Resolution of Singularities and Stochastic Complexity of Complete Bipartite Graph-Type Spin Model in Bayesian Estimation

Miki Aoyagi and Sumio Watanabe

Precision and Intelligence Laboratory, Tokyo Institute of Technology,
4259 Nagatsuda, Midori-ku, R2-5, Yokohama, 226-8503, Japan
aoyagi.m.aa@m.titech.ac.jp, swatanab@pi.titech.ac.jp

Abstract. In this paper, we obtain the main term of the average stochastic complexity for certain complete bipartite graph-type spin models in Bayesian estimation. We study the Kullback function of the spin model by using a new method of eigenvalue analysis first and use a recursive blowing up process for obtaining the maximum pole of the zeta function which is defined by using the Kullback function. The papers [1,2] showed that the maximum pole of the zeta function gives the main term of the average stochastic complexity of the hierarchical learning model.

1 Introduction

The spin model in statistical physics is also called the Boltzmann machine. In mathematics, the spin model can be regarded as the Bayesian network or the graphical model. So, the model is widely used in many fields. However, its many theoretical problems have been unsolved so far. Clarifying its stochastic complexity is one of those problems in the artificial intelligence. Stochastic complexities are used in model selection methods well. Therefore, it is an important problem to know the behavior of stochastic complexities. The fact that the spin model is a non-regular statistical model makes the problem difficult. We cannot analyze it by using classic theories of regular statistical models, since their Fisher matrix functions are singular. This is the reason why we may not apply model selection methods such as AIC[3], TIC[4], HQ[5], NIC[6], BIC[7], MDL[8] to the non-regular statistical model.

Recently, the papers [1,2] showed that the maximum pole of the zeta function of hierarchical learning models gives the main term of their average stochastic complexity. The results are for all non-regular statistical models which include not only the spin model but also the layered neural network, the reduced rank regression and the normal mixture model. It is known that the desingularization of an arbitrary polynomial can be obtained by using a blowing up process (Hironaka's Theorem [9]). Therefore, the maximum pole is obtained by a blowing up process of its Kullback function.

However, in spite of such results, it is still difficult to obtain stochastic complexities by the following two main reasons. (1) The desingularization of any

V. Torra, Y. Narukawa, and Y. Yoshida (Eds.): MDAI 2007, LNAI 4617, pp. 443–454, 2007.

polynomial in general, although it is known as a finite process, is very difficult. Furthermore, most of the Kullback functions of non-regular statistical models are degenerate (over $\mathbb{R}$) with respect to their Newton polyhedrons, singularities of the Kullback functions are not isolated, and the Kullback functions are not simple polynomials, i.e., they have parameters. Therefore, to obtain the desingularization of the Kullback functions is a new problem even in mathematics, since these singularities are very complicated and so most of them have not been investigated so far. (2) Since the main purpose is for obtaining the maximum pole, getting the desingularization is not enough for us. We need some techniques for comparing poles. However, no theorems for comparing poles have developed as far as we know.

Therefore, the exact main terms of the average stochastic complexities of spin models were unknown, while upper bounds were reported in several papers [10,11]. In this paper, we clarify explicitly the main terms of the stochastic complexities of certain complete bipartite graph-type spin models, by using a new method of eigenvalue analysis and a recursive blowing up process (Theorem 4).

We already have obtained the exact main terms of the average stochastic complexities for the three layered neural network in [12] and [13], and the reduced rank regression in [14].

There are usually direct and inverse problems to be considered. The direct problem is to solve the stochastic complexity with a known true density function. The inverse problem is to find proper learning models and learning algorithms under the condition of an unknown true density function. The inverse problem is important for practical usage, but in order to solve the inverse problem, first the direct problem has to be solved. So it is necessary and crucial to construct fundamental mathematical theories for solving the direct problem. Our standpoint comes from that direct problem.

This paper consists of five sections. In Section 2, we summary Bayesian learning models [1,2]. Section 3 contains Hironaka's Theorem [9]. In Section 4, our main results are stated. In Section 5, we conclude our paper.

2 Bayesian Learning Models

Let $x^n := \{x_i\}_{i=1}^n$ be n training samples randomly selected from a true probability density function $q(x)$. Consider a learning model $p(x|w)$, where w is a parameter. We assume that the true probability density function $q(x)$ is defined by $q(x) = p(x|w^*)$, where w^* is constant.

Let $K_n(w) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \log \dfrac{p(x^n|w^*)}{p(x^n|w)}$.

The average stochastic complexity or the free energy is defined by

$$F(n) = -E_n\{\log \int \exp(-nK_n(w))\psi(w)dw\}.$$

Let $p(w|x^n)$ be the *a posteriori* probability density function:

$p(w|x^n) = \frac{1}{Z_n}\psi(w)\prod_{i=1}^{n} p(x_i|w)$, where $\psi(w)$ is an *a priori* probability density function on the parameter set W and $Z_n = \int_W \psi(w)\prod_{i=1}^n p(x_i|w)\mathrm{d}w$.

So the average inference $p(x|x^n)$ of the Bayesian density function is given by $p(x|x^n) = \int p(x|w)p(w|x^n)\mathrm{d}w$.

Set $K(q||p) = \sum_x q(x)\log\frac{q(x)}{p(x|x^n)}$. This function represents a measure function between the true density function $q(x)$ and the predictive density function $p(x|x^n)$. It always takes a positive value and satisfies $K(q||p) = 0$ if and only if $q(x) = p(x|x^n)$.

The generalization error $G(n)$ is its expectation value over training samples:

$$G(n) = E_n\{\sum_x p(x|w^*)\log\frac{p(x|w^*)}{p(x|x^n)}\},$$

which satisfies $G(n) = F(n+1) - F(n)$ if it has an asymptotic expansion.

Define the zeta function $J(z)$ of a complex variable z for the learning model by $J(z) = \int K(w)^z\psi(w)\mathrm{d}w$, where $K(w)$ is the Kullback function: $K(w) = \sum_x p(x|w^*)\log\frac{p(x|w^*)}{p(x|w)}$. Then, for the maximum pole $-\lambda$ of $J(z)$ and its order θ, we have

$$F(n) = \lambda\log n - (\theta-1)\log\log n + O(1), \tag{1}$$

where $O(1)$ is a bounded function of n, and

$$G(n) \cong \lambda/n - (\theta-1)/(n\log n) \text{ as } n\to\infty. \tag{2}$$

Therefore, our aim is to obtain λ and θ in this paper.

We state Lemmas 2 and 3 in [14] below which are frequently used in this paper. Define the norm of a matrix $C = (c_{ij})$ by $||C|| = \sqrt{\sum_{i,j}|c_{ij}|^2}$.

Lemma 1 ([14]). *Let U be a neighborhood of $w_0 \in \mathbb{R}^d$, $C(w)$ be an analytic $H\times H'$ matrix function from U, $\psi(w)$ be a C^∞ function from U with compact support, and P, Q be any regular $H\times H$, $H'\times H'$ matrices, respectively. Then the maximum pole of $\int_U ||C(w)||^{2z}\psi(w)dw$ is the same of $\int_U ||PC(w)Q||^{2z}\psi(w)dw$.*

3 Resolution of Singularities

In this section, we introduce Hironaka's Theorem [9] on a resolution of singularities and construction of blowing up. Blowing up is a main tool in a resolution of singularities of an algebraic variety.

Theorem 1 (Hironaka [9])
Let f be a real analytic function in a neighborhood of $w = (w_1,\cdots,w_d) \in \mathbb{R}^d$ with $f(w) = 0$. There exists an open set $V \ni w$, a real analytic manifold U and a proper analytic map μ from U to V such that

(1) $\mu : U - \mathcal{E} \to V - f^{-1}(0)$ *is an isomorphism, where* $\mathcal{E} = \mu^{-1}(f^{-1}(0))$,
(2) for each $u \in U$, *there is a local analytic coordinate system* $(u_1, \cdots, u_n)$ *such that* $f(\mu(u)) = \pm u_1^{s_1} u_2^{s_2} \cdots u_n^{s_n}$, *where* $s_1, \cdots, s_n$ *are non-negative integers.*

Next we explain about blowing up along a manifold used in this paper [15]. Define a manifold $\mathcal{M}$ by gluing k open sets $U_i \cong \mathbb{R}^d$, $i = 1, 2, \cdots, k (d \geq k)$ as follows. Denote a coordinate system of U_i by $(\xi_{1i}, \cdots, \xi_{di})$.

Define an equivalence relation $(\xi_{1i}, \xi_{2i}, \cdots, \xi_{di}) \sim (\xi_{1j}, \xi_{2j}, \cdots, \xi_{dj})$ at $\xi_{ji} \neq 0$ and $\xi_{ij} \neq 0$, by $\xi_{ij} = 1/\xi_{ji}, \xi_{jj} = \xi_{ii}\xi_{ji}, \xi_{hj} = \xi_{hi}/\xi_{ji} (1 \leq h \leq k, h \neq i, j), \xi_{\ell j} = \xi_{\ell i} (k+1 \leq \ell \leq d)$, and set $\mathcal{M} = \coprod_{i=1}^{k} U_i / \sim$. Also define $\pi : \mathcal{M} \to \mathbb{R}^d$ by $U_i \ni (\xi_{1i}, \cdots, \xi_{ni}); \mapsto (\xi_{ii}\xi_{1i}, \cdots, \xi_{ii}\xi_{i-1i}, \xi_{ii}, \xi_{ii}\xi_{i+1i}, \cdots, \xi_{ii}\xi_{ki}, \xi_{k+1i}, \cdots, \xi_{di})$.

This map is well-defined and called blowing up along

$$X = \{(w_1, \cdots, w_k, w_{k+1}, \cdots, w_d) \in \mathbb{R}^d \mid w_1 = \cdots = w_k = 0\}.$$

The blowing map satisfies (1) $\pi : \mathcal{M} \to \mathbb{R}^d$ is proper and (2) $\pi : \mathcal{M} - \pi^{-1}(X) \to \mathbb{R}^d - X$ is isomorphic.

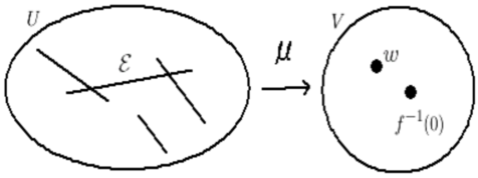

Fig. 1. Hironaka Theorem

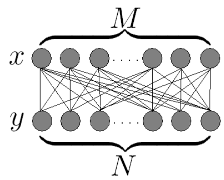

Fig. 2. A complete bipartite graph-type spin model

4 Spin Models

For simplicity, we use the notation $\mathrm{d}a$ instead of $\prod_{i=1}^{H} \prod_{j=1}^{H'} \mathrm{d}a_{ij}$ for $a = (a_{ij})$.

Let $2 \leq M \in \mathbb{N}$ and $N \in \mathbb{N}$. Consider a complete bipartite graph-type spin model

$$p(x, y|a) = \frac{\exp(\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij} x_i y_j)}{Z(a)}, Z(a) = \sum_{x_i = \pm 1, y_i = \pm 1,} \exp(\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij} x_i y_j),$$

with $x = (x_j) \in \{1, -1\}^M$ and $y = (y_j) \in \{1, -1\}^N$.

We have

$$p(x|a) = \frac{\prod_{j=1}^{N} (\prod_{i=1}^{M} \exp(a_{ij} x_i) + \prod_{i=1}^{M} \exp(-a_{ij} x_i))}{Z(a)}$$

$$= \{\prod_{j=1}^{N} (\prod_{i=1}^{M} (1 + x_i \tanh(a_{ij})) + \prod_{i=1}^{M} (1 - x_i \tanh(a_{ij})))\} \frac{\prod_{j=1}^{N} \prod_{i=1}^{M} \cosh(a_{ij})}{Z(a)}$$

$$= \frac{\prod_{j=1}^{N} \prod_{i=1}^{M} \cosh(a_{ij})}{Z(a)}$$

$$\times \prod_{j=1}^{N} (2 \sum_{0 \leq p \leq M/2} \sum_{i_1 < \cdots < i_{2p}} x_{i_1} x_{i_2} \cdots x_{i_{2p}} \tanh(a_{i_1 j}) \tanh(a_{i_2 j}) \cdots \tanh(a_{i_{2p} j})).$$

Let $B = (b_{ij}) = (\tanh(a_{ij}))$.

Denote $B^J = \prod_{i=1}^M \prod_{j=1}^N b_{ij}^{J_{ij}}$ and $x^J = \prod_{i=1}^M x_i^{\sum_{j=1}^N J_{ij}}$, where $J = (J_{ij})$ is an $M \times N$ matrix with $J_{ij} \in \{0, 1\}$.

Then we have

$$p(x|a) = \frac{2^N \prod_{j=1}^N \prod_{i=1}^M \cosh(a_{ij})}{Z(a)} \sum_{J: \sum_{i=1}^M J_{ij} \text{ even for all j}} B^J x^J.$$

Let $Z(b) = \dfrac{Z(a)}{2^N \prod_{j=1}^N \prod_{i=1}^M \cosh(a_{ij})}$. Set $\mathcal{I} = \{I \in \{0,1\}^M | \sum_{i=1}^M I_i \text{ is even }\}$, and $B^I = \sum_{\substack{J: \sum_{i=1}^M J_{ij} \text{ is even} \\ \sum_{j=1}^N J_{ij} = I_i \mod 2}} B^J$ for $I \in \mathcal{I}$. Then we have $p(x|a) = \dfrac{1}{Z(b)} \sum_{I \in \mathcal{I}} B^I x^I$ and $Z(b) = 2^N B^0$. Since $\sum_{0 \le i \le M/2} \binom{M}{2i} = ((1+1)^M + (1-1)^M)/2 = 2^{M-1}$, the number of all elements in $\mathcal{I}$ is 2^{M-1}.

Assume that a true distribution is $p(x|a^*)$ with $a^* = (a^*_{ij})$. Then the Kullback function $K(a)$ is

$$\sum_{x_i = \pm 1} p(x|a^*)(\log p(x|a^*) - \log p(x|a)) = \sum_{x_i = \pm 1} p(x|a^*) \sum_{i=2}^{\infty} \frac{(-1)^i}{i} \left(\frac{p(x|a)}{p(x|a^*)} - 1\right)^i.$$

Since we consider a neighborhood of $\frac{p(x|a)}{p(x|a^*)} = 1$, we only need to obtain the maximum pole of $J(z) = \int \Psi_0^z \mathrm{d}b$, where

$$\Psi_0 = \sum_{x_i = \pm 1} \frac{(p(x|a) - p(x|a^*))^2}{p(x|a^*)} = \sum_{x_i = \pm 1} \frac{\left(\frac{\sum_{I \in \mathcal{I}} B^I x^I}{Z(b)} - \frac{\sum_{I \in \mathcal{I}} B^{*I} x^I}{Z(b^*)}\right)^2}{p(x|a^*)}.$$

By Lemma 5 in [1], we may replace Ψ_0 by

$$\Psi_1 = \sum_{I \in \{0,1\}^M} 2^{2N} \left(\frac{B^I}{Z(b)} - \frac{B^{*I}}{Z(b^*)}\right)^2 = \sum_{I \in \{0,1\}^M} \left(\frac{B^I}{B^0} - \frac{B^{*I}}{B^{*0}}\right)^2.$$

Assume that the true distribution is $p(x|a^*)$ with $a^* = 0$. By using Lemma 1, Ψ_1 can be replaced by

$$\Psi(b) = \sum_{I \neq 0 \in \mathcal{I}} (B^I)^2, \tag{3}$$

and from now on, we consider the zeta function $J(z) = \int_V \Psi^z \mathrm{d}b$, where V is a sufficiently small neighborhood of 0.

Let $I, I', I'' \in \mathcal{I}$. We set $B_N^I = B^I$ and $b_j^I = \prod_{i=1}^M b_{ij}^{I_i}$. Also set

$$B_N = (B_N^I) = (B_N^{(0,\ldots,0)}, B_N^{(1,1,0,\ldots,0)}, B_N^{(1,0,1,0,\ldots,0)}, \ldots).$$

We have $B_N^I = \sum_{I' + I'' = I \mod 2} b_N^{I''} B_{N-1}^{I'}$.

Now consider the eigenvalues of the matrix $C_N = (c_N^{I,I'})$ where $c_N^{I,I'} = b_N^{I''}$ with $I' + I'' = I \mod 2$. Note that $B_N = C_N B_{N-1}$. Let $\ell = (\ell_1, \ldots, \ell_{2^{M-1}}) = (\ell_I) \in \{-1, 1\}^{2^{M-1}}$ with $\ell_{(0,\ldots,0)} = 1$. ℓ is an eigenvector, if and only if $\sum_{I' \in \mathcal{I}} c_N^{I,I'} \ell_{I'} = \ell_I \sum_{I' \in \mathcal{I}} c_N^{(0,\ldots,0),I'} \ell_{I'} = \ell_I \sum_{I' \in \mathcal{I}} b_N^{I'} \ell_{I'}$. That is,

ℓ is an eigenvector $\Longleftrightarrow$ if $I+I'=I''$ mod 2 ($I+I'+I''=0$ mod 2) then $\ell_{I''}=\ell_I\ell_{I'}$ ($\ell_I\ell_{I'}\ell_{I''}=1$).

Denote the number of all elements in a set K by $\#K$.

Theorem 2. *Let $K_1, K_2 \subset \{1,\ldots,M\}$, $1\in K_2$, $K_1\cap K_2=\phi$, and $K_1\cup K_2=\{1,\ldots,M\}$.*

Set $\ell_I=\begin{cases}-1, & \text{if } \#\{i\in K_1 : I_i=1\} \text{ is odd,}\\ 1, & \text{otherwise.}\end{cases}$ *If $K_1=\phi$, set $\ell=(1,\ldots,1)$.*

Then $\ell=(\ell_I)$ is an eigenvector of C_N and its eigenvalue is $\sum_{I\in\mathcal{I}}\ell_I b_N^I$.

Proof. Assume that $I'+I''+I'''=0$ mod 2. If all $\#\{i\in K_1 : I'_i=1\}$, $\#\{i\in K_1 : I''_i=1\}$ and $\#\{i\in K_1 : I'''_i=1\}$ are even, then $\ell_{I'}\ell_{I''}\ell_{I'''}=1$.

If $\#\{i\in K_1 : I'_i=1\}$ and $\#\{i\in K_1 : I''_i=1\}$ are odd, then $\#\{i\in K_1 : I'''_i=1\}$ is even and $\ell_{I'}\ell_{I''}\ell_{I'''}=1$ since $I'+I''+I'''=0$ mod 2.

If $\#\{i\in K_1 : I'_i=1\}$ is odd, then $\#\{i\in K_1 : I''_i=1\}$ or $\#\{i\in K_1 : I'''_i=1\}$ is odd, since $I'+I''+I'''=0$ mod 2. □

Since we have 2^{M-1} pairs of K_1, K_2 with $1\in K_2$, $K_1\cap K_2=\phi$ and $K_1\cup K_2=\{1,\ldots,M\}$, those eigenvectors ℓ's span the whole space $\mathbb{R}^{2^{M-1}}$ and are orthogonal to each other.

Set $\mathbf{1}=(1,\ldots,1)^t\in\mathbb{Z}^{2^{M-1}-1}$ (t denotes the transpose). Let D be the matrix by arranging the eigenvectors ℓ's such that $D=\begin{pmatrix}1 & \mathbf{1}^t\\ \mathbf{1} & D'\end{pmatrix}$ and $DD=2^{M-1}E$, where E is the unit matrix.

Since $DD=\begin{pmatrix}2^{M-1} & \mathbf{1}^tD'\\ \mathbf{1}+D'\mathbf{1} & \mathbf{1}\mathbf{1}^t+D'D'\end{pmatrix}=2^{M-1}E$, we have $D'\mathbf{1}=-\mathbf{1}$.

Theorem 3. *Let* $C'_j=DC_jD/2^{M-1}=DC_jD^{-1}=\begin{pmatrix}s_{0j} & 0 & 0\cdots & 0\\ 0 & s_{1j} & 0\cdots & 0\\ \vdots & \vdots & \vdots\ \vdots & \vdots\\ 0 & 0 & 0\cdots & s_{2^{M-1}-1,j}\end{pmatrix}$ *which is the diagonal matrix. We have the followings.*

(1) *Let* $d_{ij}=\begin{cases}1, & \text{if } i=1 \text{ or } j=1,\\ D^{I,J}, & \text{if } I=(1,0,\ldots,0,\overset{i}{1},0,\ldots,0)\\ & \text{and } J=(1,0,\ldots,0,\overset{j}{1},0,\ldots,0).\end{cases}$

Then $D^{I,J}=\prod_{i\in I,j\in J}d_{ij}$ for all $I,J\in\mathcal{I}$.

(2) $B_N=C_NB_{N-1}=C_N\cdots C_2B_1=DC'_N\cdots C'_2D^{-1}B_1=\dfrac{DC'_N\cdots C'_1\mathbf{1}}{2^{M-1}}$.

(3) *We have $2^{M-1}D'^{-1}=D'-\mathbf{1}\mathbf{1}^t$.*

(4) *Let $\tilde{B}_1=(B_1^I)_{I\neq 0}$, $\tilde{B}_N=(B_N^I)_{I\neq 0}$ and $S=$*

$$(\prod_{j=2}^N s_{0j})\begin{pmatrix}1\cdots 1\\ \vdots\ \vdots\ \vdots\\ 1\cdots 1\end{pmatrix}+\begin{pmatrix}\prod_{j=2}^N s_{1j} & 0 & 0\cdots & 0\\ 0 & \prod_{j=2}^N s_{2j} & 0\cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0\cdots & \prod_{j=2}^N s_{2^{M-1}-1,j}\end{pmatrix}.$$

We have

$$(\det S)D'^{-1}S^{-1}D'^{-1}2^{M-1}\tilde{B}_N = (\det S)\tilde{B}_1 - (\mathbf{1}\ D')\begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 1}\prod_{j=2}^N s_{ij} \\ \vdots \\ \prod_{i\neq 2^{M-1}-1}\prod_{j=2}^N s_{ij} \end{pmatrix}.$$

(5) *The corresponding element to* I *of* $(\mathbf{1}\ D')\begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 1}\prod_{j=2}^N s_{ij} \\ \vdots \\ \prod_{i\neq 2^{M-1}-1}\prod_{j=2}^N s_{ij} \end{pmatrix}$ *consists of monomials* $c_J\prod_{i=1}^M\prod_{j=2}^N b_{ij}^{J_{ij}}$, *where* $c_J\in\mathbb{R}$, $0\le J_{ij}\in\mathbb{Z}$ *and* $\sum_{j=1}^N J_{ij}=I_i \mod 2$.

Proof. (5) is obtained by

$$(C_N\cdots C_2)^{-1} = D\begin{pmatrix} 1/\prod_{j=2}^N s_{0j} & 0 & \cdots & 0 \\ 0 & 1/\prod_{j=2}^N s_{1j} & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & 1/\prod_{j=2}^N s_{2^{M-1}j} \end{pmatrix}D^{-1}.$$

We prove only (4). Let $H=2^{M-1}-1$. We have

$$2^{M-1}\tilde{B}_N = (\mathbf{1}\ D')C'_N\cdots C'_2\begin{pmatrix} 1 & \mathbf{1}^t \\ \mathbf{1} & D' \end{pmatrix}B_1$$

$$= (\mathbf{1}\ D')\begin{pmatrix} \prod_{j=2}^N s_{0j} & 0 & 0\cdots & 0 \\ 0 & \prod_{j=2}^N s_{1j} & 0\cdots & 0 \\ \vdots & \vdots & \vdots\cdots & \vdots \\ 0 & 0 & 0\cdots & \prod_{j=2}^N s_{H,j} \end{pmatrix}\begin{pmatrix} 1 & \mathbf{1}^t \\ \mathbf{1} & D' \end{pmatrix}B_1$$

$$= (\mathbf{1}\ D')\{\begin{pmatrix} \prod_{j=2}^N s_{0j} \\ \prod_{j=2}^N s_{1j} \\ \vdots \\ \prod_{j=2}^N s_{H,j} \end{pmatrix} + \begin{pmatrix} \prod_{j=2}^N s_{0j} & 0 & 0\cdots & 0 \\ 0 & \prod_{j=2}^N s_{1j} & 0\cdots & 0 \\ \vdots & \vdots & \vdots\cdots & \vdots \\ 0 & 0 & 0\cdots & \prod_{j=2}^N s_{H,j} \end{pmatrix}\begin{pmatrix} \mathbf{1}^t \\ D' \end{pmatrix}\tilde{B}_1\}$$

$$= (\mathbf{1}\ D')\begin{pmatrix} \prod_{j=2}^N s_{0j} \\ \prod_{j=2}^N s_{1j} \\ \vdots \\ \prod_{j=2}^N s_{H,j} \end{pmatrix} + D'(-\mathbf{1}\ E)\begin{pmatrix} \prod_{j=2}^N s_{0j} & 0 & 0\cdots & 0 \\ 0 & \prod_{j=2}^N s_{1j} & 0\cdots & 0 \\ \vdots & \vdots & \vdots\cdots & \vdots \\ 0 & 0 & 0\cdots & \prod_{j=2}^N s_{H,j} \end{pmatrix}\begin{pmatrix} -\mathbf{1}^t \\ E \end{pmatrix}D'\tilde{B}_1$$

$$= (\mathbf{1}\ D')\begin{pmatrix} \prod_{j=2}^N s_{0j} \\ \prod_{j=2}^N s_{1j} \\ \vdots \\ \prod_{j=2}^N s_{H,j} \end{pmatrix} + D'SD'\tilde{B}_1.$$

Therefore $D'^{-1}2^{M-1}\tilde{B}_N = \begin{pmatrix}-\mathbf{1} & E\end{pmatrix}\begin{pmatrix}\prod_{j=2}^N s_{0j} \\ \prod_{j=2}^N s_{1j} \\ \vdots \\ \prod_{j=2}^N s_{H,j}\end{pmatrix} + SD'\tilde{B}_1.$

We have

$$S^{-1}_{i_1 j_1} = (\det S)^{-1}\begin{cases}\sum_{i_2=0,i_2\neq i_1}^H \prod_{0\leq i\leq H, i\neq i_1,i_2}\prod_{j=2}^N s_{ij}, & \text{if } i_1 = j_1, \\ -\prod_{0\leq i\leq H, i\neq i_1,j_1}\prod_{j=2}^N s_{ij}, & \text{if } i_1 \neq j_1,\end{cases}$$

and $\det S = \sum_{i_2=0}^H \prod_{i\neq i_2}\prod_{j=2}^N s_{ij}$.

Let $\mathbf{s} = \begin{pmatrix}\prod_{i\neq 0}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 1}\prod_{j=2}^N s_{ij} \\ \vdots \\ \prod_{i\neq H}\prod_{j=2}^N s_{ij}\end{pmatrix}$ and $\tilde{\mathbf{s}} = \begin{pmatrix}\prod_{i\neq 1}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 2}\prod_{j=2}^N s_{ij} \\ \vdots \\ \prod_{i\neq H}\prod_{j=2}^N s_{ij}\end{pmatrix}$.

Since $(\det S)S^{-1}\begin{pmatrix}\prod_{j=2}^N s_{1j} - \prod_{j=2}^N s_{0j} \\ \vdots \\ \prod_{j=2}^N s_{H,j} - \prod_{j=2}^N s_{0j}\end{pmatrix} = \sum_{i_2=0}^H \prod_{i\neq i_2}\prod_{j=2}^N s_{ij}\mathbf{1} - 2^{M-1}\tilde{\mathbf{s}}$, we have

$$(\det S)D'^{-1}S^{-1}D'^{-1}2^{M-1}\tilde{B}_N = (\det S)\tilde{B}_1 - \sum_{i_2=0}^{H}\prod_{i\neq i_2}\prod_{j=2}^{N} s_{ij}\mathbf{1} - 2^{M-1}D'^{-1}\tilde{\mathbf{s}}$$

$$= (\det S)\tilde{B}_1 - \sum_{i_2=0}^{H}\prod_{i\neq i_2}\prod_{j=2}^{N} s_{ij}\mathbf{1} - (D' - \mathbf{1}\mathbf{1}^t)\tilde{\mathbf{s}}$$

$$= (\det S)\tilde{B}_1 - \prod_{i\neq 0}\prod_{j=2}^{N} s_{ij}\mathbf{1} - D'\tilde{\mathbf{s}} = (\det S)\tilde{B}_1 - (\mathbf{1}\ D')\mathbf{s},$$

by using (3) $2^{M-1}D'^{-1} = D' - \mathbf{1}\mathbf{1}^t$. □

Theorem 4. *The average stochastic complexity $F(n)$ in (1) and the generalization error $G(n)$ in (2) are given by using the following maximum pole $-\lambda$ of $J(z)$ and its order θ.*

(Case 1)*: If $N = 1$ then $\lambda = M/4$ and* $\theta = \begin{cases}2, & \text{if } M = 2, \\ 1, & \text{if } M \geq 3.\end{cases}$

(Case 2)*: If $M = 2$ then $\lambda = 1/2$ and* $\theta = \begin{cases}2, & \text{if } N = 1, \\ 1, & \text{if } N \geq 2.\end{cases}$

(Case 3)*: If $M = 3$ then* $\lambda = \begin{cases}3/4, & \text{if } N = 1, \\ 3/2, & \text{if } N \geq 2,\end{cases}$ *and* $\theta = \begin{cases}1, & \text{if } N = 1, \\ 3, & \text{if } N = 2, \\ 1, & \text{if } N \geq 3.\end{cases}$

(Case 4)*: If $M = 4$ then* $\lambda = \begin{cases}1, & \text{if } N = 1, \\ 2, & \text{if } N = 2,\end{cases}$ *and $\theta = 1$, if $N = 1, 2$.*

Proof. By Theorem 3 (4) and Lemma 1, we only need to consider the maximum pole of $J(z) = \int ||\Psi'||^{2z} \mathrm{d}b$, where $\Psi' = (\det S)\tilde{B}_1 - (\mathbf{1}\ D') \begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 1}\prod_{j=2}^N s_{ij} \\ \vdots \\ \prod_{i\neq H}\prod_{j=2}^N s_{ij} \end{pmatrix}$.

(Case 1): Since $B_1^I = \prod_{j\in I} b_{1j}$, we have the poles $-\dfrac{M}{4}$ and $-\dfrac{M-1}{2}$.

(Case 2): The fact $B^{11} = \sum_{k=1}^N b_{1k}b_{2k}(1+\cdots)$ yields Case 2.

(Case 3): Assume that $M = 3$.

Let $N \geq 2$. We have $D' = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$, $\begin{cases} s_{0j} = 1 + b_{1j}b_{2j} + b_{1j}b_{3j} + b_{2j}b_{3j}, \\ s_{1j} = 1 + b_{1j}b_{2j} - b_{1j}b_{3j} - b_{2j}b_{3j}, \\ s_{2j} = 1 - b_{1j}b_{2j} + b_{1j}b_{3j} - b_{2j}b_{3j}, \\ s_{3j} = 1 - b_{1j}b_{2j} - b_{1j}b_{3j} + b_{2j}b_{3j}, \end{cases}$

and $\Psi' = (\det S)\begin{pmatrix} b_{11}b_{21} \\ b_{11}b_{31} \\ b_{21}b_{31} \end{pmatrix} - (1, D')\begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 1}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 2}\prod_{j=2}^N s_{ij} \\ \prod_{i\neq 3}\prod_{j=2}^N s_{ij} \end{pmatrix}$.

Construct blowing up of Ψ' along the submanifold $\{b_{ij} = 0, 1 \leq i \leq M, 1 \leq j \leq N\}$.

Let $b_{11} = u$, $b_{ij} = ub'_{ij}$ for $(i,j) \neq (1,1)$.

Remark. By setting the general case as $b_{i_0j_0} = b'_{i_0j_0}$, $b_{ij} = b'_{i_0j_0}b'_{ij}$ for $(i,j) \neq (i_0,j_0)$, we have a manifold $\mathcal{M}$ by gluing MN open sets $U_{i_0j_0}$ with a coordinate system $(b'_{11}, b'_{12}, \cdots, b'_{MN})$ (cf. Section 3). We don't need to consider all cases since we obtain the same poles in $U_{i_0j_0}$ as those in U_{11}.

We have $\Psi'' = u^2(\det S)\begin{pmatrix} b'_{21} \\ b'_{31} \\ b'_{21}b'_{31} \end{pmatrix} + 4u^2\begin{pmatrix} \sum_{k=2}^N b'_{1k}b'_{2k} + u^2 f_1 \\ \sum_{k=2}^N b'_{1k}b'_{3k} + u^2 f_2 \\ \sum_{k=2}^N b'_{2k}b'_{3k} + u^2 f_3 \end{pmatrix}$, where f_1, f_2 and f_3 are polynomials of b'_{ij} with at least two degree.

By putting $\begin{pmatrix} b''_{21} \\ b''_{31} \end{pmatrix} = \begin{pmatrix} b'_{21} \\ b'_{31} \end{pmatrix} + 4\begin{pmatrix} \sum_{k=2}^N b'_{1k}b'_{2k} + u^2 f_1 \\ \sum_{k=2}^N b'_{1k}b'_{3k} + u^2 f_2 \end{pmatrix}/(\det S)$, we have

$$\Psi'' = \frac{u^2}{\det S} \times \begin{pmatrix} (\det S)^2 b''_{21} \\ (\det S)^2 b''_{31} \\ (b''_{21}\det S - 4\sum_{k=2}^N b'_{1k}b'_{2k} - 4u^2 f_1)(b''_{31}\det S - 4\sum_{k=2}^N b'_{1k}b'_{3k} - 4u^2 f_2) \end{pmatrix} + u^2\begin{pmatrix} 0 \\ 0 \\ 4\sum_{k=2}^N b'_{2k}b'_{3k} + 4u^2 f_3 \end{pmatrix}.$$

By using Lemma 1 again, the maximum pole of $\int ||\Psi''||^{2z}u^{3N}db$ is that of $J(z)=\int ||\Psi'''||^{2z}u^{3N}\mathrm{d}b$, where $\Psi'''=u^2\begin{pmatrix} b''_{21}\\ b''_{31}\\ g_1\end{pmatrix}$, and

$g_1=(\sum_{k=2}^N b'_{1k}b'_{2k}+u^2f_1)(\sum_{k=2}^N b'_{1k}b'_{3k}+u^2f_2)+\frac{\det S}{4}(\sum_{k=2}^N b'_{2k}b'_{3k}+u^2f_3)$.

Construct blowing up of Ψ''' along the submanifold $\{b''_{21}=0, b''_{31}=0, b'_{3k}=0, 2\leq k\leq N\}$. Then we have (I), (II) cases.

(I) Let $b'_{32}=v$, $b''_{21}=vb'''_{21}$, $b''_{31}=vb'''_{21}$, $b'_{3k}=vb''_{3k}$, for $3\leq k\leq N$. Then $\Psi'''=u^2v\begin{pmatrix} b'''_{21}\\ b'''_{31}\\ g'_1\end{pmatrix}$, where $g'_1=(\sum_{k=2}^N b'_{1k}b'_{2k}+u^2f_1)(b'_{12}+\sum_{k=3}^N b'_{1k}b''_{3k}+u^2f_2/v)+\frac{\det S}{4}(b'_{22}+\sum_{k=3}^N b'_{2k}b''_{3k}+u^2f_3/v)$.

By Theorem 3 (5), we can set $f_2=vf'_2$ and $f_3=vf'_3$, where f'_2 and f'_3 are polynomials.

We have $(\sum_{k=2}^N b'_{1k}b'_{2k})(b'_{12}+\sum_{k=3}^N b'_{1k}b''_{3k})+\frac{\det S}{4}(b'_{22}+\sum_{k=3}^N b'_{2k}b''_{3k})$

$$=(b'_{2,2},b'_{2,3},\cdots,b'_{2,N})\left(\begin{pmatrix} b'_{1,2}\\ b'_{1,3}\\ \vdots\\ b'_{1,N}\end{pmatrix}(b'_{1,2},b'_{1,3},\cdots,b'_{1,N})+\frac{\det S}{4}E\right)\begin{pmatrix}1\\ b''_{3,3}\\ \vdots\\ b''_{3,N}\end{pmatrix}.$$

Since $\begin{pmatrix} b'_{1,2}\\ b'_{1,3}\\ \vdots\\ b'_{1,N}\end{pmatrix}(b'_{1,2},b'_{1,3},\cdots,b'_{1,N})+\frac{\det S}{4}E$ is regular, we can change variables from $(b'_{2,2},b'_{2,3},\cdots,b'_{2,N})$ to $(b''_{2,2},b''_{2,3},\cdots,b''_{2,N})$ by $(b''_{2,2},b''_{2,3},\cdots,b''_{2,N})=(b'_{2,2},b'_{2,3},\cdots,b'_{2,N})\left(\begin{pmatrix} b'_{1,2}\\ b'_{1,3}\\ \vdots\\ b'_{1,N}\end{pmatrix}(b'_{1,2},b'_{1,3},\cdots,b'_{1,N})+\frac{\det S}{4}E\right)$. Moreover, let $b'''_{22}=b''_{2,2}+b''_{2,3}b''_{3,3}+\cdots+b''_{2,N}b''_{3,N}$.

Then, we have

$$\Psi'''=u^2v\begin{pmatrix} b'''_{21}\\ b'''_{31}\\ b'''_{22}+u^2f_4\end{pmatrix},$$

where f_4 is a polynomial. Therefore, we have the poles $-\frac{3N}{4},-\frac{N+1}{2},-\frac{3}{2}$.

(II) Let $b''_{21}=v$, $b''_{31}=vb'''_{21}$, $b'_{3k}=vb''_{3k}$, for $2\leq k\leq N$. Then we have the poles $-\frac{3N}{4},-\frac{N+1}{2}$.

(Case 4): Let $M = 4$. We have $D' = \begin{pmatrix} -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}$ and

$$\begin{aligned}
s_{0j} &= 1 + b_{1j}b_{2j} + b_{1j}b_{3j} + b_{1j}b_{4j} + b_{2j}b_{3j} + b_{2j}b_{4j} + b_{3j}b_{4j} + b_{1j}b_{2j}b_{3j}b_{4j}, \\
s_{1j} &= 1 + b_{2j}b_{3j} + b_{2j}b_{4j} + b_{3j}b_{4j} - b_{1j}(b_{2j} + b_{3j} + b_{4j} + b_{2j}b_{3j}b_{4j}), \\
s_{2j} &= 1 + b_{1j}b_{3j} + b_{1j}b_{4j} + b_{3j}b_{4j} - b_{2j}(b_{1j} + b_{3j} + b_{4j} + b_{1j}b_{3j}b_{4j}), \\
s_{3j} &= 1 + b_{1j}b_{3j} + b_{2j}b_{4j} + b_{1j}b_{2j}b_{3j}b_{4j} - (b_{1j} + b_{3j})(b_{2j} + b_{4j}), \\
s_{4j} &= 1 + b_{1j}b_{2j} + b_{3j}b_{4j} + b_{1j}b_{2j}b_{3j}b_{4j} - (b_{1j} + b_{2j})(b_{3j} + b_{4j}), \\
s_{5j} &= 1 + b_{1j}b_{2j} + b_{1j}b_{4j} + b_{2j}b_{4j} - b_{3j}(b_{1j} + b_{2j} + b_{4j} + b_{1j}b_{2j}b_{4j}), \\
s_{6j} &= 1 + b_{1j}b_{2j} + b_{1j}b_{3j} + b_{2j}b_{3j} - b_{4j}(b_{1j} + b_{2j} + b_{3j} + b_{1j}b_{2j}b_{3j}), \\
s_{7j} &= 1 + b_{1j}b_{4j} + b_{2j}b_{3j} + b_{1j}b_{2j}b_{3j}b_{4j} - (b_{1j} + b_{4j})(b_{2j} + b_{3j}).
\end{aligned}$$

Let $M = 4$ and $N = 2$. Then we have

$$\Psi' = \det S \begin{pmatrix} b_{11}b_{21} \\ b_{11}b_{31} \\ b_{11}b_{41} \\ b_{21}b_{31} \\ b_{21}b_{41} \\ b_{31}b_{41} \\ b_{11}b_{21}b_{31}b_{41} \end{pmatrix} - \begin{pmatrix} -b_{12}b_{22}(8 + f_1) \\ -b_{12}b_{32}(8 + f_2) \\ -b_{12}b_{42}(8 + f_3) \\ -b_{22}b_{32}(8 + f_4) \\ -b_{22}b_{42}(8 + f_5) \\ -b_{32}b_{42}(8 + f_6) \\ b_{12}b_{22}b_{32}b_{42}(40 + f_7) \end{pmatrix},$$

where f_i's are polynomials of b_{ij} with at least two degree. As space is limited, we will omit the proof in detail, but we have the poles $-\frac{8}{4}, -\frac{6}{2}, -\frac{5}{2}, -\frac{9}{4}$. □

5 Conclusion

In this paper, we obtain the main term of the average stochastic complexity for certain complete bipartite graph-type spin models in Bayesian estimation (Theorem 4). We use a new method of eigenvalue analysis and a recursive blowing up method in algebraic geometry and show that these are effective for solving the problems in the artificial intelligence. Our future purpose is to improve our methods and apply them to more general cases. Since eigenvalue analysis can be applied to general cases, we seem to formulate a new direction for solving the behavior of the spin model's stochastic complexity.

The applications of our results are as follows. The explicit values of generalization errors have been used to construct mathematical foundation for analyzing and developing the precision of the MCMC method [16]. Moreover, these values have been compared to such as the generalization error of localized Bayes estimation [17].

Acknowledgments. This research was supported by the Ministry of Education, Science, Sports and Culture in Japan, Grant-in-Aid for Scientific Research 18079007 and 16700218.

References

1. Watanabe, S.: Algebraic analysis for nonidentifiable learning machines. Neural Computation 13(4), 899–933 (2001)
2. Watanabe, S.: Algebraic geometrical methods for hierarchical learning machines. Neural Networks 14(8), 1049–1060 (2001)
3. Akaike, H.: A new look at the statistical model identification. IEEE Trans. on Automatic Control 19, 716–723 (1974)
4. Takeuchi, K.: Distribution of an information statistic and the criterion for the optimal model. Mathematical Science 153, 12–18 (1976)
5. Hannan, E.J., Quinn, B.G.: The determination of the order of an autoregression. Journal of Royal Statistical Society Series B 41, 190–195 (1979)
6. Murata, N.J., Yoshizawa, S.G., Amari, S.: Network information criterion - determining the number of hidden units for an artificial neural network model. IEEE Trans. on Neural Networks 5(6), 865–872 (1994)
7. Schwarz, G.: Estimating the dimension of a model. Annals of Statistics 6(2), 461–464 (1978)
8. Rissanen, J.: Universal coding, information, prediction, and estimation. IEEE Trans. on Information Theory 30(4), 629–636 (1984)
9. Hironaka, H.: Resolution of Singularities of an algebraic variety over a field of characteristic zero. Annals of Math. 79, 109–326 (1964)
10. Yamazaki, K., Watanabe, S.: Singularities in Complete Bipartite Graph-Type Boltzmann Machines and Upper Bounds of Stochastic Complexities. IEEE Trans. on Neural Networks 16(2), 312–324 (2005)
11. Nishiyama, Y., Watanabe, S.: Asymptotic Behavior of Free Energy of General Boltzmann Machines in Mean Field Approximation. Technical report of IEICE NC2006 38, 1–6 (2006)
12. Aoyagi, M., Watanabe, S.: Resolution of Singularities and the Generalization Error with Bayesian Estimation for Layered Neural Network. IEICE Trans. J88-D-II,10, 2112–2124 (2005) (English version: Systems and Computers in Japan John Wiley & Sons Inc. 2005) (in press)
13. Aoyagi, M.: The zeta function of learning theory and generalization error of three layered neural perceptron. RIMS Kokyuroku, Recent Topics on Real and Complex Singularities 2006 (in press)
14. Aoyagi, M., Watanabe, S.: Stochastic Complexities of Reduced Rank Regression in Bayesian Estimation. Neural Networks 18, 924–933 (2005)
15. Watanabe, S., Hagiwara, K., Akaho, S., Motomura, Y., Fukumizu, K., Okada, M., Aoyagi, M.: Theory and Application of Learning System. Morikita, p. 195, 2005 (Japanese)
16. Nagata, K., Watanabe, S.: A proposal and effectiveness of the optimal approximation for Bayesian posterior distribution. In: Workshop on Information-Based Induction Sciences, pp. 99–104 (2005)
17. Takamatsu, S., Nakajima, S., Watanabe, S.: Generalization Error of Localized Bayes Estimation in Reduced Rank Regression. In: Workshop on Information-Based Induction Sciences, pp. 81–86 (2005)