The Generalization Error of Reduced Rank Regression in Bayesian Estimation

Miki AOYAGI† and Sumio WATANABE‡

† Department of Mathematics
Sophia University,
7-1 Kioi-cho, Chiyoda-ku, Tokyo,
102–8554, JAPAN
E-mail: miki-a@sophia.ac.jp

‡ Precision and Intelligence Laboratory
Tokyo Institute of Technology,
4259 Nagatsuda, Midori-ku, Yokohama,
226–8503, JAPAN
E-mail: swatanab@pi.titech.ac.jp

Abstract

Reduced rank regression, or a three-layer neural network with linear hidden units, is an important research area, because this method picks up the essential information from examples of input-output pairs. However, those models are non-regular learning machines. Its generalization error had been left unknown because of its singularities in the parameter space. In this paper, we introduce a new computational technique of recursive blowing-ups for densingularization of a learning machine, and compute explicitly the main term in the asymptotic form of the stochastic complexity in the case of the reduced rank regression models.

1. INTRODUCTION

Hierarchical learning machines such as reduced rank regression, multi-layer perceptrons, normal mixtures and Boltzmann machines have the singular Fisher metrics. Their parameters are not identifiable. For instance, the Fisher information matrix $I(w)$ of reduced rank regression, is singular ($\det I(w) = 0$) for such a parameter w that w represents some small model. Here the small model means that the corresponding parameter matrix has the lower column or row rank than one of the learning model. That is, the set of parameters representing small models are analytic variety in the set of all parameters. These learning models are called *non-regular* or *non-identifiable* statistical models. The theory of regular statistical models, for example, model selection methods AIC[1], TIC[11], HQ[5], NIC[7], BIC[10], MDL[8], cannot be applied to the reduce rank approximation, as it is non-regular. It is necessary and crucial to construct a mathematical theory for such learning machines.

Recently, the asymptotic form of the Bayesian stochastic complexity has been obtained using the method of resolution of singularities in [12, 13, 14]. Let n be the number of arbitrary training samples. If exists, the Bayesian generalization error $G(n)$ has an asymptotic expansion, given by

$$G(n) \cong \lambda/n - (m-1)/(n \log n),$$

where λ is a positive and rational number and m is a natural number. Let $\psi(w)$ be a certain *a priori* probability density function, $q(x)$ the true probability distribution and $p(x|w)$ the learner. Also let $K(w)$ be the Kullback information

$$K(w) := \int q(x) \log\{q(x)/p(x|w)\} \mathrm{d}x.$$

By using the blowing-up process, we can calculate the poles of the zeta function

$$J(z) := \int K(w)^z \psi(w) \mathrm{d}w.$$

The maximum pole of $J(z)$ (as real numbers) is $-\lambda$ and its order is m. For regular models, $\lambda = d/2$ and $m = 1$, where d is the dimension of the parameter space. In other words, it does not depend on the true distribution. However, non-regular models have λ depending on the true distribution and it is smaller than $d/2$. Non-regular models are better learning machines than regular ones provided that the Bayes estimation is applied. In [16], the upper bound of the constant λ for reduced rank regression models was obtained. The exact value for λ had been left unknown.

In this paper, we use the inductive method to obtain the exact values λ for the reduced rank regression models, and give the asymptotic form of the stochastic complexity explicitly. The proposed method is recursive blowing-ups. This work also reveals that desingularization is effective to analyze zeta functions for learning theory.

2. BAYESIAN LEARNING MODELS

In this section, we summary the framework of Bayesian learning.

Let $x \in \mathbf{R}^M$ be an input, $y \in \mathbf{R}^N$ an output and $w \in W \subset \mathbf{R}^d$ a parameter. Consider a learning machine $p(x, y|w)$ and a fixed *a priori* probability density function $\psi(w)$. Assume that the true probability distribution $p(x, y|w_0)$ is contained in the learning model.

Let $X^n = (X_1, X_2, ..., X_n)$ be arbitrary n training samples which are independently taken from the true probability distribution $p(x, y|w_0)$. The *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

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

Then the average inference $p(x, y|X^n)$ of the Bayesian distribution is given by

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

Let $G(n)$ be the generalization error or the learning efficiency

$$G(n) := E_n\{\int p(x, y|w_0)\log\frac{p(x, y|w_0)}{p(x, y|X_n)}dxdy\},$$

where $E_n\{\}$ is the expectation value over all sets of n training samples.

Then the average stochastic complexity or the free energy

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

satisfies

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

where

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

Let $J(z)$ be the zeta function of the learning model

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

where $K(w)$ is the Kullback distance

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

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

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

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

The values λ and m can be calculated by using the blowing-up process.

3. RESOLUTION OF SINGULARITIES

In this section, we introduce Hironaka's Theorem [6] on the resolution of singularities and the construction of blowing up. The blowing up is the main tool in the resolution of singularities of an algebraic variety. We also show its application in the field of learning theory [12, 13, 14].

Theorem[Hironaka [6]]

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

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

(2) for each $u \in U$, there are local analytic coordinates $(u_1, \cdots, u_n)$ 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.

The above theorem is one of analytic versions of Hironaka's Theorem used by Atiyah[2].

Consequently, we have

Theorem [Atiyah[2], Bernstein[4], Sato & Shintani[9]]

Let $f(w)$ be an arbitrary analytic function of variables $w \in \mathbf{R}^d$. Let $g(w)$ be a $C^\infty-$ function with compact support W.

Then

$$\zeta(z) = \int_W |f(w)|^z g(w)dw,$$

is a holomorphic function in the right-half plane.

Furthermore, $\zeta(z)$ can be analytically continued to a meromorphic function on the entire complex plane. Its poles are negative rational numbers.

Apply 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 from a neighborhood V_w of w to an analytic manifold U_w, which satisfy the above (1) and (2). Then the local integration on V_w of the zeta function $J(z)$, is written by

$$J_w(z) = \int_{V_w} K(w)^z\psi(w)dw$$

$$= \int_{U_w} (\pm u_1^{s_1} u_2^{s_2} \cdots u_n^{s_n})^z \psi(\mu(u)) |\mu'(u)| du.$$

Then the values $J_w(z)$ can be obtained easily and so the poles and their orders of $J(z)$, since the parameter space W is compact.

(For $w \in W \setminus K^{-1}(0)$, there exists a neighborhood V_w of w such that $K(w') \neq 0, (w' \in V_w)$ and so $J_w(z) = \int_{V_w} K(w)^z \psi(w) dw$ has no poles.)

Next we explain the construction of blowing up. There are three kinds of blowing up, i.e., blowing up at the point, blowing up along the manifold and blowing up with respect to the coherent sheaf of ideals. The blowing up along the manifold includes blowing up at the point. The blowing up with respect to the coherent sheaf of ideals includes blowing up along the manifold.

Here let us explain only the blowing up along the manifold used in this paper. Define a manifold $\mathcal{M}$ by gluing k open sets $U_i \cong \mathbf{R}^d$, $i = 1, 2, \cdots, k (d \geq k)$ as follows.

Denote the coordinate of U_i by $(\xi_{1i}, \cdots, \xi_{di})$.

Define the equivalence relation

$$(\xi_{1i}, \xi_{2i}, \cdots, \xi_{di}) \sim (\xi_{1j}, \xi_{2j}, \cdots, \xi_{dj})$$

at $\xi_{ji} \neq 0$ and $\xi_{ij} \neq 0$, by $\xi_{ij} = 1/\xi_{ji}, \xi_{jj} = \xi_{ii}\xi_{ji}, \xi_{hj} = \xi_{hi}/\xi_{ji} (1 \leq h \leq k, h \neq i, j), \xi_{\ell j} = \xi_{\ell i} (k+1 \leq \ell \leq d)$, and set $\mathcal{M} = \coprod_{i=1}^k U_i / \sim$.

Also define $\pi : \mathcal{M} \to \mathbf{R}^d$ by

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

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

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

The blowing map satisfies

(1) $\pi : \mathcal{M} \to \mathbf{R}^d$ is proper and

(2) $\pi : \mathcal{M} - \pi^{-1}(X) \to \mathbf{R}^d - X$ is isomorphic.

4. LEARNING CURVES OF REDUCED RANK REGRESSION MODELS

In this section, we show how to obtain the poles of the zeta function of learning models in the case of certain reduced rank regression models.

Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ & \vdots & & \\ a_{H1} & a_{H2} & \cdots & a_{HM} \end{pmatrix}$ and

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1H} \\ b_{21} & b_{22} & \cdots & b_{2H} \\ & \vdots & & \\ b_{N1} & b_{N2} & \cdots & b_{NH} \end{pmatrix}.$$

We define the norm of any matrix $C = (c_{ij})$ by $||C|| = \sqrt{\sum_{i,j} |c_{ij}|^2}$.

Let us denote the input value by x. Then the output value y of the reduced rank regression model is given by

$$y = BAx + (\text{noise}).$$

Consider the statistical model

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

Assume that the *a priori* probability density function $\psi(w)$ is a $C^\infty-$ function with compact support W where $\psi(0) > 0$, and that the true parameters w are all 0, i.e., $(A, B) = (0, 0)$.

Then the Kullback information is

$$K(w) = c_0 ||BA||^2,$$

for some constant c_0. Then the zeta function is as follows:

$$J(z) = \int_W ||BA||^{2z} \mathrm{d}w.$$

Main Theorem

The maximum pole $-\lambda$ and its order m are given, case by case, as follows:

Case (1) Let $N \leq M + H$, $M \leq N + H$ and $H \leq M + N$.

(a) If $M + H + N$ is even, then $m = 1$ and

$$\lambda = \frac{2MN + 2HN + 2MH - N^2 - M^2 - H^2}{8}.$$

(b) If $M + H + N$ is odd, then $m = 2$ and

$$\lambda = \frac{1 + 2MN + 2HN + 2MH - N^2 - M^2 - H^2}{8}.$$

Case (2) Let $M + H < N$. Then $m = 1$ and

$$\lambda = \frac{MH}{2}.$$

Case (3) Let $N + H < M$. Then $m = 1$ and

$$\lambda = \frac{HN}{2}.$$

Case (4) Let $M + N < H$. Then $m = 1$ and

$$\lambda = \frac{MN}{2}.$$

Before providing Main Theorem, let us give some notation.

Since we often change the variables during the blowing-up process, it is more convenient for us to use the same symbols a_{ij} rather than a'_{ij}, a''_{ij}, $\cdots$, etc, for the sake of simplicity. For instance,

"Let $\begin{cases} a_{11} = u_{11} \\ a_{ij} = u_{11}a_{ij},\ (i,j) \neq (1,1). \end{cases}$"

instead of

"Let $\begin{cases} a_{11} = u_{11} \\ a_{ij} = u_{11}a'_{ij},\ (i,j) \neq (1,1). \end{cases}$"

Proof of Main Theorem.

Let

$$\Psi = ||BA||^2.$$

We need to calculate poles of the following function by using the blowing-up process together with an inductive method.

$$\Psi' = \sum_{i=1}^{s} ||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s} \mathbf{b}_i f_i(a_{kl}) + B_{s+1}A_{s+1}||^2, \quad (1)$$

where for $i = 1, \cdots, H$, $\mathbf{b}_i = {}^T (b_{1i} \ \cdots \ b_{Ni})$, (T denotes the transpose),

$$B_{s+1} = \begin{pmatrix} b_{1,s+1} & b_{1,s+2} & \cdots & b_{1,H} \\ b_{2,s+1} & b_{2,s+2} & \cdots & b_{2,H} \\ & \vdots & & \\ b_{N,s+1} & b_{N,s+2} & \cdots & b_{N,H} \end{pmatrix} \text{ and}$$

$$A_{s+1} = \begin{pmatrix} a_{s+1,s+1} & a_{s+1,s+2} & \cdots & a_{s+1,M} \\ a_{s+2,s+1} & a_{s+2,s+2} & \cdots & a_{s+2,M} \\ & \vdots & & \\ a_{H,s+1} & a_{H,s+2} & \cdots & a_{H,M} \end{pmatrix}.$$

$f_i(a_{kl})$ is a function of the entries of the matrix A excluding the entries of A_{s+1}. The definition of the function $f_i(a_{kl})$ will be given recursively in Equation (2) below.

Let us construct the blowing-up of Ψ along the submanifold $\{a_{ij} = 0, 1 \leq i \leq H, 1 \leq j \leq M\}$.

Let $\begin{cases} a_{11} = u_{11} \\ a_{ij} = u_{11}a_{ij},\ (i,j) \neq (1,1). \end{cases}$ Then we have

$$\Psi = u_{11}^2(||(\mathbf{b}_1 + B_2\mathbf{a}_1)||^2 + ||(\ \mathbf{b}_1 \quad B_2 \ ) \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A_2 \end{pmatrix} ||^2,$$

where $\tilde{\mathbf{a}}_1 = (a_{12} \ \cdots \ a_{1M})$ and $\mathbf{a}_1 = {}^T (a_{21} \ \cdots \ a_{H1})$.

By the symmetry of the norm function, it is enough to consider the above case.

Put $\mathbf{b}_1 = \mathbf{b}_1 + B_2\mathbf{a}_1$. Then

$$\begin{aligned} \Psi &= u_{11}^2||\mathbf{b}_1||^2 + || \begin{pmatrix} \mathbf{b}_1 - B_2\mathbf{a}_1 & B_2 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A_2 \end{pmatrix} ||^2 \\ &= u_{11}^2||\mathbf{b}_1||^2 + || \begin{pmatrix} \mathbf{b}_1 & 0 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A_2 \end{pmatrix} \\ &\quad + B_2 \begin{pmatrix} -\mathbf{a}_1 & E \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A_2 \end{pmatrix} ||^2 \\ &= u_{11}^2||\mathbf{b}_1||^2 + ||\mathbf{b}_1\tilde{\mathbf{a}}_1 + B_2 \left(-\mathbf{a}_1\tilde{\mathbf{a}}_1 + A_2\right) ||^2. \end{aligned}$$

Let $A_2 = -\mathbf{a}_1\tilde{\mathbf{a}}_1 + A_2$, then

$$\Psi = u_{11}^2||\mathbf{b}_1||^2 + ||\mathbf{b}_1\tilde{\mathbf{a}}_1 + B_2A_2||^2.$$

Therefore we have the form

$$\Psi' = ||\mathbf{b}_1||^2 + ||\mathbf{b}_1\tilde{\mathbf{a}}_1 + B_2A_2||^2,$$

of the equation (1) with $s = 1$.

Let us construct the blowing-up of Ψ' in (1) along the submanifold $\{b_{ji} = 0, 1 \leq i \leq s, 1 \leq j \leq N, a_{j\ell} = 0, s+1 \leq j \leq H, s+1 \leq \ell \leq M\}$.

Let

$$\begin{cases} b_{11} = v \\ b_{ji} = vb_{ji}, 1 \leq i \leq s, 1 \leq j \leq N, (i,j) \neq (1,1) \\ a_{j\ell} = va_{j\ell}, s+1 \leq j \leq H, s+1 \leq \ell \leq M. \end{cases}$$

We have

$$\begin{aligned} \Psi' &= v^2(1 + \sum_{i=2}^{N} b_{i1}^2 + \sum_{i=2}^{s} ||\mathbf{b}_i||^2 \\ &\quad + ||\sum_{i=1}^{s} \mathbf{b}_i f_i + B_{s+1}A_{s+1}||^2). \end{aligned}$$

Here the Jacobian is $v^{sN+(M-s)(H-s)-1}$.

Therefore we have the pole

$$-\frac{sN + (M-s)(H-s)}{2}.$$

Next let

$$\begin{cases} a_{s+1,s+1} = u \\ b_{ji} = ub_{ji}, 1 \leq i \leq s, 1 \leq j \leq N \\ a_{j\ell} = ua_{j\ell}, \quad s+1 \leq j \leq H, s+1 \leq \ell \leq M \\ \qquad\qquad (j,\ell) \neq (s+1, s+1). \end{cases}$$

By the symmetry of the norm function, this setting is the general case as $a_{j,\ell} = u$.

We have

$$\begin{aligned} \Psi' &= u^2(\sum_{i=1}^{s} ||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s} \mathbf{b}_i f_i \\ &\quad + \begin{pmatrix} \mathbf{b}_{s+1} & B_{s+2} \end{pmatrix} \begin{pmatrix} 1 & \tilde{\mathbf{a}}_{s+1} \\ \mathbf{a}_{s+1} & A_{s+2} \end{pmatrix} ||^2) \\ &= u^2(\sum_{i=1}^{s} ||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s} \mathbf{b}_i f_i \end{aligned}$$

$$+\left(\begin{array}{cc}\mathbf{b}_{s+1}+B_{s+2}\mathbf{a}_{s+1} & 0\end{array}\right)$$
$$+\left(\begin{array}{cc}\mathbf{b}_{s+1} & B_{s+2}\end{array}\right)\left(\begin{array}{cc}0 & \tilde{\mathbf{a}}_{s+1}\\ & A_{s+2}\end{array}\right)||^2),$$

where $\tilde{\mathbf{a}}_{s+1} = (a_{s+1,s+2} \ \cdots \ a_{s+1,M})$ and $\mathbf{a}_{s+1} =^T (a_{s+2,s+1} \ \cdots \ a_{H,s+1})$.

Denote the first column of f_i by $\mathbf{f}_i$. Let $f_i = (\mathbf{f}_i \ \ f_i')$.

Put $\mathbf{b}_{s+1} = \mathbf{b}_{s+1} + B_{s+2}\mathbf{a}_{s+1} + \sum_{i=1}^{s}\mathbf{b}_i\mathbf{f}_i$, then

$$\begin{aligned}
&\Psi'/u^2\\
=\ &\sum_{i=1}^{s}||\mathbf{b}_i||^2 + ||\mathbf{b}_{s+1}||^2 + ||\sum_{i=1}^{s}\mathbf{b}_i f_i'\\
&+\left(\begin{array}{cc}\mathbf{b}_{s+1} - B_{s+2}\mathbf{a}_{s+1} - \sum_{i=1}^{s}\mathbf{b}_i\mathbf{f}_i & B_{s+2}\end{array}\right)\\
&\left(\begin{array}{c}\tilde{\mathbf{a}}_{s+1}\\ A_{s+2}\end{array}\right)||^2\\
&=\sum_{i=1}^{s+1}||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s}\mathbf{b}_i f_i' + \left(\mathbf{b}_{s+1} - \sum_{i=1}^{s}\mathbf{b}_i\mathbf{f}_i \ \ 0\right)\\
&\left(\begin{array}{c}\tilde{\mathbf{a}}_{s+1}\\ A_{s+2}\end{array}\right) + \left(\begin{array}{cc}-B_{s+2}\mathbf{a}_{s+1} & B_{s+2}\end{array}\right)\left(\begin{array}{c}\tilde{\mathbf{a}}_{s+1}\\ A_{s+2}\end{array}\right)||^2\\
=\ &\sum_{i=1}^{s+1}||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s}\mathbf{b}_i(f_i' - \mathbf{f}_i\tilde{\mathbf{a}}_{s+1})\\
&+\mathbf{b}_{s+1}\tilde{\mathbf{a}}_{s+1} + B_{s+2}(-\mathbf{a}_{s+1}, E)\left(\begin{array}{c}\tilde{\mathbf{a}}_{s+1}\\ A_{s+2}\end{array}\right)||^2\\
=\ &\sum_{i=1}^{s+1}||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s}\mathbf{b}_i(f_i' - \mathbf{f}_i\tilde{\mathbf{a}}_{s+1})\\
&+\mathbf{b}_{s+1}\tilde{\mathbf{a}}_{s+1} + B_{s+2}(-\mathbf{a}_{s+1}\tilde{\mathbf{a}}_{s+1} + A_{s+2})||^2.
\end{aligned}$$

Now let $A_{s+2} = -\mathbf{a}_{s+1}\tilde{\mathbf{a}}_{s+1} + A_{s+2}$. Then

$$\begin{aligned}
\Psi'/u^2 \ =\ &\sum_{i=1}^{s+1}||\mathbf{b}_i||^2 + ||\sum_{i=1}^{s}\mathbf{b}_i(f_i' - \mathbf{f}_i\tilde{\mathbf{a}}_{s+1})\\
&+\mathbf{b}_{s+1}\tilde{\mathbf{a}}_{s+1} + B_{s+2}A_{s+2}||^2.
\end{aligned}$$

Repeat this whole process by letting

$$f_i = f_i' - \mathbf{f}_i\tilde{\mathbf{a}}_{s+1}, \qquad f_{s+1} = \tilde{\mathbf{a}}_{s+1}. \tag{2}$$

Then, s will be replaced by $s+1$ in (1) and so on.

Finally, we have $\Psi' = \sum_{i=1}^{s}||\mathbf{b}_i||^2$ with $s = \min\{H, M\}$.

Q.E.D.

Therefore the poles are

$$-\frac{(N+M)r - r^2 + s(N-r) + (M-r-s)(H-r-s)}{2},$$

for $s = 0, \cdots, \min\{H-r, M-r\}$.

(i) If $\frac{M+H-N-r}{2} < 0$, the maximum pole at $s = 0$ is

$$-\frac{HM - Hr + Nr}{2},$$

and its order m is 1.

(ii) If $0 \le \frac{M+H-N-r}{2} \le \min\{H-r, M-r\}$ and $M+H-N-r$ is even, the maximum pole at $s = \frac{M+H-N-r}{2}$ is

$$-\frac{-(H+r)^2 - M^2 - N^2}{8} - \frac{2(H+r)M + 2(H+r)N + 2MN}{8},$$

and its order m is 1.

(iii) If $0 \le \frac{M+H-N-r}{2} \le \min\{H-r, M-r\}$ and $M+H-N-r$ is odd, the maximum pole at $s = \frac{M+H-N+1-r}{2}$ and $\frac{M+H-N-1-r}{2}$ is

$$-\frac{-(H+r)^2 - M^2 - N^2}{8} - \frac{2(H+r)M + 2(H+r)N + 2MN + 1}{8},$$

and its order m is 2.

(iv) If $\frac{M+H-N-r}{2} > \min\{H-r, M-r\}$ and $H \le M$, the maximum pole at $s = H-r$ is

$$-\frac{HN - Hr + Mr}{2},$$

and its order m is 1.

(v) If $\frac{M+H-N-r}{2} > \min\{H-r, M-r\}$ and $M < H$, the maximum pole at $s = M-r$ is

$$-\frac{MN}{2},$$

and its order m is 1.

So Main Theorem follows.

5. DISCUSSION AND CONCLUSION

In this paper, we show the computational method to obtain the poles of the zeta functions of the certain reduced rank regression models.

The Figure 1 shows the graphs of the maximum poles λ with λ-values in y-axis and H-values in x-axis, when $M = N = 5$. It is clear that the curve is not linear.

Significance of the obtained result from the viewpoint of learning theory is as follows.

First, using our result, we can construct model selection methods for Bayesian estimation.

Second, we can construct mathematical foundation for analyzing and developing the precision of the MCMC method. By the MCMC method, the estimated values of marginal likelihoods had been calculated for hyper-parameter estimation and model selection methods of complex learning models, but the theoretical

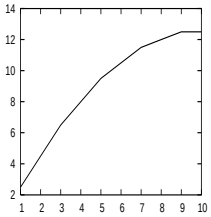

Figure 1: The curve of λ-values in y-axis and H-values in x-axis, when $M = N = 5$.

values were not known. Now, the theoretical value of marginal likelihoods is given in this paper, when the true function is zero.

Desingularization of an arbitrary polynomial is achieved by using the blowing-up process (Hironaka's Theorem[6]). This is a finite process. However our algebraic calculation cannot be done by the computer aid, since the number of rows and columns of the parametric matrix (i.e., N, M, H) corresponding to the Kullback information is not constant, which computers concretely require.

Our conclusion is that the algebraic method will lead us to solve the difficult problems of learning theory.

The method would be useful to calculate the asymptotic form for not only reduced rank regression models but also other cases. Our aim is to develop a mathematical theory in that context.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. on Automatic Control.* **19** 716-723.

[2] Atiyah, M. F. (1970). Resolution of singularities and division of distributions. *Comm. Pure and Appl. Math.* **13** 145-150.

[3] Baldi, P. and Hornik, K. (1995). Learning in Linear Networks: a Survey. *IEEE Transactions on Neural Networks.* **6** (4) 837-858.

[4] Bernstein, I. N. (1972). The analytic continuation of generalized functions with respect to a parameter. *Functional Analysis Applications.* **6** 26-40.

[5] Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of Royal Statistical Society, Series B.* **41** 190-195.

[6] Hironaka, H. (1964). Resolution of Singularities of an algebraic variety over a field of characteristic zero. *Annals of Math.* **79** 109-326.

[7] Murata, N. J., Yoshizawa, S. G. and Amari, S. (1994). Network information criterion - determining the number of hidden units for an artificial neural network model. *IEEE Trans. on Neural Networks.* **5** (6) 865-872.

[8] Rissanen, J. (1984). Universal coding, information, prediction, and estimation. *IEEE Trans. on Information Theory.* **30** (4) 629-636.

[9] Sato, M. and Shintani, T. (1974). On zeta functions associated with prehomogeneous vector space. *Annals of Math.* **100** 131-170.

[10] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics.* **6** (2) 461-464.

[11] Takeuchi, K. (1976). Distribution of an information statistic and the criterion for the optimal model. *Mathematical Science.* **153** 12-18 (In Japanese).

[12] Watanabe, S. (1999). Algebraic analysis for singular statistical estimation. *Lecture Notes on Computer Science.* **1720** 39-50.

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

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

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

[16] Watanabe, K. and Watanabe, S. (2003). Upper Bounds of Bayesian Generalization Errors in Reduced Rank Regression. *IEICE Trans.* **J86-A** (3) 278-287 (In Japanese).