Resolution of Singularities and the Generalization Error with Bayesian Estimation for Layered Neural Network

Miki Aoyagi *and Sumio Watanabe †

Abstract

Hierarchical learning models such as layered neural networks have singular Fisher metrics, since their parameters are not identifiable. These are called non-regular learning models. The stochastic complexities of non-regular learning models in Bayesian estimation are asymptotically obtained by using poles of their zeta functions which are the integrals of their Kullback distances and their priori probability density functions [1, 2, 3]. However, for several examples, upper bounds of the main terms in asymptotic forms of the stochastic complexities were obtained but not the exact values, because of their computational complexities. In this paper, we show a computational way for obtaining the exact value of the layered neural network and we give the asymptotic form of its stochastic complexity explicitly.

Key Words

*Department of Mathematics, Faculty of Science and Technology, Sophia University, 7–1, Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan, miki-a@sophia.ac.jp

†Precision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama, 226–8503 Japan, swatanab@pi.titech.ac.jp

Stochastic complexity, layered neural networks, non-regular learning models, Bayesian estimate, resolution of singularities

1 Introduction

Learning models such as a layered neural network, a normal mixture and a Boltzmann machine have their singular Fisher matrix functions $I(w)$. Their parameters w are not identifiable. For example, $I(w_0)$ of a three layered neural network is singular ($\det I(w_0) = 0$), when w_0 represents a small model with less hidden units than those of the three layered neural network. The subset which consists of the parameters representing the small model is an analytic variety in all parameter space. Such a learning model is called a non-regular model. Theories of regular statistical models, for example, model selection theories AIC[4], TIC[5], HQ[6], NIC[7], BIC[8], MDL[9] cannot be applied to analyzing such non-regular models. So rapid construction of mathematical theories is necessary, since non-regular models have been applied practically to many information technology fields.

Recently, a close connection between the Bayesian stochastic complexities of non-regular learning models and resolution of singularities has been revealed in [1, 2, 3] as follows. Let n be the number of training samples of a non-regular learning model. Its average stochastic complexity (Its free energy) $F(n)$ is asymptotically equal to

$$F(n) = \lambda \log n - (\ell - 1) \log\log n + O(1),$$

where λ is a positive rational number, ℓ is a natural number and $O(1)$ is a bounded function of n. Its Bayesian generalization error $G(n)$ is the average

Kullback distance of the inference of the non-regular learning model from its true distribution. Since it has been known that $G(n) = F(n+1) - F(n)$ if it has an asymptotic expansion, we have

$$G(n) \cong \frac{\lambda}{n} - \frac{\ell - 1}{n \log n}.$$

The values λ and ℓ are obtained by using the poles of the learning model's zeta function, based on a resolution method of its singularities. The zeta function is defined by the integral of the Kullback distance and the *a priori* probability density function of the learning model. For regular models, $\lambda = d/2$ and $\ell = 1$, where d is the dimension of their parameter space. Non-regular models have smaller value λ than $d/2$, so they are more effective learning models than regular ones in the Bayesian estimation.

In spite of those mathematical foundations, the values λ were not obtained since it was difficult to calculate them. Only their upper bounds were obtained in the case of the three layered neural network [11] and the reduced rank regression [12] in the past. To overcome this difficulty, the paper [10] proposed a probabilistic calculation method for λ, but the method could not obtain the exact values λ, also.

In fact, poles of zeta functions have been investigated well only in special cases, for example in the prehomogeneous spaces, but Kullback distances do not occur in the prehomogeneous spaces. So, to investigate the poles of Kullback distances is a new problem even in mathematics.

Moreover, by Hironaka's Theorem [13], it is known that desingularization of an arbitrary polynomial can be obtained by using a blowing-up process. However desingularization of any polynomial in general, although it is known as a finite process, is very difficult. Furthermore, Kullback distances are not simple polynomials, i.e., they have parameters, for example p which is the number of hidden units of the three layered neural networks in Eq.(3).

In this paper, we propose a new method for obtaining the exact asymptotic form of the stochastic complexity and its main term λ for hieratical learning models. Our method uses a recursive blowing-up, which yields a complete desingularization.

By applying it to the three layered neural network, we show the effectiveness of the method.

Our method in this paper first clarifies the asymptotic behavior of the stochastic complexity in Bayesian estimation for the three layered neural network. So, we can compare asymptotic behaviors of regular models and non-regular models.

One of applications of our result in view of the learning theory is as follows.

By the MCMC method, estimated values of marginal likelihoods were calculated for hyper-parameter estimations and model selection methods of non-regular learning models, but theoretical values were not known. The theoretical values of the marginal likelihoods are given in this paper. This enables us to construct a mathematical foundation for analyzing and developing the precision of the MCMC method.

We explain Bayesian learning theory in section 2 and resolution of singularities in section 3. The main term λ and its order ℓ for the three layered neural network are obtained in section 4.

2 Bayesian learning theory

In this section, we give a framework of Bayesian learning obtained in [1, 2, 3].

Let $\mathbb{R}^N$ be an input-output space and $W \subset \mathbb{R}^d$ a parameter space. Take $x \in \mathbb{R}^N$ and $w \in W$. Let $p(x|w)$ be a learning model and $\psi(w)$ an *a priori* probability density function. Assume that its true probability distribution $p(x|w_0)$ is included in the learning model. Let $X^n = (X_1, X_2, ..., X_n)$ be arbitrary n training samples. X_i's are randomly selected from the true probability distribution $p(x|w_0)$. Then, its *a posteriori* probability density function $p(w|X^n)$ is written by

$$p(w|X^n) = \frac{1}{Z_n}\psi(w)\prod_{i=1}^{n} p(X_i|w),$$

where

$$Z_n = \int_W \psi(w)\prod_{i=1}^{n} p(X_i|w)dw.$$

So the average inference $p(x|X^n)$ of its Bayesian distribution is given by

$$p(x|X^n) = \int p(x|w)p(w|X^n)dw.$$

Then, its generalization error $G(n)$ is written as

$$G(n) = E_n\{\int p(x|w_0)\log\frac{p(x|w_0)}{p(x|X^n)}dx\}, \tag{1}$$

where $E_n\{\cdot\}$ is the expectation value.

Let

$$K_n(w) = 1/n\sum_{i=1}^{n}\log(p(X_i|w_0)/p(X_i|w)).$$

Its average stochastic complexity (the free energy)

$$F(n) = -E_n\{\log\int \exp(-nK_n(w))\psi(w)dw\}, \tag{2}$$

satisfies

$$G(n) = F(n+1) - F(n),$$

if it has an asymptotic expansion.

Define the zeta function $J(z)$ of the learning model by

$$J(z) = \int K(w)^z \psi(w) dw,$$

where $K(w)$ is the Kullback distance of the learning model:

$$K(w) = \int p(x|w_0) \log \frac{p(x|w_0)}{p(x|w)} dx.$$

Then, for the maximum pole $-\lambda$ of $J(z)$ and its order ℓ, we have

$$F(n) = \lambda \log n - (\ell - 1) \log \log n + O(1),$$

and

$$G(n) \cong \frac{\lambda}{n} - \frac{\ell - 1}{n \log n},$$

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

The values λ and ℓ can be calculated by using a blowing-up process.

3 Resolution of singularities

In this section, we introduce resolution of singularities. A blowing-up process is a main tool in resolution of singularities of algebraic varieties.

The following theorem is the analytic version of Hironaka's theorem[13] used by Atiyah[14].

Theorem 1

Let $f(x)$ be a real analytic function in a neighborhood of $0 \in \mathbb{R}^n$. There exist an open set $V \ni 0$, a real analytic manifold U and a proper analytic map μ (proper means that μ's inverse images of compact sets are compact) from U to V such that

(1) $\mu : U \setminus E \to V \setminus f^{-1}(0)$ is an isomorphism, where $E = \mu^{-1}(f^{-1}(0))$,

(2) for each $u \in U$, there is a local analytic coordinate $(u_1, \cdots, u_n)$ 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 distance $K(w)$ 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 Theorem 1 (1) and (2). Then the local integration on V_w of the zeta function $J(z)$ of the learning model is

$$\begin{aligned} J_w(z) &= \int_{V_w} K(w)^z \psi(w) dw \\ &= \int_{U_w} |u_1^{s_1} u_2^{s_2} \cdots u_n^{s_n}|^z \psi(\mu_w(u)) |\mu'_w(u)| du. \end{aligned}$$

Therefore, we can obtain the value $J_w(z)$. 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 $J_w(z) = \int_{V_w} K(w)^z \psi(w) dw$ has no poles. Since the parameter set W is compact, the poles and their orders of $J(z)$ are computable.

Next we explain construction of the blowing-up along a manifold used in this paper.

Define the manifold $\mathcal{M}$ by gluing k open sets $U_i \cong \mathbb{R}^n$, $i = 1, 2, \cdots, k$ $(n \geq k)$ as follows. Denote the coordinate of U_i by $(\xi_{1i}, \cdots, \xi_{ni})$.

Set the equivalence relation

$$(\xi_{1i}, \xi_{2i}, \cdots, \xi_{ni}) \sim (\xi_{1j}, \xi_{2j}, \cdots, \xi_{nj})$$

at $\xi_{ji} \neq 0$ and $\xi_{ij} \neq 0$, by

$$\begin{aligned}
&\xi_{ij} = 1/\xi_{ji},\ \ \xi_{jj} = \xi_{ii}\xi_{ji}, \\
&\xi_{hj} = \xi_{hi}/\xi_{ji}, 1 \leq h \leq k, h \neq i, j, \\
&\xi_{lj} = \xi_{li}, k+1 \leq \ell \leq n.
\end{aligned}$$

Put $M = \coprod_{i=1}^{k} U_i / \sim$.

Then the blowing map $\pi : M \to \mathbb{R}^n$ is defined by

$$(\xi_{1i}, \cdots, \xi_{ni}) \mapsto (\xi_{ii}\xi_{1i}, \cdots, \xi_{ii}\xi_{i-1i}, \xi_{ii}, \xi_{ii}\xi_{i+1i}, \cdots, \xi_{ii}\xi_{ki}, \xi_{k+1i}, \cdots, \xi_{ni}),$$

for each $(\xi_{1i}, \cdots, \xi_{ni}) \in U_i$.

This map is well-defined and is called the blowing-up along

$N = \{(x_1, \cdots, x_k, x_{k+1}, \cdots, x_n) \in \mathbb{R}^n \mid x_1 = \cdots = x_k = 0\}$,

The blowing map satisfies

(1) $\pi : M \to \mathbb{R}^n$ is proper,

(2) $\pi : M \setminus \pi^{-1}(N) \to \mathbb{R}^n \setminus N$ is an isomorphism.

4 The learning curve of a three layered neural network

In this section, we obtain the maximum pole of the zeta function of a three layered neural network, by using a blowing-up process.

Consider the three layered neural network of one input unit, p hidden units and one output unit. Denote an input value by x, and an output values by y.

Let

$$f(x, w) = \sum_{m=1}^{p} a_m \tanh(b_m x),$$

where $w = \{a_m, b_m | m = 1, \cdots, p\}$ is the parameter vector.

Consider the statistical model

$$p(y|x,w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(y - f(x,w))^2).$$

Assume that the probability density function $q(x)$ of the input x is the uniform distribution on the interval $[-1,1]$ and that the a priori probability density function $\psi(w)$ of w is a $C^\infty-$ function with compact support W, satisfying $\psi(0) > 0$.

Let the true distribution be $p(y|x,w_0) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)$. That is, the true parameter set which gives the true distribution contains the case of $a_1 = \cdots = a_p = b_1 = \cdots = b_p = 0$.

Then the Kullback distance is

$$K(w) = \frac{1}{2}\int_{-1}^{1} f(x,w)^2 \mathrm{d}x.$$

By using Taylor expansion of $K(w)$, the maximum pole λ and its order ℓ of $\int_W K(w)^z \psi \mathrm{d}w$ are equal to those of

$$\int_W \{\sum_{n=1}^{p}(\sum_{m=1}^{p} a_m b_m^{2n-1})^2\}^z \prod_{m=1}^{p} \mathrm{d}a_m \mathrm{d}b_m. \tag{3}$$

In the paper [15], it is shown that

$$\sum_{m=1}^{p} \frac{1}{4m-2} \leq \lambda \leq \frac{\sqrt{p}}{2}.$$

Let Ψ be the differential form such as

$$\Psi = \{\sum_{n=1}^{p}(\sum_{m=1}^{p} a_m b_m^{2n-1})^2\}^z \prod_{m=1}^{p} \mathrm{d}a_m \mathrm{d}b_m.$$

Main Theorem

We have

$$\lambda = \frac{p+i^2+i}{4i+2},\ \ell = \begin{cases} 2, & i^2 = p, \\ 1, & i^2 < p, \end{cases}$$

where i is the maximum integer satisfying $i^2 \leq p$.

We prove Main Theorem by using a blowing-up process of Ψ.

Before the proof of Main Theorem, let us give some notation.

Since we often change variables by using a blowing-up process, it is more convenient for us to use the same symbols b_m rather than b'_m, b''_m, $\cdots$, etc, for the sake of simplicity. For instance,

"Let $b_1 = v_1, b_m = v_1 b_m,$

instead of

"Let $b_1 = v_1, b_m = v_1 b'_m.$

We divide the proof into two parts. One part is for obtaining desingularization and the other is for comparing the absolute values of poles.

Proof of Main Theorem: Part 1

Construct the blowing-up along the submanifold $\{b_1 = 0, \cdots, b_p = 0\}$.

Let M be the manifold $M = \coprod_{i=1}^p U_i / \sim$. Set the coordinate on U_1 by $(a_1, \cdots, a_p, v_1, b_2, \cdots, b_p)$. After the blowing-up by the transform $\{b_1 = v_1, b_m = v_1 b_m,\ m = 2, \cdots, p\}$ on U_1, we have

$$\begin{aligned}\Psi &= \{\sum_{n=1}^{p} v_1^{4n-2}(a_1 + \sum_{m=2}^{p} a_m b_m^{2n-1})^2\}^z v_1^{p-1} \\ &\qquad\qquad\qquad\qquad \mathrm{d}a_1 \mathrm{d}v_1 \prod_{m=2}^{p} \mathrm{d}a_m \mathrm{d}b_m.\end{aligned}$$

By the symmetry of b_1, ..., b_p, this setting is considered as the general case $\{b_i = v_i, b_m = v_i b_m,\ m = 1, 2, \cdots, p, m \neq i\}$.

Let $d_1 = a_1 + \sum_{m=2}^{p} a_m b_m$. Then we have

$$\begin{aligned}\Psi = \{&v_1^2(d_1^2 + \sum_{n=2}^{p} v_1^{4n-4}\\ &(d_1 - \sum_{m=2}^{p} a_m b_m + \sum_{m=2}^{p} a_m b_m^{2n-1})^2)\}^z\\ &v_1^{p-1} \mathrm{d}d_1 \mathrm{d}v_1 \prod_{m=2}^{p} \mathrm{d}a_m \mathrm{d}b_m.\end{aligned}$$

Put the auxiliary function $f_{n,l}$ by

$$f_{n,l}(x) = \sum_{j_2+\cdots+j_l=0}^{n-l} b_2^{2j_2} \cdots b_{l-1}^{2j_{l-1}} x^{2j_l} = 1 + \cdots \geq 1.$$

This function satisfies

$$f_{n,l}(b_m) - f_{n,l}(b_l) = (b_m^2 - b_l^2) f_{n,l+1}(b_m).$$

Set

$$c_2 = \sum_{m=i}^{p} a_m b_m (b_m^2 - 1),$$

and

$$c_i = \sum_{m=i}^{p} a_m b_m (b_m^2 - 1)(b_m^2 - b_2^2) \cdots (b_m^2 - b_{i-1}^2),$$

for $i \geq 3$. By using $f_{n,l}$, we have

$$\begin{aligned}\Psi = \{&v_1^2(d_1^2 + \sum_{n=2}^{p} v_1^{4n-4}(d_1 + f_{n,2}(b_2)c_2 + \cdots\\ &+ f_{n,i}(b_i)c_i + \cdots + f_{n,n}(b_n)c_n)^2)\}^z\\ &v_1^{p-1} \mathrm{d}d_1 \mathrm{d}v_1 \prod_{m=2}^{p} \mathrm{d}a_m \mathrm{d}b_m. \qquad (4)\end{aligned}$$

We have the condition $d_1^2 + \sum_{n=2}^{p} v_1^{4n-4}(d_1 + f_{n,2}(b_2)c_2 + \cdots + f_{n,i}(b_i)c_i + \cdots + f_{n,n}(b_n)c_n)^2 = 0$ if and only if $d_1 = c_2 = \cdots = c_p = 0$.

For the sake of simplicity, we use an abbreviation $f_{n,i}$ instead of $f_{n,i}(b_i)$.

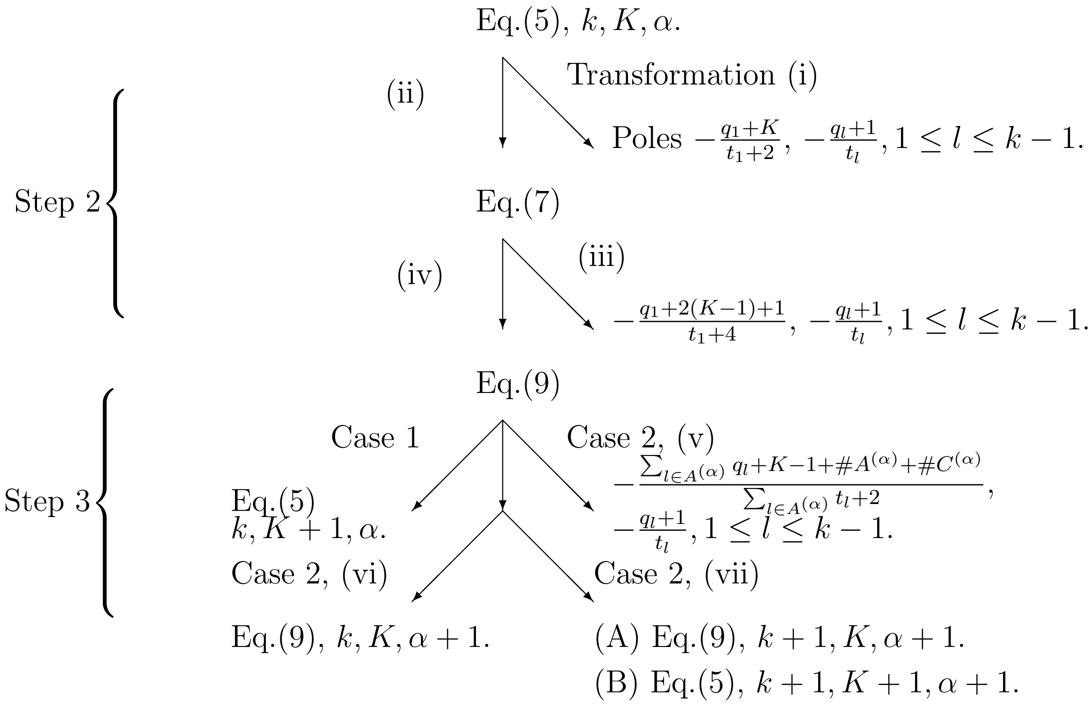

Figure 1: The flow chart of the Main Theorem's proof.

Take $J^{(\alpha)}$ or $J \in \mathbb{R}^\alpha$. Denote $J^{(\alpha)} = (J^{(\alpha')}, *)$ and $\alpha \ge \alpha'$ by $J^{(\alpha)} > J^{(\alpha')}$. Also denote $J^{(\alpha)} = (0, \cdots, 0)$ by $J^{(\alpha)} = 0^{(\alpha)}$ or $J^{(\alpha)} = 0$.

We need the following inductive statements of k, K, α for calculating poles by using a blowing-up process.

Figure 1 is the flow chart of the Main Theorem's proof.

Inductive statements

Set $s(J) = \#\{m; k \le m \le p, J_m^{(\alpha)} = J\}$,
$s(i, J) = \#\{m; k \le m \le i-1, J_m^{(\alpha)} = J\}$
for any $J \in \mathbb{R}^\alpha$, where $\#$ implies the number of elements.

(a) $K \ge k$,

(b)

$$
\begin{aligned}
&\Psi = \{v_1^{t_1} v_2^{t_2} v_3^{t_3} \cdots v_{k-1}^{t_{k-1}} \Big(d_1^2 + (d_1 v_1^2 + d_2)^2 \\
&+ \cdots + (d_1 v_1^{2K-4} + d_2 v_1^{2K-6} f_{K-1,2} + \cdots \\
&+ d_{K-1} f_{K-1,K-1})^2 + \sum_{n=K}^{p} v_1^{4n-4K+4} (d_1 v_1^{2K-4} \\
&+ d_2 v_1^{2K-6} f_{n,2} + \cdots + d_{K-1} f_{n,K-1} \\
&+ \sum_{i=K}^{p} f_{n,i} c_i)^2 \Big) \}^z \prod_{m=1}^{k-1} v_m^{q_m} \\
&\prod_{m=1}^{K-1} \mathrm{d}d_m \prod_{m=K}^{p} \mathrm{d}a_m \prod_{m=1}^{k-1} \mathrm{d}v_m \prod_{m=k}^{p} \mathrm{d}b_m.
\end{aligned} \tag{5}
$$

Here, t_l, $q_m \in \mathbb{Z}_+$. Also, there exist $RJ^{(\alpha)} \subset \mathbb{R}^\alpha$, $t(i, J, l) \in \mathbb{Z}_+$, and functions $g(i, m) \neq 0$, $(K \leq i \leq p, 2 \leq l \leq k-1, i \leq m \leq p)$ such that

$$
\begin{aligned}
c_i &= v_2^{t(i,0,2)} v_3^{t(i,0,3)} \cdots v_{k-1}^{t(i,0,k-1)} \\
&\sum_{\substack{i \leq m \leq p \\ J_m^{(\alpha)} = 0}} g(i,m) a_m b_m \prod_{\substack{k \leq i' < i \\ J_{i'}^{(\alpha)} = 0}} (b_m^2 - b_{i'}^2) \\
&+ \sum_{J \in RJ^{(\alpha)}} v_2^{t(i,J,2)} v_3^{t(i,J,3)} \cdots v_{k-1}^{t(i,J,k-1)} \\
&\sum_{\substack{i \leq m \leq p \\ J_m^{(\alpha)} = J}} g(i,m) a_m b_m \prod_{\substack{k \leq i' < i \\ J_{i'}^{(\alpha)} = J}} (b_m - b_{i'}) \\
&+ \sum_{J \notin RJ^{(\alpha)}, J \neq 0} v_2^{t(i,J,2)} v_3^{t(i,J,3)} \cdots v_{k-1}^{t(i,J,k-1)} \\
&\sum_{\substack{i \leq m \leq p \\ J_m^{(\alpha)} = J}} g(i,m) a_m \prod_{\substack{k \leq i' < i \\ J_{i'}^{(\alpha)} = J}} (b_m - b_{i'}).
\end{aligned}
$$

(c) $J_{i'}^{(\alpha)} \neq J_i^{(\alpha)}$ for $k \leq i' < i < K$. $J_i^{(\alpha)} \notin RJ^{(\alpha)} \cup \{0\}$ for $k \leq i < K$.

(d) For $J \in \mathbb{R}^\alpha$, $K \le i \le p, 2 \le l \le k-1$, set $\tilde{t}(i,J,l) = t_l/2 + t(i,J,l)$. There exist $D_{J^{(\mu)},l} \in \mathbb{Z}_+$ such that

$$\begin{aligned}\tilde{t}(i,J,l) &= \sum_{J>0^{(\mu)}} D_{0^{(\mu)},l}(2s(i,0^{(\mu)})+1) \\ &+ \sum_{\substack{J>J^{(\mu)} \\ J^{(\mu)} \in RJ^{(\mu)}}} D_{J^{(\mu)},l}(s(i,J^{(\mu)})+1) \\ &+ \sum_{\substack{J>J^{(\mu)} \\ J^{(\mu)} \notin RJ^{(\mu)}, J^{(\mu)} \ne 0}} D_{J^{(\mu)},l} s(i,J^{(\mu)}).\end{aligned}$$

(e) There exist $g_l \ge 0$, $\eta^{(\xi)}_{k',l} \ge 0$ for $2 \le k' \le K-1, 1 \le \xi \le g_l, 2 \le l \le k-1$, such that

$$\begin{aligned}&\frac{t_l}{2} = \sum_{\xi=1}^{g_l}(1+\eta^{(\xi)}_{2,l} + \cdots + \eta^{(\xi)}_{K-1,l}), \\ &0 \le \eta^{(\xi)}_{2,l} \le 2, 0 \le \eta^{(\xi)}_{2,l} + \eta^{(\xi)}_{3,l} \le 4, \\ &\vdots \\ &0 \le \eta^{(\xi)}_{2,l} + \eta^{(\xi)}_{3,l} + \cdots + \eta^{(\xi)}_{K-1,l} \le 2(K-2).\end{aligned}$$

(f) Set $\varphi^{(\xi)}_l = p + 2\eta^{(\xi)}_{2,l} + \cdots + (K-1)\eta^{(\xi)}_{K-1,l}$. There exist $\phi_l \in \mathbb{Z}_+$ for $2 \le l \le k-1$, such that $g_l \le \sum_{J^{(\alpha)}_m > J^{(\mu)}} D_{J^{(\mu)},l}$ and

$$\begin{aligned}q_l + 1 &= \sum_{\xi=1}^{g_l} \varphi^{(\xi)}_l + \phi_l \\ &\quad + \sum_{m=k}^{p}(-g_l + \sum_{J^{(\alpha)}_m > J^{(\mu)}} D_{J^{(\mu)},l}).\end{aligned}$$

The end of inductive statements

Statements (d), (e) and (f) are needed to compare poles. The definitions of all variables will be given later on in the proof.

If $J^{(\alpha)}_m = 0$ for all m and α, we have $\alpha = k-1$ and the followings:

(a') $k = K$.

(b')

$$c_i = v_2^{t(i,0,2)} v_3^{t(i,0,3)} \cdots v_{k-1}^{t(i,0,k-1)}$$
$$\sum_{i \leq m \leq p} a_m b_m \prod_{k \leq i' < i} (b_m^2 - b_{i'}^2).$$

(d') $D_{0^{(l-1)},l} = 1$ and the others $D_{J,l}$ are 0.

$$\begin{aligned} \tilde{t}(i, 0^{(k-1)}, l) &= D_{0^{(l-1)},l}(2(i-l)+1) \\ &= 2(i-l)+1. \end{aligned}$$

(e') $\frac{t_l}{2} = 1 + \eta_{2,l}^{(1)} + \cdots + \eta_{k-1,l}^{(1)}$,

$$\begin{aligned} &\eta_{2,l}^{(1)} = 0, \ldots, \eta_{l-1,l}^{(1)} = 0, \\ &\eta_{l,l}^{(1)} = 2, \ldots, \eta_{k-2,l}^{(1)} = 2, \\ &0 \leq \eta_{k-1,l}^{(1)} \leq 2, t(k, 0^{(k-1)}, l) + \eta_{k-1,l}^{(1)} = 2. \end{aligned}$$

(f') Set $\varphi_l^{(1)} = p + 2\eta_{2,l}^{(1)} + \cdots + (k-1)\eta_{k-1,l}^{(1)}$. Then $q_l + 1 = \varphi_l^{(1)}$.

4.1 Step 1

Set $k = K = 2$. For any numbers $j_2^{(1)}, \cdots, j_p^{(1)} \in \mathbb{R}$, take a neighborhood V of $d_1 = 0, b_m = j_m^{(1)}, m = 2, \cdots, p$ such that we have $|b_m| \neq |b_{m'}|$ if $|j_m^{(1)}| \neq |j_{m'}^{(1)}|$ and we have $|b_m| \neq 0, 1$, if $|j_m^{(1)}| \neq 0, 1$.

Assume that $d_1, b_m \in V$. Then we have

$$c_i = \sum_{\substack{i \leq m \leq p, \\ j_m^{(1)} = 0}} g(i,m) a_m b_m \prod_{\substack{2 \leq i' < i, \\ j_{i'}^{(1)} = 0}} (b_m^2 - b_{i'}^2)$$

$$+\sum_{\substack{i\le m\le p,\\ |j^{(1)}_m|=1}} g(i,m)a_m b_m \prod_{\substack{2\le i'<i,\\ |j^{(1)}_{i'}|=1}} (b_m-b_{i'})$$

$$+\sum_{\substack{i\le m\le p,\\ |j^{(1)}_m|\neq 0,1}} g(i,m)a_m \prod_{\substack{2\le i'<i,\\ |j^{(1)}_{i'}|=|j^{(1)}_m|}} (b_m-b_{i'}),$$

with the functions $g(i,m)\neq 0$.

Let $\begin{cases} \alpha=1, & J^{(1)}_m=((j^{(1)}_m)^2),\\ t_1=2, & q_1=p-1,\\ RJ^{(1)}=\{J=((j^{(1)}_m)^2);|j^{(1)}_m|=1\}. \end{cases}$

$RJ^{(1)}$ is the set of J as

$$\sum_{\substack{i\le m\le p\\ J^{(1)}_m=J}} g(i,m)a_m b_m \prod_{\substack{2\le i'<i\\ J^{(1)}_{i'}=J}} (b_m-b_{i'}),$$

in the formulas c_i.

Those new parameters defined in Step 1 satisfy Statements (a)$\sim$(c)

Set the other parameters such as $t(i,J,l)$ and q_l by 0, since these parameters do not appear.

Then we have Statements (d)$\sim$(f).

4.2 Step 2

We assume the case k,K.

Construct the blowing-up of (5) along the submanifold $\{v_1=d_1=\cdots=d_{K-1}=0\}$.

(i) Let $\{d_1=u_1, d_l=u_1d_l, 2\le l\le K-1, v_1=u_1v_1\}$ in (5). Then we have the poles

$$-\frac{q_1+K}{t_1+2},$$
$$-\frac{q_l+1}{t_l}, 1\le l\le k-1. \tag{6}$$

Another blowing-up is not necessary in this neighborhood, anymore.

(ii) Let $\{d_l = u_l d_l, 1 \le l \le K-1, v_1 = v_1\}$ in (5) and we have

$$\begin{aligned}\Psi = \{&v_1^{t_1+2} v_2^{t_2} v_3^{t_3} \cdots v_{k-1}^{t_{k-1}} \Big(d_1^2 + (d_1 v_1^2 + d_2)^2 \\ &+ \cdots + (d_1 v_1^{2K-4} + d_2 v_1^{2K-6} f_{K-1,2} + \cdots \\ &+ d_{K-1} f_{K-1,K-1})^2 + \sum_{n=K}^{p} v_1^{4n-4K+2} (d_1 v_1^{2K-3} \\ &+ d_2 v_1^{2K-5} f_{n,2} + \cdots + d_{K-1} f_{n,K-1} \\ &+ \sum_{i=K}^{p} f_{n,i} c_i)^2 \Big) \}^z v_1^{q_1+K-1} \prod_{m=1}^{k-1} v_m^{q_m} \\ &\prod_{m=1}^{K-1} \mathrm{d}d_m \prod_{m=K}^{p} \mathrm{d}a_m \prod_{m=1}^{k-1} \mathrm{d}v_m \prod_{m=k}^{p} \mathrm{d}b_m. \qquad (7)\end{aligned}$$

In addition, let us construct the blowing-up of (7) along the submanifold $\{v_1 = d_1 = \cdots = d_{K-1} = 0\}$.

(iii) Let $\{d_1 = u_1, d_l = u_l d_l, 2 \le l \le K-1, v_1 = u_1 v_1\}$ in (7). Then we have the poles

$$\begin{aligned}&-\frac{q_1 + 2(K-1) + 1}{t_1 + 4}, \\ &-\frac{q_l + 1}{t_l}, 1 \le l \le k-1. \qquad (8)\end{aligned}$$

(iv) Let $\{d_l = u_l d_l, 1 \le l \le K-1, v_1 = v_1\}$ in (7) and we have

$$\begin{aligned}\Psi = \{&v_1^{t_1+4} v_2^{t_2} v_3^{t_3} \cdots v_{k-1}^{t_{k-1}} \Big(d_1^2 + (d_1 v_1^2 + d_2)^2 + \cdots \\ &+ (d_1 v_1^{2K-4} + d_2 v_1^{2K-6} f_{K-1,2} + \cdots + d_{K-1})^2 \\ &+ \sum_{n=K}^{p} v_1^{4n-4K} (d_1 v_1^{2K-2} + d_2 v_1^{2K-4} f_{n,2} + \cdots + \\ &d_{K-1} v_1^2 f_{n,K-1} + \sum_{i=K}^{p} f_{n,i} c_i)^2 \Big) \}^z v_1^{q_1+2(K-1)} \\ &\prod_{m=1}^{k-1} v_m^{q_m} \prod_{m=1}^{K-1} \mathrm{d}d_m \prod_{m=K}^{p} \mathrm{d}a_m \prod_{m=1}^{k-1} \mathrm{d}v_m \prod_{m=k}^{p} \mathrm{d}b_m. \qquad (9)\end{aligned}$$

4.3 Step 3

Let us concentrate on c_K. Put

$$\begin{aligned}
JB^{(\alpha)} &= \{J \in \mathbb{R}^{\alpha}; \exists l, t(K,J,l) > 0\}, \\
\Omega &= \{A' \subset \{2 \le l \le k-1\}; \forall J \in JB^{(\alpha)}, \\
&\quad \exists l \in A' \text{ s.t. } t(K,J,l) > 0\} \\
C^{(\alpha)} &= \{m \ge k; t(K, J_m^{(\alpha)}, l) = 0 \text{ for all } l\}, \\
JC^{(\alpha)} &= \{J \in \mathbb{R}^{\alpha}; t(K,J,l) = 0 \text{ for all } l\}.
\end{aligned}$$

Fix $A^{(\alpha)} \in \Omega$ whose number of elements is minimum in Ω:

$$\#A^{(\alpha)} = \min_{A \in \Omega} \#A.$$

Let

$$T_{i,J} = \begin{cases} \sum_{l \in A^{(\alpha)}} t(i,J,l) + 2s(i,J), \\ \quad \text{if } J = 0, J \in JC^{(\alpha)}, \\ \sum_{l \in A^{(\alpha)}} t(i,J,l) + s(i,J), \\ \quad \text{if } J \in RJ^{(\alpha)} \cap JC^{(\alpha)}, \\ \sum_{l \in A^{(\alpha)}} t(i,J,l) + s(i,J) - 1, \\ \quad \text{if } J \in JC^{(\alpha)} \setminus (RJ^{(\alpha)} \cup \{0\}), \\ \sum_{l \in A^{(\alpha)}} t(i,J,l) - 1, \\ \quad \text{otherwise,} \end{cases} \tag{10}$$

$$T = \sum_{l \in A^{(\alpha)}} t_l + 2, \tag{11}$$

$$\begin{aligned} Q = \sum_{l \in A^{(\alpha)}} q_l \\ +K - 1 + \#A^{(\alpha)} + \#C^{(\alpha)} - 1. \end{aligned} \tag{12}$$

In addition, let $C_*^{(\alpha)} = \{m \in C^{(\alpha)} \mid J_m^{(\alpha)} \notin RJ^{(\alpha)}, J_m^{(\alpha)} \neq 0, J_m^{(\alpha)} \neq J_{i'}^{(\alpha)}$ for all $k \le i' < K\}$.

Case 1 $C_*^{(\alpha)} \neq \phi$.

4.3.1 Transformation in Case 1

We can assume that $J_K^{(\alpha)} \notin RJ^{(\alpha)}, J_K^{(\alpha)} \neq 0$ and $K \in C^{(\alpha)}$. Then we obtain $c_K = a_K g(K,K) + \cdots$. Change the variable from a_K to d_K by $d_K = c_K$.

Set

$K \to K+1, t_1 \to t_1 + 4, q_1 \to q_1 + 2(K-1),$

and we have the inductive statements of $K \to K+1$.

If all $J_m^{(\alpha)} = 0$, Case 1 does not appear.

Case 2 $C_*^{(\alpha)} = \phi$.

Construct the blowing-up of (9) along the submanifold $\{d_1 = \cdots = d_{K-1} = v_{k"} = b_m = 0, k" \in A^{(\alpha)}, m \in C^{(\alpha)}\}$.

4.3.2 Transformation(v)

(v) Let $\{d_1 = u_1, d_l = u_l d_l, 2 \leq l \leq K-1, v_{k"} = u_1 v_{k"}, k" \in A^{(\alpha)}, b_m = u_1 b_m, m \in C^{(\alpha)}\}$ in (9). Then we have the poles

$$-\frac{\sum_{l\in A^{(\alpha)}} q_l + K - 1 + \#A^{(\alpha)} + \#C^{(\alpha)}}{\sum_{l\in A^{(\alpha)}} t_l + 2}, \tag{13}$$

$$-\frac{q_l+1}{t_l}, 1 \leq l \leq k-1. \tag{14}$$

If all $J_m^{(\alpha)} = 0$, then we have $\#A^{(\alpha)} \neq \phi$ and the poles

$$-\frac{q_{k"}+k}{t_{k"}+2}, -\frac{q_l+1}{t_l}, 1 \leq l \leq k-1,$$

are obtained, since $\#A^{(\alpha)} = 1$ and $\#C^{(\alpha)} = 0$.

4.3.3 Transformation(vi)

Fix $k' \in A^{(\alpha)}$.

(vi) Let $\{d_l = v_{k'} d_l, 1 \le l \le K-1, v_{k'} = v_{k'}, v_{k"} = v_{k'} v_{k"}, k" \in A^{(\alpha)} - \{k'\}, b_m = v_{k'} b_m, m \in C^{(\alpha)}\}$ in (9).

By using Eq.(10), (11), (12), set

$t_{k'} \to T, t(i, J, k') \to T_{i,J}, q_{k'} \to Q, c_i \to c_i / v_{k'}$.

Then, $t_{k'}/2 = (\sum_{l \in A^{(\alpha)}} t_l + 2)/2 = \sum_{l \in A^{(\alpha)}} \sum_{\xi=1}^{g_l} (1 + \eta_{2,l}^{(\xi)} + \cdots + \eta_{K-1,l}^{(\xi)}) + 1$, and

$$\begin{aligned}
q_{k'} + 1 &= \sum_{l \in A^{(\alpha)}} (q_l + 1) + K - 1 + \# C^{(\alpha)} \\
&= \sum_{l \in A^{(\alpha)}} \left\{ \sum_{\xi=1}^{g_l} \varphi_l^{(\xi)} + \sum_{m=k}^{p} (-g_l + \right. \\
&\left. \sum_{J_m^{(\alpha)} > J^{(\mu)}} D_{J^{(\mu)}, l}) + \phi_l \right\} + K - 1 \\
&+ \sum_{J_m^{(\alpha)} \in JC^{(\alpha)}} 1 \\
&= \sum_{l \in A^{(\alpha)}} \sum_{\xi=1}^{g_l} \varphi_l^{(\xi)} + \sum_{\substack{m=k \\ J_m^{(\alpha)} \notin JC^{(\alpha)}}}^{p} (- \sum_{l \in A^{(\alpha)}} g_l \\
&+ \sum_{J_m^{(\alpha)} > J^{(\mu)}} \sum_{l \in A^{(\alpha)}} D_{J^{(\mu)}, l}) + \sum_{\substack{m=k \\ J_m^{(\alpha)} \in JC^{(\alpha)}}}^{p} \\
&(- \sum_{l \in A^{(\alpha)}} g_l + (\sum_{J_m^{(\alpha)} > J^{(\mu)}} \sum_{l \in A^{(\alpha)}} D_{J^{(\mu)}, l}) + 1) \\
&+ \sum_{l \in A^{(\alpha)}} \phi_l + K - 1.
\end{aligned}$$

(I) Assume that there exist $l' \in A^{(\alpha)}$ and ξ' such that $0 \le \eta_{2,l'}^{(\xi')} + \eta_{3,l'}^{(\xi')} + \cdots + \eta_{K-1,l'}^{(\xi')} < 2(K-2)$. Let

$$g_{k'} \to \sum_{l \in A^{(\alpha)}} g_l, \quad \phi_{k'} \to \sum_{l \in A^{(\alpha)}} \phi_l,$$

$$D_{J^{(\alpha)},k'} \to \sum_{l\in A^{(\alpha)}} D_{J^{(\alpha)},l} + 1 \text{ if } J^{(\alpha)} \in JC^{(\alpha)},$$

$$D_{J^{(\mu)},k'} \to \sum_{l\in A^{(\alpha)}} D_{J^{(\mu)},l} \text{ if } J^{(\mu)} \notin JC^{(\alpha)}.$$

Let

$$\varphi_{k'}^{(1)}, \cdots, \varphi_{k'}^{(g_{k'})}$$

be $\varphi_l^{(\xi)}$, $l \in A^{(\alpha)}$, $\xi = 1 \cdots, g_l$ and

$$\eta_{\ell,k'}^{(1)}, \cdots, \eta_{\ell,k'}^{(g_{k'})}$$

be $\eta_{\ell,l}^{(\xi)}$, $l \in A^{(\alpha)}$, $\xi = 1 \cdots, g_l$ by numbering in the same order for any ℓ.

Then we have

$$\begin{aligned} t_{k'}/2 &= \sum_{\xi=1}^{g_{k'}} (1 + \eta_{2,k'}^{(\xi)} + \cdots + \eta_{K-1,k'}^{(\xi)}) + 1, \\ q_{k'} + 1 &= \sum_{\xi=1}^{g_{k'}} \varphi_{k'}^{(\xi)} + K - 1 \\ &+ \sum_{m=k}^{p} (-g_{k'} + \sum_{J_m^{(\alpha)} > J^{(\mu)}} D_{J^{(\mu)},k'}) + \phi_{k'}. \end{aligned}$$

By the assumption (I), there exists ξ' such that $0 \le \eta_{2,k'}^{(\xi')} + \eta_{3,k'}^{(\xi')} + \cdots + \eta_{K-1,k'}^{(\xi')} < 2(K-2)$. Therefore, by putting

$$\begin{aligned} &\varphi_{k'}^{(\xi')} \to p + 2\eta_{2,k'}^{(\xi')} + \cdots + (K-1)(\eta_{K-1,k'}^{(\xi')} + 1), \\ &\eta_{K-1,k'}^{(\xi')} \to \eta_{K-1,k'}^{(\xi')} + 1, \end{aligned}$$

we have Statements (d), (f). **The end of (I)**

(II) Assume $0 \le \eta_{2,l}^{(\xi)} + \eta_{3,l}^{(\xi)} + \cdots + \eta_{K-1,l}^{(\xi)} = 2(K-2)$ for all $l \in A^{(\alpha)}$ and ξ.

By the assumption (d), for any $J \in \mathbb{R}^{\alpha}$, we have

$$
\begin{aligned}
\frac{t_l}{2} &= \sum_{\xi=1}^{g_l}(1+\eta_{2,l}^{(\xi)}+\cdots+\eta_{K-1,l}^{(\xi)}) \\
&= g_l(1+2(K-2)) \\
&\leq \sum_{J>0^{(\mu)}} D_{0^{(\mu)},l}(2s(i,0^{(\mu)})+1) \\
&\quad + \sum_{\substack{J>J^{(\mu)} \\ J^{(\mu)}\in RJ^{(\mu)}}} D_{J^{(\mu)},l}(s(i,J^{(\mu)})+1) \\
&\quad + \sum_{\substack{J>J^{(\mu)} \\ J^{(\mu)}\notin RJ^{(\mu)},J^{(\mu)}\neq 0}} D_{J^{(\mu)},l}s(i,J^{(\mu)}).
\end{aligned}
$$

In particular, if some $J^{(\mu)}$ are not equal to 0 then $s(K,J^{(\mu)}) \leq K-2$. Therefore we have

$$
g_l < \sum_{J>J^{(\mu)}} D_{J^{(\mu)},l}. \tag{15}
$$

Let

$$
D_{J^{(\alpha)},k'} \to \sum_{l\in A^{(\alpha)}} D_{J^{(\alpha)},l}+1 \text{ if } J^{(\alpha)} \in JC^{(\alpha)},
$$

$$
D_{J^{(\mu)},k'} \to \sum_{l\in A^{(\alpha)}} D_{J^{(\mu)},l} \text{ if } J^{(\mu)} \notin JC^{(\alpha)},
$$

$$
g_{k'} \to \sum_{l\in A^{(\alpha)}} g_l+1,\ \phi_{k'} \to \sum_{l\in A^{(\alpha)}} \phi_l+K-k,
$$

and $\varphi_{k'}^{(g_{k'})}=p$. In addition, let

$$
\varphi_{k'}^{(1)},\cdots,\varphi_{k'}^{(g_{k'}-1)}
$$

be $\varphi_l^{(\xi)}$, $l \in A^{(\alpha)}$, $\xi = 1\cdots,g_l$, and

$$
\eta_{\ell,k'}^{(1)},\cdots,\eta_{\ell,k'}^{(g_{k'}-1)}
$$

be $\eta^{(\xi)}_{\ell,l}$, $l \in A^{(\alpha)}$, $\xi = 1 \cdots, g_l$ by numbering in the same order for any ℓ. Put $\eta^{(g_{k'})}_{\ell,k'} = 0$ for $\ell \geq 1$. Then we have

$$\begin{aligned}
q_{k'} + 1 = & \sum_{l \in A^{(\alpha)}} \sum_{\xi=1}^{g_l} \varphi^{(\xi)}_l + p + \sum_{m=k}^{p} (- \sum_{l \in A^{(\alpha)}} g_l \\
& -1 + \sum_{J^{(\alpha)}_m > J^{(\mu)}} \sum_{l \in A^{(\alpha)}} D_{J^{(\mu)},l}) \\
& + \sum_{l \in A^{(\alpha)}} \phi_l + p - k + 1 - p + K - 1 + \#C^{(\alpha)} \\
= & \sum_{\xi=1}^{g_{k'}} \varphi^{(\xi)}_{k'} + \sum_{m=k}^{p} (-g_{k'} + \sum_{J^{(\alpha)}_m > J^{(\mu)}} D_{J^{(\mu)},k'}) + \phi_{k'}.
\end{aligned}$$

By Eq.(15) we have $g_{k'} \leq \sum_{J^{(\alpha)}_m > J^{(\mu)}} D_{J^{(\mu)},k'}$.

The end of (II)

In particular, assume that all $J^{(\alpha)}_m = 0$. Let

$$\begin{aligned}
& t_{k'} \rightarrow t_{k'} + 2, t(i, 0, k') \rightarrow t(i, 0, k') - 1, \\
& q_{k'} \rightarrow q_{k'} + k - 1, c_i \rightarrow c_i / v_{k'}.
\end{aligned}$$

Then $t_{k'}/2 = 1 + \eta^{(1)}_{2,k'} + \cdots + \eta^{(1)}_{k-1,k'} + 1$. By putting

$$\eta^{(1)}_{k-1,k'} \rightarrow \eta^{(1)}_{k-1,k'} + 1,$$

we have

$$\begin{aligned}
&(\mathrm{d}') \qquad \tilde{t}(i, 0^{(k-1)}, k') = D_{0^{(k'-1)},k'}(2(i-k')+1), \\
&(\mathrm{e}') \qquad t(k, 0^{(k-1)}, k') + \eta^{(1)}_{k-1,k'} = 2, \\
&(\mathrm{f}') \qquad q_{k'} + 1 = \varphi^{(1)}_{k'} \\
&\qquad\qquad = p + 2\eta^{(1)}_{2,k'} + \cdots + (k-1)\eta^{(1)}_{k-1,k'}.
\end{aligned}$$

If $JC^{(\alpha)}$is not empty, we need the following Step($*$).

Step(*)

Let $j_m^{(\alpha+1)}$ be any real number for each m with $J_m^{(\alpha)} \in JC^{(\alpha)}$. For m with $J_m^{(\alpha)} \notin JC^{(\alpha)}$, let $j_m^{(\alpha+1)} = j_m^{(\alpha)}$ where $J_m^{(\alpha)} = (J', j_m^{(\alpha)})$, $J' \in \mathbb{R}^{\alpha-1}$ and $j_m^{(\alpha)} \in \mathbb{R}$. Consider a sufficiently small neighborhood of $b_m = j_m^{(\alpha+1)}$ and fix it. Put

$$J_m^{(\alpha+1)} = \begin{cases} (J_m^{(\alpha)}, {j_m^{(\alpha+1)}}^2) & \text{if } J_m^{(\alpha)} = 0, \\ (J_m^{(\alpha)}, j_m^{(\alpha+1)}) & \text{if } J_m^{(\alpha)} \neq 0. \end{cases}$$

Let $t(i, J, (l, \tau)) = t(i, J', (l, \tau))$ for $J = (J', *) \in \mathbb{R}^{\alpha+1}$, where $J' \in \mathbb{R}^{\alpha}$. Change $g(i, m) \neq 0$ and b_m properly, taking into account that the neighborhood of $b_m = j_m^{(\alpha+1)}$. Let $RJ^{(\alpha+1)}$ be the set of J satisfying

$$\sum_{\substack{i \leq m \leq p \\ J_m^{(\alpha+1)} = J}} g(i, m) a_m b_m \prod_{\substack{k+1 \leq i' < i \\ J_{i'}^{(\alpha+1)} = J}} (b_m - b_{i'}),$$

in the formulas c_i. That is,

$$RJ^{(\alpha+1)} = \{J \in \mathbb{R}^{\alpha+1} \mid J = (J', 0), J' \in RJ^{(\alpha)}\}.$$

Let $\alpha \to \alpha + 1$. **The end of Step(*)**

Those new parameters defined in Step 2 satisfy Statements (a)∼(f) with Eq.(9) instead of Eq.(5), since K does not increase.

4.3.4 Transformation(vii)

Assume that $C^{(\alpha)} \neq \phi$. Fix $k' \in C^{(\alpha)}$. (vii) Let $\{d_l = v_k d_l, 1 \leq l \leq K - 1, b_{k'} = v_k, v_{k"} = v_k v_{k"}, k" \in A^{(\alpha)}, b_m = v_k b_m, m \in C^{(\alpha)} - \{k'\}\}$ in Eq.(9).

(A) If $k < k' \leq K$, then we can assume $k' = k + 1$. For $k \leq i \leq p$, let

$$t_k \to T, t(i, J, k) \to T_{i,J}, q_k \to Q, c_i \to c_i / v_k,$$

by using Eq.(10), (11), (12).

(III) If $A^{(\alpha)} \neq \phi$, do the same procedure (I) or (II) by substituting k' into k. Adding it, set $\phi_l \to (-g_l + \sum_{J_k^{(\alpha)} > J^{(\mu)}} D_{J^{(\mu)},l}) + \phi_l$ for all l, since k is replaced by $k+1$ later.

If $A^{(\alpha)} = \phi$, then we have all $J_m^{(\alpha)} \in JC^{(\alpha)}$. Therefore, $\#C = p - (k-1)$, $q_k + 1 = p + K - k$, $t_k = 2$. Put

$$D_{J_m^{(\alpha)},k} = 1, D_{J_m^{(\alpha)},l} = 0 (l \neq k),$$
$$\eta_{\ell,k}^{(1)} = 0, \varphi_k^{(1)} = p, \phi_k = K - k.$$

The end of (III)

Do the procedure Step($*$) and let $k \to k+1$.

Those new parameters satisfy Statements (a)$\sim$(f) with Eq.(9) instead of Eq.(5). If we have all $J_m^{(\alpha)} = 0$, the above case does not appear.

(B)

Next consider the case $K \leq k' < p$. We can assume $k' = K$. If $J_K^{(\alpha)} \notin RJ^{(\alpha)}, J_K^{(\alpha)} \neq 0$ then there exists $i' < K$ such that $J_{i'}^{(\alpha)} = J_K^{(\alpha)}$ because of the Case 2 assumption. So the case results in (A).

Consider the case $\tilde{J} := J_K^{(\alpha)} \in RJ$ or $\tilde{J} := J_K^{(\alpha)} = 0$. By the transformation (vii), we have $c_K = v_k(a_K g(K,K) + \cdots)$. Now, there is no a_K in the formulas c_i, $i \geq K+1$. So, change the variable from a_K to d_K by $d_K = a_K g(K,K) + \cdots : c_K = v_k d_K$. Since a_K, b_K have disappeared in all formulas c_i, we change the variables from $b_k, \cdots, b_{K-1}$ to $b_{k+1}, \cdots, b_K$. Also we change $J_i^{(\alpha)}$ and $RJ^{(\alpha)}$, properly.

Let

$$t_k \to T, t(i,J,k) \to T_{i,J}, q_k \to Q, c_i \to c_i/v_k,$$

for $K+1 \leq i \leq p$, by using Eq.(10), (11), (12).

Proceed Step (III) and Step($*$). Let

$$k \to k+1, K \to K+1,$$
$$t_1 \to t_1+4, q_1 \to q_1+2(K-1),$$

and we have Statements (a)$\sim$(f).

Assume that all $J_m^{(\alpha)}$ are 0. Then we have $A^{(k)}=\phi$. Let

$$t_k=2, t(i,0^{(k-1)},k)=2(i-k), q_k=p-1,$$
$$c_i \to c_i/v_k, D_{0^{(k-1)},k}=1, D_{0^{(k-1)},l}=0 (l \neq k),$$
$$\eta_{\ell,k}^{(1)}=0, \varphi_k^{(1)}=p.$$

By $t(k,0^{(k-1)},l)=0 (2 \leq l \leq k-1)$ and (e'), we have $\eta_{k-1,l}^{(1)}=2$. Also we obtain $t(k+1,0^{(k-1)},l)=\tilde{t}(k+1,0^{(k-1)},l)-t_l/2=2$, by using (d') and (e'). Therefore before Step($*$), we have

$$
\begin{aligned}
&(\mathrm{d}') \qquad \tilde{t}(i,0^{(k-1)},k)=D_{0^{(k-1)},k}(2(i-k)+1)\\
&\qquad\qquad =2(i-k)+1,\\
&(\mathrm{e}') \qquad t(k+1,0^{(k-1)},l)+\eta_{k,k}^{(1)}=2+0,\\
&\qquad\qquad\qquad (2 \leq l \leq k),\\
&(\mathrm{f}') \qquad q_k+1=\varphi_k^{(1)}=p\\
&\qquad\qquad =p+2\eta_{2,k}^{(1)}+\cdots+(k-1)\eta_{k-1,k}^{(1)}.
\end{aligned}
$$

After Step($*$), we have $\tilde{t}(i,0^{(k-1)},k)=\tilde{t}(i,0^{(k)},k)$ and $t(i,0^{(k-1)},k)=t(i,0^{(k)},k)$. Let

$$k \to k+1, K \to K+1,$$
$$t_1 \to t_1+4, q_1 \to q_1+2(K-1).$$

Those new parameters satisfy Statements (a')∼(f').

By repeating Step 3, the conditions Case 2 (vi) and (vii)(A) disappear, whose transformations (vi) and (vii)(A) in Case 2 satisfy Eq.(5) instead of Eq.(9). So, K is increased with these finite steps. $K = p + 1$ completes the blowing-up process.

Proof of Main Theorem: Part 2

To obtain the maximum pole and its order, we prepare the following four lemmas.

Lemma 1

If all $J_m^{(\alpha)} = 0$, then for each $1 \le k \le p$, we have the poles,

$$\begin{array}{c} -\frac{p}{2}, \\ -\frac{p+k}{4}, -\frac{p+2k}{6}, \\ -\frac{p+2k+k+1}{8}, -\frac{p+2k+2(k+1)}{10}, \\ \vdots \\ -\frac{p+(i-1)(2k-2+i)+k+i-1}{4i}, \\ -\frac{p+(i-1)(2k-2+i)+2(k+i-1)}{4i+2}, \\ \vdots \\ -\frac{(p-k-1)(k-2+p)+p-1}{4(p-k)}, \\ -\frac{p+(p-k-1)(k-2+p)+2(p-2)}{4(p-k)+2}. \end{array}$$

which are related to v_k.

Proof The proof is obtained by Statements (a')∼(f') in Part 1. Q.E.D.

Lemma 2 If $a_m, b_m > 0$, $m = 1, \cdots, M$, then we have $\frac{\sum_{m=1}^M a_m}{\sum_{m=1}^M b_m} \ge \min\{\frac{a_m}{b_m} \mid m = 1, \ldots, M\}$.

Lemma 3

Let $K \in \mathbb{N}$. Assume that $\eta_{k'} \in \mathbb{Z}_+$, $2 \le k' \le K-1$ satisfy $0 \le \eta_2 \le 2$, $\cdots$, $0 \le \eta_2 + \eta_3 + \cdots + \eta_{K-1} \le 2(K-2)$. Let

$$t := 1 + \eta_2 + \cdots + \eta_{K-1},$$

$$\varphi := p + 2\eta_2 + \cdots + (K-1)\eta_{K-1},$$

and

$$t = 2i + m, \quad i \in \mathbb{N}, \quad m = 0 \text{ or } 1.$$

Then we have

$$\frac{\varphi}{2t} > \frac{p + i^2 + im}{4i + 2m} =$$

$$\frac{p + 1 + 1 + 2 + 2 + \cdots + (i-1) + (i-1) + i + im}{2t}$$

Proof It is necessary only to compare those numerators. Q.E.D.

Remark The poles

$$-\frac{p + i^2 + im}{4i + 2m}$$

in Lemma 3 are equal to those for v_1 $(k = 1)$ in Lemma 1.

Lemma 4 If some $J_m^{(\alpha)}$ are not equal to 0, then the poles of (6), (8), (13) and (14) are smaller than one of those in Lemma 1.

Proof We use the same notations in Part 1 of the Main Theorem's proof.

By using Statements (e) and (f), we have

$$\frac{q_l + 1}{t_l} \geq \frac{\sum_{\xi=1}^{g_l} \varphi_l^{(\xi)}}{2\sum_{\xi=1}^{g_l}(1 + \eta_{2,l}^{(\xi)} + \cdots + \eta_{K-1,l}^{(\xi)})},$$

where $\frac{q_l+1}{t_l}$ is in Eq.(6), Eq.(8) or Eq.(14) for $2 \leq l \leq k-1$. By Lemma 2, we have

$$\frac{q_l + 1}{t_l} \geq \min_{1 \leq \xi \leq g_l} \{\frac{\varphi_l^{(\xi)}}{2(1 + \eta_{2,l}^{(\xi)} + \cdots + \eta_{K-1,l}^{(\xi)})}\}.$$

Therefore, by Lemma 3, one of the poles for v_1 in Lemma 1 is greater than $-\frac{q_l+1}{t_l}$.

Next consider the poles in Eq.(13).

(I') Assume that there exist $l' \in A^{(\alpha)}$ and ξ' such that $0 \leq \eta^{(\xi')}_{2,l'} + \eta^{(\xi')}_{3,l'} + \cdots + \eta^{(\xi')}_{K-1,l'} < 2(K-2)$. Then

$$\frac{\sum_{l\in A^{(\alpha)}} q_l + K - 1 + \#A^{(\alpha)} + \#C^{(\alpha)}}{\sum_{l\in A^{(\alpha)}} t_l + 2}$$

$$\geq \frac{\sum_{l\in A^{(\alpha)}}(q_l+1) + K - 1}{\sum_{l\in A^{(\alpha)}} t_l + 2}$$

$$\geq \frac{\sum_{l\in A^{(\alpha)}, l\neq l'}(q_l+1) + (q_{l'}+1) + K - 1}{\sum_{l\in A^{(\alpha)}, l\neq l'} t_l + t_{l'} + 2}$$

$$\geq \min\{\frac{q_l+1}{t_l}, \frac{q_{l'}+1+K-1}{t_{l'}+2} \mid l \in A^{(\alpha)}, l \neq l'\}$$

$$\geq \min\{\frac{\sum_\xi \varphi^{(\xi)}_l}{t_l}, \frac{\sum_\xi \varphi^{(\xi)}_{l'} + K - 1}{t_{l'}+2} \mid l \in A^{(\alpha)}, l \neq l'\}$$

$$\geq \min\{\min_{l\in A^{(\alpha)}, l\neq l', \xi} \frac{\varphi^{(\xi)}_l}{2(1+\eta^{(\xi)}_{2,l} + \cdots + \eta^{(\xi)}_{K-1,l})},$$

$$\min_{\xi\neq\xi'} \frac{\varphi^{(\xi)}_{l'}}{2(1+\eta^{(\xi)}_{2,l'} + \cdots + \eta^{(\xi)}_{K-1,l'})},$$

$$\frac{\varphi^{(\xi')}_{l'} + K - 1}{2(1+\eta^{(\xi')}_{2,l'} + \cdots + (\eta^{(\xi')}_{K-1,l'} + 1))}\}.$$

Therefore, by Lemma 3, we have the poles which are related to v_1 in Lemma 1 and are greater than the pole in Eq.(13).

(II') Assume that for all $l \in A^{(\alpha)}$ and ξ, we have $0 \leq \eta^{(\xi)}_{2,l} + \eta^{(\xi)}_{3,l} + \cdots + \eta^{(\xi)}_{K-1,l} = 2(K-2)$ and some $J^{(\mu)}$ are not equal to 0.

Then $g_l < \sum_{J > J^{(\mu)}} D_{J^{(\mu)},l}$ and

$$\frac{\sum_{l\in A^{(\alpha)}} q_l + K - 1 + \#A^{(\alpha)} + \#C^{(\alpha)}}{\sum_{l\in A^{(\alpha)}} t_l + 2}$$

$$\geq \frac{\sum_{l\in A^{(\alpha)}}(q_l+1) + K - 1}{\sum_{l\in A^{(\alpha)}} t_l + 2}$$

$$\geq \min_{l\in A^{(\alpha)}}\{\frac{\sum_\xi \varphi_l^{(\xi)}}{t_l}, \frac{\sum_\xi \varphi_l^{(\xi)}+p-k+1+K-1}{t_l+2}\}$$

$$\geq \min_{l\in A^{(\alpha)}}\{\frac{\sum_\xi \varphi_l^{(\xi)}}{t_l}, \frac{p}{2}\}.$$

So by Lemma 3, we have a greater pole with respect to v_1 in Lemma 1 than $-\frac{\sum_{l\in A^{(\alpha)}} q_l+K-1+\#A^{(\alpha)}+\#C^{(\alpha)}}{\sum_{l\in A^{(\alpha)}} t_l+2}$ in Eq.(13).

Q.E.D.

Therefore the maximum pole is chosen from the followings:

$$-\frac{p}{2}, -\frac{p+1}{4}, -\frac{p+2}{6},$$

$$-\frac{p+4}{8}, -\frac{p+6}{10},$$

$$\vdots$$

$$-\frac{p+i^2}{4i}, -\frac{p+i^2+i}{4i+2},$$

$$\vdots$$

$$-\frac{p+(p-1)^2}{4(p-1)}, -\frac{p+(p-1)^2+(p-1)}{4(p-1)+2}.$$

Some computations show that

$$\frac{p+i^2+2}{4i+2},$$

where $i = \max\{j \in \mathbb{N} \mid j^2 \leq p\}$ is the maximum pole. If $i^2 = p$ then we have $-\frac{p+i^2}{4i} = -\frac{p+i^2+i}{4i+2}$. So the order of the maximum pole is $\begin{cases} 2 & \text{if } i^2 = p, \\ 1 & \text{if } i^2 \neq p. \end{cases}$

The end of proof of Main Theorem

5 Conclusion

In this paper, we introduce our inductive method for obtaining the poles of the zeta function for the three layered neural network. The blowing-up process of the method enables us to obtain the stochastic complexity of the three layered neural network asymptotically. So, we show that this algebraic method can be effectively used to solve our problem in the learning theory.

Also, the purpose of obtaining the maximum poles of zeta functions for the leaning theory may be considered as a new problem in mathematics, since most of Kullback functions' singularities have not been investigated, yet. The method in this paper would be useful for calculating asymptotic forms for not only the layered neural network model but also other cases. Our aim is to develop the mathematical methods in that context.

Finally, we remark that λ in Main Theorem 1 becomes the same value as the upper bound $\frac{\sqrt{p}}{2}$ in the paper [15], if $p = i^2$.

References

[1] Watanabe, S., "Algebraic analysis for singular statistical estimation," Lecture Notes on Computer Science, 1720, pp. 39-50, 1999.

[2] Watanabe, S.,"Algebraic analysis for nonidentifiable learning machines,"Neural Computation, 13 (4), pp. 899-933, 2001.

[3] Watanabe, S.,"Algebraic geometrical methods for hierarchical learning machines,"Neural Networks, 14 (8), pp. 1049-1060, 2001.

[4] Akaike, H.,"A new look at the statistical model identification," IEEE Trans. on Automatic Control, vol.19, pp.716-723, 1974.

[5] Takeuchi, K.,"Distribution of an information statistic and the criterion for the optimal model," Mathematical Science, No.153, pp.12-18, 1976 (In Japanese).

[6] Hannan, E.J. and Quinn, B.G.," The determination of the order of an autoregression," Journal of Royal Statistical Society, Series B. vol.41, pp.190-195, 1979.

[7] Murata, N., Yoshizawa, S. and Amari, S., " Network information criterion - determining the number of hidden units for an artificial neural network model," IEEE Trans. on Neural Networks, vol.5, no.6, pp.865-872, 1994.

[8] Schwarz, G.," Estimating the dimension of a model," Annals of Statistics, vol.6, no.2, pp.461-464, 1978.

[9] Rissanen, J.," Universal coding, information, prediction, and estimation," IEEE Trans. on Information Theory, vol.30, no.4, pp.629-636, 1984.

[10] Yamazaki, K. and Watanabe, S., " A probabilistic algorithm to calculate the learning curves of hierarchical learning machines with singularities," IEICE Trans., J85-DII, no.3, pp.363-372, 2002.

[11] Watanabe, S., "Learning efficiency of redundant neural networks in Bayesian estimation," IEEE Transactions on Neural Networks, vol.12, no.6, pp.1475-1486, 2001.

[12] Watanabe, K. and Watanabe, S., " Upper bounds of Bayesian generalization errors in reduced rank regression," IEICE Trans., J86-A, no.3, pp.278-287, 2003.

[13] Hironaka, H., "Resolution of Singularities of an algebraic variety over a field of characteristic zero," Annals of Math., vol. 79, pp.109-326, 1964.

[14] Atiyah, M. F., " Resolution of singularities and division of distributions," Comm. Pure and Appl. Math., vol.13, pp.145-150, 1970.

[15] Watanabe, S.,"On the generalization error by a layered statistical model with Bayesian estimation," IEICE Trans., J81-A, pp. 1442-1452, 1998, (English version : Elect. and Comm. in Japan., John Wiley and Sons, 83 (6), pp.95-106, 2000.

Miki Aoyagi received the Doctor's degree of Mathematics from Graduate School of Kyushu University, Fukuoka, Japan, in 1995 and 1997, respectively. From 1997 until 2001, she was a PD Research Fellow of the Japan Society for the Promotion of Science for Young Scientists. Since 2001, she has been an Assistant at Department of Mathematics, Sophia University, Tokyo, Japan. Her research interests are focused on algebraic geometry, complex analysis and learning theory. (Doctor of Mathematics) She is a member of Japan

Mathematical Society.

Sumio Watanabe received the Master's degree of Mathematics from Graduate School of Kyoto University, Kyoto, Japan, in 1982. He is now a Professor of Tokyo Institute of Technology. His research interests are focused on learning theory. (Doctor of Engineering) He is a member of IEEE and JNNS.