matrices (operators)
X^*	:	operator adjoint (Hermitian conjugate) of the operator X
I	:	identity matrix or operator
Tr	:	trace of a matrix or an operator
$\mathcal{M}$	:	a family of probability (density) functions or quantum states (density operators)
$\lvert\mathcal{X}\rvert$	:	cardinality of a set $\mathcal{X}$
$\mathcal{P} = \mathcal{P}(\mathcal{X})$	:	the set of positive distributions on a set $\mathcal{X}$
$\lVert x \rVert$	:	norm of x
$D(q\lVert p)$	:	Kullback-Leibler information divergence of q with respect to p
$D(\rho\lVert\sigma)$	:	quantum relative entropy of ρ with respect to σ
$S(\rho)$	:	von Neumann entropy of ρ
$\mathbb{E}[x]$	:	expectation of a random variable x

Chapter 1

Introduction

1.1 Overview

It is often both important and interesting to study the problems in the interface of the fields such as physics, geometry, information theory, and statistics. This paves the way to the advancement of such fields too. In this thesis, we study a similar interdisciplinary problem which lies in the interface of such fields.

Information geometry as a new field has been found very useful in understanding the deep mathematical structures in the interface of many fields [Ama85, AN00]. This theory has been used in many situations to study the geometrical properties of the space of parametrized probability distributions or quantum states.

On the other hand, the (classical) Boltzmann machine (CBM) [AHS85] is a well-known stochastic neural network model. Physically, it can be viewed as a classical spin system. The statistical aspects of CBM have been elucidated in [NK95]. The information geometrical structure of the space of CBMs has been discussed in [AKN92].

Stimulated by the above works, we find a possible quantum extension which correspond to the equilibrium distribution of CBM and is called *quantum Boltzmann machine* (QBM*). We can consider QBM as a quantum spin system. Similar to CBM, each QBM has two kinds of real-valued parameters, namely h, the thresholds in neural network contexts and the external fields in physical contexts, and w, the coupling coefficients (to be defined in Section 3.1). We remark that these parameters are arbitrary but deterministic in our model, while the coupling coefficients are usu-

*We note that QBM, at present, lacks any notions corresponding to the stochastic dynamics of CBM which determine its equilibrium distribution. This means that our approach does not suggest how to quantize the neural aspects of CBM.

ally considered to be random variables in case of spin glasses in physics. We study the approximation process[†] for QBM similar to that of CBM.

We regard the set of QBMs parametrized by (h, w) as a *quantum exponential family*, a smooth manifold which has a similar form to an exponential family in statistics, on which a Riemannian metric and a couple of affine connections are naturally defined. These differential geometrical structures turn out to have a characteristic property called the dually flatness and are closely related to the quantum relative entropy (see Chapter 4).

A restricted class of QBMs called the strongly separable QBMs (SSQBMs) is defined. The totality of SSQBMs and that of CBMs are shown to be geometrically equivalent. For SSQBMs, we define a state renewal rule based on that for CBM. The approximation process is also discussed similarly.

As mentioned above, we view the QBM as a spin system with finite but large number of spins connected in some way. When we are working with such multi-particle systems with mutual interactions, the calculation of characteristic quantities such as expectations and correlations of relevant observables with respect to the density operator of the system is very important. This is, in general, computationally a very time-consuming problem since the required time increases exponentially with the number of elements in the system. One may find many obstacles due to the high complexity and such problems are not easy to tackle with exact methods. Thus, it is inevitable to employ an approximation method to get rid of this difficulty.

The so-called *mean-field approximation* was originated in statistical physics. The basic idea of the mean-field approximation is to use a simple tractable family of probability distributions (or density operators) to calculate characteristic quantities with respect to a probability distribution (or a density operator) including mutual interactions. It has been widely used both in classical and quantum physics as well as in other fields such as information theory, statistics, etc. In particular, it has been employed to get rid of the high computational cost in stochastic neural networks like CBMs. Such approximations have been found indispensable also to the study of probabilistic graphical models [JGJS99, OS01] in general where CBM is one example. T. Tanaka [Tan96, Tan00] has studied the mean-field approximation for a general class of classical Ising spin models, which are identified with the equilibrium distributions of CBMs, from the viewpoint of information geometry; see also [BK00, AIS01] for

[†]The "approximation process" and the "mean-field approximation" considered in this thesis correspond to m-projection and e-projection respectively and should clearly be distinguished from one another

related works.

Motivated by the above works, we study the mean-field approximation for QBMs from the viewpoint of quantum information geometry. Using the above mentioned geometrical setup on the space of QBMs, we find the information geometrical interpretation of mean-field approximation for QBMs. We elucidate the geometrical essence of the naive mean-field approximation for the QBMs as well as the higher-order extensions based on the idea of Plefka expansion in statistical physics.

1.2 Summary of results

In this section, we summarize the results obtained in this thesis. The results can be divided into two main parts:

(1) A quantum extension of classical Boltzmann machine (CBM). The relevant results are contained in Chapter 3 and Chapter 4

- Definition of quantum Boltzmann machine (QBM): Section 3.1.
- Definition of a restricted class of QBMs called strongly separable QBMs (SSQBMs) and proposal of a state renewal rule based on that for CBM: Section 3.2
- Information geometrical structure of the space of QBMs: Chapter 4.
- The approximation processes for QBMs and SSQBMs and the parameter estimation of SSQBM are discussed in: Section 4.5

(2) Information geometry of mean-field approximation for QBMs. The relevant results are contained in Chapter 5

- Derivation of naive mean-field equation for QBMs: Section 5.1
- Information geometrical understanding of higher-order approximations: Section 5.2

The above mentioned results have been published in [YN05, YN06, YN08].

1.3 Organization of the thesis

The structure of this thesis is as follows. In the next chapter, we briefly review the CBM including dynamics, approximation process and information geometrical interpretation of mean-field approximation. We define QBM in Chapter 3 corresponding

to the equilibrium distribution of CBM. Furthermore, we define a restricted class of QBMs called SSQBMs. The information geometrical structure of the space of QBMs is discussed in Chapter 4. We devote Chapter 5 to study the mean-field approximation for QBMs from an information geometrical point of view. Concluding remarks and open problems are presented in Chapter 6.

Chapter 2

A review of (classical) Boltzmann machine (CBM)

In this chapter, we briefly review the properties of the Boltzmann machine (which we call CBM in this thesis) including the dynamics, approximation process and learning, elucidating the fact that the totality of CBMs forms an exponential family. Next, we discuss the information geometrical interpretation of the mean-field approximation for CBM following the formulation originally presented in [Tan00].

2.1 Definition, state renewal rule and some properties

A CBM [AHS85, AK89] can be considered as 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 CBM can be represented 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 $h_i \in \mathbb{R}$. The CBM also has a real-valued parameter w_{ij} for each pair of elements $\{i, j\}$, which is called the coupling coefficient or weight between i and j. These parameters are assumed to satisfy conditions $w_{ij} = w_{ji},\ w_{ii} = 0$. When the CBM is in the state $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the input to the element i from the other elements is defined by

$$I_i(\hat{\boldsymbol{x}}) = \sum_{j=1}^{n} w_{ij} x_j + h_i,$$

where $\hat{\boldsymbol{x}} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. The state renewal rule $\boldsymbol{x} \to \boldsymbol{x}' = (x'_1, \ldots, x'_n)$ of the CBM is as follows: we choose an element, say the element i, randomly and the

value of the element is set to 1 according to the probability

$$\mathrm{Prob}\{x_i' := 1\} = \frac{1}{1+\exp(-2I_i(\hat{\boldsymbol{x}})/T)}, \tag{2.1}$$

where the real-valued parameter T is called the temperature of the network.

Remark 1. *In view of parametrization of probability distributions (or density operators in the quantum setting), the temperature T is redundant. Therefore, in the sequel, we assume $T = 1$.*

During this state renewal process, the value of the elements $j(\neq i)$ does not change. That means the state update of the CBM is sequential. This process defines a Markov chain on the product set $\{-1,+1\}^n$, whose equilibrium (stationary) distribution is given by

$$p_{h,w}(\boldsymbol{x}) = \exp\Big\{\sum_i h_i x_i + \sum_{i<j} w_{ij}x_i x_j - \psi(h,w)\Big\}, \tag{2.2}$$

where

$$\psi(h,w) = \log\sum_{\boldsymbol{x}} \exp\Big\{\sum_i h_i x_i + \sum_{i<j} w_{ij}x_i x_j\Big\}.$$

Thus, noting that the correspondence $p_{h,w} \leftrightarrow (h,w)$ is one to one, we can, at least mathematically, identify each CBM with its equilibrium probability distribution.

Remark 2. *Neural network community usually considers $x_i \in \{0,1\}$ while physicists use $x_i \in \{-1,+1\}$. Each can be replaced by the other without changing the essence of the argument.*

Many good properties of the CBM are consequences of the fact that the equilibrium distributions form an exponential family. Here, we discuss this important aspect of the CBM [NK95] briefly. Let $\mathcal{X}$ be a finite set or, more generally, a measurable space with an underlying measure $\mathrm{d}\mu$. We denote the set of positive probability distributions (probability mass functions for a finite $\mathcal{X}$ and probability density functions for a general $(\mathcal{X},\mathrm{d}\mu)$) on $\mathcal{X}$ by $\mathcal{P} = \mathcal{P}(\mathcal{X})$. When a family of distributions, say

$$\mathcal{M} = \{p_\theta \mid \theta = (\theta^i);\ i = 1,\dots,n\} \subset \mathcal{P}, \tag{2.3}$$

is represented in the form

$$p_\theta(x) = \exp\Big\{c(x) + \sum_{i=1}^n \theta^i f_i(x) - \psi(\theta)\Big\}, \quad x \in \mathcal{X}, \tag{2.4}$$

$\mathcal{M}$ is called an *exponential family*. Here, θ^i; $i = 1, \ldots, n$ are real-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

$$\eta_i = \partial_i \psi(\theta) \qquad \Big(\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \theta^i}\Big). \tag{2.5}$$

The set that consists of equilibrium probability distributions of Boltzmann machine (2.2) is one example of exponential family. In addition, threshold values and coupling coefficients (weights) become the natural coordinates while $\mathbb{E}_\theta[x_i]$ and $\mathbb{E}_\theta[x_i x_j]$ become expectation coordinates. The notion of exponential family is very important in statistics and information geometry, and is also useful in studying properties of CBMs with their mean-field approximations.

Let us now consider a hierarchy of exponential families and show that the totality of CBMs is included in it. Let $\mathcal{P} = \mathcal{P}(\{-1, +1\}^n)$. For $k \in \{1, \ldots, n\}$, let $\mathcal{P}_k$ be the set of probability distributions of the form

$$\begin{aligned} p_\theta(\boldsymbol{x}) &= \exp\Big\{\sum_i \theta_i^{(1)} x_i + \sum_{i<j} \theta_{ij}^{(2)} x_i x_j + \cdots + \sum_{i_1<\cdots<i_k} \theta_{i_1 \ldots i_k}^{(k)} x_{i_1} \cdots x_{i_k} - \psi(\theta)\Big\} \\ &= \exp\Big\{\sum_{j=1}^{k} \sum_{i_1<\cdots<i_j} \theta_{i_1 \ldots i_j}^{(j)} x_{i_1} \cdots x_{i_j} - \psi(\theta)\Big\} \end{aligned} \tag{2.6}$$

with

$$\psi(\theta) = \log \sum_{\boldsymbol{x}} \exp\Big\{\sum_{j=1}^{k} \sum_{i_1<\cdots<i_j} \theta_{i_1 \ldots i_j}^{(j)} x_{i_1} \cdots x_{i_j}\Big\}, \tag{2.7}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{-1, +1\}^n$. Then $\mathcal{P}_k$ also turns out to be an exponential family of dimension $\sum_{j=1}^{k} {}_nC_j$. Thus, we have a hierarchical structure of exponential families $\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_n = \mathcal{P}$. In particular, $\mathcal{P}_1$ and $\mathcal{P}_2$ can be represented by

$$\begin{aligned} \mathcal{P}_1 &= \{p_1(x_1) \cdots p_n(x_n)\} = \{\text{product distributions}\} \quad \text{and} \\ \mathcal{P}_2 &= \{\text{equilibrium distributions of CBM}\} \quad (\text{see (2.2)}) \end{aligned}$$

respectively. Note also that $\mathcal{P}$ itself is an exponential family of dimension $2^n - 1$.

2.2 Approximation process and learning

Suppose that we are given an exponential family $\mathcal{M} \subset \mathcal{P}$ of the form (2.4) and an arbitrary probability distribution $q \in \mathcal{P}$ outside $\mathcal{M}$. Let us consider the problem of approximating q by an element $p_\theta \in \mathcal{M}$. We adopt the Kullback-Leibler information divergence

$$D(q\|p_\theta) \stackrel{\text{def}}{=} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p_\theta(\boldsymbol{x})} \tag{2.8}$$

as the criterion of approximation. Our interest is to find θ which minimizes $D(q\|p_\theta)$, that is, $\arg\min_\theta D(q\|p_\theta)$. Geometrically, this can be understood as the m-projection defined in Subsection 2.3.2. Then, it is shown that

$$\theta^* = \arg\min_\theta D(q\|p_\theta) \Longleftrightarrow \eta_i(\theta^*) = \mathbb{E}_q[f_i],\ \forall i, \tag{2.9}$$

where $\mathbb{E}_q$ denotes the expectation with respect to the probability distribution q. Actually, for any θ, we have

$$D(q\|p_\theta) = D(q\|p_{\theta^*}) + D(p_{\theta^*}\|p_\theta),$$

which is an example of the Pythagorean relation for D.

When a sequence of data $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)$ is given, the approximation

$$\min_\theta D(\hat{q}\|p_\theta)$$

of the empirical distribution

$$\hat{q} = \frac{1}{N} \sum_{t=1}^{N} \delta_{\boldsymbol{x}(t)}$$

(where $\delta_{\boldsymbol{x}}$ denotes the probability distribution satisfying $\delta_{\boldsymbol{x}}(\boldsymbol{x}) = 1$) can be regarded as a process of *learning* from the data. In particular, when the data $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)$ are assumed to be drawn from an unknown distribution in $\mathcal{M}$, the learning turns out to be the *maximum likelihood estimation (MLE)*.

An algorithm for computing this approximation is the *gradient method*. In this method, we assume that a positive-definite symmetric matrix $[\gamma^{ij}(\theta)] \in \mathbb{R}^{n\times n}$ is specified for each point $\theta \in \mathbb{R}^n$, and a small positive constant ε is given in advance. Then, starting from an arbitrary initial value θ, this process recurrently updates θ for sufficiently many times according to

$$\Delta\theta^i \stackrel{\text{def}}{=} -\varepsilon \sum_j \gamma^{ij}(\theta) \partial_j D(q\|p_\theta), \tag{2.10}$$

$$= \varepsilon \sum_j \gamma^{ij}(\theta) \{\mathbb{E}_q[f_j] - \eta_j(\theta)\}, \tag{2.11}$$

$$\theta^i \stackrel{\text{def}}{=} \theta^i + \Delta\theta^i,$$

until $||\Delta\theta||$ becomes sufficiently small, which implies the convergence of p_θ to p_{θ^*}. Here, taking the value of ε smaller, the accuracy of approximation becomes better but by that much the rate of convergence becomes slower.

As a special case of this method, we obtain the learning algorithm for the CBM proposed by Ackley et al. in [AHS85]. They showed that, setting the initial value

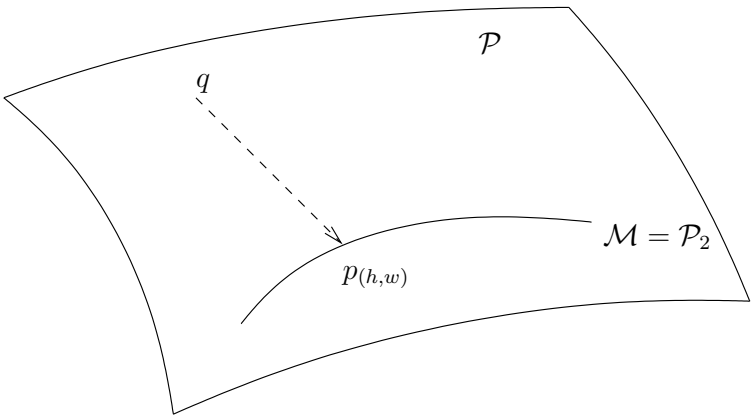

Figure 2.1: Approximation process for CBM.

of $\theta = \{h_i, w_{ij}\}$ arbitrarily and going recursively updating this according to the following equations, θ which minimizes $D(q\|p_\theta)$ can be found approximately after sufficient number of updates:

$$\begin{aligned} \Delta h_i &\stackrel{\text{def}}{=} \varepsilon\left(\sum_{\boldsymbol{x}} x_i q(\boldsymbol{x}) - \sum_{\boldsymbol{x}} x_i p_\theta(\boldsymbol{x})\right) \qquad (2.12) \\ &= \varepsilon\Big(\mathbb{E}_q[x_i] - \mathbb{E}_\theta[x_i]\Big) \\ \Delta w_{ij} &\stackrel{\text{def}}{=} \varepsilon\left(\sum_{\boldsymbol{x}} x_i x_j q(\boldsymbol{x}) - \sum_{\boldsymbol{x}} x_i x_j p_\theta(\boldsymbol{x})\right) \qquad (2.13) \\ &= \varepsilon\Big(\mathbb{E}_q[x_i x_j] - \mathbb{E}_\theta[x_i x_j]\Big) \end{aligned}$$

and

$$\begin{aligned} h_i &\stackrel{\text{def}}{=} h_i + \Delta h_i \\ w_{ij} &\stackrel{\text{def}}{=} w_{ij} + \Delta w_{ij}, \end{aligned}$$

where $\mathbb{E}_q[\cdot]$ and $\mathbb{E}_\theta[\cdot]$ represent the expectation values with respect to q and p_θ respectively. Equations (2.12) and (2.13) are nothing but (2.11) applied to $\mathcal{M} = \mathcal{P}_2$ with

$[\gamma^{ij}(\theta)] = [\delta^{ij}]$ (the Kronecker delta). On the other hand, a geometrically natural choice of $[\gamma^{ij}(\theta)]$ is $[g^{ij}(\theta)]$, which is the *Fisher information matrix*

$$g^{ij}(\theta) \stackrel{\text{def}}{=} \mathbb{E}_\theta[\partial^i \log p_\theta \; \partial^j \log p_\theta], \qquad \partial^i = \frac{\partial}{\partial \eta_i} \tag{2.14}$$

for (η_i) or, equivalently is the inverse of the Fisher information matrix

$$g_{ij}(\theta) = \mathbb{E}_\theta[\partial_i \log p_\theta \; \partial_j \log p_\theta] \tag{2.15}$$

for (θ^i) [AKN92]. This is an example of the so-called *natural gradient method* [Ama98].

2.3 Mean-field approximation for CBMs

In this section, first we give a brief introduction to the mean-field approximation in statistical physics. Then we discuss some important points of the information geometrical aspects of the mean-field approximation including both naive and higher-order approximations for CBMs following [Tan00, AIS01]. We also briefly describe the information geometrical viewpoint of the so-called Plefka expansion.

2.3.1 Introduction to mean-field approximation

Mean-field theory, dating back to Curie, Weiss and Ginzburg-Landau, is one of the most common approaches to the study of complex physical systems. The basic underlying idea is that the complicated interactions that each element of a complex system is subjected to by its neighbors can be replaced by the interaction with an effective (or mean) field. While mean-field approximation arose primarily in the field of statistical mechanics, it provides an alternative perspective on the probabilistic inference problem. For instance, it has more recently been applied for doing inference in graphical models in artificial intelligence [JGJS99].

In magnetic materials the microscopic state of the system is supposed to be defined by the values of local spin magnetizations. In many magnetic materials the electrons responsible for the magnetic behavior are localized near the atoms of the crystal lattice, and the force which tends to orient the spins is the (short range) exchange interaction.

The most popular models, which describe this situation qualitatively are called the Ising models. The microscopic variables in these systems are the Ising spins x_i which by definition can take only two values -1 or +1. The microscopic energy or the

Hamiltonian H as a function of all the Ising spins is given by

$$H = -\sum_{i<j} J_{ij} x_i x_j - h \sum_i x_i, \tag{2.16}$$

where J_{ij} are the values of the spin-spin interactions and h is the external magnetic field. In particular, physicists usually consider nearest neighbor models for example, the ferromagnetic Ising model with $J_{ij} > 0$ and the anti-ferromagnetic Ising model with $J_{ij} < 0$. It is a well-known fact that in statistical physics, the thermal equilibrium state of a system is given by the Boltzmann distribution

$$P(\boldsymbol{x}) = \frac{1}{\mathcal{Z}} \exp\Big(-\frac{H(\boldsymbol{x})}{T}\Big), \tag{2.17}$$

where T is the temperature* and

$$\mathcal{Z} = \sum_{\boldsymbol{x}} \exp\Big(-\frac{H(\boldsymbol{x})}{T}\Big)$$

is a normalizing factor called the partition function. In spite of the apparent simplicity of the Ising model, an exact solution (which means the calculation of the partition function $\mathcal{Z}$, expectations and the correlation functions) for the lattice system has been found only for the one- and the two-dimensional cases in the zero external magnetic field. In the higher dimensions one needs to use approximate methods. One of the simplest methods is called the *mean-field approximation* (see for example

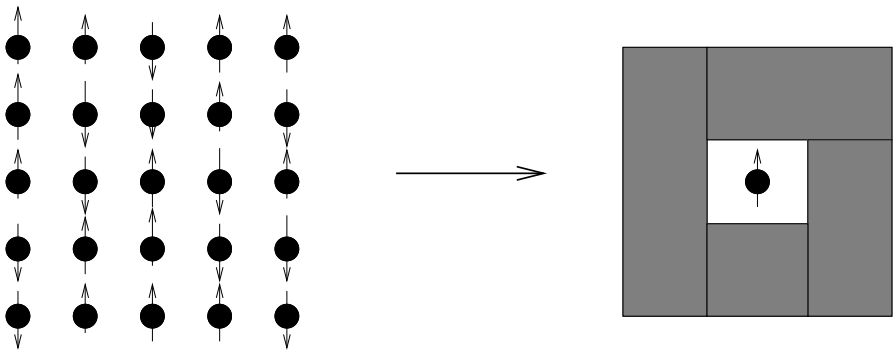

Figure 2.2: Mean-field approximation

[PA87, HKP91] for application to neural networks). In many, but not all, cases this method gives the results which are not too far from the correct ones, and very often

*As mentioned in Section 2.1, in the sequel, we assume $T = 1$.

it makes possible to get some qualitative understanding of what is going on in the system under consideration. The starting point of the mean-field approximation is the fact that the joint distribution function of the non-interacting system can be factorized as the product of the independent distribution functions in the individual sites of the spin system.

2.3.2 Naive and higher-order mean-field approximations for CBMs

In this subsection, we derive the naive mean-field equation and discuss briefly the higher-order approximations based on the Plefka expansion from an information geometrical point of view. The complete description of the information geometrical interpretation of Plefka expansion is presented in the Chapter 5 for the quantum setting.

Let us now recall $\mathcal{P}_2$, the manifold of CBMs. Here, we treat the case that each element is subjected to different external magnetic fields (in the physical interpretation) h_i which were considered equal to h in (2.16). Suppose that we are interested in obtaining expectations m_i with respect to a probability distribution in $\mathcal{P}_2$. However, when the system size is large, the partition function $\exp(\psi(h, w))$ 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 $p_{h,w} \in \mathcal{P}_2$.

First, we consider the subspace $\mathcal{P}_1$ of $\mathcal{P}_2$. We parametrize each distribution in $\mathcal{P}_1$ by $\bar{h}$ in order to distinguish from the parametrization in (2.2) and write as

$$p_{\bar{h}}(\boldsymbol{x}) = \exp\Big\{ \sum_i \bar{h}_i x_i - \psi(\bar{h}) \Big\}, \tag{2.18}$$

where

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

Then, $\mathcal{P}_1$ forms a submanifold of $\mathcal{P}_2$ specified by $w_{ij} = 0$ 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

$$\bar{m}_i = \frac{\partial \psi(\bar{h})}{\partial \bar{h}_i} = \frac{\exp(\bar{h}_i) - \exp(-\bar{h}_i)}{\exp(\bar{h}_i) + \exp(-\bar{h}_i)} = \tanh(\bar{h}_i), \tag{2.19}$$

from which we obtain

$$\bar{h}_i = \frac{1}{2}\log\left(\frac{1+\bar{m}_i}{1-\bar{m}_i}\right). \tag{2.20}$$

The simple idea behind the mean field approximation for a $p_{h,w} \in \mathcal{P}_2$ 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. For the present purpose, we adopt the Kullback-Leibler divergence (2.8). Given $p_{h,w} \in \mathcal{P}_2$, its e-(exponential) and m-(mixture) projections (see [AN00]) onto $\mathcal{P}_1$ are defined by

$$\bar{p}^{(e)} = p_{\bar{h}^{(e)}} \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{p_{\bar{h}} \in \mathcal{P}_1} D(p_{\bar{h}} \| p_{h,w}) \tag{2.21}$$

and

$$\bar{p}^{(m)} = p_{\bar{h}^{(m)}} \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{p_{\bar{h}} \in \mathcal{P}_1} D(p_{h,w} \| p_{\bar{h}}) \tag{2.22}$$

respectively, where

$$\bar{h}^{(e)} = \operatorname*{arg\,min}_{\bar{h}=(\bar{h}_i)} D(p_{\bar{h}} \| p_{h,w}) \tag{2.23}$$

and

$$\bar{h}^{(m)} = \operatorname*{arg\,min}_{\bar{h}=(\bar{h}_i)} D(p_{h,w} \| p_{\bar{h}}). \tag{2.24}$$

As necessary conditions, we have

$$\frac{\partial}{\partial \bar{h}_i} D(p_{\bar{h}} \| p_{h,w}) = 0 \tag{2.25}$$

and

$$\frac{\partial}{\partial \bar{h}_i} D(p_{h,w} \| p_{\bar{h}}) = 0, \tag{2.26}$$

which are weaker than (2.23) and (2.24). But sometimes (2.25) and (2.26) 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 $\mathbb{E}_{(h,w)}[x_i] = \mathbb{E}_{\bar{h}}[x_i]$ for $p_{\bar{h}} = \bar{p}^{(m)}$ (see (2.9)). It should be noted that the approximation process in Section 2.2 can be considered as the m-projection from $\mathcal{P}$ onto $\mathcal{P}_2$. On the other hand, the e-projection $\bar{p}^{(e)}$ from $\mathcal{P}_2$ onto $\mathcal{P}_1$ gives the naive mean-field approximation (see [Tan00]) which is explicitly given by

$$\bar{m}_i = \tanh\Big(\sum_j w_{ij}\bar{m}_j + h_i\Big), \tag{2.27}$$

where $\bar{m}_i = \mathbb{E}_{\bar{h}}[x_i]$ for $p_{\bar{h}} = \bar{p}^{(e)}$. From (2.27) and (2.19), it is clear that in the mean-field approximation the effective fields of other elements $j(\neq i)$ are replaced by external fields called mean fields (see Figure 2.2).

Now we derive the naive mean-field equation for CBM. Recall that the equilibrium distribution for CBM (2.2) is given by

$$p \stackrel{\text{def}}{=} p_{h,w}(\boldsymbol{x}) = \exp\Big\{\sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j - \psi(h,w)\Big\}, \tag{2.28}$$

where

$$\psi(p) \stackrel{\text{def}}{=} \psi(h,w) = \log \sum_{\boldsymbol{x}} \exp\Big\{\sum_i h_i x_i + \sum_{i<j} w_{ij} x_i x_j\Big\}.$$

Now we define another function

$$\phi(p) = \phi(h,w) \stackrel{\text{def}}{=} \sum_{i<j} w_{ij}\, \mathbb{E}_p[x_i x_j] + \sum_i h_i\, \mathbb{E}_p[x_i] - \psi(p), \tag{2.29}$$

which coincides with the negative entropy:

$$\phi(p) = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \log p(\boldsymbol{x}). \tag{2.30}$$

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

$$\phi(p_{\bar{h}}) = \sum_i \left[\left(\frac{1+\bar{m}_i}{2}\right)\log\left(\frac{1+\bar{m}_i}{2}\right) + \left(\frac{1-\bar{m}_i}{2}\right)\log\left(\frac{1-\bar{m}_i}{2}\right)\right]. \tag{2.31}$$

The KL divergence between $p \in \mathcal{P}_2$ and $p_{\bar{h}} \in \mathcal{P}_1$ can be expressed in the following form:

$$\begin{aligned} D(p_{\bar{h}}\|p) &= \psi(p) + \phi(p_{\bar{h}}) - \sum_{i<j} w_{ij}\, \mathbb{E}_{\bar{h}}[x_i x_j] - \sum_i h_i\, \mathbb{E}_{\bar{h}}[x_i] \\ &= \psi(p) + \phi(p_{\bar{h}}) - \sum_{i<j} w_{ij}\bar{m}_i\bar{m}_j - \sum_i h_i \bar{m}_i. \\ &= \psi(p) + \frac{1}{2}\sum_i \left[(1+\bar{m}_i)\log\left(\frac{1+\bar{m}_i}{2}\right) + (1-\bar{m}_i)\log\left(\frac{1-\bar{m}_i}{2}\right)\right] \\ &\quad - \sum_{i<j} w_{ij}\bar{m}_i\bar{m}_j - \sum_i h_i\bar{m}_i. \end{aligned} \tag{2.32}$$

Now consider the e-projection (2.25) from $p \in \mathcal{P}_2$ onto $p_{\bar{h}} \in \mathcal{P}_1$, i.e.

$$\frac{\partial}{\partial \bar{h}_i} D(p_{\bar{h}}\|p) = 0. \tag{2.33}$$

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

$$\frac{\partial}{\partial \bar{m}_i} D(p_{\bar{h}} \| p) = 0. \tag{2.34}$$

Since $\psi(p)$ does not depend on $\bar{m}_i$, we obtain from (2.32) that

$$\begin{aligned} 0 &= \frac{1}{2} \log\left(\frac{1+\bar{m}_i}{1-\bar{m}_i}\right) - \sum_{j \neq i} w_{ij} \bar{m}_j - h_i \\ &= \bar{h}_i - \sum_{j \neq i} w_{ij} \bar{m}_j - h_i, \end{aligned} \tag{2.35}$$

where the second equality is from (2.20). Thus the naive mean-field equation is obtained from (2.19) and (2.35) as

$$\tanh^{-1}(\bar{m}_i) = \sum_{j \neq i} w_{ij} \bar{m}_j + h_i \tag{2.36}$$

and this is usually written in the form (2.27).

Let us briefly discuss the higher-order mean-field approximations for CBMs using the information geometrical interpretation of the so-called Plefka expansion. Recall that the elements of $\mathcal{P}_2$ are parametrized as $p_{h,w}$ by $h = (h_i)$ and $w = (w_{ij})$. This means that (h, w) forms a coordinate system of the manifold $\mathcal{P}_2$. When $\mathcal{P}_2$ is viewed as an exponential family, (h, w) turns out to be a natural coordinate system, while the corresponding expectation coordinate system is given by (m, η) where $m = (m_i)$ and $\eta = (\eta_{ij})$ with $m_i = \mathbb{E}_{(h,w)}[x_i])$ and $\eta_{ij} = \mathbb{E}_{(h,w)}[x_i x_j])$ respectively. We now define a third (or hybrid) coordinate system (m, w). The elements of $\mathcal{P}_2$ are then parametrized by (m, w), which we denote by $\hat{p}_{m,w}$ to avoid any confusion with $p_{h,w}$. Note that

$$\begin{aligned} \mathcal{P}_2 &= \{p_{h,w} \mid (h, w) : \text{free}\} \\ &= \{\hat{p}_{m,w} \mid (m, w) : \text{free}\} \end{aligned} \tag{2.37}$$

and that

$$\hat{p}_{m,w} = p_{h,w} \iff \forall i,\ m_i = \mathbb{E}_{(h,w)}[x_i]. \tag{2.38}$$

One may think of obtaining the expectations m_i by solving the equation

$$\frac{\partial}{\partial m_i} D(\hat{p}_{m,w} \| p_{h,w}) = 0. \tag{2.39}$$

The difficulty in this approach is that it is hard to derive the stationary condition explicitly because $\hat{p}_{m,w}$ is complex enough so that we cannot obtain an explicit expression for $D(\hat{p}_{m,w} \| p_{h,w})$ in terms of $\{m_i\}$. On the other hand, when $w = 0$ in the

first argument of $D(\hat{p}_{m,w}\|p_{h,w})$, the equation (2.39) turns out to be

$$\frac{\partial}{\partial m_i} D(\hat{p}_{m,0}\|p_{h,w}) = 0 \tag{2.40}$$

which is nothing but the naive mean-field approximation discussed earlier.

The Plefka expansion was originally presented as a Taylor expansion of the Gibbs potential [Ple82, Ple06]. The information geometrical interpretation of it is that to use the Taylor expansion of the difference $D(\hat{p}_{m,0}\|p_{h,w}) - D(\hat{p}_{m,w}\|p_{h,w})$ up to a required order of w to derive more precise approximations of $\{m_i\}$. For the detailed discussion of the information geometrical viewpoint of the Plefka expansion, see Chapter 5.

We mention, for example, the second-order mean-field equation

$$\bar{m}_i = \tanh\left(h_i + \sum_{j\neq i} w_{ij}\bar{m}_j - \bar{m}_i \sum_{j\neq i} w_{ij}^2 (1-\bar{m}_j^2) \right) \tag{2.41}$$

which was originally derived by a different method in [TAP77]. Some other higher-order terms have been calculated in [NT97].

Chapter 3

A quantum extension of CBM

We present a possible quantum extension of CBM which is called the quantum Boltzmann machine (QBM). Furthermore, a subfamily of the totality of QBMs is defined which is called the set of strongly separable QBMs (SSQBMs). A state renewal rule is proposed for these states based on that for CBM. Note that, in this thesis, we use the concepts of density operator, quantum measurement etc. without explicitly describing their physical meaning and the interested reader is referred, for example, to [NC02].

3.1 Definition of quantum Boltzmann machine (QBM)

Let us consider an n-element system of quantum Ising spins. Each element is represented as a quantum bit (qubit) or quantum spin-$\frac{1}{2}$ 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 faithful (strictly positive) states on $\mathcal{H}$;

$$\mathcal{S} = \{\rho \,|\, \rho = \rho^* > 0 \text{ and } \operatorname{Tr}\rho = 1\}. \tag{3.1}$$

Here, each ρ is a $2^n \times 2^n$ matrix; $\rho = \rho^* > 0$ means that ρ is Hermitian and strictly positive or positive definite respectively; and $\operatorname{Tr}\rho = 1$ shows that the trace of the density matrix ρ is unity. Now corresponding to (2.6), an element of $\mathcal{S}$ is said to have *at most kth-order interactions* if it is written as

$$\begin{aligned}\rho_\theta \;=\; & \exp\Big\{\sum_{i,s}\theta^{(1)}_{is}X_{is} + \sum_{i<j}\sum_{s,t}\theta^{(2)}_{ijst}X_{is}X_{jt} + \cdots \\ & + \sum_{i_1<\cdots<i_k}\sum_{s_1\ldots s_k}\theta^{(k)}_{i_1\ldots i_k s_1\ldots s_k}X_{i_1s_1}\cdots X_{i_ks_k} - \psi(\theta)\Big\}\end{aligned}$$

$$= \exp\Big\{\sum_{j=1}^{k}\sum_{i_1<\cdots<i_j}\sum_{s_1\ldots s_j}\theta^{(j)}_{i_1\ldots i_j s_1\ldots s_j}X_{i_1 s_1}\cdots X_{i_j s_j} - \psi(\theta)\Big\} \tag{3.2}$$

with

$$\psi(\theta) = \log \mathrm{Tr}\exp\Big\{\sum_{j=1}^{k}\sum_{i_1<\cdots<i_j}\sum_{s_1\ldots s_j}\theta^{(j)}_{i_1\ldots i_j s_1\ldots s_j}X_{i_1 s_1}\cdots X_{i_j s_j}\Big\}, \tag{3.3}$$

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

$$X_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad X_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

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

Our main concern in the present chapter is $\mathcal{S}_2$. In the sequel, we let $h_{is}=\theta^{(1)}_{is}$ and $w_{ijst}=\theta^{(2)}_{ijst}$ to rewrite (3.2) for $k=2$ as

$$\rho_{h,w} = \exp\Big\{\sum_{i,s}h_{is}X_{is} + \sum_{i<j}\sum_{s,t}w_{ijst}X_{is}X_{jt} - \psi(h,w)\Big\}, \tag{3.4}$$

where $h=(h_{is})$, $w=(w_{ijst})$. The real dimension of $\mathcal{S}_2$ is $3n(3n-1)/2$ which gives the number of parameters to specify a density operator $\rho_{h,w}$.

Corresponding to the classical case, an element of $\mathcal{S}_2$ of the form (3.4) is called a *quantum Boltzmann machine* or a QBM in this thesis (see also [YN05, YN06]), although we have no quantum dynamics corresponding to the stochastic state change of a CBM at present. Physically, a QBM simply means a general quantum state for n-fold spins with at most second-order interactions which are arbitrary and deterministic, not random as in a spin glass.

Now, here we briefly discuss the set $\mathcal{S}_1$ for later use. The elements of $\mathcal{S}_1$ are represented as $\rho_{h,0}$ by letting $w=0$ in (3.4). In the sequel, we write them as

$$\tau_{\bar h} = \exp\Big\{\sum_{i,s}\bar h_{is}X_{is} - \psi(\bar h)\Big\} \tag{3.5}$$

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 $\mathcal{S}_2$. We have

$$\tau_{\bar h} = \bigotimes_{i=1}^{n}\exp\Big\{\sum_{s}\bar h_{is}X_s - \psi_i(\bar h_i)\Big\}, \tag{3.6}$$

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

$$\begin{aligned} \psi_i(\bar{h}_i) &= \log \mathrm{Tr} \exp\Big\{ \sum_s h_{is} X_s \Big\} \\ &= \log\{\exp(||\bar{h}_i||) + \exp(-||\bar{h}_i||)\} \end{aligned} \tag{3.7}$$

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

$$\psi(\bar{h}) = \sum_i \psi_i(\bar{h}_i). \tag{3.8}$$

3.2 Strongly separable QBM (SSQBM)

3.2.1 Separability criterion and a state renewal rule

First, we review the state renewal process of the CBM. In Chapter 2, Section 2.1, we mentioned that the probability distribution (2.2) is obtained as the equilibrium distribution for the stochastic state renewal rule (2.1). The significance of this process is its *locality*. That is, the random generation of data is simple and carried out by each element based on the data sent from the other elements. The mathematical essence of (2.1) is that it is the conditional distribution defined from (2.2). That is,

$$\begin{aligned} \frac{1}{1+\exp(-2I_i(\hat{\boldsymbol{x}}))} &= \frac{P(x_1,\ldots,x_i=1,\ldots,x_n)}{\sum_{x_i'} P(x_1,\ldots,x_i',\ldots,x_n)} \\ &= P(x_i = 1 \,|\, x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n). \end{aligned}$$

This method to generate a random sequence $X_1, X_2, \ldots, X_n$ subject to $P(x_1, \ldots, x_n)$ by the use of conditional distributions is generally called the *Gibbs sampler* (see [GG84]) in classical mathematical statistics. The same idea is applied to a limited class of QBMs.

Separability is a well-known concept in quantum information theory. For a state ρ on the Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, it is called separable if there exist $\{\lambda_i\}_{i=1}^m \subset [0,1]$ and $\{\tau_i^{(j)}\}_{1\le i\le m, 1\le j\le n} \subset \mathcal{S}(\mathbb{C}^2)$, such that $\sum_{i=1}^m \lambda_i = 1$ and

$$\rho = \sum_{i=1}^m \lambda_i \, \tau_i^{(1)} \otimes \cdots \otimes \tau_i^{(n)}. \tag{3.9}$$

Separability is equivalent to the existence of finite sets $\mathcal{X}_1, \ldots, \mathcal{X}_n$, a probability distribution $P = P(x_1, \ldots, x_n)$ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and $\{\tau_{x_i}^{(i)} \,|\, x_i \in \mathcal{X}_i\} \subset \mathcal{S}(\mathbb{C}^2), i = 1, \cdots, n$, such that

$$\rho = \sum_{\boldsymbol{x}} P(\boldsymbol{x}) \tau_{\boldsymbol{x}}, \tag{3.10}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\tau_{\boldsymbol{x}} = \tau_{x_1}^{(1)} \otimes \cdots \otimes \tau_{x_n}^{(n)}$.

We can apply the method of Gibbs sampler to generate ρ in the form (3.10) as follows:

State renewal rule:

(i) Choose i randomly.

(ii) Update $x_i \in \mathcal{X}_i$ according to $P(x_i \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

(iii) Set the state of i to $\tau_{x_i}^{(i)}$.

It is obvious that, according to the Gibbs sampler, the state converges to ρ by repeating this process for sufficiently many times.

3.2.2 Strongly separable states and SSQBM

Definition 1. *A separable state ρ of the form (3.10) is called strongly separable if*

$$[\tau_{x_i}^{(i)}, \tau_{x_i'}^{(i)}] = 0 \quad \forall i, x_i, x_i' \in \mathcal{X}_i. \tag{3.11}$$

This is equivalent to the existence of $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$, where each $\boldsymbol{u}_i$ is a unit vector in $\mathbb{R}^3$, and a probability distribution P on $\{-1, +1\}^n$ such that

$$\rho = \sum_{x_1, \ldots, x_n} P(x_1, \ldots, x_n) \pi_{x_1}^{\boldsymbol{u}_1} \otimes \cdots \otimes \pi_{x_n}^{\boldsymbol{u}_n}, \tag{3.12}$$

where for any $\boldsymbol{u} = (u_1, u_2, u_3)$ we define

$$\begin{aligned} \pi_1^{\boldsymbol{u}} &= \frac{1}{2} \begin{bmatrix} 1 + u_3 & u_1 - iu_2 \\ u_1 + iu_2 & 1 - u_3 \end{bmatrix} \\ &= \frac{1}{2}(I + u_1 X_1 + u_2 X_2 + u_3 X_3) \\ \pi_{-1}^{\boldsymbol{u}} &= \pi_1^{-\boldsymbol{u}} = I - \pi_1^{\boldsymbol{u}}. \end{aligned}$$

Note that $\pi^{\boldsymbol{u}} = (\pi_1^{\boldsymbol{u}}, \pi_{-1}^{\boldsymbol{u}})$ represents a projective measurement (projection-valued measure, PVM) where $\pi_1^{\boldsymbol{u}}$ and $\pi_{-1}^{\boldsymbol{u}}$ are mutually orthogonal projection operators satisfying $\pi_1^{\boldsymbol{u}} + \pi_{-1}^{\boldsymbol{u}} = I$. For a spin-$\frac{1}{2}$ particle, this PVM can be physically realized by a Stern-Gerlach measurement of direction $\boldsymbol{u}$. We call $\mathcal{U} = \{\boldsymbol{u}_i\}$ a *frame* of a strongly separable state ρ.

Combination of steps (ii) and (iii) in the state renewal algorithm is equivalent to setting the quantum state of ith element to

$$\begin{aligned} \tau^{(i)} &= P(1 \mid \hat{\boldsymbol{x}}) \pi_1^{\boldsymbol{u}_i} + P(-1 \mid \hat{\boldsymbol{x}}) \pi_{-1}^{\boldsymbol{u}_i} \\ &= \mathrm{Tr}^{(i)}[\rho'_{\hat{\boldsymbol{x}}}], \end{aligned}$$

where $(\hat{\boldsymbol{x}} = (x_1, \ldots, \hat{x}_i, \ldots, x_n))$, $\mathrm{Tr}^{(i)}$ is the partial trace over $n-1$ qubit systems except for ith system,

$$\rho'_{\hat{\boldsymbol{x}}} \stackrel{\mathrm{def}}{=} \frac{\tilde{\pi}_{\hat{\boldsymbol{x}}} \rho \tilde{\pi}_{\hat{\boldsymbol{x}}}}{\mathrm{Tr}[\tilde{\pi}_{\hat{\boldsymbol{x}}} \rho \tilde{\pi}_{\hat{\boldsymbol{x}}}]}$$

and $\tilde{\pi}_{\hat{\boldsymbol{x}}} = \pi_{x_1}^{\boldsymbol{u}_1} \otimes \cdots \otimes I \otimes \cdots \otimes \pi_{x_n}^{\boldsymbol{u}_n}$ where I is in the ith position. Note that $\rho'_{\hat{\boldsymbol{x}}}$ is the post measurement state of the target state ρ for the measurement results $\hat{\boldsymbol{x}}$.

Next theorem gives the necessary and sufficient conditions for a state of the form (3.2) to be strongly separable.

Theorem 1. *A state ρ represented in the form (3.2) in terms of the parameters $(\theta^{(j)}_{i_1 \ldots i_j s_1 \ldots s_j})$ is strongly separable with frame $\mathcal{U} = \{\boldsymbol{u}_i\}$ iff there exists $(\theta^{(j)\prime}_{i_1 \ldots i_j})$ such that $\forall j, \forall i_1 < \cdots < \forall i_j, \forall s_1, \ldots, \forall s_j$,*

$$\theta^{(j)}_{i_1 \ldots i_j s_1 \ldots s_j} = \theta^{(j)\prime}_{i_1 \ldots i_j} u_{i_1 s_1} \cdots u_{i_j s_j}, \tag{3.13}$$

where $\boldsymbol{u}_i = (u_{i1}, u_{i2}, u_{i3})$.

Proof. First assume that (3.13) holds. Substituting, (3.13) to (3.2), we have

$$\begin{aligned} \rho &= \exp\Bigg[\sum_{j=1}^{k} \sum_{i_1 < \cdots < i_j} \sum_{s_1 \ldots s_j} \theta^{(j)}_{i_1 \ldots i_j s_1 \ldots s_j} X_{i_1 s_1} \ldots X_{i_j s_j} - \psi(\theta)\Bigg] && (3.14) \\ &= \exp\Bigg[\sum_{j=1}^{k} \sum_{i_1 < \cdots < i_j} \theta^{(j)\prime}_{i_1 \cdots i_j} X_{i_1 \boldsymbol{u}_{i_1}} \ldots X_{i_j \boldsymbol{u}_{i_j}} - \psi(\theta)\Bigg], && (3.15) \end{aligned}$$

where for any i and $\boldsymbol{u} = (u_s)_{s=1}^3$ we define

$$\begin{aligned} X_{i\boldsymbol{u}} &\stackrel{\mathrm{def}}{=} \sum_s u_s X_{is} \\ &= I \otimes \cdots \otimes \Big(\sum_s u_s X_s\Big) \otimes \cdots \otimes I \\ &= I \otimes \cdots \otimes X_{\boldsymbol{u}} \otimes \cdots \otimes I \end{aligned}$$

with $X_{\boldsymbol{u}} \stackrel{\mathrm{def}}{=} \sum_s u_s X_s = \pi_1^{\boldsymbol{u}} - \pi_{-1}^{\boldsymbol{u}}$. Now, let for $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{-1, +1\}^n$,

$$f(\boldsymbol{x}) = \sum_{j=1}^{k} \sum_{i_1 < \cdots < i_j} \theta^{(j)\prime}_{i_1, \ldots, i_j} x_{i_1} \ldots x_{i_j}. \tag{3.16}$$

Then

$$\begin{aligned} \sum_{\boldsymbol{x}} f(\boldsymbol{x}) \pi_{x_1}^{\boldsymbol{u}_1} \otimes \cdots \otimes \pi_{x_n}^{\boldsymbol{u}_n} &= \sum_{j=1}^{k} \sum_{i_1 < \cdots < i_j} \theta^{(j)\prime}_{i_1, \cdots, i_j} \sum_{\boldsymbol{x}} x_{i_1} \cdots x_{i_j} \pi_{x_1}^{\boldsymbol{u}_1} \otimes \cdots \otimes \pi_{x_n}^{\boldsymbol{u}_n} \\ &= \sum_{j=1}^{k} \sum_{i_1 < \cdots < i_j} \theta^{(j)\prime}_{i_1 \ldots i_j} X_{i_1 \boldsymbol{u}_{i_1}} \ldots X_{i_j \boldsymbol{u}_{i_j}}. \end{aligned} \tag{3.17}$$

Therefore, from (3.15) and (3.17)

$$\begin{aligned}\rho &= \exp\Big[\sum_{\boldsymbol{x}} f(\boldsymbol{x})\pi_{x_1}^{\boldsymbol{u}_1}\otimes\cdots\otimes\pi_{x_n}^{\boldsymbol{u}_n}-\psi(\theta)\Big]\\ &= \exp\Big[\sum_{\boldsymbol{x}}\Big\{f(\boldsymbol{x})-\psi(\theta)\Big\}\pi_{x_1}^{\boldsymbol{u}_1}\otimes\cdots\otimes\pi_{x_n}^{\boldsymbol{u}_n}\Big]\\ &= \sum_{\boldsymbol{x}}\exp\Big[f(\boldsymbol{x})-\psi(\theta)\Big]\pi_{x_1}^{\boldsymbol{u}_1}\otimes\cdots\otimes\pi_{x_n}^{\boldsymbol{u}_n}\\ &= \sum_{\boldsymbol{x}} P(\boldsymbol{x})\pi_{x_1}^{\boldsymbol{u}_1}\otimes\cdots\otimes\pi_{x_n}^{\boldsymbol{u}_n}, \qquad (3.18)\end{aligned}$$

where

$$\begin{aligned}P(\boldsymbol{x}) &= \exp\Big[f(\boldsymbol{x})-\psi(\theta)\Big]\\ &= \exp\Big[\sum_{j=1}^{k}\sum_{i_1,\ldots,i_j}\theta^{(j)\prime}_{i_1<\cdots<i_j}x_{i_1}\ldots x_{i_j}-\psi(\theta)\Big]. \qquad (3.19)\end{aligned}$$

Therefore ρ is strongly separable with frame $\mathcal{U}$. Conversely, let us assume that ρ of the form (3.2) is strongly separable with frame $\mathcal{U}$. Then, by definition, ρ is represented as (3.12) by a probability distribution P on $\{-1,+1\}^n$. Since ρ is assumed to be faithful, $P(\boldsymbol{x})$ is positive for all $\boldsymbol{x}$, and therefore P can be represented as (3.19) with $k=n$ for some $\theta'=(\theta^{(j)\prime}_{i_1,\ldots,i_j})$. Then, tracing back the previous argument in the reverse order, we have

$$\begin{aligned}\rho &= \sum_{\boldsymbol{x}} P(\boldsymbol{x})\pi_{x_1}^{\boldsymbol{u}_1}\otimes\cdots\otimes\pi_{x_n}^{\boldsymbol{u}_n}\\ &= \exp\Big[\sum_{\boldsymbol{x}}\sum_{j=1}^{n}\sum_{i_1<\cdots<i_j}\theta^{(j)\prime}_{i_1\ldots i_j}x_{i_1}\ldots x_{i_j}\pi_{x_1}^{\boldsymbol{u}_1}\ldots\pi_{x_n}^{\boldsymbol{u}_n}-\psi(\theta)\Big]\\ &= \exp\Big[\sum_{j=1}^{n}\sum_{i_1<\cdots<i_j}\theta^{(j)\prime}_{i_1\ldots i_j}X_{i_1\boldsymbol{u}_{i_1}}\ldots X_{i_j\boldsymbol{u}_{i_j}}-\psi(\theta)\Big]\\ &= \exp\Big[\sum_{j=1}^{n}\sum_{i_1<\cdots<i_j}\sum_{s_1,\cdots,s_j}\theta^{(j)\prime}_{i_1\ldots i_j}u_{i_1s_1}\ldots u_{i_js_j}X_{i_1s_1}\ldots X_{i_js_j}-\psi(\theta)\Big]. \qquad (3.20)\end{aligned}$$

Comparing (3.20) with (3.14), we have

$$\theta^{(j)}_{i_1\ldots i_j s_1\ldots s_j}=\theta^{(j)\prime}_{i_1\ldots i_j}u_{i_1s_1}\ldots u_{i_js_j}\ \ (1\le\forall j\le k)$$

and

$$\theta^{(j)\prime}_{i_1\ldots i_j}=0\ \text{ if }\ j>k.$$

which establishes the result. □

Let us now define $\mathcal{S}'(\mathcal{U})$ as the set of strongly separable states $\rho \in \mathcal{S}$ with frame $\mathcal{U}$ and $\mathcal{S}'_k(\mathcal{U}) \stackrel{\text{def}}{=} \mathcal{S}'(\mathcal{U}) \cap \mathcal{S}_k$. Then we define the elements of the $\mathcal{S}'_2(\mathcal{U})$ to be the SSQBMs. First, we give a corollary of Theorem 1.

Corollary 1. *A QBM $\rho_\theta, \theta = (h_{is}, w_{ijst})$, of the form (3.4) is strongly separable with frame $\mathcal{U} = \{\boldsymbol{u}_i\}$ iff there exists $\theta' = (h'_i, w'_{ij})$ such that $\forall i, j, \forall s, t$,*

$$h_{is} = h'_i u_{is}, \quad w_{ijst} = w'_{ij} u_{is} u_{jt}, \tag{3.21}$$

with $\boldsymbol{u}_i = (u_{i1}, u_{i2}, u_{i3})$, which means that ρ_θ is represented as (3.12) with

$$P(x_1, \ldots, x_n) = \exp\Big\{ \sum_i h'_i x_i + \sum_{i<j} w'_{ij} x_i x_j - \psi(h', w') \Big\}. \tag{3.22}$$

In particular, if $\boldsymbol{h}_i \neq 0, \forall i$, the necessary and sufficient condition for a QBM ρ_θ to be strongly separable (with some frame) is that

$$\forall i, j, \quad W_{ij} \propto \boldsymbol{h}_i \boldsymbol{h}_j^{\mathsf{T}}, \tag{3.23}$$

where

$$\boldsymbol{h}_i = \begin{bmatrix} h_{i1} \\ h_{i2} \\ h_{i3} \end{bmatrix}, \qquad W_{ij} = \begin{bmatrix} w_{ij11} & w_{ij12} & w_{ij13} \\ w_{ij21} & w_{ij22} & w_{ij23} \\ w_{ij31} & w_{ij32} & w_{ij33} \end{bmatrix}.$$

and T denotes the matrix transpose.

Proof. Obvious from the Theorem 1. □

Next, we give a state renewal rule for the SSQBM in the following theorem.

Theorem 2. *When the target state of a SSQBM is represented by (3.4) and (3.21), the state renewal rule in Subsection 3.2.1 is carried out, starting from an arbitrary initial state $\rho \in \mathcal{S}$ and arbitrary initial data $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{-1, +1\}^n$, by the following procedure.*

(i) Choose i randomly.

(ii) Using data $\hat{\boldsymbol{x}} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ available at the time, renew the state of the ith element to

$$\sigma_i = \frac{1}{2}(I + \sum_{s=1}^{3} v_{is} X_s) \tag{3.24}$$

with

$$\boldsymbol{v}_i = (v_{i1}, v_{i2}, v_{i3}) = \tanh\Big(\sum_j w'_{ij} x_j + h'_i \Big) \boldsymbol{u}_i.$$

(iii) Perform the measurement $\pi^{\boldsymbol{u}_i} = \{\pi_1^{\boldsymbol{u}_i}, \pi_{-1}^{\boldsymbol{u}_i}\}$ *to the* i*th element and update* x_i *by the measurement outcome.*

Proof. For a SSQBM, the steps (ii) and (iii) of the state renewal rule in Subsection 3.2.1 and the following steps (ii)′ and (iii)′ are same:

(ii)′ Update x_i according to

$$x_i := \begin{cases} 1 & \text{with probability } \frac{1}{1+\exp(-2I_i(\hat{\boldsymbol{x}}))} =: \lambda, \\ -1 & \text{with probability } 1-\lambda. \end{cases} \tag{3.25}$$

(iii)′ Set the state of the ith element to $\pi_1^{\boldsymbol{u}_i}$ if $x_i = 1$ or to $\pi_{-1}^{\boldsymbol{u}_i}$ if $x_i = -1$.

On the other hand, for the state renewal rule in the statement of the theorem, we can show by some calculations that the updated state is represented as (see Figure 3.1)

$$\sigma_i = \lambda\pi_1^{\boldsymbol{u}_i} + (1-\lambda)\pi_{-1}^{\boldsymbol{u}_i}. \tag{3.26}$$

Noting that, according to the von Neumann postulate for the projective measurement $\pi^{\boldsymbol{u}_i} = \{\pi_1^{\boldsymbol{u}_i}, \pi_{-1}^{\boldsymbol{u}_i}\}$, the post measurement state $\pi_{x_i}^{\boldsymbol{u}_i} (\propto \pi_{x_i}^{\boldsymbol{u}_i}\sigma_i\pi_{x_i}^{\boldsymbol{u}_i})$ depending on the measurement result x_i, we see the equivalence of (ii)′ and (iii)′ to (ii) and (iii) of the present state renewal rule. □

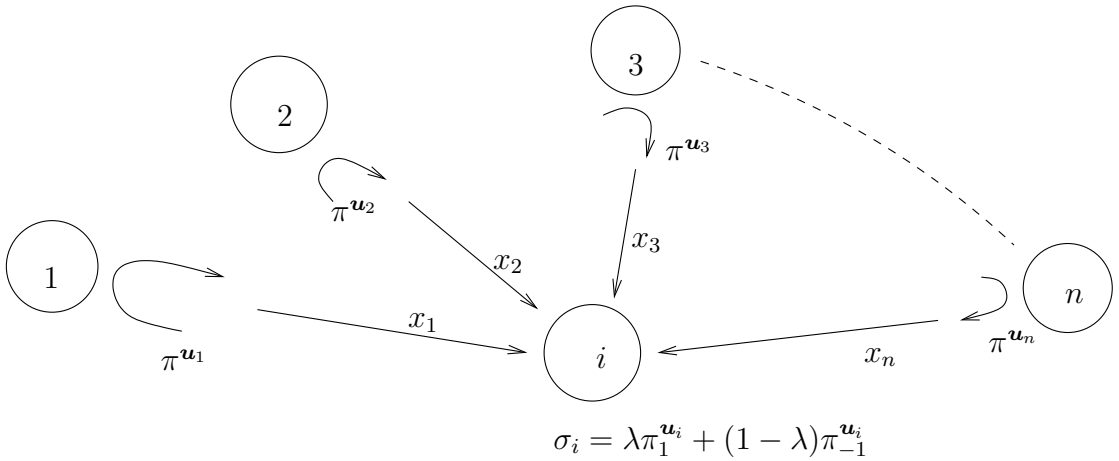

Figure 3.1: State transition process for SSQBM

Chapter 4

Geometry of quantum exponential family and the space of QBMs

This chapter presents a discussion on the information geometrical structure of the space of QBMs. First we introduce a quantum exponential family and discuss its geometrical structure in general. Then it is shown that the space of QBMs can be understood as one family in a hierarchical structure of exponential families. Furthermore, the present chapter discusses the approximation processes for QBMs and SSQBMs. Finally, we study the parameter estimation of SSQBM.

4.1 Quantum exponential families

We introduce a quantum version of exponential family (2.4) in the following. Let $\mathcal{H}$ be a finite dimensional Hilbert space and denote the totality of faithful states on $\mathcal{H}$ by

$$\mathcal{S} = \{\rho \,|\, \rho = \rho^* > 0 \text{ and } \operatorname{Tr}\rho = 1\}.$$

Suppose that a parametric family

$$\mathcal{M} = \{\rho_\theta \,|\, \theta = (\theta^i);\ i = 1, \ldots, m\} \subset \mathcal{S} \tag{4.1}$$

is represented in the form

$$\rho_\theta = \exp\Big\{C + \sum_{i=1}^{m} \theta^i F_i - \psi(\theta)\Big\}, \tag{4.2}$$

where $F_i\,(i = 1, \ldots, m), C$ are Hermitian operators and $\psi(\theta)$ is a real-valued function. We assume in addition that the operators $\{F_1, \ldots, F_m, I\}$, 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* (or *canonical*) *coordinates* $\theta = (\theta^i)$. It is easy to see that $\mathcal{S}$ is a QEF of dimension $(\dim \mathcal{H})^2 - 1$. Note also that for any $1 \leq k \leq n$ the set $\mathcal{S}_k$ of states (3.2) forms a QEF, including $\mathcal{S}_2$ of QBMs and $\mathcal{S}_1$ of product states.

If we let

$$\eta_i(\theta) \stackrel{\text{def}}{=} \mathrm{Tr}\big[\rho_\theta F_i\big], \tag{4.3}$$

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 $\mathcal{S}_2$ are given by $(h, w) = (h_{is}, w_{ijst})$ in (3.4), while the expectation coordinates are $(m, \mu) = (m_{is}, \mu_{ijst})$ defined by

$$m_{is} = \mathrm{Tr}[\rho_{h,w} X_{is}] \quad \text{and} \quad \mu_{ijst} = \mathrm{Tr}[\rho_{h,w} X_{is} X_{jt}]. \tag{4.4}$$

On the other hand, the natural coordinates of $\mathcal{S}_1$ are $\bar{h} = (\bar{h}_{is})$ in (3.5), while the expectation coordinates are $\bar{m} = (\bar{m}_{is})$ defined by

$$\bar{m}_{is} = \mathrm{Tr}[\tau_{\bar{h}} X_{is}]. \tag{4.5}$$

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

$$\bar{m}_{is} = \frac{\partial \psi_i(\bar{h}_i)}{\partial \bar{h}_{is}} = \frac{\bar{h}_{is}}{||\bar{h}_i||} \tanh(||\bar{h}_i||) \tag{4.6}$$

or as

$$\bar{h}_{is} = \frac{\bar{m}_{is}}{||\bar{m}_i||} \tanh^{-1}(||\bar{m}_i||), \tag{4.7}$$

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

4.2 A metric and affine connections

In classical information geometry (see [AN00]), a Riemannian metric, called the Fisher metric, and a one-parameter family of affine connections, called the α-connections

*It should be noted, however, (4.2) is merely one of the possible definitions of quantum exponential family. Our definition has the advantage that it is closely related to the quantum relative entropy (4.25) and is completely analogous to the classical exponential family from a purely geometrical point of view. On the other hand, it does not well fit to the framework of quantum estimation theory, which needs another definition of QEF such as the one based on symmetric logarithmic derivatives (see Section 7.4 of [AN00]).

($\alpha \in \mathbb{R}$), are canonically defined on an arbitrary manifold of probability distributions. In particular, the ($\alpha = 1$)-connection and the ($\alpha = -1$)-connection, which are also called the e-connection and m-connection respectively, together with the Fisher metric have been shown very useful in many problems in statistics and other fields. In the quantum case, on the other hand, we have infinitely many mathematical equivalents of the Fisher metric and the α-connections including the e- and m-connections defined on a manifold of quantum states. We introduce, in the present Section, an example of quantum Fisher metric and e-, m-connections, and describe their properties, mainly following [AN00]. Our choice of information geometrical structure has the advantage that it is naturally linked with QEF and quantum relative entropy. For general terms of differential geometry such as manifold, Riemannian metric and affine connection, refer, for example, to [KN63].

Let $\mathcal{H}$ be a finite dimensional Hilbert space. We consider a d-dimensional parametric family

$$\mathcal{M} = \{\rho_\theta \mid \theta = (\theta^1, \ldots, \theta^d) \in \Theta\}, \quad \Theta \subset \mathbb{R}^d$$

of faithful states on $\mathcal{H}$. Then, $\theta = (\theta^i);\ i = 1, \ldots, d$ can be considered as a coordinate system and $\mathcal{M}$ becomes a submanifold of the manifold of faithful states $\mathcal{S}$ on $\mathcal{H}$. In the following, we discuss the information geometrical structure of $\mathcal{M}$ including the case $\mathcal{M} = \mathcal{S}$. As a first step, a Riemannian metric $g = [g_{ij}]$ is defined on $\mathcal{M}$ by

$$\begin{aligned} g_{ij}(\theta) &= g(\partial_i, \partial_j), \quad \text{where} \quad \partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \theta^i} \\ &= \int_0^1 \mathrm{Tr}\Big[\rho_\theta^\lambda (\partial_i \log \rho_\theta) \rho_\theta^{1-\lambda} (\partial_j \log \rho_\theta)\Big] \mathrm{d}\lambda \\ &= \mathrm{Tr}[(\partial_i \rho_\theta)(\partial_j \log \rho_\theta)]. \end{aligned} \tag{4.8}$$

This is a quantum version of the Fisher information metric and is called the *BKM (Bogoliubov-Kubo-Mori) metric*. Next, two torsion-free affine connections, the *exponential connection* (or *e-connection* for short) $\nabla^{(\mathrm{e})}$ and the *mixture connection* (or *m-connection* for short) $\nabla^{(\mathrm{m})}$, are defined on $\mathcal{M}$ as follows:

$$\Gamma^{(\mathrm{e})}_{ij,k}(\theta) \stackrel{\text{def}}{=} g(\nabla^{(\mathrm{e})}_{\partial_i} \partial_j, \partial_k) = \mathrm{Tr}[(\partial_i \partial_j \log \rho_\theta)(\partial_k \rho_\theta)] \tag{4.9}$$

and

$$\Gamma^{(\mathrm{m})}_{ij,k}(\theta) \stackrel{\text{def}}{=} g(\nabla^{(\mathrm{m})}_{\partial_i} \partial_j, \partial_k) = \mathrm{Tr}[(\partial_i \partial_j \rho_\theta)(\partial_k \log \rho_\theta)], \tag{4.10}$$

where g is the BKM metric. Note that both $\nabla^{(\mathrm{e})}$ and $\nabla^{(\mathrm{m})}$ are mappings (covariant derivatives) which map two vector fields X, Y to $\nabla^{(\mathrm{e})}_X Y$ and to $\nabla^{(\mathrm{m})}_X Y$ respectively.

The coefficients $\Gamma^{(\mathrm{e})}_{ij,k}$ give a coordinate representation of the connection $\nabla^{(\mathrm{e})}$ relative to the metric g, while $\Gamma^{(\mathrm{e})k}_{ij}$ defined by

$$\nabla^{(\mathrm{e})}_{\partial_i}\partial_j = \sum_k \Gamma^{(\mathrm{e})k}_{ij}\partial_k \tag{4.11}$$

purely represent $\nabla^{(\mathrm{e})}$. They are related to each other by

$$\Gamma^{(\mathrm{e})}_{ij,k} = \sum_l \Gamma^{(\mathrm{e})l}_{ij} g_{kl}.$$

Similarly, we have $\Gamma^{(\mathrm{m})k}_{ij}$ for $\nabla^{(\mathrm{m})}$ such that

$$\nabla^{(\mathrm{m})}_{\partial_i}\partial_j = \sum_k \Gamma^{(\mathrm{m})k}_{ij}\partial_k \tag{4.12}$$

and

$$\Gamma^{(\mathrm{m})}_{ij,k} = \sum_l \Gamma^{(\mathrm{m})l}_{ij} g_{kl}.$$

These two connections $\nabla^{(\mathrm{e})}$ and $\nabla^{(\mathrm{m})}$ are dual with respect to the BKM metric (4.8) in the sense that, for any vector fields X, Y, Z,

$$Xg(Y,Z) = g(\nabla^{(\mathrm{e})}_X Y, Z) + g(Y, \nabla^{(\mathrm{m})}_X Z), \tag{4.13}$$

or equivalently in the component form

$$\partial_i g_{jk} = \Gamma^{(\mathrm{e})}_{ij,k} + \Gamma^{(\mathrm{m})}_{ik,j}. \tag{4.14}$$

This kind of duality for affine connections plays a key role in classical and quantum information geometry. Another notable relation between the two connections is

$$\Gamma^{(\mathrm{m})}_{ij,k} - \Gamma^{(\mathrm{e})}_{ij,k} = T_{ijk}, \tag{4.15}$$

where

$$T_{ijk}(\theta) \stackrel{\mathrm{def}}{=} 2\,\mathrm{Re} \iint\limits_{0\le\nu\le\lambda\le 1} \mathrm{Tr}\Big[\rho_\theta^\nu(\partial_i \log\rho_\theta)\rho_\theta^{\lambda-\nu}(\partial_j\log\rho_\theta)\rho_\theta^{1-\lambda}(\partial_k\log\rho_\theta)\Big]\mathrm{d}\nu\,\mathrm{d}\lambda. \tag{4.16}$$

4.3 Geometrical structure of the space of QBMs

We have shown in Section 4.1 that the totality of QBMs form a QEF. In this section, we describe the geometrical structure of QEF including the space of QBMs. Let us now consider the case when $\mathcal{M}$ is a QEF (4.2) with natural coordinates $\theta =$

(θ^i). It is then easy to check from (4.9) that the coefficients $\Gamma^{(\mathrm{e})}_{ij,k}$ or $\Gamma^{(\mathrm{e})\mathrm{k}}_{ij}$ of the e-connection are all zero. In the context of differential geometry, this means that $\mathcal{M}$ is *flat* with respect to the connection $\nabla^{(\mathrm{e})}$ (*e-flat*, for short) and $\theta = (\theta^i)$ forms an *affine coordinate system* for $\nabla^{(\mathrm{e})}$ (*e-affine coordinate system*, for short). On the other hand, the coefficients $\Gamma^{(\mathrm{m})}_{ij,k}$ of the m-connection do not vanish with respect to the natural coordinates $\theta = (\theta^i)$. However, one of the remarkable consequences of the duality (4.13) is that, if one of the two connections $\nabla^{(\mathrm{e})}$ and $\nabla^{(\mathrm{m})}$ is flat, then the other is also flat, which is referred to as the *dually flatness* of the manifold with respect to the information geometrical structure $(g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$. In the present case, the connection coefficients of $\nabla^{(\mathrm{m})}$ with respect to the expectation coordinates $\eta = (\eta_i)$ defined by (4.3) turns out to identically vanish. This means that $\mathcal{M}$ is *m-flat* with an *m-affine* coordinate system (η_i). Moreover, we have

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_i^j \quad (= 1 \text{ if } i = j, \quad 0 \text{ otherwise}) \tag{4.17}$$

and

$$\eta_i = \frac{\partial\psi}{\partial\theta^i}, \qquad \theta^i = \frac{\partial\phi}{\partial\eta_i}, \tag{4.18}$$

where ψ given in (4.2), is regarded as a function $\mathcal{M} \to \mathbb{R}$ by $\psi(\rho_\theta) = \psi(\theta)$, and $\phi : \mathcal{M} \to \mathbb{R}$ is defined by the relation

$$\phi(\rho) + \psi(\rho) = \sum_i \eta_i(\rho)\theta^i(\rho), \quad \forall\rho \in \mathcal{M}. \tag{4.19}$$

Note that equation (4.6) is an example of the first equation in (4.18). It can also be shown that

$$\phi(\rho) = -\operatorname{Tr}[\rho\, C] - S(\rho), \tag{4.20}$$

where

$$S(\rho) \stackrel{\mathrm{def}}{=} -\operatorname{Tr}[\rho \log \rho] \tag{4.21}$$

is the von Neumann entropy. In particular, for the QEF $\mathcal{S}_k$ of states (3.2), we have $C = 0$ and hence $\phi(\rho) = -S(\rho)$. We note that the existence of m-affine coordinates $\eta = (\eta_i)$ and functions ψ, ϕ satisfying the relations (4.17), (4.18) and (4.19) is ensured as a general property of dually flat space (see Theorem 3.6 in [AN00]), although it is not difficult to directly verify these relations for a QEF (4.2).

Finally, let us rewrite (4.17) into a form which will be useful in later arguments. Noting that (4.17) is written as

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \frac{\partial\eta_i}{\partial\eta_j} \tag{4.22}$$

and that $\left\{ \left(\frac{\partial}{\partial \eta_j} \right)_\rho \right\}$ form a basis of the tangent space $T_\rho(\mathcal{M})$, we have

$$g\left(\left(\frac{\partial}{\partial \theta^i} \right)_\rho, \partial' \right) = \partial' \eta_i \qquad \forall \partial' \in T_\rho(\mathcal{M}). \tag{4.23}$$

Similarly, we have

$$g\left(\left(\frac{\partial}{\partial \eta_i} \right)_\rho, \partial' \right) = \partial' \theta^i \qquad \forall \partial' \in T_\rho(\mathcal{M}), \tag{4.24}$$

although we use only (4.23) in this thesis.

4.4 Geometry of quantum relative entropy

In this section, we focus on the quantum relative entropy

$$D(\rho\|\sigma) \stackrel{\text{def}}{=} \mathrm{Tr}[\rho(\log\rho - \log\sigma)] \tag{4.25}$$

for two density operators $\rho, \sigma \in \mathcal{M}$ and describe its properties related to the dually flat structure $(g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ of a QEF $\mathcal{M}$ as the continuation of the previous section. First, we note that the relation

$$D(\rho\|\sigma) = \phi(\rho) + \psi(\sigma) - \sum_i \eta_i(\rho)\theta^i(\sigma) \tag{4.26}$$

holds for any $\rho, \sigma \in \mathcal{M}$ with ψ and ϕ defined in (4.18) and (4.19). From this, we have

$$D(\rho\|\sigma) + D(\sigma\|\tau) - D(\rho\|\tau) = \sum_i \{\eta_i(\rho) - \eta_i(\sigma)\}\{\theta^i(\tau) - \theta^i(\sigma)\}. \tag{4.27}$$

Moreover, it can be shown that (4.27), with the positivity

$$D(\rho\|\sigma) \geq 0, \qquad D(\rho\|\sigma) = 0 \quad \text{iff} \quad \rho = \sigma, \tag{4.28}$$

completely characterizes the quantum relative entropy D.

Let us clarify the geometrical meaning of the right hand side of (4.27). In general, given a coordinate system ξ^i and an affine connection ∇ with coefficients Γ^k_{ij}, a geodesic with respect to ∇ is defined by the second order ordinary differential equation

$$\ddot{\xi}^k + \sum_{i,j} \Gamma^k_{ij} \dot{\xi}^i \dot{\xi}^j = 0. \tag{4.29}$$

If, in addition, ξ^i is an affine coordinate system with respect to a flat ∇, the equation becomes $\ddot{\xi}^k = 0$ or equivalently

$$\xi_t^i = t\xi_0^i + (1-t)\xi_1^i. \tag{4.30}$$

In particular, an e-geodesic in the QEF (4.2) is given by

$$\theta_t^i = t\theta_0^i + (1-t)\theta_1^i. \tag{4.31}$$

This turns out to be equivalent to

$$\log \rho_t = t \log \rho_0 + (1-t) \log \rho_1 - \psi(t),$$

where $\psi(t)$ is the normalization constant. In other words, an e-geodesic of a QEF is itself a one dimensional QEF. On the other hand, an m-geodesic is represented as

$$\eta_{ti} = t\eta_{0i} + (1-t)\eta_{1i}. \tag{4.32}$$

If we consider the case $\mathcal{M} = \mathcal{S}$, the m-geodesic can be written as

$$\rho_t = t\rho_0 + (1-t)\rho_1. \tag{4.33}$$

Such a family of states $\{\rho_t\}$ is called a (one dimensional) mixture family, which is related to the origin of the name "mixture connection", but note that (4.32) is not generally represented as (4.33) unless $\mathcal{M} = \mathcal{S}$.

Let $\gamma : [0,1] \to \mathcal{M}$ be an m-geodesic such that $\gamma(0) = \sigma, \gamma(1) = \rho$ and $\delta : [0,1] \to \mathcal{M}$ be an e-geodesic such that $\delta(0) = \sigma, \delta(1) = \tau$. Then, from (4.31) and (4.32) we obtain

$$\dot{\gamma}(0) = \sum_i \{\eta_i(\rho) - \eta_i(\sigma)\} \Big(\frac{\partial}{\partial \eta_i}\Big)_\sigma \in T_\sigma(\mathcal{M}) \tag{4.34}$$

and

$$\dot{\delta}(0) = \sum_i \{\theta^i(\tau) - \theta^i(\sigma)\} \Big(\frac{\partial}{\partial \theta^i}\Big)_\sigma \in T_\sigma(\mathcal{M}). \tag{4.35}$$

Hence, from (4.17) we have

$$g(\dot{\gamma}(0), \dot{\delta}(0)) = \sum_i \{\eta_i(\rho) - \eta_i(\sigma)\}\{\theta^i(\tau) - \theta^i(\sigma)\} \tag{4.36}$$

which coincides with the right hand side of (4.27). We thus obtain the following theorem:

Theorem 3. (Pythagorean relation) *Let ρ, σ, τ be three points in the manifold $\mathcal{M}$ such that the m-geodesic connecting ρ and σ is orthogonal at σ to the e-geodesic connecting σ and τ with respect to the BKM Riemannian metric (see Figure 4.1). Then, the generalized Pythagorean relation*

$$D(\rho\|\sigma) + D(\sigma\|\tau) = D(\rho\|\tau) \tag{4.37}$$

holds.

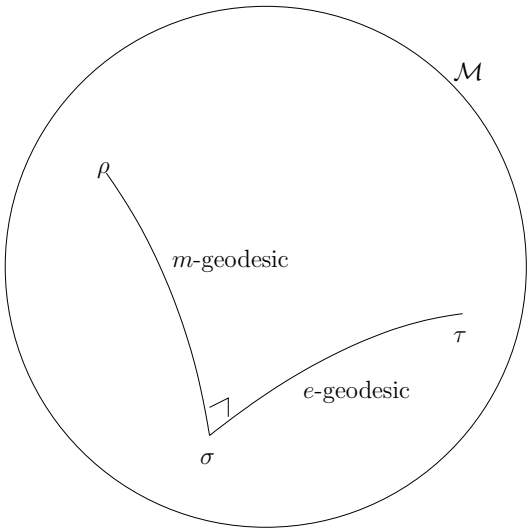

Figure 4.1: Pythagorean theorem $D(\rho\|\sigma) + D(\sigma\|\tau) = D(\rho\|\tau)$

Next, we define the m- and e-projections. Let $\mathcal{M}$ be a QEF of the form (4.2), and $\mathcal{N}$ be a smooth submanifold of $\mathcal{M}$. For an arbitrary point $\rho \in \mathcal{M}$, let $D(\rho\|\cdot)|_{\mathcal{N}}$ be a function on $\mathcal{N}$ defined by $\mathcal{N} \ni \sigma \mapsto D(\rho\|\sigma)$. When this function is stationary (i.e., the derivative is zero for every direction in $\mathcal{N}$) at a point $\sigma \in \mathcal{N}$, we say that σ is an *m-projection* of ρ onto $\mathcal{N}$. Similarly, when $D(\cdot\|\rho)|_{\mathcal{N}}$ is stationary at $\sigma \in \mathcal{N}$, we say that σ is an *e-projection* of ρ onto $\mathcal{N}$. Then, we have the following two theorems which are closely related to Theorem 3.

Theorem 4. *The necessary and sufficient condition for σ to be an m-projection (resp. e-projection) of ρ onto $\mathcal{N}$ is that the m-geodesic (resp. e-geodesic) connecting ρ and σ is orthogonal to $\mathcal{N}$ at σ (see Figure 4.2).*

Theorem 5. *If $\mathcal{N}$ is e-autoparallel (resp. m-autoparallel) in $\mathcal{M}$ in the sense that $\mathcal{N}$ forms an affine subspace in e-affine coordinates (θ^i) (resp. m-affine coordinates) of*

$\mathcal{M}$, then an m-projection (resp. e-projection) is unique and attains the minimum of $D(\rho\|\cdot)|_{\mathcal{N}}$ (resp. $D(\cdot\|\rho)|_{\mathcal{N}}$). Note that an e-autoparallel submanifold is nothing but a QEF.

Remark 3. *For proofs of the Theorem* 3*, Theorem* 4 *and Theorem* 5*, refer to [AN00].*

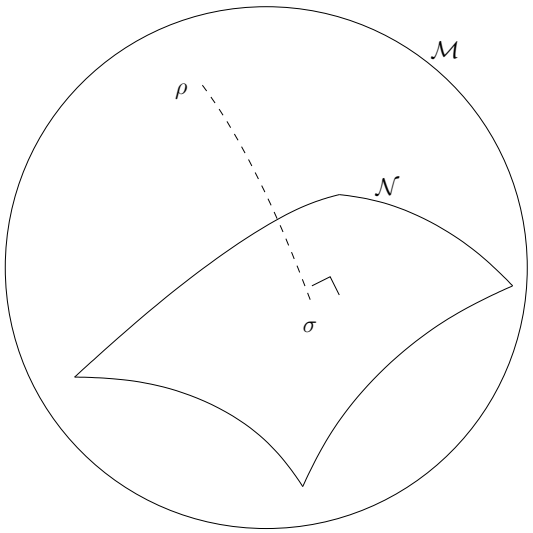

Figure 4.2: Projection σ from $\rho \in \mathcal{M}$ to $\mathcal{N}$

Finally, we note another property of D for later use. We have the Taylor expansion of $D(\rho\|\sigma)$ (see [AN00], p 55) as

$$D(\rho\|\sigma) \stackrel{\text{def}}{=} \frac{1}{2}\sum_{ij} g_{ij}(\rho)\Delta\theta^i\Delta\theta^j + \frac{1}{6}\sum_{ijk} h_{ijk}(\rho)\Delta\theta^i\Delta\theta^j\Delta\theta^k + \cdots, \tag{4.38}$$

where $\Delta\theta^i \stackrel{\text{def}}{=} \theta^i(\sigma) - \theta^i(\rho)$. Here, the second-order coefficients g_{ij} are the components of the BKM metric and the third-order coefficients h_{ijk} are determined from g_{ij} and the connection coefficients by

$$h_{ijk} \stackrel{\text{def}}{=} \partial_i g_{jk} + \Gamma^{(\mathrm{e})}_{jk,i} = \Gamma^{(\mathrm{e})}_{ij,k} + \Gamma^{(\mathrm{m})}_{ik,j} + \Gamma^{(\mathrm{e})}_{jk,i}, \tag{4.39}$$

where the second equality is due to (4.14).

4.5 Approximation processes for QEF and SSQBM

The purpose of this section is to discuss the approximation processes for QEF and SSQBM corresponding to that for CBM. We also study the parameter estimation of

SSQBM. Recalling that the approximation process for the CBM is well understood within the general framework of exponential families, as mentioned in Section 2.2, we consider the corresponding problem† for QEF (see Figure 4.3). Then the following theorem is obtained as in the classical case.

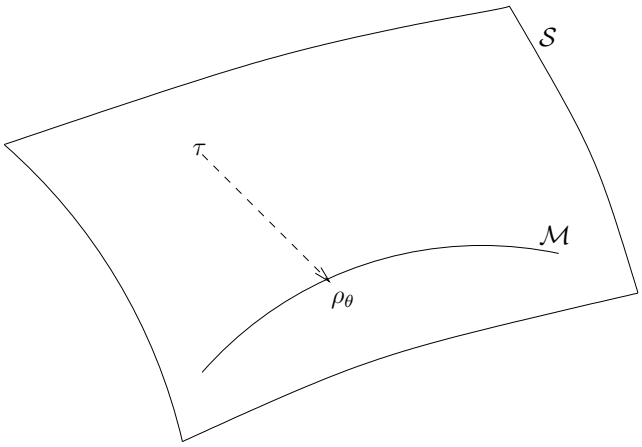

Figure 4.3: Approximation process for QEF.

Theorem 6. *Suppose that we are given a density operator $\tau \in \mathcal{S}$ and a QEF $\mathcal{M} = \{\rho_\theta\} \subset \mathcal{S}$ of the form (4.2), and consider the minimization $\min_\theta D(\tau\|\rho_\theta)$, where D denotes the quantum relative entropy defined in (4.25). Then,*

$$\theta^* = \arg\min_\theta D(\tau\|\rho_\theta) \tag{4.40}$$

iff

$$\eta_i(\theta^*) = \mathrm{Tr}[\tau F_i], \quad \forall i. \tag{4.41}$$

The gradient algorithm for computing θ^* is represented as

$$\theta^i \stackrel{\mathrm{def}}{=} \theta^i + \Delta\theta^i,$$

where

$$\begin{aligned} \Delta\theta^i &\stackrel{\mathrm{def}}{=} -\varepsilon \sum_\beta \gamma^{ij}(\theta)\partial_j D(\tau\|\rho_\theta) \\ &= \varepsilon \sum_j \gamma^{ij}(\theta)\{\mathrm{Tr}[\tau F_j] - \eta_j(\theta)\}. \end{aligned} \tag{4.42}$$

†It should be noted that the approximation process for QBM can be considered as the m-projection from $\mathcal{S}$ onto $\mathcal{S}_2$. On the other hand, the naive mean-field approximation discussed in Chapter 5 can be understood as the e-projection from $\mathcal{S}_2$ onto $\mathcal{S}_1$

Let us now consider the geometrical structure of the space of SSQBMs. Following the definition of SSQBM in Section 3.2, we have natural diffeomorphisms $\mathcal{S}'(\mathcal{U}) \simeq \mathcal{P}$ and $\mathcal{S}'_k(\mathcal{U}) \simeq \mathcal{P}_k$. We give the following theorem which describes the information geometrical structure of $\mathcal{S}'_k(\mathcal{U})$.

Theorem 7. *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 (e-autoparallel). The induced information geometrical structure $(g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ on $\mathcal{S}'_k(\mathcal{U})$ is dually flat and is equivalent to that of $\mathcal{P}_k$.*

Proof. Obvious from the fact that an element of $\mathcal{S}'_k(\mathcal{U})$ is represented as (see the proof of Theorem 1 in Subsection 3.2.2)

$$\begin{aligned} \rho &= \sum_{x_1,\ldots,x_n} P(x_1,\ldots,x_n)\pi_{x_1}^{\boldsymbol{u}_1} \otimes \cdots \otimes \pi_{x_n}^{\boldsymbol{u}_n} \\ &= \exp\left[\sum_{j=1}^{n} \sum_{i_1<\cdots<i_j} \theta^{(j)\prime}_{i_1\ldots i_j} X_{i_1\boldsymbol{u}_{i_1}} \ldots X_{i_j\boldsymbol{u}_{i_j}} - \psi(\theta')\right] \end{aligned} \tag{4.43}$$

by $P \in \mathcal{P}_k$ and $\theta' = (\theta^{(j)\prime}_{i_1,\ldots,i_j})$. □

We now revisit the approximation process for SSQBMs (see Figure 4.4). Suppose that we are given a density operator $\tau \in \mathcal{S}$, and 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. Let

$$\tau' \stackrel{\mathrm{def}}{=} \sum_{\boldsymbol{x}} \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}}\, \tau\, \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} \mathrm{Tr}\left[\tau\tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}}\right] \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} \in \mathcal{S}'(\mathcal{U}),$$

where $\tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} = \pi_{x_1}^{\boldsymbol{u}_1} \otimes \cdots \otimes \pi_{x_n}^{\boldsymbol{u}_n}$ for $\boldsymbol{x} = (x_1,\ldots,x_n)$. Then we obtain

$$\mathrm{Tr}[\tau\tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}}] = \mathrm{Tr}[\tau'\tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}}], \ \ \forall\, \boldsymbol{x} \tag{4.44}$$

from which it implies

$$\mathrm{Tr}[\tau(\log\sigma - \log\tau')] = \mathrm{Tr}[\tau'(\log\sigma - \log\tau')], \ \ \forall\, \sigma \in \mathcal{S}'(\mathcal{U}). \tag{4.45}$$

The above equation is rewritten as $\mathrm{Tr}(\tau - \tau')(\log\sigma - \log\tau') = 0$, which equals to

$$D(\tau\|\tau') + D(\tau'\|\sigma) - D(\tau\|\sigma) = \mathrm{Tr}(\tau - \tau')(\log\sigma - \log\tau') = 0. \tag{4.46}$$

Thus we obtain the Pythagorean relation

$$D(\tau\|\sigma) = D(\tau\|\tau') + D(\tau'\|\sigma) \tag{4.47}$$

for any $\sigma \in \mathcal{S}'(\mathcal{U})$, which implies

$$\tau' = \arg\min_{\sigma \in \mathcal{S}'(\mathcal{U})} D(\tau \| \sigma). \tag{4.48}$$

Let us now consider a SSQBM

$$\rho_{\theta'} = \sum P_{\theta'}(\boldsymbol{x}) \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} \qquad (P_{\theta'} \in \mathcal{P}_2) \tag{4.49}$$

and

$$\tau' = \sum q'(\boldsymbol{x}) \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} \qquad (q' \in \mathcal{P}). \tag{4.50}$$

By replacing σ by $\rho_{\theta'}$ in (4.47), we obtain the following Pythagorean relation

$$D(\tau \| \rho_{\theta'}) = D(\tau \| \tau') + D(\tau' \| \rho_{\theta'}).$$

This implies

$$\arg\min_{\theta'} D(\tau \| \rho_{\theta'}) = \arg\min_{\theta'} D(\tau' \| \rho_{\theta'})$$

where $D(\tau' \| \rho_{\theta'})$ is nothing but the KL divergence $D(q' \| P_{\theta'})$. This leads to the following theorem.

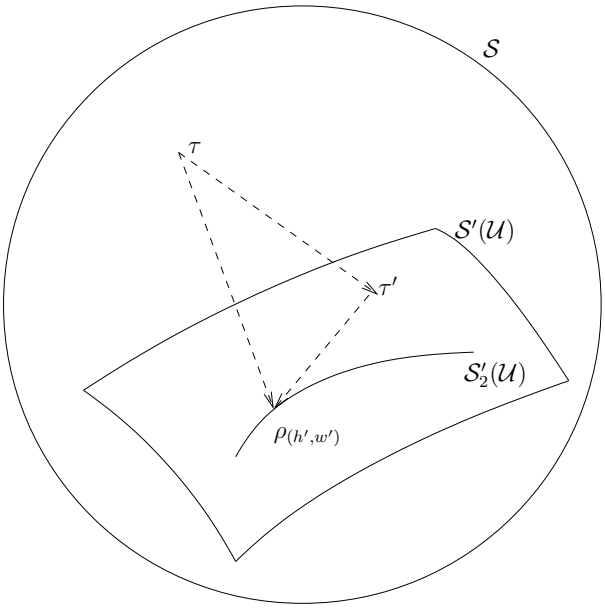

Figure 4.4: Approximation process for SSQBM.

Theorem 8. *The approximation process to find*

$$\theta'^* = \arg\min_{\theta'} D(\tau \| \rho_{\theta'}),$$

where $\theta' = (h', w')$, 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 to be equivalent to the approximation problem for the CBM addressed in Section 2.2 by $\mathcal{S}'_2(\mathcal{U}) \simeq \mathcal{P}_2$.

Finally, in this section, we consider the parameter estimation of SSQBMs. Suppose that we are given N independent quantum systems, each of which is a SSQBM $\rho_{\theta'}$ ($\theta' = (h'_i, w'_{ij})$) with frame $\mathcal{U}$ and that we perform a projective measurement $\pi_{\mathcal{U}}$ on each SSQBM given by the PVM

$$\pi_{\mathcal{U}} \stackrel{\text{def}}{=} \{\tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \{-1,+1\}^n}, \tag{4.51}$$

where $\tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} = \pi^{\boldsymbol{u_1}}_{x_1} \otimes \pi^{\boldsymbol{u_2}}_{x_2} \otimes \pi^{\boldsymbol{u_n}}_{x_n}$ for $\boldsymbol{x} = (x_1, \ldots, x_n)$, to yield a set of data $\{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)\}$.

Let $\hat{P}$ denote the empirical distribution of $\{\boldsymbol{x}(t)\}_{t=1}^N$ and let

$$\hat{\tau} \stackrel{\text{def}}{=} \sum_{\boldsymbol{x}} \hat{P}(\boldsymbol{x}) \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}}. \tag{4.52}$$

Then, recalling that an SSQBM $\rho_{\theta'} \in \mathcal{S}'_2(\mathcal{U})$ is represented as

$$\rho_{\theta'} = \sum_{\boldsymbol{x}} P_{\theta'}(\boldsymbol{x}) \tilde{\pi}^{\mathcal{U}}_{\boldsymbol{x}} \tag{4.53}$$

by a CBM $P_{\theta'}$, we have

$$\begin{aligned}
\hat{\theta'} &\stackrel{\text{def}}{=} \hat{\theta}'_{\text{MLE}}(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)) \\
&= \arg\max_{\theta'} P_{\theta'}(\boldsymbol{x}(1)) P_{\theta'}(\boldsymbol{x}(2)) \cdots P_{\theta'}(\boldsymbol{x}(N)) \\
&= \arg\min_{\theta'} D(\hat{P} \| P_{\theta'}) \\
&= \arg\min_{\theta'} D(\hat{\tau} \| \rho_{\theta'}).
\end{aligned} \tag{4.54}$$

This means that the MLE for the data is given by the approximation process for $\hat{\tau}$.

Chapter 5

Information geometry of mean-field approximation for QBMs

This chapter discusses the information geometrical interpretation of the mean-field approximation for QBMs. First we derive the naive mean-field equation for QBMs using the concepts of e- and m- projections. Then the higher-order corrections are discussed noting the correspondence of the coefficients of Plefka expansion to the information geometrical quantities of a Taylor expansion of the quantum relative entropy.

5.1 The e-, m-projections and naive mean-field approximation

In this section, we derive the naive mean-field equation for QBMs explicitly from the viewpoint of information geometry. Suppose that we are interested in calculating the expectations $m_{is} = \mathrm{Tr}[\rho_{h,w} X_{is}]$ 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 $\rho_{h,w} \in \mathcal{S}_2$ is to use quantities obtained in the form of expectation with respect to some relevant $\tau_{\bar{h}} \in \mathcal{S}_1$. As explained in the Chapter 1, T. Tanaka [Tan96, Tan00] 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.

In the following arguments, we regard $\mathcal{S}_2$ as a QEF with the natural coordinates

$(\theta^\alpha) = (h_{is}, w_{ijst})$ and the expectation coordinates $(\eta_\alpha) = (m_{is}, \mu_{ijst})$ (see (4.4)), where α is an index denoting $\alpha = (i,s)$ or $\alpha = (i,j,s,t)$. First, let us consider the m-projection onto $\mathcal{S}_1$ and show that it preserves the expectations m_{is}. Note that since $\mathcal{S}_1$ is e-autoparallel in $\mathcal{S}_2$, the m-projection is unique and attains the minimum of D by Theorem 5. Suppose $\rho = \rho_{h,w} \in \mathcal{S}_2$ is given and $\tau \in \mathcal{S}_1$ be its m-projection. Let γ be the m-geodesic such that $\gamma(1) = \rho, \gamma(0) = \tau$. Then, from (4.32) we have

$$\eta_\alpha(\gamma(t)) = t\,\eta_\alpha(\rho) + (1-t)\,\eta_\alpha(\tau), \tag{5.1}$$

and, in particular,

$$m_{is}(\gamma(t)) = t\,m_{is}(\rho) + (1-t)\,m_{is}(\tau). \tag{5.2}$$

Hence, substituting $\theta^i := h_{is}, \eta_i := m_{is}, \partial' := \dot\gamma(0)$ and $\rho := \tau$ into (4.23) we get

$$g\left(\left(\frac{\partial}{\partial h_{is}}\right)_\tau, \dot\gamma(0)\right) = \left.\frac{\mathrm{d}m_{is}(\gamma(t))}{\mathrm{d}t}\right|_{t=0} = m_{is}(\rho) - m_{is}(\tau). \tag{5.3}$$

Since

$$T_\tau(\mathcal{S}_1) = \mathrm{span}\left\{\left(\frac{\partial}{\partial h_{is}}\right)_\tau\right\}, \tag{5.4}$$

it follows from (5.3) and Theorem 4 that $m_{is}(\rho) = m_{is}(\tau)$. This means that the expectation values do not change if we use the m-projection.

Next, we show that the naive mean-field equation is derived by considering e-projection. Suppose that $\tau = \tau_{\bar h} \in \mathcal{S}_1$ is an e-projection of $\rho = \rho_{h,w} \in \mathcal{S}_2$ onto $\mathcal{S}_1$, and let γ be the e-geodesic such that $\gamma(1) = \rho, \gamma(0) = \tau$. Note that from (4.31)

$$\left.\frac{\mathrm{d}\theta^\alpha(\gamma(t))}{\mathrm{d}t}\right|_{t=0} = \theta^\alpha(\rho) - \theta^\alpha(\tau), \tag{5.5}$$

i.e.,

$$\left.\frac{\mathrm{d}h_{is}(\gamma(t))}{\mathrm{d}t}\right|_{t=0} = h_{is}(\rho) - h_{is}(\tau) = h_{is} - \bar h_{is} \tag{5.6}$$

$$\left.\frac{\mathrm{d}w_{ijst}(\gamma(t))}{\mathrm{d}t}\right|_{t=0} = w_{ijst}(\rho) = w_{ijst} \quad \text{since} \quad w_{ijst}(\tau) = 0. \tag{5.7}$$

Now, recall that $(\bar m_{is})$ defined by (4.5) form a coordinate system of $\mathcal{S}_1$, so that

$$T_\tau(\mathcal{S}_1) = \mathrm{span}\left\{\left(\frac{\partial}{\partial \bar m_{is}}\right)_\tau\right\}_{is}. \tag{5.8}$$

Hence it follows from Theorem 4 that, $\forall i, s$,

$$\begin{aligned}
0 &= g\left(\dot{\gamma}(0), \left(\frac{\partial}{\partial \bar{m}_{is}}\right)_\tau\right) \\
&= \sum_\alpha \left.\frac{\mathrm{d}\theta^\alpha(\gamma(t))}{\mathrm{d}t}\right|_{t=0} g\left(\left(\frac{\partial}{\partial \theta^\alpha}\right)_\tau, \left(\frac{\partial}{\partial \bar{m}_{is}}\right)_\tau\right) \\
&= \sum_\alpha \left.\frac{\mathrm{d}\theta^\alpha(\gamma(t))}{\mathrm{d}t}\right|_{t=0} \left(\frac{\partial \eta_\alpha}{\partial \bar{m}_{is}}\right)_\tau \qquad \text{(from (4.23))} \\
&= \sum_{j,t} (h_{jt} - \bar{h}_{jt}) \frac{\partial m_{jt}}{\partial \bar{m}_{is}} + \sum_{j<k} \sum_{t,u} w_{jktu} \frac{\partial \mu_{jktu}}{\partial \bar{m}_{is}} \\
&= h_{is} - \bar{h}_{is} + \sum_{(i<)k,u} w_{iksu} \bar{m}_{ku} + \sum_{(i>)j,t} w_{jits} \bar{m}_{jt} \\
&\quad \text{(since } m_{jt} = \bar{m}_{jt} \text{ and } \mu_{jktu} = \bar{m}_{jt}\bar{m}_{ku} \text{ on } \mathcal{S}_1\text{)} \\
&= h_{is} - \bar{h}_{is} + \sum_{j,t} w_{ijst} \bar{m}_{jt},
\end{aligned}$$

where the last equality follows by letting

$$w_{iist} \stackrel{\text{def}}{=} 0 \ \text{ and } \ w_{ijst} \stackrel{\text{def}}{=} w_{jits} \quad (i > j). \tag{5.9}$$

We thus obtain

$$\bar{h}_{is} = h_{is} + \sum_{j,t} w_{ijst} \bar{m}_{jt}. \tag{5.10}$$

Both (4.6) (or (4.7)) and (5.10) together give the naive mean-field equation for QBMs. It should be remarked that this naive mean-field equation may have several solutions $\{\bar{h}_{is}\}$ for a given set of $\{h_{is}, w_{ijst}\}$ as in the classical case, which correspond to the fact that e-projection onto an e-autoparallel submanifold is not unique in general.

In this section, we have shown some properties of m- and e-projections based on the geometrical characterization given in Theorem 4, but note that the same properties can also be derived in several different ways; for instance, we can use the relations (4.18)-(4.19) and (4.26) for ψ and ϕ to derive them as was shown for the classical case in Subsection 2.3.2.

5.2 Plefka expansion and higher-order mean-field approximations

Although the naive mean-field approximation is used extensively as a common tool to compute characteristic quantities of multi-particle systems, it is necessary

to consider higher-order mean-field approximations to improve the accuracy in some situations. In this section, we discuss a method to derive higher-order mean-field approximations which utilizes a Taylor expansion of the quantum relative entropy. This coincides with the *Plefka expansion* of the Gibbs potential as pointed out at the end of the Section. We elucidate the correspondence of the coefficients of the Taylor expansion to the information geometrical quantities such as the metric and the e-, m-connections.

We start the discussion of higher-order approximations by recalling that the elements of $\mathcal{S}_2$ are parametrized as $\rho_{h,w}$ by $h = (h_{is})$ and $w = (w_{ijst})$. This means that $(\theta^\alpha) = (h, w)$ forms a coordinate system of the manifold $\mathcal{S}_2$. In viewing $\mathcal{S}_2$ as a QEF, (h, w) is a natural coordinate system, while the corresponding expectation coordinate system is given by $(\eta_\alpha) = (m, \mu)$ with $m = (m_{is})$ and $\eta = (\eta_{ijst})$. Let us now define a third (or hybrid) coordinate system $(\xi^\alpha) \stackrel{\text{def}}{=} (m, w)$. The elements of $\mathcal{S}_2$ are then parametrized by (m, w), which we denote by $\hat{\rho}_{m,w}$ to avoid confusion with $\rho_{h,w}$. Note that

$$\mathcal{S}_2 = \{\rho_{h,w} \mid (h, w) : \text{free}\} = \{\hat{\rho}_{m,w} \mid (m, w) : \text{free}\} \tag{5.11}$$

and that

$$\hat{\rho}_{m,w} = \rho_{h,w} \iff \forall i, \forall s,\ m_{is} = \mathrm{Tr}[\rho_{h,w} X_{is}]. \tag{5.12}$$

For an arbitrarily fixed w, a submanifold of $\mathcal{S}_2$ is defined by

$$\mathcal{F}(w) \stackrel{\text{def}}{=} \{\rho_{h,w} \mid h : \text{free}\} = \{\hat{\rho}_{m,w} \mid m : \text{free}\}. \tag{5.13}$$

As a special case we have

$$\mathcal{F}(0) = \{\rho_{h,0} \mid h : \text{free}\} = \{\hat{\rho}_{m,0} \mid m : \text{free}\} = \mathcal{S}_1$$

which is the manifold of product states. We see that the family $\{\mathcal{F}(w)\}_w$ forms a foliation of $\mathcal{S}_2$ as

$$\mathcal{S}_2 = \bigcup_w \mathcal{F}(w). \tag{5.14}$$

Similarly, for an arbitrarily fixed m we define

$$\begin{aligned} \mathcal{A}(m) &\stackrel{\text{def}}{=} \{\hat{\rho}_{m,w} \mid w : \text{free}\} \\ &= \{\rho \in \mathcal{S}_2 \mid \forall i, \forall s,\ m_{is} = \mathrm{Tr}[\rho X_{is}]\}, \end{aligned} \tag{5.15}$$

which yields another foliation of $\mathcal{S}_2$ as

$$\mathcal{S}_2 = \bigcup_m \mathcal{A}(m). \tag{5.16}$$

These foliations have several special properties. First, for any w, $\mathcal{F}(w)$ is defined by fixing w which is a part of e-affine coordinates $(\theta^\alpha) = (h, w)$ of $\mathcal{S}_2$. This implies that $\mathcal{F}(w)$ is e-autoparallel in $\mathcal{S}_2$ in the sense mentioned in Theorem 5. On the other hand, each $\mathcal{A}(m)$ is m-autoparallel in $\mathcal{S}_2$. Furthermore, $\forall w, \forall m$,

$$\mathcal{F}(w) \perp \mathcal{A}(m) \ \text{ at } \ \sigma \in \mathcal{F}(w) \cap \mathcal{A}(m). \tag{5.17}$$

To see this, we note that the tangent spaces of $\mathcal{F}(w)$ and $\mathcal{A}(m)$ at σ are given by

$$T_\sigma(\mathcal{F}(w)) = \mathrm{span}\left\{ \left(\frac{\partial}{\partial h_{is}} \right)_\sigma \right\}_{is} \tag{5.18}$$

and

$$T_\sigma(\mathcal{A}(m)) = \mathrm{span}\left\{ \left(\frac{\partial}{\partial \mu_{jktu}} \right)_\sigma \right\}_{jktu}. \tag{5.19}$$

The inner product $g\left(\frac{\partial}{\partial h_{is}}, \frac{\partial}{\partial \mu_{jktu}}\right)$ is a special case of $g\left(\frac{\partial}{\partial \theta^\alpha}, \frac{\partial}{\partial \eta_\beta}\right)$ with $\alpha \neq \beta$, hence is zero from (4.17), which proves (5.17). These properties mean that $\{\mathcal{F}(w)\}_w$ and $\{\mathcal{A}(m)\}_m$ jointly give an example of mutually dual foliations (see Figure 5.1) defined in [AN00] (pp 75-76). It is now easy to see from Theorem 3 that for any points $\rho \in \mathcal{A}(m)$

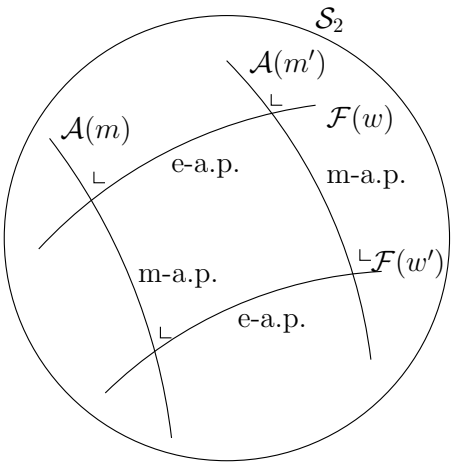

Figure 5.1: Mutually dual foliations of $\mathcal{S}_2$. Here, "e-a.p." and "m-a.p." stand for "e-autoparallel" and "m-autoparallel", respectively.

and $\tau \in \mathcal{F}(w)$ with the intersecting point $\sigma \in \mathcal{A}(m) \cap \mathcal{F}(w)$ the Pythagorean relation (4.37) holds. We also note that, for any w and m, both $\mathcal{F}(w)$ and $\mathcal{A}(m)$ are dually flat with respect to their e-, m-connections and the BKM metrics. This is obvious for $\mathcal{F}(w)$ because $\mathcal{F}(w)$ itself is a QEF. On the other hand, since $\mathcal{A}(m)$ is m-autoparallel

in $\mathcal{S}_2$ which is m-flat, we can easily see that $\mathcal{A}(m)$ is also m-flat, and hence is dually flat as mentioned in Section 4.3. Actually, (μ_{ijst}) and (w_{ijst}) restricted to $\mathcal{A}(m)$ turn out m-affine and e-affine coordinate systems respectively.

Let us now restate the problem which motivates both the naive mean-field approximation and its higher-order extension. Given $h = (h_{is})$ and $w = (w_{ijst})$ arbitrarily, consider the problem of calculating the expectations $\mathrm{Tr}[\rho_{h,w} X_{is}]$ or their approximations from (h, w). From

$$\begin{aligned} \forall i, \forall s,\ \frac{\partial}{\partial m_{is}} D(\hat{\rho}_{m,w} \| \rho_{h,w}) = 0 \qquad &\iff D(\hat{\rho}_{m,w} \| \rho_{h,w}) = \min_{m'} D(\hat{\rho}_{m',w} \| \rho_{h,w}) \\ &\iff \hat{\rho}_{m,w} = \arg\min_{\sigma \in \mathcal{F}(w)} D(\sigma \| \rho_{h,w}) = \rho_{h,w} \\ &\iff \forall i, \forall s,\ m_{is} = \mathrm{Tr}[\rho_{h,w} X_{is}], \end{aligned}$$

where the last equivalence follows from (5.12), we have the expectations m_{is} as the solution of the equation

$$\frac{\partial}{\partial m_{is}} D(\hat{\rho}_{m,w} \| \rho_{h,w}) = 0. \tag{5.20}$$

Of course, this method is practical only when the relative entropy $D(\hat{\rho}_{m,w} \| \rho_{h,w})$ is not too complicated as a function of the variables $m = (m_{is})$, which cannot be expected in general when n, the number of elements in the system, is large. On the other hand, if we let $w = 0$ in the first argument of $D(\hat{\rho}_{m,w} \| \rho_{h,w})$, then the resulting $D(\hat{\rho}_{m,0} \| \rho_{h,w})$ becomes the sum of simple functions of $m = (m_{is})$, and hence the equation

$$\frac{\partial}{\partial m_{is}} D(\hat{\rho}_{m,0} \| \rho_{h,w}) = 0 \tag{5.21}$$

is much more tractable than the original one in (5.20). When $\|w\|$ is sufficiently small so that $D(\hat{\rho}_{m,w} \| \rho_{h,w})$ is well approximated by $D(\hat{\rho}_{m,0} \| \rho_{h,w})$, the solution of (5.21) will give a good approximation for the true expectations. This is nothing but the idea of naive mean-field approximation. Actually, equation (5.21) means that $\hat{\rho}_{m,0}$ is an e-projection of $\rho_{h,w}$ onto $\mathcal{F}(0) = \mathcal{S}_1$, which turns out to be equivalent to (5.10) as shown in the previous Section.

Now that the accuracy of the naive mean-field approximation depends on how close the function $D(\hat{\rho}_{m,w} \| \rho_{h,w})$ is to its substitute $D(\hat{\rho}_{m,0} \| \rho_{h,w})$. Therefore, it is natural to expect that the approximation can be improved by properly retrieving the difference $D(\hat{\rho}_{m,0} \| \rho_{h,w}) - D(\hat{\rho}_{m,w} \| \rho_{h,w})$ up to a certain order of w. This is the information geometrical interpretation of the idea due to Plefka [Ple82, Ple06], and we call the expansion of the difference with respect to w the *Plefka expansion* following Tanaka [Tan00] who originally gave a similar interpretation in the classical case.

From the information geometrical viewpoint, the gist of this approach is the fact that the Pythagorean relation (4.37) holds for the three points $\rho_{h,w}, \hat{\rho}_{m,w}$ and $\hat{\rho}_{m,0}$ (see Figure 5.2) so that we have

$$D(\hat{\rho}_{m,0}\|\rho_{h,w}) - D(\hat{\rho}_{m,w}\|\rho_{h,w}) = D(\hat{\rho}_{m,0}\|\hat{\rho}_{m,w}). \tag{5.22}$$

The problem is thus reduced to expansion of $D(\hat{\rho}_{m,0}\|\hat{\rho}_{m,w})$ with respect to w. Noting

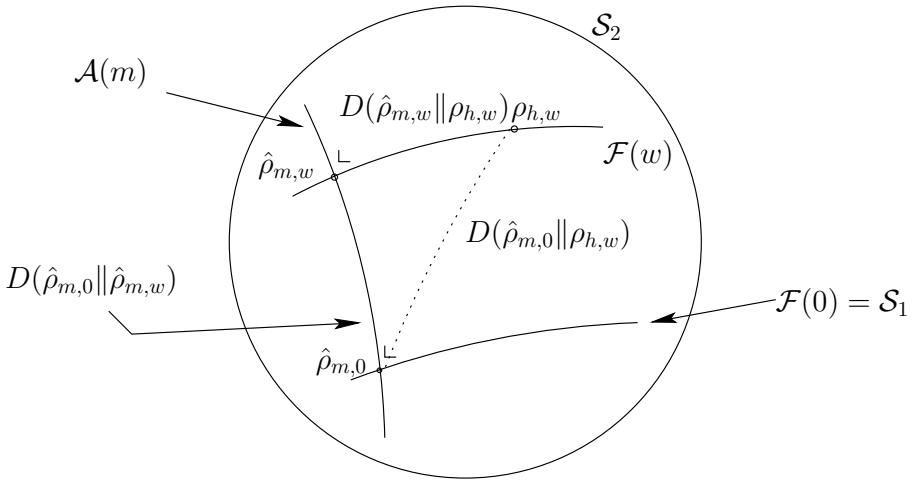

Figure 5.2: Pythagorean relation $D(\hat{\rho}_{m,0}\|\rho_{h,w}) = D(\hat{\rho}_{m,0}\|\hat{\rho}_{m,w}) + D(\hat{\rho}_{m,w}\|\rho_{h,w})$

that $\hat{\rho}_{m,0}$ and $\hat{\rho}_{m,w}$ are points on the manifold $\mathcal{A}(m)$ for which the coupling coefficients $w = (w_{ijst})$ form a coordinate system, the expansion formula (4.38) with (4.39) is applied to yield the Plefka expansion

$$D(\hat{\rho}_{m,0}\|\hat{\rho}_{m,w}) = \frac{1}{2}\sum_{IJ} g_{IJ}\, w_I\, w_J + \frac{1}{6}\sum_{IJK} h_{IJK}\, w_I\, w_J\, w_K + \cdots \tag{5.23}$$

with

$$h_{IJK} = \hat{\partial}_I g_{JK} + \Gamma^{(\mathrm{e})}_{JK,I} = \Gamma^{(\mathrm{e})}_{IJ,K} + \Gamma^{(\mathrm{m})}_{IK,J} + \Gamma^{(\mathrm{e})}_{JK,I}, \tag{5.24}$$

where the indices I, J, K represent quadruplets of indices such as (i, j, s, t). Here, g_{IJ}, $\Gamma^{(\mathrm{e})}_{IJ,K}$ and $\Gamma^{(\mathrm{m})}_{IJ,K}$ are respectively the components of the BKM metric, the e-connection and the m-connection of the manifold $\mathcal{A}(m)$, and $\hat{\partial}_I$ denotes $\frac{\partial}{\partial w_I}$ for the coordinates (w_I) of $\mathcal{A}(m)$, all evaluated at the point $\hat{\rho}_{m,0}$.

More specifically, it follows from (4.8) that

$$g_{IJ} = \int_0^1 \mathrm{Tr}\Big[\hat{\rho}^{\lambda}(\hat{\partial}_I \log\hat{\rho})\hat{\rho}^{1-\lambda}(\hat{\partial}_J \log\hat{\rho})\Big]\mathrm{d}\lambda, \tag{5.25}$$

where $\hat{\rho} = \hat{\rho}_{m,0}$ and

$$\hat{\partial}_I \log \hat{\rho} = \frac{\partial}{\partial w_I} \log \hat{\rho}_{m,w}\Big|_{w=0}. \tag{5.26}$$

We have the following relation

$$\hat{\partial}_I \log \hat{\rho} = (X_{is} - m_{is})\,(X_{jt} - m_{jt}) \tag{5.27}$$

for $I = (i, j, s, t)$ which can be calculated as shown below. It follows from (5.37) that

$$\Big(\frac{\partial \chi}{\partial w_{ijst}}\Big)_\xi = \mu_{ijst} = m_{is}m_{jt} \quad (\text{when } w = 0), \tag{5.28}$$

$$\Big(\frac{\partial \chi}{\partial m_{is}}\Big)_\xi = -h_{is}, \tag{5.29}$$

where $(\cdot)_\xi$ means that the partial differentiations are those with respect to the coordinate system $\xi = (m, w)$. Note that $\hat{\partial}_I$ in (5.27) for $I = (i, j, s, t)$ is $(\frac{\partial}{\partial w_{ijst}})_\xi$ evaluated at $w = 0$. Now, from equations (3.4) and (5.34), we obtain

$$\begin{aligned} \log \hat{\rho}_{m,w} &= \sum_{ku} h_{ku} X_{ku} + \sum_{kluv} w_{kluv} X_{ku} X_{lv} - \psi \\ &= \sum_{ku} h_{ku}(X_{ku} - m_{ku}) + \sum_{kluv} w_{kluv} X_{ku} X_{lv} - \chi \end{aligned} \tag{5.30}$$

and

$$\hat{\partial}_I \log \hat{\rho}_{m,w} = \sum_{ku} \Big(\frac{\partial h_{ku}}{\partial w_I}\Big)_\xi (X_{ku} - m_{ku}) + X_{is}X_{jt} - \Big(\frac{\partial \chi}{\partial w_I}\Big)_\xi. \tag{5.31}$$

Here,

$$\begin{aligned} \Big(\frac{\partial h_{ku}}{\partial w_I}\Big)_\xi &= -\Big(\frac{\partial^2 \chi}{\partial w_I \partial m_{ku}}\Big)_\xi \qquad (\text{from equation (5.29)}) \\ &= -\Big(\frac{\partial}{\partial m_{ku}}\Big(\frac{\partial \chi}{\partial w_I}\Big)_\xi\Big)_\xi \\ &= -\frac{\partial\, m_{is}m_{jt}}{\partial m_{ku}} \qquad (\text{from equation (5.28)}) \\ &= -\delta_{ik}\delta_{su} m_{jt} - \delta_{jk}\delta_{tu} m_{is}. \end{aligned}$$

Substituting this and (5.28) in (5.31), we have

$$\begin{aligned} \hat{\partial}_I \log \hat{\rho}_{m,w} &= -(X_{is} - m_{is})m_{jt} - (X_{jt} - m_{jt})m_{is} + X_{is}X_{jt} - m_{is}m_{jt} \\ &= (X_{is} - m_{is})(X_{jt} - m_{jt}), \end{aligned}$$

which completes the derivation. As for the third-order coefficients h_{IJK} in (5.24), we first note that $w = (w_I)$ is an e-affine coordinate system of $\mathcal{A}(m)$ as mentioned before

and hence $\Gamma^{(\mathrm{e})}_{IJ,K} = 0$. As a consequence, we have

$$\begin{aligned} h_{IJK} &= \hat{\partial}_I g_{JK} = \Gamma^{(\mathrm{m})}_{IK,J} \\ &= 2\,\mathrm{Re} \iint\limits_{0\le\nu\le\lambda\le 1} \mathrm{Tr}\Big[\hat{\rho}^{\nu}(\hat{\partial}_I \log\hat{\rho})\hat{\rho}^{\lambda-\nu}(\hat{\partial}_J \log\hat{\rho})\hat{\rho}^{1-\lambda}(\hat{\partial}_K \log\hat{\rho})\Big]\mathrm{d}\nu\,\mathrm{d}\lambda, \end{aligned} \tag{5.32}$$

where we have invoked (4.15) and (4.16).

If we succeed in obtaining explicit expressions for g_{IJ}, h_{IJK} and if they are not too complicated as functions of the mean variables $m = (m_{is})$, we can take

$$D(\hat{\rho}_{m,0}\|\rho_{h,w}) - \frac{1}{2}\sum_{IJ} g_{IJ}\, w_I\, w_J$$

or

$$D(\hat{\rho}_{m,0}\|\rho_{h,w}) - \frac{1}{2}\sum_{IJ} g_{IJ}\, w_I\, w_J - \frac{1}{6}\sum_{IJK} h_{IJK}\, w_I\, w_J\, w_K$$

as a substitute of

$$D(\hat{\rho}_{m,w}\|\rho_{h,w}) = D(\hat{\rho}_{m,0}\|\rho_{h,w}) - D(\hat{\rho}_{m,0}\|\hat{\rho}_{m,w}) \tag{5.33}$$

in equation (5.20) to improve the naive mean-field approximation (5.21).

Before closing this section, we verify the equivalence between our discussion and the original formulation of Plefka for expansion of the Gibbs potential. Let us define a function $\chi : \mathcal{S}_2 \to \mathbb{R}$ by

$$\chi(\rho) \stackrel{\mathrm{def}}{=} \psi(\rho) - \sum_{i,s} m_{is}(\rho) h_{is}(\rho) \tag{5.34}$$

$$= S(\rho) + \sum_{i<j,s,t} \mu_{ijst}(\rho) w_{ijst}(\rho), \qquad \forall \rho \in \mathcal{S}_2 \tag{5.35}$$

where the second equality follows from (4.19) and $\phi(\rho) = -S(\rho)$. Noting that equation (4.18) yields

$$\mathrm{d}\psi = \sum_{\alpha} \eta_\alpha \mathrm{d}\theta^\alpha = \sum_{i,s} m_{is}\mathrm{d}h_{is} + \sum_{i<j,s,t} \mu_{ijst}\mathrm{d}w_{ijst}, \tag{5.36}$$

we obtain from equation (5.34) that

$$\begin{aligned} \mathrm{d}\chi &= \mathrm{d}\psi - \sum_{i,s} m_{is}\mathrm{d}h_{is} - \sum_{i,s} h_{is}\mathrm{d}m_{is} \\ &= \sum_{i<j,s,t} \mu_{ijst}\mathrm{d}w_{ijst} - \sum_{i,s} h_{is}\mathrm{d}m_{is}. \end{aligned} \tag{5.37}$$

This shows that it is natural to represent χ as a function of independent variables (m, w) by $\chi(\hat{\rho}_{m,w})$, which corresponds to what is called the Gibbs potential in [Ple82, Ple06]. Now, it is immediate from (4.26) and (5.34) that, for any $(m, w) = (m_{is}, w_{ijst})$,

$$\chi(\hat{\rho}_{m,w}) = S(\hat{\rho}_{m,0}) + \sum_{i<j,s,t} m_{is} m_{jt} w_{ijst} + D(\hat{\rho}_{m,0} \| \hat{\rho}_{m,w}). \tag{5.38}$$

This implies that the expansions of $\chi(\hat{\rho}_{m,w})$ and $D(\hat{\rho}_{m,0} \| \hat{\rho}_{m,w})$ with respect to w are equivalent except for 0th and first-order terms.

5.3 Some possible extensions

The present information geometrical formulation can be applied to models other than the QBMs. For instance, an immediate application is to employ this method for the mean-field approximation of higher-order QBMs. The higher-order QBMs (3.2) can be represented as

$$\rho = \exp\left\{ \sum_{i,s} h_{is} X_{is} + \sum_{i<j} \sum_{s,t} w_{ijst} X_{is} X_{jt} + \sum_{i<j<k} \sum_{s,t,u} v_{ijkstu} X_{is} X_{jt} X_{ku} + \cdots - \psi(\theta) \right\}. \tag{5.39}$$

Another important extension is to consider the q-state quantum spin model where each element in the system has a Hilbert space $\mathbb{C}^q$ and the whole system corresponds to a Hilbert space $(\mathbb{C}^q)^{\otimes n}$, where n is the number of elements. Obviously, we obtain the present model for $q = 2$. It would be useful to find many other applications of this framework for quantum statistical models beyond the given two examples.

Chapter 6

Concluding remarks

6.1 Conclusions

In this thesis, we have found a possible quantum extension of the CBM and investigated some properties from the information geometrical point of view. We outline the conclusions of this work:

(1) A quantum extension of the CBM has been found which we call QBM. We have also defined a restricted class of it called SSQBM.

(2) A state renewal rule for SSQBM has been proposed based on that for CBM. The convergence of this state renewal rule is guaranteed by the Gibbs sampler.

(3) It has been shown that the totality of QBMs forms a quantum exponential family. Moreover, the information geometrical structure of the space of QBMs has also been discussed.

(4) The approximation process for QBMs has been studied.

(5) The geometrical structure of the totality of SSQBMs has been shown to be equivalent to that of CBMs. The approximation process for SSQBM and the estimation of parameters has also been studied in this framework.

(6) Mean-field approximation for QBMs has been studied from the information geometrical point of view. Naive mean-field equation for QBMs has been explicitly derived utilizing the concept of e-projection.

(7) Higher-order mean-field approximations based on the Plefka expansion have also been discussed in this geometrical framework.

6.2 Future work

We mention several open problems remaining for further research:

(1) To study the dynamical and neural aspects of QBMs including the computational capabilities. It may be interesting to apply such a model to optimization problems (possibly from the point of view of adiabatic quantum computation and quantum annealing).

(2) A more general study concerning the statistical and probabilistic aspects of QBMs (as a special case of quantum Markov random field).

(3) To investigate the deeper relations to quantum statistical mechanics.

(4) To establish the higher-order mean-field approximations explicitly and to study higher-order QBMs.

Bibliography

[AK89] E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines*. Wiley-Interscience, New York, 1989.

[AHS85] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science* vol. 9, 6, pp.147–169, 1985.

[Ama85] S. Amari. *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics 28, Springer-Verlag, 1985.

[Ama98] S. Amari. Natural gradient works efficiently in learning. *Neural Computation* **10**, pp. 251–276, 1998.

[AIS01] S. Amari, S. Ikeda and H. Shimokawa. Information Geometry of mean field approximation: α-divergence approach in [OS01], pp.241–257.

[AKN92] 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.

[AN00] S. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society and Oxford University Press, 2000.

[BK00] C. Bhattacharyya and S. S. Keerthi. Information geometry and Plefka's mean-field theory. *J.Phys. A: Math. Gen.* **33**, pp.1307–1312, 2000.

[GG84] S. Geman, D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis & Machine Intelligence* PAMI-6, 6, pp.721–741, 1984.

[HKP91] J. Hertz, A. Krogh and R. G. Palmer. *Introduction to the theory of neural computation*. Addison-Wesley Publishing Company, 1991.

[JGJS99] M. Jordan, Z. Ghahramani, T. S. Jakkola and L. Saul. An introduction to variational methods for graphical models in *Learning in Graphical Models*, Cambridge, MA: MIT Press, 1999, pp. 105–161.

[KN63] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry* vol 1. Wiley-Interscience, New York, 1963.

[OS01] Manfred Opper and David Saad. *Advanced Mean Field Methods - Theory and Practice*, MIT Press, Cambridge, MA, 2001.

[NC02] Michael A. Nielsen and Issac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2002.

[NK95] 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).

[NT97] K. Nakanishi and H. Takayama. Mean-field theory for a spin-glass model of neural networks: TAP free energy and the paramagnetic to spin-glass transition. *J. Phys. A: Math. Gen.* **30**, pp. 8085–8094, 1997.

[PA87] C. Peterson and J. Anderson. A mean-field theory learning algorithm for neural networks. *Complex Systems* **1**, pp. 995–1019, 1987.

[Ple82] T. Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *J. Phys. A: Math. Gen.* **15**, pp.1971–1978, 1982.

[Ple06] T. Plefka. Expansion of the Gibbs potential for quantum many-body systems: General formalism with applications to the spin glass and the weakly non-ideal Bose gas. *Phys. Rev. E* **73**, 016129, 2006.

[Tan96] T. Tanaka. Information Geometry of mean field Approximation. *IEICE Trans. Fundamentals* E79-A **5**, pp.709–715, 1996.

[Tan00] T. Tanaka, Information Geometry of mean field Approximation. *Neural Computation*, **12** pp.1951–1968, 2000.

[TAP77] D. J. Thouless, P. W. Anderson and R. G. Palmer. Solution of a 'Solvable Model of a Spin Glass'. *Phil. Mag.* **35**, pp.593–601, 1977.

[YN05] N. Yapage, H. Nagaoka. A Quantum Extension of Boltzmann machine: An information geometrical viewpoint. *Proc. of the ERATO Conference on Quantum Information Science (EQIS'05)*, (Tokyo, Japan), pp.204–205, 2005.

[YN06] N. Yapage, H. Nagaoka. Information geometry of mean field approximation for quantum Boltzmann machines. *Proc. of the Asian Conference on Quantum Information Science (AQIS'06)*, (Beijing, China), pp.143–144, 2006.

[YN08] 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.

Author Biography

Nihal Yapage was born in Galle, Sri Lanka. He received the B. Sc. (Special) degree in Mathematics from the University of Ruhuna, Matara, Sri Lanka, in May 1998. Then he joined the Department of Mathematics of the same university as an assistant teacher. He won a Japanese government (Monbukagakusho) scholarship in 2002 for postgraduate studies and now on study leave from the University of Ruhuna. He received the M. Eng. degree in quantum information theory from the Graduate School of Information Systems, University of Electro-Communications, Tokyo, Japan, in March 2005. He has been with the same working towards the Ph.D. degree since April 2005. Mr. Yapage is a student member of the American Physical Society.

List of Publications Related to the Thesis

(1) N. Yapage, H. Nagaoka. A Quantum Extension of Boltzmann machine: An information geometrical viewpoint. *Proceedings of the ERATO Conference on Quantum Information Science (EQIS'05)*, (Tokyo, Japan), pp.204–205, 2005.

(2) N. Yapage, H. Nagaoka. Information geometry of mean field approximation for quantum Boltzmann machines. *Proceedings of the Asian Conference on Quantum Information Science (AQIS'06)*, (Beijing, China), pp.143–144, 2006.

(3) 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.