Journal of Machine Learning Research 11 (2010) 1201-1230 Submitted 5/08; Revised 12/09; Published 4/10

Stochastic Complexity and Generalization Error of a Restricted Boltzmann Machine in Bayesian Estimation

Miki Aoyagi AOYAGI.MIKI@NIHON-U.AC.JP
Department of Mathematics, College of Science & Technology
Nihon University
1-8-14, Surugadai, Kanda, Chiyoda-ku
Tokyo 101-8308, Japan

Editor: Tommi Jaakola

Abstract

In this paper, we consider the asymptotic form of the generalization error for the restricted Boltzmann machine in Bayesian estimation. It has been shown that obtaining the maximum pole of zeta functions is related to the asymptotic form of the generalization error for hierarchical learning models (Watanabe, 2001a,b). The zeta function is defined by using a Kullback function. We use two methods to obtain the maximum pole: a new eigenvalue analysis method and a recursive blowing up process. We show that these methods are effective for obtaining the asymptotic form of the generalization error of hierarchical learning models.

Keywords: Boltzmann machine, non-regular learning machine, resolution of singularities, zeta function

1. Introduction

A learning system consists of data, a learning model and a learning algorithm. The purpose of such a system is to estimate an unknown true density function from data distributed by the true density function. The data associated with image or speech recognition, artificial intelligence, the control of a robot, genetic analysis, data mining, time series prediction, and so on, are very complicated and usually not generated by a simple normal distribution, as they are influenced by many factors. Learning models for analyzing such data should likewise have complicated structures. Hierarchical learning models such as the Boltzmann machine, layered neural network, reduced rank regression and the normal mixture model are known to be effective learning models. They are, however, non-regular statistical models, which cannot be analyzed using the classic theories of regular statistical models (Hartigan, 1985; Sussmann, 1992; Hagiwara, Toda, and Usui, 1993; Fukumizu, 1996).

For example, consider a simple restricted Boltzmann machine that has two observable units and one hidden unit with binary variables (Fig. 1). The model is expressed by the probability form of two observable units $x = (x_1, x_2) \in \{1, -1\}^2$ with a parameter $a = (a_1, a_2) \in \mathbb{R}^2$:

$$p(x|a) = \sum_{y=\pm 1} p(x, y|a) = \frac{\exp(a_1x_1 + a_2x_2) + \exp(-a_1x_1 - a_2x_2)}{Z(a)},$$

where $y \in \{1, -1\}$ is the hidden variable,

$$p(x, y|a) = \frac{\exp(a_1x_1y + a_2x_2y)}{Z(a)}, \text{ and } Z(a) = \sum_{x_i=\pm 1, y=\pm 1,} \exp(a_1x_1y + a_2x_2y).$$

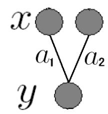

Figure 1: Simple restricted Boltzmann machine model: Two observable units and one hidden unit. The learning model is $p(x|a) \approx \exp(a_1x_1 + a_2x_2) + \exp(-a_1x_1 - a_2x_2)$.

We have

$$p(x|a) = \{(\prod_{i=1}^{2}(1 + x_i \tanh(a_i)) + \prod_{i=1}^{2}(1 - x_i \tanh(a_i)))\}\frac{\prod_{i=1}^{2}\cosh(a_i)}{Z(a)}$$
$$= \frac{\prod_{i=1}^{2}\cosh(a_i)}{Z(a)}(2 + 2x_1x_2 \tanh(a_1)\tanh(a_2)) = \frac{1 + x_1x_2 \tanh(a_1)\tanh(a_2)}{4}.$$

Assume that the true density function is $p(x|a^*)$ with $a^* = 0$. Then the true parameter set is $\{a = (a_1, a_2) \in \mathbb{R}^2 | p(x|a^*) = p(x|a)\} = \{a_1 = 0\} \cup \{a_2 = 0\}$. This set does not consist of only one point, resulting in a non-positive definite Fisher matrix function. On the other hand, the true parameter set of regular models should be one point and its Fisher matrix function is positive definite. Usually, the true parameter set of non-regular models is an analytic set with complicated singularities. Consequently, the many theoretical problems, such as clarifying generalization errors in learning theory, have remained unsolved.

The generalization error measures the difference between the true density function $q(x)$ and the predictive density function $p(x|x^n)$ obtained using n distributed training samples $x^n = (x_1, \ldots, x_n)$ of x from the true density function $q(x)$. We define it as the average Kullback distance between $q(x)$ and $p(x|x^n)$:

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

where E_n is the expectation value over n training samples. This function clarifies precisely how $p(x|x^n)$ can approximate $q(x)$. Thus, $G(n)$ is also called a learning curve or a learning efficiency. For an arbitrary fixed parameter w^* in a parameter space W, we have

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

The first and second terms are called the function approximation error and the statistical estimation error, respectively. The asymptotic form of the generalization error is important for model selection methods. The optimal model balances the function approximation error with the statistical estimation error. Since the Fisher matrix function is singular, non-regular models cannot be analyzed using the classic model selection methods of regular statistical models such as AIC (Akaike, 1974), TIC (Takeuchi, 1976), HQ (Hannan and Quinn, 1979), NIC (Murata, Yoshizawa, and Amari, 1994), BIC (Schwarz, 1978), and MDL (Rissanen, 1984). Therefore, it is important to construct a mathematical foundation for clarifying the generalization error of non-regular models.

In this paper, we clarify the generalization error of certain restricted Boltzmann machines, explicitly (Theorem 2 and Theorem 3), and give new bounds for the generalization error of the other types (Theorem 4), using both a new method of eigenvalue analysis and a recursive blowing up process. The restricted Boltzmann machine is one of the non-regular models and a complete bipartite graph type model that does not allow connections between hidden units (Hinton, 2004; Salakhutdinov, Mnih, and Hinton, 2007). It has been applied efficiently in recognizing hand-written digits and faces.

Several papers (Yamazaki and Watanabe, 2005; Nishiyama and Watanabe, 2006) have reported upper bounds for the asymptotic form of the generalization error for the Boltzmann machine model, but not the exact main terms.

We usually consider the generalization error in terms of a direct and an inverse problem. The direct problem involves solving the generalization error with a known true density function. The inverse problem is finding proper learning models and learning algorithms to minimize the generalization error under the condition of an unknown true density function. The inverse problem is important for practical usage, but in order to solve it, we first need to solve the direct problem. In this paper, we consider the direct problem of the restricted Boltzmann machine model.

We have already obtained the exact asymptotic forms of the generalization errors for the three layered neural network (Aoyagi and Watanabe, 2005a; Aoyagi, 2006), and for the reduced rank regression (Aoyagi and Watanabe, 2005b). In addition, Rusakov and Geiger (2005) obtained the same for Naive Bayesian networks (cf. Remark 1).

This paper consists of four sections. In Section 2, we summarize the framework of Bayesian learning models. In Section 3, we explain the restricted Boltzmann machine and show our main results, and we give our conclusions in Section 4.

2. Stochastic Complexity and Generalization Error in Bayesian Estimation

It is well known that Bayesian estimation is more appropriate than the maximum likelihood method when a learning machine is non-regular (Akaike, 1980; Mackay, 1992). In this paper, we consider the stochastic complexity and the generalization error in Bayesian estimation.

Let $q(x)$ be a true probability density function and $x^n := \{x_i\}_{i=1}^n$ be n training samples randomly selected from $q(x)$. Consider a learning model which is written by a probability form $p(x|w)$, where w is a parameter. The purpose of the learning system is to estimate $q(x)$ from x^n by using $p(x|w)$.

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

$$p(x|x^n) = \int p(x|w)p(w|x^n)\mathrm{d}w,$$

which is the predictive density function.

Set

$$K(q||p) = \sum_x q(x) \log \frac{q(x)}{p(x|x^n)}.$$

This function always has 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 E_n over n training samples:

$$G(n) = E_n\{\sum_x q(x) \log \frac{q(x)}{p(x|x^n)}\}.$$

Let

$$K_n(w) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{q(x)}{p(x^n|w)}.$$

The average stochastic complexity or the free energy is defined by

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

Then we have $G(n) = F(n+1) - F(n)$ for an arbitrary natural number n (Levin, Tishby, and Solla, 1990; Amari, Fujita, and Shinomoto, 1992; Amari and Murata, 1993). $F(n)$ is known as the Bayesian criterion in Bayesian model selection (Schwarz, 1978), stochastic complexity in universal coding (Rissanen, 1986; Yamanishi, 1998), Akaike's Bayesian criterion in optimization of hyperparameters (Akaike, 1980) and evidence in neural network learning (Mackay, 1992). In addition, $F(n)$ is an important function for analyzing the generalization error.

It has recently been proved that the maximum pole of a zeta function gives the generalization error of hierarchical learning models asymptotically, assuming that the function approximation error is negligible compared to the statistical estimation error (Watanabe, 2001a,b). This assumption is natural for the model selection problem. To compare various models of different parameter's dimension, we assume that the true distribution is a certain dimensional model. If the parameter's dimension of the true distribution is larger than that of the learning model, clarifying the behavior of the generalization error is rather easy. We assume, therefore, that the true density distribution $q(x)$ is included in the learning model, that is, $q(x) = p(x|w^*)$ for $w^* \in W$, where W is the parameter space.

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 if $G(n)$ has an asymptotic expansion,

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

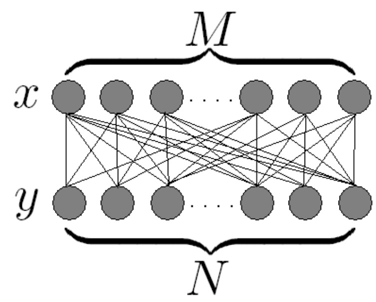

Figure 2: A restricted Boltzmann machine: M is the number of binary observable units x and N is the number of binary hidden units y. The learning model is $p(x,y|a) \propto \exp(\sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}x_iy_j)$, where a_{ij} is a parameter between x_i and y_j.

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

To assist in achieving this aim, we use the desingularization in algebraic geometry (Watanabe, 2009). It is, however, a new problem, even in mathematics, to obtain the desingularization of Kullback functions, since the singularities of these functions are very complicated and as such most of them have not yet been investigated (Appendix A). We, therefore, need a new method of eigenvalue analysis and a recursive blowing up process.

3. Restricted Boltzmann Machine

From now on, for simplicity, we denote

$$\{\{n\}\} = \begin{cases} 0, & \text{if } n = 0 \mod 2, \\ 1, & \text{if } n = 1 \mod 2, \end{cases} \quad \{\{(n_1, \cdots, n_m)\}\} = (\{\{n_1\}\}, \cdots, \{\{n_m\}\}),$$

and 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 $N \in \mathbb{N}$. Set

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

where

$$Z(a) = \sum_{x_i=\pm 1, y_j=\pm 1,} \exp(\sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}x_iy_j),$$

$x = (x_i) \in \{1,-1\}^M$ and $y = (y_j) \in \{1,-1\}^N$ (Fig. 2).

Consider a restricted Boltzmann machine

$$p(x|a) = \sum_{y_j=\pm 1} p(x,y|a) = \frac{\prod_{j=1}^N (\prod_{i=1}^M \exp(a_{ij}x_i) + \prod_{i=1}^M \exp(-a_{ij}x_i))}{Z(a)}$$
$$= \{\prod_{j=1}^N (\prod_{i=1}^M (1+x_i \tanh(a_{ij})) + \prod_{i=1}^M (1-x_i\tanh(a_{ij})))\} \frac{\prod_{j=1}^N \prod_{i=1}^M \cosh(a_{ij})}{Z(a)}$$
$$= \frac{\prod_{j=1}^N \prod_{i=1}^M \cosh(a_{ij})}{Z(a)}$$
$$\times \prod_{j=1}^N (2 \sum_{0\le p\le M/2} \sum_{i_1<\cdots<i_{2p}} x_{i_1}x_{i_2}\cdots x_{i_{2p}} \tanh(a_{i_1 j})\tanh(a_{i_2 j})\cdots\tanh(a_{i_{2p} j})).$$

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

$$p(x|a) = \frac{2^N \prod_{j=1}^N \prod_{i=1}^M \cosh(a_{ij})}{Z(a)} \sum_{J:\{\{\sum_{i=1}^M J_{ij}\}\}=0 \text{ for all } j} B^J x^J.$$

Let

$$Z(b) = \frac{Z(a)}{2^N \prod_{j=1}^N \prod_{i=1}^M \cosh(a_{ij})}.$$

Set $\mathcal{I} = \{I=(I_i)\in\{0,1\}^M | \{\{\sum_{i=1}^M I_i\}\}=0\}$, and $B^I = \sum_{\substack{J:\{\{\sum_{i=1}^M J_{ij}\}\}=0 \\ \{\{\sum_{j=1}^N J_{ij}\}\}=I_i}} B^J$ for $I\in\mathcal{I}$. Then we have

$$p(x|a) = \frac{1}{Z(b)} \sum_{I\in\mathcal{I}} B^I x^I$$

and $Z(b)=2^N B^0$. Since $\sum_{0\le i\le M/2} \begin{pmatrix} M \\ 2i \end{pmatrix} = ((1+1)^M + (1-1)^M)/2 = 2^{M-1}$, the number of elements in $\mathcal{I}$ is 2^{M-1}.

Remark 1 Rusakov and Geiger (2005) obtained λ and θ for the following class of Naive Bayesian networks with two hidden states and binary features:

$$p(x|c,d,t) = t\prod_{i=1}^M c_i^{(1+x_i)/2}(1-c_i)^{(1-x_i)/2} + (1-t)\prod_{i=1}^N d_i^{(1+x_i)/2}(1-d_i)^{(1-x_i)/2}.$$

where $x\in\{1,-1\}^M$, $c=\{c_i\}_{i=1}^M \in \mathbb{R}^M$, $d=\{d_i\}_{i=1}^M\in\mathbb{R}^M$ and $0\le t\le 1$. Our models with one hidden unit ($N=1$) are obtained by setting $t=1/2$, $\tanh(a_i)=2c_i-1$ and $d_i=-c_i$. The relation $d_i=-c_i$ creates a parameter space different from that of our models.

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

$$\sum_{x_i=\pm1} p(x|a^*)(\log p(x|a^*) - \log p(x|a)) = \sum_{x_i=\pm1} p(x|a^*)\sum_{k=2}^\infty \frac{(-1)^k}{k}(\frac{p(x|a)}{p(x|a^*)}-1)^k$$
$$= \sum_{x_i=\pm1} \frac{(p(x|a)-p(x|a^*))^2}{p(x|a^*)}(1+\sum_{k=1}^\infty \frac{(-1)^k}{k+2}(\frac{p(x|a)}{p(x|a^*)}-1)^k).$$

Lemma 1 **Watanabe, 2001c** *If analytic functions K_1, K_2 satisfy $\gamma_1|K_2| \leq |K_1| \leq \gamma_2|K_2|$ for some positive constants γ_1 and γ_2, then the maximum pole and its order of $\int |K_1|^z dw$ are those of $\int |K_2|^z dw$.*

By Lemma 1, 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

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

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

By Lemma 1 again, we can replace Ψ_0 by

$$\Psi = \sum_{I\in\{0,1\}^M} 2^{2N}\left(\frac{B^I}{Z(b)} - \frac{B^{*I}}{Z(b^*)}\right)^2 = \sum_{I\in\{0,1\}^M}\left(\frac{B^I}{B^0} - \frac{B^{*I}}{B^{*0}}\right)^2. \tag{3}$$

4. Main Results

Consider the zeta function $J(z) = \int_V \Psi^z \mathrm{d}b$, where V is a sufficiently small neighborhood of a^*.

From the eigenvalue analysis method, we obtain the following theorem.

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

(Case 1)*: If $M = 2$ then $\lambda = 1/2$ and* $\theta = \begin{cases} 2, & \textit{if } N = 1, b^* = 0 \\ 1, & \textit{otherwise.} \end{cases}$

(Case 2)*: If $M = 3$ then* $\lambda = \begin{cases} 3/4, & \textit{if } N = 1, b^* = 0 \\ 1/2, & \textit{if } N = 1, b^* \neq 0, \prod_{i=1}^3 b_{i1}^* = 0 \\ 3/2, & \textit{if } N = 1, \prod_{i=1}^3 b_{i1}^* \neq 0 \\ 3/2, & \textit{if } N \geq 2, \end{cases}$

and $\theta = \begin{cases} 3, & \textit{if } N = 2, b^* = 0, \\ 2, & \textit{if } N = 2, b^* \neq 0, b_{i_0 j}^* = b_{i_1 j}^* = 0 \textit{ for } 1 \leq j \leq N, \\ 2, & \textit{if } N = 2, b_{i_0 j_0}^* b_{i_1 j_0}^* \neq 0, b_{i_2 j_0} = b_{ij}^* = 0 \textit{ for } 1 \leq i \leq 3, 1 \leq j \leq N, j \neq j_0, \\ 1, & \textit{otherwise}, \end{cases}$ *where* $i_0, i_1, i_2 \in \{1,2,3\}$ *are different from each other and* $1 \leq j_0 \leq N$.

For its proof, we use the eigenvalues and the eigenvectors of the matrix $C_j = (c_j^{I,I'})$ where $b_j^I = \prod_{i=1}^M b_{ij}^{I_i}$, and $c_j^{I,I'} = b_j^{I''}$ with $\{\{I' + I''\}\} = I$, for $I, I', I'' \in I$. Its proof appears in Appendix B.

We obtain λ and θ in Eqs. (1) and (2) for $M > N$ using a recursive blowing up.

Theorem 3 *Assume that $M > N$ and $a^* = 0$. The average stochastic complexity $F(n)$ in Eq. (1) and the generalization error $G(n)$ in Eq. (2) are given by using the maximum pole $-\lambda = -\frac{MN}{4}$ of $J(z)$ and its order* $\theta = \begin{cases} 1, & \textit{if } M > N+1, \\ M, & \textit{if } M = N+1. \end{cases}$

We also bound values of λ for other cases.

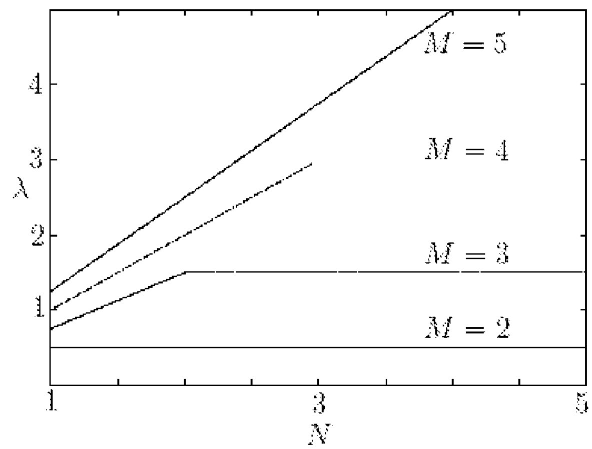

Figure 3: The curve of λ along the y-axis and N along the x-axis, when $M = 2,3,4,5$ and $a^* = 0$.

Theorem 4 *Let* $(a_{1j}, a_{2j}, \cdots, a_{Mj}) \neq 0$ *for* $j = 1,\ldots,N_0$ *and* $(a_{1j}, a_{2j}, \cdots, a_{Mj}) = 0$ *for* $j = N_0 + 1,\ldots,N$ *in* V*, where* V *is a sufficiently small neighborhood of* a^**.*

Then we have

$$\frac{M(N-N_0)}{4} \leq \lambda \leq \frac{M(N-N_0)}{4} + \frac{MN_0}{2}, \qquad \text{if } M > N - N_0$$
$$\frac{M(M-1)}{4} + \frac{MN_0}{2} \leq \lambda \leq \frac{2N_0+(M-1)(M-2)}{4} + \frac{MN_0}{2} \left(< \frac{MN_0}{2} + \frac{M(N-N_0)}{4}\right), \qquad \text{if } M \leq N - N_0.$$

The proofs for these two theorems appear in Appendix C.

5. Conclusion

In this paper, we obtain the generalization error of restricted Boltzmann machines asymptotically (Fig. 3).

We use a new method of eigenvalue analysis and a recursive blowing up in algebraic geometry and show that these are effective for solving the problem in learning theory.

We have not used the eigenvalue analysis method where $M > N$, which is usually the case in applications. Eigenvalue analysis seems to be necessary for solving the behavior of the restricted Boltzmann machine model's generalization error for $M \leq N$.

In this paper, we clarify the generalization error for (i) $M = 3$ (Theorem 2) and (ii) $M > N$, $a^* = 0$ (Theorem 3) explicitly and give new bounds for the generalization error of the other types (Theorem 4). The case (i) shows that λ is independent of a^* for $M - 1 = 2 \leq N$, and so implies that we need more careful consideration for obtaining the exact values λ for the case of Theorem 4.

Our future research aims to improve our methods, and to apply them to the case of Theorem 4 and to obtain the generalization error of the general Boltzmann machine, which is also known as the Bayesian network, the graphical model and the spin model, as such models are widely used in many fields. We believe that extending our results would provide a mathematical foundation for the analysis of various graphical models.

This study involves applying techniques of algebraic geometry to learning theory and it seems that we can contribute to the development of both these fields in the future.

The application of our results is as follows. The results of this paper introduce a mathematical measure of preciseness for numerical calculations such as the Markov Chain Monte Carlo. Using the Markov Chain Monte Carlo (MCMC) method, estimated values for marginal likelihoods had previously been calculated for hyper-parameter estimations and model selection methods of complex learning models, but the theoretical values were not known. The theoretical values of marginal likelihoods have been given in this paper. This enables us to construct a mathematical foundation for analyzing and developing the precision of the MCMC method (Nagata and Watanabe, 2005). Moreover, Nagata and Watanabe (2007) studied the setting of temperatures for the exchange MCMC method and proved the mathematical relation between the symmetrized Kullback function and the exchange ratio, from which an optimal setting of temperatures could be devised. Our theoretical results will be helpful in these numerical experiments. Furthermore, these values have been compared with those of the generalization error of a localized Bayes estimation (Takamatsu, Nakajima, and Watanabe, 2005).

Acknowledgments

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

Appendix A. Hironaka's Theorem

We introduce Hironaka's Theorem about the desingularization.

Theorem 5 [Desingularization (Fig. 4)] (Hironaka, 1964)
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 exist 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.

Applying Hironaka's theorem to the Kullback function $K(w)$, for each $w \in K^{-1}(0) \cap W$, we have a proper analytic map μ_w from an analytic manifold U_w to a neighborhood V_w of w satisfying Hironaka's Theorem (1) and (2). Then the local integration on V_w of the zeta function $J(z)$ of the learning model is

$$\begin{aligned} J_w(z) &= \int_{V_w} K(w)^z \psi(w) dw \\ &= \int_{U_w} \sum_u (u_1^{2s_1} u_2^{2s_2} \cdots u_d^{2s_d})^z \psi(\mu_w(u)) |\mu_w'(u)| du. \end{aligned} \tag{4}$$

Therefore, the poles of $J_w(z)$ can be obtained. For example, the function

$$\int_{U_0} (u_1^{2s_1} u_2^{2s_2} \cdots u_d^{2s_d})^z u_1^{t_1} u_1^{t_2} \cdots u_1^{t_d} du$$

has the poles $-(t_1+1)/(2s_1), \cdots, -(t_d+1)/(2s_d)$, where U_0 is a small neighborhood of 0. For each $w \in W \setminus K^{-1}(0)$, there exists a neighborhood V_w such that $K(w') \neq 0$, for all $w' \in V_w$. So

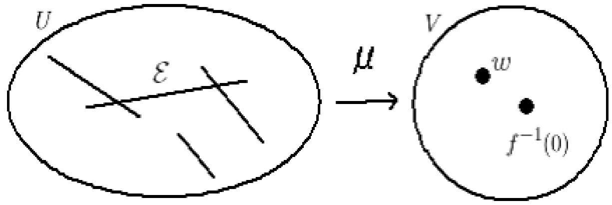

Figure 4: Hironaka's Theorem: This is the picture of a desingularization μ of f : $\mathcal{E}$ maps to $f^{-1}(0)$. $U-\mathcal{E}$ is isomorphic to $V-f^{-1}(0)$ by μ, where V is a small neighborhood of w with $f(w)=0$.

$J_w(z)=\int_{V_w}K(w)^z\psi(w)dw$ has no poles. It is known that μ of an arbitrary polynomial in Hironaka's Theorem can be obtained by using a blowing up process. Note that the exponents in the integral are $2s_i$ instead of s_i as shown in Eq. (4), since the Kullback function is positive.

In spite of such results, it is still difficult to obtain the generalization error mainly for the following two reasons. (a) The desingularization of any polynomial is in general very difficult, although it is known to be a finite process. Furthermore, most of the Kullback functions of non-regular statistical models are degenerate (over $\mathbb{R}$) with respect to their Newton polyhedrons, which is the condition for using a toric resolution (Fulton, 1993; Watanabe, Hagiwara, Akaho, Motomura, Fukumizu, Okada, and Aoyagi, 2005). Also, points in the singularity set $\{K=\partial K/\partial w=0\}$ of Kullback functions $K(w)$ are not isolated, and Kullback functions are not simple polynomials, as their number of variables and number of terms grow with parameters, for example, M and N in Eq. (3). It is therefore, a new problem, even in mathematics, to obtain desingularizations of such Kullback functions, since their singularities are very complicated and as such most of them have not yet been investigated. (b) Since our main purpose is to obtain the maximum pole, obtaining a desingularization is not enough. We need techniques for choosing the maximum one from all poles. However, to the best of our knowledge, no theorems for such a purpose have been developed.

We give below Lemmas 2 and 3 in (Aoyagi and Watanabe, 2005b), as they 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 6 (Aoyagi and Watanabe, 2005b) *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 and Q be any regular $H\times H$ and $H'\times H'$ matrices, respectively. Then the maximum pole of $\int_U\|C(w)\|^{2z}\psi(w)dw$ and its order are those of $\int_U\|PC(w)Q\|^{2z}\psi(w)dw$.*

Lemma 7 *Assume that $p(x|a)=\frac{\prod_{j=1}^N W_j(x,a)}{\sum_x\prod_{j=1}^N W_j(x,a)}$ for $x\in X$. Then the maximum pole of $\int_U\{\sum_{x\in X}(p(x|a)-p(x|a^*))^2\}^z\psi(w)da$ and its order are those of*

$$\int_U\{\sum_{x,x'\in X}(\sum_j^N(\log W_j(x,a)-\log W_j(x,a^*)-\log W_j(x',a)+\log W_j(x',a^*))\}^2\psi(w)dw.$$

(Proof)

Consider the ideal I generated by $p(x|a) - p(x|a^*)$ for $x \in X$.

Then I is generated by $\frac{\prod_{j=1}^N W_j(x,a)}{\prod_{j=1}^N W_j(x,a^*)} - \frac{\sum_x \prod_{j=1}^N W_j(x,a)}{\sum_x \prod_{j=1}^N W_j(x,a^*)}$, and so by $\frac{\prod_{j=1}^N W_j(x,a)}{\prod_{j=1}^N W_j(x,a^*)} - \frac{\prod_{j=1}^N W_j(x',a)}{\prod_{j=1}^N W_j(x',a^*)}$ for $x, x' \in X$.

Since $|x-1|/2 \leq |\log x| \leq 2|x-1|$ for $|x-1| < 1/2$, we have

$$\sum_{x,x' \in X} \left(\frac{\prod_{j=1}^N W_j(x,a)}{\prod_{j=1}^N W_j(x,a^*)} \frac{\prod_{j=1}^N W_j(x',a^*)}{\prod_{j=1}^N W_j(x',a)} - 1\right)^2/4$$

$$\leq \sum_{x,x' \in X} \left(\sum_{j}^{N} (\log(W_j(x,a)) - \log(W_j(x,a^*)) + \log(W_j(x',a^*)) - \log(W_j(x',a)))\right)^2$$

$$\leq \sum_{x,x' \in X} \left(\frac{\prod_{j=1}^N W_i(x,a)}{\prod_{j=1}^N W_j(x,a^*)} \frac{\prod_{j=1}^N W_j(x',a^*)}{\prod_{j=1}^N W_j(x',a)} - 1\right)^2 4$$

Q.E.D.

Appendix B. Eigenvalue Analysis

The purpose of eigenvalue analysis is to simplify the blowing up process.

Hierarchical learning machines often have Kullback functions involving a matrix product such as $K(w) = \|D_1 D_2 \cdots D_N\|^2$, where D_i is a parameter matrix. Therefore, analyzing the eigenvalues of these matrices and applying Lemma 6 sometimes results in an easier function to handle. For example, the restricted Boltzmann machine has a Kullback function $\|\tilde{B}_N\|^2 = \|\begin{pmatrix} \mathbf{0} & E \end{pmatrix} C_N \cdots C_2 C_1 (1,0,\ldots,0)^t\|^2$, where E is the identity matrix (t denotes the transpose). Theorem 9 (4) below shows that analyzing the eigenvalues of C_N makes an easier function $\|R\tilde{B}_N\|^2$ to blow up, where R is a certain regular matrix. This is the main point of this method.

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

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

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

For convenience, we denote the "(I,I')th" element of a $2^{M-1} \times 2^{M-1}$ matrix C by $c^{I,I'}$.

Now consider the eigenvalues of the matrix $C_N = (c_N^{I,I'})$ where $c_N^{I,I'} = b_N^{I''}$ with $\{\{I'+I''\}\} = I$.

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

$$\sum_{I' \in \mathcal{I}} c_N^{I,I'} \ell_{I'} = \ell_I \sum_{I' \in \mathcal{I}} c_N^{(0,\ldots,0),I'} \ell_{I'} = \ell_I \sum_{I' \in \mathcal{I}} b_N^{I'} \ell_{I'}, \text{ for all } I \in \mathcal{I}.$$

That is,

ℓ is an eigenvector $\iff$ if $\{\{I+I'\}\} = I''$ ($\{\{I+I'+I''\}\} = 0$)
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 8 *Let $K_1 \subset \{2,\dots,M\}$. Set $\ell_I = \begin{cases} -1, & \text{if } \#\{i \in K_1 : I_i = 1\} \text{ is odd,} \\ 1, & \text{otherwise.} \end{cases}$ Then $\ell = (\ell_I)$ is an eigenvector of C_N and its eigenvalue is*

$$\sum_{I \in I} \ell_I b_N^I = \frac{\prod_{i=1}^M (1+x_i b_i) + \prod_{i=1}^M (1-x_i b_i)}{2}, \text{ where } x_i = -1 \text{ if } i \in K_1, \text{ and } x_i = 1 \text{ if } i \notin K_1.$$

Note that $\sum_{I \in I} \ell_I b_N^I > 0$ since $b_i = \tanh(a_i)$.

(Proof)

Assume that $\{\{I' + I'' + I'''\}\} = 0$. 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 $\#\{i \in K_1 : I_i' = 1\}$ is odd, then $\#\{i \in K_1 : I_i'' = 1\}$ or $\#\{i \in K_1 : I_i''' = 1\}$ is odd, since $\{\{I' + I'' + I'''\}\} = 0$.

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 $\{\{I' + I'' + I'''\}\} = 0$.

Q.E.D.

We have 2^{M-1} eigenvectors ℓ. Moreover, they are orthogonal to each other, since the eigenvectors of a symmetric matrix are orthogonal. These eigenvectors ℓ's, therefore, span the whole space $\mathbb{R}^{2^{M-1}}$.

Set $\mathbf{1} = (1,\dots,1)^t \in \mathbb{Z}^{2^{M-1}-1}$ (t denotes the transpose). Let $D = (D^{I,I'})$ be a symmetric matrix formed 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 identity matrix and $D^{I,I'}$ is "(I,I')th" element of D.

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

Let $C_N' = DC_N D/2^{M-1} = DC_N D^{-1} = \begin{pmatrix} s_N^0 & 0 & 0 & \cdots & 0 \\ 0 & s_N^1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & s_N^{2^{M-1}-1} \end{pmatrix}$.

We use s_N^i or s_N^I ($I \in I$), depending on the situation.

Since $C_N = D^{-1}C_N' D$, we have $b_N^{\{\{I+K\}\}} = \sum_{J \in I} D^{I,J} s_N^J D^{J,K}/2^{M-1}$.

B.1 Example

Let $M = 4$.

We have the matrix by arranging the eigenvectors of C_N,

$$D = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix} \text{ and the eigenvalues}$$

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

Theorem 9 *Let* $H = 2^{M-1}-1$.

(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 $D^{I,J} = \prod_{i,j:I_i=1,J_j=1} d_{ij}$ *for all* $I,J \in I$.

(2) $B_N = C_N B_{N-1} = C_N \cdots C_2 B_1 = DC'_N \cdots C'_2 D^{-1} B_1 = \dfrac{DC'_N \cdots C'_1 \mathbf{1}}{2^{M-1}}$.

(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*

$$S = -\frac{1}{H+1}\begin{pmatrix} \prod_{j=1}^N s^{*1}_j - \prod_{j=1}^N s^{*0}_j \\ \vdots \\ \prod_{j=1}^N s^{*H}_j - \prod_{j=1}^N s^{*0}_j \end{pmatrix} \begin{pmatrix} \prod_{j=2}^N s_j^1 - \prod_{j=2}^N s_j^0 & \cdots & \prod_{j=2}^N s_j^H - \prod_{j=2}^N s_j^0 \end{pmatrix}$$

$$+B^{*0}_N \prod_{j=2}^N s_j^0 \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} + B^{*0}_N \begin{pmatrix} \prod_{j=2}^N s_j^1 & 0 & 0 & \cdots & 0 \\ 0 & \prod_{j=2}^N s_j^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \prod_{j=2}^N s_j^H \end{pmatrix}.$$

We have

$$(\det S)D'^{-1}S^{-1}D'^{-1}2^{M-1}(\tilde{B}_N B^{*0}_N - \tilde{B^*}_N B^0_N)$$

$$= (\det S)\tilde{B}_1 - (B^{*0}_N)^{H-1}(\mathbf{1}\ D')\begin{pmatrix} \prod_{j=1}^N s^{*0}_j \prod_{i\neq 0}\prod_{j=2}^N s^i_j \\ \vdots \\ \prod_{j=1}^N s^{*H}_j \prod_{i\neq H}\prod_{j=2}^N s^i_j \end{pmatrix}.$$

(5) *The corresponding element to* I *of* $(\mathbf{1}\ \ D')\begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s^i_j \\ \vdots \\ \prod_{i\neq H}\prod_{j=2}^N s^i_j \end{pmatrix}$ *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* $\{\{\sum_{j=1}^N J_{ij}\}\} = I_i$.

(Proof)

(1) Fix $K_1\subset\{2,\dots,M\}$.

Consider the eigenvector ℓ defined by using K_1.

Set $d'_1 = 1$ and $d'_i = \ell_I$ for $I = \begin{array}{ccccccccc} & & & & i & & & \\ (1, & 0, & \dots, & 0, & 1, & 0, & \dots, & 0) \end{array}$, $i\ge 2$.

Since $\ell_I = \prod_{i\in K_1: I_i=1}(-1) = \prod_{i:I_i=1} d'_i$ and D is symmetric, we have statement (1).

(2) is obvious.

(3) Since $DD = \begin{pmatrix} 2^{M-1} & \mathbf{1}^t D' \\ \mathbf{1}+D'\mathbf{1} & \mathbf{1}\mathbf{1}^t + D'D' \end{pmatrix} = 2^{M-1}E$, we have $D'D' = 2^{M-1}E' - \mathbf{1}\mathbf{1}^t$ and $D'(D' - \mathbf{1}\mathbf{1}^t) = 2^{M-1}E' - \mathbf{1}\mathbf{1}^t - D'\mathbf{1}\mathbf{1}^t = 2^{M-1}E' - \mathbf{1}\mathbf{1}^t + \mathbf{1}\mathbf{1}^t = 2^{M-1}E'$, where E' is the identity matrix.

(4)

$$2^{M-1}(\tilde{B}_N B^{*0}_N - \tilde{B^*}_N B^0_N) = 2^{M-1}\begin{pmatrix} -\tilde{B^*}_N & B^{*0}_N E \end{pmatrix} B_N$$

$$= \begin{pmatrix} -\tilde{B^*}_N & B^{*0}_N E \end{pmatrix} D \begin{pmatrix} \prod_{j=2}^N s^0_j & 0 & 0 & \cdots & 0 \\ 0 & \prod_{j=2}^N s^1_j & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \prod_{j=2}^N s^H_j \end{pmatrix} DB_1$$

$$= (-\tilde{B^*}_N \begin{pmatrix} 1 & \mathbf{1}^t \end{pmatrix} + B^{*0}_N \begin{pmatrix} \mathbf{1} & D' \end{pmatrix})$$
$$\begin{pmatrix} \prod_{j=2}^N s^0_j & 0 & 0 & \cdots & 0 \\ 0 & \prod_{j=2}^N s^1_j & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \prod_{j=2}^N s^H_j \end{pmatrix} DB_1$$

$$= (\frac{-\begin{pmatrix} \mathbf{1} & D' \end{pmatrix}}{H+1} \begin{pmatrix} \prod_{j=1}^N s^{*0}_j \\ \prod_{j=1}^N s^{*1}_j \\ \vdots \\ \prod_{j=1}^N s^{*H}_j \end{pmatrix} \begin{pmatrix} 1 & \mathbf{1}^t \end{pmatrix} + B^{*0}_N \begin{pmatrix} \mathbf{1} & D' \end{pmatrix})$$
$$\begin{pmatrix} \prod_{j=2}^N s^0_j & 0 & 0 & \cdots & 0 \\ 0 & \prod_{j=2}^N s^1_j & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \prod_{j=2}^N s^H_j \end{pmatrix} DB_1$$

$$= \begin{pmatrix} \mathbf{1} & D' \end{pmatrix}(-\frac{1}{H+1}\begin{pmatrix} \prod_{j=1}^N s^{*0}_j \\ \prod_{j=1}^N s^{*1}_j \\ \vdots \\ \prod_{j=1}^N s^{*H}_j \end{pmatrix}\begin{pmatrix} \prod_{j=2}^N s_j^0 & \cdots & \prod_{j=2}^N s_j^H \end{pmatrix}$$

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

$$= D'(-T^0 \begin{pmatrix} \prod_{j=1}^N s^{*1}_j - \prod_{j=1}^N s^{*0}_j \\ \vdots \\ \prod_{j=1}^N s^{*H}_j - \prod_{j=1}^N s^{*0}_j \end{pmatrix} + B^{*0}_N \begin{pmatrix} \prod_{j=2}^N s_j^1 - \prod_{j=2}^N s_j^0 \\ \vdots \\ \prod_{j=2}^N s_j^H - \prod_{j=2}^N s_j^0 \end{pmatrix}) + D'SD'\tilde{B}_1,$$

where $T^0 = \frac{\prod_{j=2}^N s_j^0 + \cdots + \prod_{j=2}^N s_j^H}{H+1}$.

Also we have $S^{-1}_{i_1 j_1} = (\det S)^{-1} \times$

$$\begin{cases} \frac{(B^{*0}_N)^{H-2}}{H+1} \sum_{i_2=0, i_2 \neq i_1}^H (\prod_{j=1}^N s^{*i_1}_j + H \prod_{j=1}^N s^{*i_2}_j) \prod_{0 \leq i \leq H, i \neq i_1, i_2} \prod_{j=2}^N s_j^i, & \text{if } i_1 = j_1, \\ \frac{(B^{*0}_N)^{H-2}}{H+1} \sum_{0 \leq i_2 \leq H, i_2 \neq i_1, j_1} (\prod_{j=1}^N s^{*i_1}_j - \prod_{j=1}^N s^{*i_2}_j) \prod_{0 \leq i \leq H, i \neq i_1, i_2} \prod_{j=2}^N s_j^i & \\ -\frac{(B^{*0}_N)^{H-2}}{H+1} (H \prod_{j=1}^N s^{*i_1}_j + \prod_{j=1}^N s^{*j_1}_j) \prod_{0 \leq i \leq H, i \neq i_1, j_1} \prod_{j=2}^N s_j^i, & \text{if } i_1 \neq j_1 \end{cases}$$

and $\det S = (B^{*0}_N)^{H-1} \sum_{i_2=0}^H \prod_{j=1}^N s^{*i_2}_j \prod_{i \neq i_2} \prod_{j=2}^N s_j^i$.

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

Since $(\det S)S^{-1}(-T^0 \begin{pmatrix} \prod_{j=1}^N s^{*1}_j - \prod_{j=1}^N s^{*0}_j \\ \vdots \\ \prod_{j=1}^N s^{*H}_j - \prod_{j=1}^N s^{*0}_j \end{pmatrix} + B^{*0}_N \begin{pmatrix} \prod_{j=2}^N s_j^1 - \prod_{j=2}^N s_j^0 \\ \vdots \\ \prod_{j=2}^N s_j^H - \prod_{j=2}^N s_j^0 \end{pmatrix})$

$$= (B^{*0}_N)^{H-1} \{\sum_{i_2=0}^H \prod_{j=1}^N s^{*i_2}_j \prod_{i \neq i_2} \prod_{j=2}^N s_j^i \mathbf{1} - (H+1) \begin{pmatrix} \prod_{j=1}^N s^{*1}_j \prod_{i \neq 1} \prod_{j=2}^N s_j^i \\ \vdots \\ \prod_{j=1}^N s^{*H}_j \prod_{i \neq H} \prod_{j=2}^N s_j^i \end{pmatrix}\},$$

we have

$$\begin{aligned} & D'^{-1} S^{-1} D'^{-1} 2^{M-1} (\tilde{B}_N B^{*0}_N - \tilde{B}^*_N B^0_N) \\ &= (\det S)\tilde{B}_1 - (B^{*0}_N)^{H-1} \sum_{i_2=0}^H \prod_{j=1}^N s^{*i_2}_j \prod_{i \neq i_2} \prod_{j=2}^N s_j^i \mathbf{1} - (H+1)(B^{*0}_N)^{H-1} D'^{-1} \tilde{\mathbf{s}} \\ &= (\det S)\tilde{B}_1 - (B^{*0}_N)^{H-1} \sum_{i_2=0}^H \prod_{j=1}^N s^{*i_2}_j \prod_{i \neq i_2} \prod_{j=2}^N s_j^i \mathbf{1} - (B^{*0}_N)^{H-1} (D' - \mathbf{1}\mathbf{1}^t) \tilde{\mathbf{s}} \\ &= (\det S)\tilde{B}_1 - (B^{*0}_N)^{H-1} \prod_{j=1}^N s^{*0}_j \prod_{i \neq 0} \prod_{j=2}^N s_j^i \mathbf{1} - (B^{*0}_N)^{H-1} D' \tilde{\mathbf{s}} \\ &= (\det S)\tilde{B}_1 - (B^{*0}_N)^{H-1} (\mathbf{1}, D')\mathbf{s}, \end{aligned}$$

by using (3).
(5) Since $b_j^{\{\{I+K\}\}} = \sum_{J\in I} D^{I,J} s_j^J D^{J,K}/2^{M-1}$, we have for $I' \in I$,

$$\begin{aligned} \sum_{J\in I} D^{I,J} s_j^{\{\{J+I'\}\}} D^{J,K} &= D^{I,I'} D^{I',K} \sum_{J\in I} D^{I,\{\{J+I'\}\}} s_j^{\{\{J+I'\}\}} D^{\{\{J+I'\}\},K} \\ &= 2^{M-1} D^{I,I'} D^{I',K} b_j^{\{\{I+K\}\}}, \end{aligned}$$

by using (1).

Let $I_0 = (0,\ldots,0), I_1 = (1,1,0,\ldots,0), I_2 = (1,0,1,0,\ldots,0),\ldots$.

The fact that

$$\begin{aligned} & D\begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s_j^i \\ \prod_{i\neq 1}\prod_{j=2}^N s_j^i \\ \vdots \\ \prod_{i\neq H}\prod_{j=2}^N s_j^i \end{pmatrix} \\ = & -D\begin{pmatrix} \prod_{i\neq 0}\prod_{j=2}^N s_j^i & 0 & 0 & \cdots & 0 \\ 0 & \prod_{i\neq 1}\prod_{j=2}^N s_j^i & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \prod_{i\neq H}\prod_{j=2}^N s_j^i \end{pmatrix} D^{-1} 2^{M-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ = & -\left\{ \prod_{j=2}^N \prod_{0\neq I'\in I} D \begin{pmatrix} s_j^{\{\{I_0+I'\}\}} & 0 & 0 & \cdots & 0 \\ 0 & s_j^{\{\{I_1+I'\}\}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & s_j^{\{\{I_H+I'\}\}} \end{pmatrix} D^{-1} \right\} 2^{M-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \end{aligned}$$

and $\sum_{J\in I} D^{I,J} s_j^{\{\{J+I'\}\}} D^{J,K} = 2^{M-1} D^{I,I'} D^{I',K} b_j^{\{\{I+K\}\}}$ yields statement (5).

Q.E.D.

Proof of Theorem 2

By Theorem 9 (4) and Lemma 6, we only need to consider the maximum pole of $J(z) = \int \|\Psi'\|^{2z} \mathrm{d}b$, where $\Psi' = (\det S)\tilde{B}_1 - ({B^*}_N^0)^{H-1}(\mathbf{1}\ D') \begin{pmatrix} \prod_{j=1}^N {s^*}_j^0 \prod_{i\neq 0}\prod_{j=2}^N s_j^i \\ \vdots \\ \prod_{j=1}^N {s^*}_j^H \prod_{i\neq H}\prod_{j=2}^N s_j^i \end{pmatrix}$.

(Case 1): The fact that $B^{11} = \sum_{k=1}^N b_{1k}b_{2k} + \cdots$ provides Case 1.
(Case 2): Assume that $M = 3$.

We have $D' = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$, $\begin{cases} s_j^0 = 1 + b_{1j}b_{2j} + b_{1j}b_{3j} + b_{2j}b_{3j}, \\ s_j^1 = 1 + b_{1j}b_{2j} - b_{1j}b_{3j} - b_{2j}b_{3j}, \\ s_j^2 = 1 - b_{1j}b_{2j} + b_{1j}b_{3j} - b_{2j}b_{3j}, \\ s_j^3 = 1 - b_{1j}b_{2j} - b_{1j}b_{3j} + b_{2j}b_{3j}, \end{cases}$

and $\Psi' = (\det S)\begin{pmatrix} b_{11}b_{21} \\ b_{11}b_{31} \\ b_{21}b_{31} \end{pmatrix} - \prod_{i=0}^3 \prod_{j=2}^N s_j^i ({B^*}_N^0)^2 (1, D') \begin{pmatrix} \prod_{j=1}^N {s^*}_j^0 / \prod_{j=2}^N s_j^0 \\ \prod_{j=1}^N {s^*}_j^1 / \prod_{j=2}^N s_j^1 \\ \prod_{j=1}^N {s^*}_j^2 / \prod_{j=2}^N s_j^2 \\ \prod_{j=1}^N {s^*}_j^3 / \prod_{j=2}^N s_j^3 \end{pmatrix}$.

Let $N=1$. The fact that $\Psi' = 4(B^{*0}_N)^2\tilde{B}_1 - 4(B^{*0}_N)^2\begin{pmatrix} b^*_{11}b^*_{21} \\ b^*_{11}b^*_{31} \\ b^*_{21}b^*_{31}\end{pmatrix}$ yields the statement for $N=1$.

Assume that $N\geq 2$, $b^*\neq 0$ and $b^*_{11}\neq 0, b^*_{21}\neq 0, b^*_{31}\neq 0$. Set $b'_{21}=b_{11}b_{21}$, $b'_{31}=b_{11}b_{31}$ and $b'_{11}=b_{21}b_{31}=b'_{21}b'_{31}/b^2_{11}$. Then

$$\Psi' = (\det S)\begin{pmatrix} b'_{21} \\ b'_{31} \\ b'_{11}\end{pmatrix} - \prod_{i=0}^{3}\prod_{j=2}^{N} s^i_j (B^{*0}_N)^2(\mathbf{1},D')\begin{pmatrix} \prod_{j=1}^N s^{*0}_j/\prod_{j=2}^N s^0_j \\ \prod_{j=1}^N s^{*1}_j/\prod_{j=2}^N s^1_j \\ \prod_{j=1}^N s^{*2}_j/\prod_{j=2}^N s^2_j \\ \prod_{j=1}^N s^{*3}_j/\prod_{j=2}^N s^3_j \end{pmatrix}$$

and its maximum pole is $3/2$ and its order is 1.

Assume that $N\geq 2$, $b^*\neq 0$, $b^*_{11}\neq 0$ and $\prod_{i=1}^3 b^*_{ij}=0$ for all j. Let $\psi = \begin{pmatrix}\psi_1\\ \psi_2 \\ \psi_3\end{pmatrix} = (\mathbf{1},D')\begin{pmatrix} \prod_{j=1}^N s^{*0}_j/\prod_{j=2}^N s^0_j \\ \prod_{j=1}^N s^{*1}_j/\prod_{j=2}^N s^1_j \\ \prod_{j=1}^N s^{*2}_j/\prod_{j=2}^N s^2_j \\ \prod_{j=1}^N s^{*3}_j/\prod_{j=2}^N s^3_j \end{pmatrix}$. By setting $\begin{pmatrix} b'_{21} \\ b'_{31}\end{pmatrix} = (\det S)\begin{pmatrix} b_{11}b_{21} \\ b_{11}b_{31}\end{pmatrix}$ $-\prod_{i=0}^3\prod_{j=2}^N s^i_j(B^{*0}_N)^2\begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1\end{pmatrix}\begin{pmatrix} \prod_{j=1}^N s^{*0}_j/\prod_{j=2}^N s^0_j \\ \prod_{j=1}^N s^{*1}_j/\prod_{j=2}^N s^1_j \\ \prod_{j=1}^N s^{*2}_j/\prod_{j=2}^N s^2_j \\ \prod_{j=1}^N s^{*3}_j/\prod_{j=2}^N s^3_j \end{pmatrix}$ and

$$\Psi'' = \begin{pmatrix}\Psi''_1 \\ \Psi''_2 \\ \Psi''_3\end{pmatrix} = \begin{pmatrix} b'_{21} \\ b'_{31} \\ \frac{(\prod_{i=0}^3\prod_{j=2}^N s^i_j (B^{*0}_N)^2)^2\psi_1\psi_2}{b^2_{11}\det S} - (B^{*0}_N)^2\prod_{i=0}^3\prod_{j=2}^N s^i_j\psi_3\end{pmatrix},$$

and by using Lemma 6, we need the maximum pole of $\int \|\Psi''\|^{2z}db$. Ψ'' is singular in the following cases: (i) $b^*_{11}b^*_{21}=b^*_{2j}=b^*_{3j}=0$ for all j, (ii) $b^*_{11}b^*_{21}\neq 0$, $b^*_{1j}=b^*_{2j}=b^*_{3j}=0$ for all j, since we have $\frac{\partial\psi}{\partial b_j}|_{b^*}$ $=-(\mathbf{1},D')\begin{pmatrix} s^{*0}_1/s^{*0}_j & 0 & 0 & 0 \\ 0 & s^{*1}_1/s^{*1}_j & 0 & 0 \\ 0 & 0 & s^{*2}_1/s^{*2}_j & 0 \\ 0 & 0 & 0 & s^{*3}_1/s^{*3}_j\end{pmatrix}\begin{pmatrix} b^*_{2j}+b^*_{3j} & b^*_{1j}+b^*_{3j} & b^*_{1j}+b^*_{2j} \\ b^*_{2j}-b^*_{3j} & b^*_{1j}-b^*_{3j} & -b^*_{1j}-b^*_{2j} \\ -b^*_{2j}+b^*_{3j} & -b^*_{1j}-b^*_{3j} & b^*_{1j}-b^*_{2j} \\ -b^*_{2j}-b^*_{3j} & -b^*_{1j}+b^*_{3j} & -b^*_{1j}+b^*_{2j}\end{pmatrix}$. If Ψ'' is not singular, its maximum pole is $3/2$ and its order is 1. Assume that Ψ'' is singular that is, (i) $b^*_{11}b^*_{21}=b^*_{2j}=b^*_{3j}=0$ for all j and (ii) $b^*_{11}b^*_{21}\neq 0, b^*_{1j}=b^*_{2j}=b^*_{3j}=0$ for all j. Construct the blow-up of Ψ'' along the submanifold $\{b_{3j}=0, 2\leq j\leq N\}$. Let $b_{32}=u$ and $b_{3j}=ub'_{3j}$ for $j\geq 2$. In the case (i), the coefficient of b_{2j_0} is around $4ub_{1j_0}\sum_{j=2}^N b_{1j}b'_{3j}(1/b^2_{11}-1)+4ub'_{3j_0}(1-b^2_{1j_0})$, since $\prod_{i\neq 0}\prod_{j=2}^N s^i_j \cong \prod_{j=2}^N(1-b_{1j}b_{2j}-ub_{1j}b'_{3j}-ub_{2j}b'_{3j}+2ub^2_{1j}b_{2j}b'_{3j}) \cong 1+\sum_{j=2}^N(-b_{1j}b_{2j}-ub_{1j}b'_{3j}-ub_{2j}b'_{3j}+2ub^2_{1j}b_{2j}b'_{3j})+\sum_{j\neq j'}ub_{1j}b_{1j'}b_{2j}b'_{3j'}$, $\prod_{i=0}\prod_{j=2}^N s^i_j\psi_1 \cong -4\sum_{j=2}^N b_{1j}b_{2j}$, $\prod_{i=0}\prod_{j=2}^N s^i_j\psi_2 \cong -4u\sum_{j=2}^N b_{1j}b'_{3j}$, and $\prod_{i=0}\prod_{j=2}^N s^i_j\psi_3 \cong 4\sum_{j=2}^N(-ub_{2j}b'_{3j}+2ub^2_{1j}b_{2j}b'_{3j})+4\sum_{j\neq j'}ub_{1j}b_{1j'}b_{2j}b'_{3j'}$. If $4ub_{1j_0}\sum_{j=2}^N b_{1j}b'_{3j}(1/b^2_{11}-1)+$

$4ub'_{3j_0}(1-b_{1j_0}^2)=0$ for all j_0 then $b'_{3j_0}=0$ for all j_0 since $|b_{1j}|<1$. It contradicts $b'_{32}=1$. So $(\frac{(\prod_{i=0}^3\prod_{j=2}^N s_j^i(B^{*0}_N)^2)^2\psi_1\psi_2}{b_{11}^2\det S}-(B^{*0}_N)^2\prod_{i=0}^3\prod_{j=2}^N s_j^i\psi_3)/u$ is smooth.

In the case (ii), the coefficient b_{2j_0} is around $4u(1-{b_{21}^*}^2)b'_{3j_0}$ since $s^{*0}_1\prod_{i\neq 0}\prod_{j=2}^N s_j^i\cong 4(1+b_{11}^*b_{21}^*)\prod_{j=2}^N(1-ub_{2j}b'_{3j})\cong 4(1+b_{11}^*b_{21}^*)(1-u\sum_{j=2}^N b_{2j}b'_{3j})$, $(1+b_{11}^*b_{21}^*)\prod_{i=0}\prod_{j=2}^N s_j^i\psi_1\cong 4b_{11}^*b_{21}^*$, $\prod_{i=0}\prod_{j=2}^N s_j^i\psi_2\cong -4ub_{11}^*b_{21}^*\sum_{j=2}^N b_{2j}b'_{3j}$, and $\prod_{i=0}\prod_{j=2}^N s_j^i\psi_3\cong -4u\sum_{j=2}^N b_{2j}b'_{3j}$. So $(\frac{(\prod_{i=0}^3\prod_{j=2}^N s_j^i(B^{*0}_N)^2)^2\psi_1\psi_2}{b_{11}^2\det S}-(B^{*0}_N)^2\prod_{i=0}^3\prod_{j=2}^N s_j^i\psi_3)/u$ is smooth.

We have $\Psi''=\begin{pmatrix} b'_{21}\\ b'_{31}\\ ub'_{22}\end{pmatrix}$, for a variable b'_{22} for both cases (i) and (ii) and we have the statement for $N\geq 2$, $b^*\neq 0$, $b_{11}^*\neq 0$ and $\prod_{i=1}^3 b_{ij}^*=0$ for all j.
Let $N\geq 2$ and $b^*=0$.

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

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

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

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

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

By using Lemma 6 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 the blow-up of Ψ''' along the submanifold $\{b''_{21}=0, b''_{31}=0, b'_{3k}=0, 2\leq k\leq N\}$. Then we have cases (I) and (II).

(I) Let $b'_{32}=v$, $b''_{21}=vb'''_{21}$, $b''_{31}=vb'''_{21}$ and $b'_{3k}=vb''_{3k}$ 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 9 (5), we can set $f_2 = v f_2'$ and $f_3 = v f_3'$, 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

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

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

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

Q.E.D.

Appendix C.

Definition 10 *(1) Let $R = (r_{ij})$ be an $H \times H'$ matrix, I an element of $\{0,1\}^H$, and $f(R,r')$ an analytic function of $r_{11}, r_{21}, \ldots, r_{HH'}, r'_1, \ldots, r'_k$, where $r' = (r'_1, \ldots, r'_k)$. $f(R,r')$ is an I-type function of $(r_{ij})_{i' \le i \le H, 1 \le j \le H'}$, if for any $I_{i_0} = 1$ with $i_0 \ge i'$,*

$$f(r_{11},\cdots,r_{1N},r_{21},\cdots,r_{i_0-1,N},ur_{i_0,1},\cdots,ur_{i_0,N},r_{i_0+1,1},\cdots,r_{M,N},r')/u,$$

is an analytic function of u, where u is a variable.

(2) Let $I, I' \in \{0,1\}^H$. We denote $I \le I'$ if $I_i \le I'_i$ for all $i = 1, \ldots, H$, and denote $I < I'$ if $I \le I'$ and $I \ne I'$.

For example, $B^I = B^I_N$ is an I'-type function of B for all $I' \leq I$ ($I' \in \{0,1\}^M$).

Let $I_{ij} = \overset{\qquad\qquad\qquad i \qquad\qquad\qquad\qquad\quad j}{(0, \;\ldots,\; 0,\; 1,\; 0,\; \ldots,\; 0,\; 1,\; 0,\; \ldots,\; 0)}$, for $i < j$.

Proof of Theorem 3

Assume that $a^* = 0$.

Let $B'^{I_{ij}} = B^{I_{ij}} - \sum_{k=1}^N b_{ik}b_{jk}$, which is a polynomial of at least degree four.

For $I \in \mathcal{I}$, let $I^{(s)} \in \{0,1\}^M$ be $I_i^{(s)} = \begin{cases} 0, & \text{if } i \leq s, \\ I_i, & \text{if } i > s. \end{cases}$

We set $\mathcal{I}' = \mathcal{I} - \{I_{ij} : 1 \leq i < j \leq M\}$.

By using a blowing up process together with an inductive method of s, we have functions (5) and (6) below.

$$\int \{ \sum_{1 \leq i < j \leq M} (B^{I_{ij}}_{(s)})^2 + \sum_{I \in \mathcal{I}'} (B^I_{(s)})^2 \}^z$$
$$u_1^{MN-1} u_2^{(M-1)N-1} \cdots u_s^{(M-s+1)N-1} \prod_{i=1}^s v_i^{N-i-1} \mathrm{d}u \mathrm{d}b^{(s)} \mathrm{d}v, \tag{5}$$

where

$$B^{I_{ij}}_{(s)} = \begin{cases} u_1^2 u_2^2 \cdots u_i^2 u_{i+1} \cdots u_j \{ f^{(s)}_{ij} + b^{(s)}_{ji} + u_1^2 B'^{I_{ij}}_{(s)} \}, & i < j \leq s, \\ u_1^2 u_2^2 \cdots u_i^2 u_{i+1} \cdots u_s \{ f^{(s)}_{ij} + b^{(s)}_{ji} + u_1^2 B'^{I_{ij}}_{(s)} \}, & i \leq s < j, \\ u_1^2 u_2^2 \cdots u_s^2 \{ f^{(s)}_{ij} + \sum_{k=s+1}^N b^{(s)}_{ik} b^{(s)}_{jk} + u_1^2 B'^{I_{ij}}_{(s)} \}, & s < i < j, \end{cases}$$

$$B^I_{(s)} = \prod_{k=1}^s u_k^{\sum_{k'=k}^M I_i} B'^I_{(s)}, \text{ for } I \in \mathcal{I}',$$

$f^{(s)}_{ij}$ is an $I^{(s)}_{ij}$-type function of $(b^{(s)}_{kl})_{s+1 \leq k \leq M, 1 \leq l \leq N}$,

$$f^{(s)}_{ij}|_{b^{(s)}_{i1} = \cdots = b^{(s)}_{i,\min\{i-1,s\}} = b^{(s)}_{j1} = \cdots = b^{(s)}_{j,\min\{i-1,s\}} = 0} = 0,$$

$B'^{I_{ij}}_{(s)}$ is an $I^{(s)}_{ij}$-type function of $(b^{(s)}_{kl})_{s+1 \leq k \leq M, 1 \leq l \leq N}$, and $B'^I_{(s)}$ ($I \in \mathcal{I}'$) is an $I^{(s)}$-type function of $(b^{(s)}_{kl})_{s+1 \leq k \leq M, 1 \leq l \leq N}$.

For $s+1 \leq \ell \leq M$ and $1 \leq \ell' \leq s$,

$$\int \{ \sum_{\{i<j\leq s\} \cup \{i \leq s < j, i < \ell'\} \cup \{i = \ell', j = \ell\}} (B^{I_{ij}}_{(s)})^2 \}^z u_1^{MN-1} u_2^{(M-1)N-1} \cdots u_s^{(M-s+1)N-1} u_{s+1}^{(M-s)N-1} \mathrm{d}u \mathrm{d}\tilde{b}, \tag{6}$$

where

$$B^{I_{ij}}_{(s)} = \begin{cases} u_1^2 u_2^2 \cdots u_i^2 u_{i+1} \cdots u_j \tilde{b}_{ji}, & \text{if } i < j \leq s, \\ u_1^2 u_2^2 \cdots u_i^2 u_{i+1} \cdots u_{s+1} \tilde{b}_{ji}, & \text{if } i \leq s < j, i < \ell', \\ u_1^2 u_2^2 \cdots u_{\ell'}^2 u_{\ell'+1} \cdots u_{s+1}, & \text{if } i = \ell', j = \ell. \end{cases}$$

C.1 Step 1

Construct the blow-up of function (3) along the submanifold $\{b_{ij}=0, 1\le i\le M, 1\le j\le N\}$.

Let $b_{11}=u_1$, $b_{ij}=u_1 b'_{ij}$, $(i,j)\neq(1,1)$.

Then we have $B^{I_{1j}}=u_1^2(b'_{j1}+\sum_{k=2}^N b'_{1k}b'_{jk}+B'^{I_{1j}}/u_1^2)$ for $j\ge 2$ and $B^{I_{ij}}=u_1^2(\sum_{k=1}^N b'_{ik}b'_{jk}+B'^{I_{ij}}/u_1^2)$ for $2\le i<j$.

Let $b''_{j1}=b'_{j1}+\sum_{k=2}^N b'_{1k}b'_{jk}$ for $j\ge 2$.

Then for $2\le i<j$,

$$\begin{aligned}
\sum_{k=1}^N b'_{ik}b'_{jk} &= (b''_{i1}-\sum_{k=2}^N b'_{1k}b'_{ik})(b''_{j1}-\sum_{k=2}^N b'_{1k}b'_{jk})+\sum_{k=2}^N b'_{ik}b'_{jk}\\
&= b''_{i1}(b''_{j1}-\sum_{k=2}^N b'_{1k}b'_{jk})-(\sum_{k=2}^N b'_{1k}b'_{ik})b''_{j1}+(\sum_{k=2}^N b'_{1k}b'_{ik})(\sum_{k=2}^N b'_{1k}b'_{jk})+\sum_{k=2}^N b'_{ik}b'_{jk}\\
&= f_{ij}^{(1)}+(b'_{i2},\ldots,b'_{iN})\begin{pmatrix} b'_{12}\\ b'_{13}\\ \vdots\\ b'_{1N}\end{pmatrix}(b'_{12},\ldots,b'_{1N})\begin{pmatrix} b'_{j2}\\ b'_{j3}\\ \vdots\\ b'_{jN}\end{pmatrix}\\
&\quad +(b'_{i2},\ldots,b'_{iN})\begin{pmatrix} b'_{j2}\\ b'_{j3}\\ \vdots\\ b'_{jN}\end{pmatrix},
\end{aligned}$$

where $f_{ij}^{(1)}=b''_{i1}(b''_{j1}-\sum_{k=2}^N b'_{1k}b'_{jk})-(\sum_{k=2}^N b'_{1k}b'_{ik})b''_{j1}$ is an $I_{ij}^{(1)}$-type function of $\begin{pmatrix} b''_{21} & b'_{22} & \cdots & b'_{2N}\\ b''_{31} & b'_{32} & \cdots & b'_{3N}\\ \vdots & \vdots & \vdots & \vdots\\ b''_{M1} & b'_{M2} & \cdots & b'_{MN}\end{pmatrix}$ with $f_{ij}^{(1)}|_{b''_{i1}=b''_{j1}=0}=0$.

Next, construct the blow-up along the submanifold $\{b'_{12}=b'_{13}=\cdots=b'_{1N}=0\}$.

Let $b'_{12}=v_1$, $b'_{13}=v_1 b''_{13}$, $\cdots$, $b'_{1N}=v_1 b''_{1N}$.

Then we have, for $2\le i<j$,

$$\begin{aligned}
&(b'_{i2},\ldots,b'_{iN})\begin{pmatrix} b'_{12}\\ b'_{13}\\ \vdots\\ b'_{1N}\end{pmatrix}(b'_{12},\ldots,b'_{1N})\begin{pmatrix} b'_{j2}\\ b'_{j3}\\ \vdots\\ b'_{jN}\end{pmatrix}+(b'_{i2},\ldots,b'_{iN})\begin{pmatrix} b'_{j2}\\ b'_{j3}\\ \vdots\\ b'_{jN}\end{pmatrix}\\
=\ &(b'_{i2},\ldots,b'_{iN})\left(v_1^2\begin{pmatrix} 1\\ b''_{13}\\ \vdots\\ b''_{1N}\end{pmatrix}(1,b''_{13},\cdots,b''_{1N})+\begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}\right)\begin{pmatrix} b'_{j2}\\ b'_{j3}\\ \vdots\\ b'_{jN}\end{pmatrix}.
\end{aligned}$$

Let $Q_i = \sqrt{1 + {b''}^2_{13} + \cdots + {b''}^2_{1i}}$ and

$$G = \begin{pmatrix} \frac{1}{Q_N} & \frac{-b''_{13}}{Q_3} & \cdots & \frac{-b''_{1i}}{Q_{i-1}Q_i} & \cdots & \frac{-b''_{1N}}{Q_{N-1}Q_N} \\ \frac{b''_{13}}{Q_N} & \frac{1}{Q_3} & \cdots & \frac{-b''_{13}b''_{1i}}{Q_{i-1}Q_i} & \cdots & \frac{-b''_{13}b''_{1N}}{Q_{N-1}Q_N} \\ \frac{b''_{14}}{Q_N} & 0 & \cdots & \frac{-b''_{14}b''_{1i}}{Q_{i-1}Q_i} & \cdots & \frac{-b''_{14}b''_{1N}}{Q_{N-1}Q_N} \\ \vdots & \vdots & & \vdots & & \vdots \\ & & & \frac{-b''_{1,i-1}b''_{1i}}{Q_{i-1}Q_i} & & \\ & & & \frac{Q_{i-1}}{Q_i} & & \\ & & & 0 & & \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \frac{b''_{1N}}{Q_N} & 0 & \cdots & 0 & \cdots & \frac{Q_{N-1}}{Q_N} \end{pmatrix}.$$

Then we have

$$\begin{aligned} & v_1^2 \begin{pmatrix} 1 \\ b''_{13} \\ \vdots \\ b''_{1N} \end{pmatrix} (1, b''_{13}, \cdots, b''_{1N}) + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ = \; & v_1^2 G \begin{pmatrix} 1 + {b''}^2_{13} + \cdots + {b''}^2_{1N} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} G^t + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ = \; & G \left(v_1^2 \begin{pmatrix} 1 + {b''}^2_{13} + \cdots + {b''}^2_{1N} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right) G^t \\ = \; & G \begin{pmatrix} 1 + v_1^2(1 + {b''}^2_{13} + \cdots + {b''}^2_{1N}) & 0 & \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} G^t. \end{aligned}$$

Therefore, we can change the variables from $(b'_{i2}, b'_{i3}, \cdots, b'_{iN})$ to $(b''_{i2}, b''_{i3}, \cdots, b''_{iN})$ by

$$\begin{pmatrix} b''_{i2} \\ b''_{i3} \\ \vdots \\ b''_{iN} \end{pmatrix} = \begin{pmatrix} \sqrt{1 + v_1^2(1 + {b''}^2_{13} + \cdots + {b''}^2_{1N})} & 0 & \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} G^t \begin{pmatrix} b'_{i2} \\ b'_{i3} \\ \vdots \\ b'_{iN} \end{pmatrix}.$$

We have

$$
\begin{aligned}
&(b'_{i2}, b'_{i3}, \cdots, b'_{iN}) \left(v_1^2 \begin{pmatrix} 1 \\ b'_{13} \\ \vdots \\ b'_{1N} \end{pmatrix} (1, b'_{13}, \cdots, b'_{1N}) + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right) \begin{pmatrix} b'_{j2} \\ b'_{j3} \\ \vdots \\ b'_{jN} \end{pmatrix} \\
= &\sum_{k=2}^{N} b''_{ik} b''_{jk}.
\end{aligned}
$$

Let $b_{ik}^{(1)} = b''_{ik}$ for $1 \leq i \leq M, 1 \leq k \leq N$ and $(i,k) \neq (1,1)$, $B_{(1)}^{\prime I_{ij}} = B^{\prime I_{ij}}/u_1^4$ for $1 \leq i < j \leq M$ and $B_{(1)}^{I} = B^{I}/u_1^{\sum_{k=1}^{M} I_i}$ for $I \in I'$.

We have $B^{I_{1j}} = u_1^2(b_{j1}^{(1)} + u_1^2 B_{(1)}^{\prime I_{1j}})$ for $1 < j$, $B^{I_{ij}} = u_1^2(f_{ij}^{(1)} + \sum_{k=2}^{N} b_{ik}^{(1)} b_{jk}^{(1)} + u_1^2 B_{(1)}^{\prime I_{ij}})$ for $1 < i < j$ and $B^{I} = u_1^{\sum_{k=1}^{M} I_i} B_{(1)}^{I}$ for $I \in I'$, where $f_{ij}^{(1)}$ is an $I_{ij}^{(1)}$-type function of $(b_{kl}^{(1)})_{2 \leq k \leq M, 1 \leq l \leq N}$ with $f_{ij}^{(1)}|_{b_{i1}^{(1)} = b_{j1}^{(1)} = 0} = 0$, $B_{(1)}^{\prime I_{ij}}$ is an $I_{ij}^{(1)}$-type function of $(b_{kl}^{(1)})_{2 \leq k \leq M, 1 \leq l \leq N}$, and $B_{(1)}^{\prime I}$ is an $I^{(1)}$-type function of $(b_{kl}^{(1)})_{2 \leq k \leq M, 1 \leq l \leq N}$.

C.2 Step 2

Assume Eq. (5). Construct the blow-up of function (5) along the submanifold $\{b_{ij}^{(s)} = 0, s+1 \leq i \leq M, 1 \leq j \leq N\}$.

Let $b_{ij}^{(s)} = u_{s+1} b'^{(s)}_{ij}$.

We have

$$
B_{(s)}^{I_{ij}} = \begin{cases} u_1^2 u_2^2 \cdots u_i^2 u_{i+1} \cdots u_j \{ f_{ij}^{(s)} + b_{ji}^{(s)} + u_1^2 B_{(s)}^{\prime I_{ij}} \}, & i < j \leq s, \\ u_1^2 u_2^2 \cdots u_i^2 u_{i+1} \cdots u_s u_{s+1} \{ f_{ij}^{(s)}/u_{s+1} + b'^{(s)}_{ji} + u_1^2 B_{(s)}^{\prime I_{ij}}/u_{s+1} \}, & i \leq s < j, \\ u_1^2 u_2^2 \cdots u_s^2 u_{s+1}^2 \{ f_{ij}^{(s)}/u_{s+1}^2 + \sum_{k=s+1}^{N} b'^{(s)}_{ik} b'^{(s)}_{jk} + u_1^2 B_{(s)}^{\prime I_{ij}}/u_{s+1}^2 \}, & s < i < j, \end{cases}
$$

and

$$
B_{(s)}^{I} = \prod_{k=1}^{s+1} u_k^{\sum_{k'=k}^{N} I_i} B_{(s)}^{\prime I} / u_{s+1}^{\sum_{k'=s+1}^{N} I_i}, I \in I'.
$$

We may consider $b'^{(s)}_{s+1,s+1} = 1$ or $b'^{(s)}_{\ell,\ell'} = 1$ for some $s+1 \leq \ell \leq M$ and $1 \leq \ell' \leq s$.

If $b'^{(s)}_{\ell,\ell'} = 1$ for some $s+1 \leq \ell \leq M$ and $1 \leq \ell' \leq s$, then we have function (6) by using

$$
f_{ij}^{(s)}|_{b_{i1}^{(s)} = \cdots = b_{i,\min\{i-1,s\}}^{(s)} = b_{j1}^{(s)} = \cdots = b_{j,\min\{i-1,s\}}^{(s)} = 0} = 0,
$$

and Lemma 6.

Let $b'^{(s)}_{s+1,s+1} = 1$.

Set $b''^{(s)}_{j,s+1} = b'^{(s)}_{j,s+1} + \sum_{k=s+2}^{N} b'^{(s)}_{s+1,k} b'^{(s)}_{jk}$ for $j \geq s+2$.

Then for $s+2 \le i < j$,

$$
\begin{aligned}
& f_{ij}^{(s)}/u_{s+1}^2 + \sum_{k=s+1}^{N} {b'}_{ik}^{(s)} {b'}_{jk}^{(s)} \\
= \quad & f_{ij}^{(s)}/u_{s+1}^2 + ({b''}_{i,s+1}^{(s)} - \sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{ik}^{(s)})({b''}_{j,s+1}^{(s)} - \sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{jk}^{(s)}) + \sum_{k=s+2}^{N} {b'}_{ik}^{(s)} {b'}_{jk}^{(s)} \\
= \quad & f_{ij}^{(s)}/u_{s+1}^2 + {b''}_{j,s+1}^{(s)}({b''}_{i,s+1}^{(s)} - \sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{ik}^{(s)}) - {b''}_{i,s+1}^{(s)}(\sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{jk}^{(s)}) \\
& + (\sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{ik}^{(s)})(\sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{jk}^{(s)}) + \sum_{k=s+2}^{N} {b'}_{ik}^{(s)} {b'}_{jk}^{(s)}.
\end{aligned}
$$

Let $f_{ij}^{(s+1)} = f_{ij}^{(s)}/u_{s+1}^2 + {b''}_{j,s+1}^{(s)}({b''}_{i,s+1}^{(s)} - \sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{ik}^{(s)}) - {b''}_{i,s+1}^{(s)}(\sum_{k=s+2}^{N} {b'}_{s+1,k}^{(s)} {b'}_{jk}^{(s)})$. Then $f_{ij}^{(s+1)}$ is an $I_{ij}^{(s+1)}$-type function of $\begin{pmatrix} {b''}_{s+2,1}^{(s)} & {b'}_{s+2,2}^{(s)} & \cdots & {b'}_{s+2,N}^{(s)} \\ \vdots & \vdots & \vdots & \vdots \\ {b''}_{M1}^{(s)} & {b'}_{M2}^{(s)} & \cdots & {b'}_{MN}^{(s)} \end{pmatrix}$ with

$$
f_{ij}^{(s+1)}|_{{b''}_{i1}^{(s)}=\cdots={b''}_{i,s+1}^{(s)}={b''}_{j1}^{(s)}=\cdots={b''}_{j,s+1}^{(s)}=0} = 0.
$$

Next, construct the blow-up along the submanifold $\{{b'}_{s+1,s+2}^{(s)} = {b'}_{s+1,s+3}^{(s)} = \cdots = {b'}_{s+1,N}^{(s)} = 0\}$.
Let ${b'}_{s+1,s+2}^{(s)} = v_s$, ${b'}_{s+1,s+3}^{(s)} = v_s {b''}_{s+1,s+3}^{(s)}$, $\cdots$,
${b'}_{s+1,N}^{(s)} = v_s {b''}_{s+1,N}^{(s)}$.
Let $Q_i^{(s)} = \sqrt{1 + {b''}_{s+1,s+3}^{(s)\,2} + \cdots + {b''}_{s+1,i}^{(s)\,2}}$ and

$$
G^{(s)} = \begin{pmatrix}
\frac{1}{Q_N} & \frac{-{b''}_{s+1,s+3}^{(s)}}{Q_{s+3}^{(s)}} & \cdots & \frac{-{b''}_{s+1,i}^{(s)}}{Q_{i-1}^{(s)} Q_i^{(s)}} & \cdots & \frac{-{b''}_{s+1,N}^{(s)}}{Q_{N-1}^{(s)} Q_N^{(s)}} \\
\frac{{b''}_{s+1,s+3}^{(s)}}{Q_N^{(s)}} & \frac{1}{Q_{s+3}^{(s)}} & \cdots & \frac{-{b''}_{s+1,s+3}^{(s)} {b''}_{s+1,i}^{(s)}}{Q_{i-1}^{(s)} Q_i^{(s)}} & \cdots & \frac{-{b''}_{s+1,s+3}^{(s)} {b''}_{s+1,N}^{(s)}}{Q_{N-1}^{(s)} Q_N^{(s)}} \\
\frac{{b''}_{s+1,s+4}^{(s)}}{Q_N^{(s)}} & 0 & \cdots & \frac{-{b''}_{s+1,s+4}^{(s)} {b''}_{s+1,i}^{(s)}}{Q_{i-1}^{(s)} Q_i^{(s)}} & \cdots & \frac{-{b''}_{s+1,s+4}^{(s)} {b''}_{s+1,N}^{(s)}}{Q_{N-1}^{(s)} Q_N^{(s)}} \\
\vdots & \vdots & & \vdots & & \vdots \\
 & & & \frac{-{b''}_{s+1,i-1}^{(s)} {b''}_{s+1,i}^{(s)}}{Q_{i-1}^{(s)} Q_i^{(s)}} & & \\
 & & & \frac{Q_{i-1}^{(s)}}{Q_i^{(s)}} & & \\
 & & & 0 & & \\
\vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
\frac{{b''}_{s+1,N}^{(s)}}{Q_N^{(s)}} & 0 & \cdots & 0 & \cdots & \frac{Q_{N-1}^{(s)}}{Q_N^{(s)}}
\end{pmatrix}.
$$

Change variables from $(b'^{(s)}_{i,s+2}, b'^{(s)}_{i,s+3}, \cdots, b'^{(s)}_{iN})$ to $(b''^{(s)}_{i,s+2}, b''^{(s)}_{i,s+3}, \cdots, b''^{(s)}_{iN})$ by

$$\begin{pmatrix} b''^{(s)}_{i,s+2} \\ b''^{(s)}_{i,s+3} \\ \vdots \\ b''^{(s)}_{iN} \end{pmatrix} = \begin{pmatrix} \sqrt{1+v^2(1+{b''^{(s)}_{s+1,s+3}}^2+\cdots+{b''^{(s)}_{s+1,N}}^2)} & 0 & \cdots & 0 \\ 0 & 1 & 0\cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} {G^{(s)}}^t \begin{pmatrix} b'^{(s)}_{i,s+2} \\ b'^{(s)}_{i,s+3} \\ \vdots \\ b'^{(s)}_{iN} \end{pmatrix}.$$

We have

$$\begin{aligned} &(\sum_{k=s+2}^{N} b'^{(s)}_{s+1,k} b'^{(s)}_{ik})(\sum_{k=s+2}^{N} b'^{(s)}_{s+1,k} b'^{(s)}_{jk}) + \sum_{k=s+2}^{N} b^{(s)}_{ik} b^{(s)}_{jk} \\ = &\sum_{k=s+2}^{N} b''^{(s)}_{ik} b''^{(s)}_{jk}. \end{aligned}$$

Let $b^{(s+1)}_{ji} = b^{(s)}_{ji}$ for $i, j \le s$ and $b^{(s+1)}_{ji} = b''^{(s)}_{ji}$ for $i, j > s$ and $(i, j) \ne (s+1, s+1)$. Also let $f^{(s+1)}_{ij} = f^{(s)}_{ij}$ for $i < j \le s$, $f^{(s+1)}_{ij} = f^{(s)}_{ij}/u_{s+1}$ for $i \le s < j$, $f^{(s+1)}_{s+1,j} = f^{(s)}_{s+1,j}/u^2_{s+1}$ for $j > s+1$, $B'^{I_{ij}}_{(s+1)} = B'^{I_{ij}}_{(s)}$ for $i < j \le s$, $B'^{I_{ij}}_{(s+1)} = B'^{I_{ij}}_{(s)}/u_{s+1}$ for $i \le s < j$, $B'^{I_{ij}}_{(s+1)} = B'^{I_{ij}}_{(s)}/u^2_{s+1}$ for $s < i < j$ and $B'^{I}_{(s+1)} = B'^{I}_{(s)}/u_{s+1}^{\sum_{k'=s+1}^{N} I_i}$ for $I \in I'$. Then we have Eq. (5) with $s+1$.

C.3 Step 3

From the above induction $(1 \le s \le N+1)$, we finally have Eq. (6) since we assume that $N < M$.

Note that we have the same inductive results for

$$\int \{\sum_{1\le i<j\le M} (\sum_{k=1}^{N} b_{ik} b_{jk})^2\}^z \mathrm{d}b, \tag{7}$$

instead of the function in Eq. (3). This means that the maximum pole, and its order, of the function in Eq. (3) are those of the function in Eq. (7).

Now we again consider the maximum pole of the function in Eq. (7) and its order.

In Step3, we use the same symbol b rather than $b^{(s)}$ for the sake of simplicity.

We need to consider the following function with the inductive method with s.

$$\begin{aligned} \int \{u_1^4 u_2^4 \cdots u_s^4 \sum_{1\le i<j\le M, i\le s} b^2_{ji} + u_1^4 u_2^4 \cdots u_s^4 \sum_{s+1\le i<j\le M} (\sum_{k=s+1}^{N} b_{ik} b_{jk})^2\}^z \\ \prod_{k=1}^{s} u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} \mathrm{d}u \mathrm{d}b. \end{aligned} \tag{8}$$

First, we set the variables the same as in Step 1. Then we have

$$\int \{u_1^4 \sum_{2\le j\le M} b^2_{j1} + u_1^4 \sum_{2\le i<j\le M} (f^{(1)}_{ij} + \sum_{k=2}^{N} b_{ik} b_{jk})^2\}^z u_1^{MN-1} \mathrm{d}u \mathrm{d}b.$$

By using Lemma 6 again, we need to consider

$$\int \{u_1^4 \sum_{2\le j\le M} b_{j1}^2 + u_1^4 \sum_{2\le i<j\le M} (\sum_{k=2}^{N} b_{ik}b_{jk})^2\}^z u_1^{MN-1} \mathrm{d}u\mathrm{d}b.$$

Assume Eq. (8). Construct the blow-up of function (8) along the submanifold $\{b_{ji}=0, 1\le i< j\le M, i\le s, b_{kl}=0, s+1\le k\le M, s+1\le l\le N\}$.

Then we have

$$\int \{u_1^4u_2^4\cdots u_s^4u_{s+1}^2 \sum_{1\le i<j\le M, i\le s} b_{ji}^2 + u_1^4u_2^4\cdots u_s^4u_{s+1}^4 \sum_{s+1\le i<j\le M} (\sum_{k=s+1}^{N} b_{ik}b_{jk})^2\}^z$$
$$u_{s+1}^{(M-s)(N-s)+(2M-1-s)s/2-1} \prod_{k=1}^{s} u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} \mathrm{d}u\mathrm{d}b,$$

where we can set $b_{21}=1$ or $b_{s+1,s+1}=1$.

If $b_{21}=1$, we have the poles

$$\frac{(M-k+1)(N-k+1)+(2M-k)(k-1)}{4}, k=1,\ldots,s$$

and

$$\frac{(M-s)(N-s)+(2M-1-s)s/2}{2}.$$

If $b_{s+1,s+1}=1$, then by setting the variables the same as in Step 2 and by using Lemma 6, we have

$$\int \{u_1^4u_2^4\cdots u_s^4u_{s+1}^2 \sum_{1\le i<j\le M, i\le s} b_{ji}^2$$
$$+u_1^4u_2^4\cdots u_s^4u_{s+1}^4 (\sum_{s+1<j\le M} b_{j,s+1}^2 + \sum_{s+2\le i<j\le M} (\sum_{k=s+2}^{N} b_{ik}b_{jk})^2)\}^z$$
$$u_{s+1}^{(M-s)(N-s)+(2M-1-s)s/2-1} \prod_{k=1}^{s} u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} \mathrm{d}u\mathrm{d}b. \tag{9}$$

Construct the blow-up of function (9) along the submanifold $\{b_{ji}=0, 1\le i<j\le M, i\le s, u_{s+1}=0\}$.

Then we have Eq. (8) with $s+1$, that is,

$$\int \{u_1^4u_2^4\cdots u_s^4u_{s+1}^4 \sum_{1\le i<j\le M, i\le s} b_{ji}^2$$
$$+u_1^4u_2^4\cdots u_s^4u_{s+1}^4 (\sum_{s+1<j\le M} b_{j,s+1}^2 + \sum_{s+2\le i<j\le M} (\sum_{k=s+2}^{N} b_{ik}b_{jk})^2)\}^z$$
$$u_{s+1}^{(M-s)(N-s)+(2M-1-s)s-1} \prod_{k=1}^{s} u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} \mathrm{d}u\mathrm{d}b.$$

or

$$\int \{u_1^4 u_2^4 \cdots u_s^4 u_{s+1}^2 b_{21}^4 (1 + \sum_{1\le i<j\le M, i\le s, (i,j)\neq(1,2)} b_{ji}^2)$$
$$+ u_1^4 u_2^4 \cdots u_s^4 u_{s+1}^4 b_{21}^4 (\sum_{s+1<j\le M} b_{j,s+1}^2 + \sum_{s+2\le i<j\le M} (\sum_{k=s+2}^{N} b_{ik} b_{jk})^2)\}^z$$
$$u_{s+1}^{(M-s)(N-s)+(2M-1-s)s/2-1} b_{21}^{(M-s)(N-s)+(2M-1-s)s-1}$$
$$\prod_{k=1}^{s} u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} \mathrm{d}u\mathrm{d}b,$$

which have the poles

$$\frac{(M-k+1)(N-k+1)+(2M-k)(k-1)}{4}, k = 1, \ldots, s+1,$$

and

$$\frac{(M-s)(N-s)+(2M-1-s)s/2}{2}.$$

Finally, we have

$$\int \{u_1^4 u_2^4 \cdots u_N^4 \sum_{1\le i<j\le M, i\le N} b_{ji}^2\}^z \prod_{k=1}^{N} u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} \mathrm{d}u\mathrm{d}b,$$

and obtain the poles

$$\frac{(M-k+1)(N-k+1)+(2M-k)(k-1)}{4}, k = 1, \ldots, N,$$

and

$$\frac{(2M-1-N)N}{4}.$$

Therefore, since we assume that $M > N$, we have the maximum pole $-\lambda = -\frac{MN}{4}$ and its order $\theta = \begin{cases} 1, & \text{if } M > N+1, \\ M, & \text{if } M = N+1. \end{cases}$ Q.E.D.

Proof of Theorem 4

Assume that $a^* = 0$

By the proof of Theorem 3, the maximum pole of $\int \{\sum_{1\le i<j\le M} (B^{I_{ij}})^2\}^z db$ is that of $\int \{\sum_{1\le i<j\le M} (\sum_{k=1}^N b_{ik} b_{jk})^2\}^z db$ even for $M \le N$. If $M \le N$ then the maximum pole of $\int \{\sum_{1\le i<j\le M} (\sum_{k=1}^N b_{ik} b_{jk})^2\}^z db$ is $-M(M-1)/4$. Therefore the maximum pole $-\lambda$ of $\int \{\sum_{I\neq 0\in I} (B^I)^2\}^z db$ satisfies $\lambda \ge M(M-1)/4$, since $\sum_{1\le i<j\le M} (B^{I_{ij}})^2 \le \sum_{I\neq 0\in I} (B^I)^2$.

Next let us prove that $\lambda \le \frac{2N+(M-1)(M-2)}{4}$. Consider Eq. (6) with $\ell = M$, $\ell' = M-1$ and $s = M-1$.

Let $\tilde{b}_{ji} = u_M \tilde{b}'_{ji}$ for $i < j < M$. Then we have the pole

$$\frac{N+(M-2)(M-1)/2}{2}.$$

For $a^* \neq 0$, Lemma 7 yields the statement.

Q.E.D.

References

H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19:716–723, 1974.

H. Akaike. Likelihood and bayes procedure. In J. M. Bernald, editor, *Bayesian Statistics*, pages 143–166. University Press, Valencia, Spain, 1980.

S. Amari and N. Murata. Statistical theory of learning curves under entropic loss. *Neural Computation*, 5:140–153, 1993.

S. Amari, N. Fujita, and S. Shinomoto. Four types of learning curves. *Neural Computation*, 4(4): 608–618, 1992.

M. Aoyagi. The zeta function of learning theory and generalization error of three layered neural perceptron. *RIMS Kokyuroku, Recent Topics on Real and Complex Singularities*, 1501:153–167, 2006.

M. Aoyagi and S. Watanabe. Stochastic complexities of reduced rank regression in bayesian estimation. *Neural Networks*, 18:924–933, 2005b.

M. Aoyagi and S. Watanabe. Resolution of singularities and the generalization error with bayesian estimation for layered neural network. *IEICE Trans. J88-D-II*, 10:2112–2124, 2005a.

K. Fukumizu. A regularity condition of the information matrix of a multilayer perceptron network. *Neural Networks*, 9(5):871–879, 1996.

W. Fulton. *Introduction to toric varieties, Annals of Mathematics Studies*. Princeton University Press, 1993.

K. Hagiwara, N. Toda, and S. Usui. On the problem of applying aic to determine the structure of a layered feed-forward neural network. In *Proceedings of IJCNN Nagoya Japan*, volume 3, pages 2263–2266, 1993.

E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of Royal Statistical Society Series B*, 41:190–195, 1979.

J. A. Hartigan. A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of J.Neyman and J.Kiefer*, volume 2, pages 807–810, 1985.

G. E. Hinton. Reinforcement learning with factored states and actions. *Brian Sallans*, 5:1063–1088, 2004.

H. Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. *Annals of Math*, 79:109–326, 1964.

E. Levin, N. Tishby, and S. A. Solla. A statistical approaches to learning and generalization in layered neural networks. In *Proceedings of IEEE*, volume 78, pages 1568–1674, 1990.

D. J. Mackay. Bayesian interpolation. *Neural Computation*, 4(2):415–447, 1992.

N. J. Murata, S. G. Yoshizawa, and S. Amari. 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.

K. Nagata and S. Watanabe. A proposal and effectiveness of the optimal approximation for bayesian posterior distribution. In *Workshop on Information-Based Induction Sciences*, pages 99–104, 2005.

K. Nagata and S. Watanabe. Analysis of exchange ratio for exchange monte carlo method. In *Proceedings of The First IEEE Symposium on Foundations of Computational Intelligence*, pages 434–439, 2007.

Y. Nishiyama and S. Watanabe. Asymptotic behavior of free energy of general boltzmann machines in mean field approximation. *Technical report of IEICE, NC2006*, 38:1–6, 2006.

J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Trans. on Information Theory*, 30(4):629–636, 1984.

J. Rissanen. Stochastic complexity and modeling. *Annals of Statistics*, 14:1080–1100, 1986.

D. Rusakov and D. Geiger. Asymptotic model selection for naive bayesian networks. *Journal of Machine Learning Research*, 6:1–35, 2005.

R. Salakhutdinov, A. Mnih, and G. E. Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th International Conference on Machine Learning*, pages 791–798. Omni Press, Zoubin Ghahramani, 2007.

G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

H. J. Sussmann. Uniqueness of the weights for minimal feed-forward nets with a given input-output map. *Neural Networks*, 5:589–593, 1992.

S. Takamatsu, S. Nakajima, and S. Watanabe. Generalization error of localized bayes estimation in reduced rank regression. In *Workshop on Information-Based Induction Sciences*, pages 81–86, 2005.

K. Takeuchi. Distribution of an information statistic and the criterion for the optimal model. *Mathematical Science*, 153:12–18, 1976.

S. Watanabe. Algebraic analysis for nonidentifiable learning machines. *Neural Computation*, 13 (4):899–933, 2001a.

S. Watanabe. Algebraic geometrical methods for hierarchical learning machines. *Neural Networks*, 14(8):1049–1060, 2001b.

S. Watanabe. Algebraic geometry of learning machines with singularities and their prior distributions. *Journal of Japanese Society of Artificial Intelligence*, 16(2):308–315, 2001c.

S. Watanabe. *Algebraic Geometry and Statistical Learning Theory*, volume 25. Cambridge Monographs on Applied and Computational Mathematics, 2009.

S. Watanabe, K. Hagiwara, S. Akaho, Y. Motomura, K. Fukumizu, M. Okada, and M. Aoyagi. *Theory and Application of Learning System*. Morikita Press, 2005.

K. Yamanishi. A decision-theoretic extension of stochastic complexity and its applications to learning. *IEEE Trans. on Information Theory*, 44(4):1424–1439, 1998.

K. Yamazaki and S. Watanabe. Singularities in complete bipartite graph-type boltzmann machines and upper bounds of stochastic complexities. *IEEE Trans. on Neural Networks*, 16(2):312–324, 2005.