

Stochastic complexities of reduced rank regression in Bayesian estimation

Miki Aoyagi^{a,*}, Sumio Watanabe^b

^aDepartment of Mathematics, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554, Japan

^bPrecision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama 226-8503, Japan

Received 17 May 2004; revised 16 March 2005; accepted 16 March 2005

Abstract

Reduced rank regression extracts an essential information from examples of input–output pairs. It is understood as a three-layer neural network with linear hidden units. However, reduced rank approximation is a non-regular statistical model which has a degenerate Fisher information matrix. Its generalization error had been left unknown even in statistics. In this paper, we give the exact asymptotic form of its generalization error in Bayesian estimation, based on resolution of learning machine singularities. For this purpose, the maximum pole of the zeta function for the learning theory is calculated. We propose a new method of recursive blowing-ups which yields the complete desingularization of the reduced rank approximation.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Stochastic complexity; Generalization error; Reduced rank regression models; Non-regular learning machines; Bayesian estimate; Resolution of singularities; Kullback information; Zeta function

1. Introduction

Hierarchical learning machine such as reduced rank regression, multi-layer perceptron, normal mixture and Boltzmann machine has its singular Fisher matrix function $I(w)$ for a parameter w . Specifically, $\det I(w_0) = 0$ for a particular parameter w_0 , representing some small model. The parameter w_0 is not identifiable, since the subset, which consists of parameters representing the small model is an analytic variety in all parameter space. Such a learning model is called a *non-regular (non-identifiable)* statistical model. For example, consider a learning machine

$$p(y|x, A, B) = \frac{1}{(\sqrt{2\pi})^2} \exp\left(-\frac{1}{2}\|y - BAx\|^2\right),$$

of the reduced rank approximation with a 2×2 matrix $A = (a_{ij})$ and a 2×2 matrix $B = (b_{ij})$. Assume that this machine estimates the true distribution $p(y|x, A_0, B_0)$, where $B_0 A_0 = 0$. Denote the subset of the parameters representing the small

model by

$$W_0 = \{(A, B); p(y|x, A, B) = p(y|x, A_0, B_0)\}.$$

Then we have

$$W_0 \supset \{(A, B); A = 0\} \cup \{(A, B); B = 0\}.$$

Recently, the asymptotic form of the Bayesian stochastic complexity has been obtained, using the method of resolution of singularities by Watanabe (1999, 2001a,b). Let n be the number of any training samples. The average stochastic complexity (the free energy) $F(n)$ is asymptotically equal to

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

where λ is a positive rational number, m is a natural number and $O(1)$ is a bounded function of n . Hence, if exists, the Bayesian generalization error $G(n)$ has an asymptotic expansion given by

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

Let $\psi(w)$ be a certain a priori probability density function, $q(x, y)$ the true simultaneous distribution of input and output, and $p(x, y|w)$ the learning model. The Kullback information $K(w)$ can be formulated as

$$K(w) = \int q(x, y) \log\{q(x, y)/p(x, y|w)\} dx dy.$$

* Corresponding author. Tel./fax: +81 3 5386 5878.

E-mail addresses: miki-a@sophia.ac.jp (M. Aoyagi), swatanab@pi.titech.ac.jp (S. Watanabe).

Mathematical Symbols

x	M -dimensional input	T, T_1, T_2 , matrix
y	N -dimensional output	$\ T\ = \sqrt{\sum_{i,j} t_{ij} ^2}$ norm of any matrix $T = (t_{ij})$
w, w_0	d -dimensional parameter	$q(x)$ probability density function of x
$I(w)$	Fisher information matrix	B_0, A_0 true parameter
W_0	subset of parameter space	r rank of $B_0 A_0$
n	number of any training samples	$\mathcal{X} = (\int x_i x_j q(x) dx)$ matrix
$G(n)$	generalization error	$c_1, c_2 > 0$ positive numbers
$F(n)$	average stochastic complexity	$S = BA - B_0 A_0$ matrix
λ	positive and rational number	Q orthogonal matrix
m	natural number	$-\lambda(f, g)$ maximum pole of $\zeta(z)$
$O(1)$	bounded function of n	$a \in \mathbf{R} - \{0\}$ number
$\psi(w)$	a priori probability distribution on the parameter space	$\text{Mat}(N', M')$ set of $N' \times M'$ matrices
$p(x, y w)$	learning machine	α, β, γ positive numbers
$q(x, y)$	true simultaneous distribution of input and output	P_0, Q_0 regular matrices
$K(w)$	Kullback information	Φ function of A and B
$J(z)$	zeta function for learning theory	$C_1 = (c_{ij}^{(1)})$ $r \times r$ matrix
w_0	true parameter	$C_2 = (c_{ij}^{(2)})$ $(N-r) \times r$ matrix
$X^n = (X_1, X_2, \dots, X_n)$	set of training samples	$C_3 = (c_{ij}^{(3)})$ $r \times (M-r)$ matrix
$p(w X^n)$	a posteriori probability density function	A' $H \times M$ matrix
$p(x, y X^n)$	average inference of the Bayesian distribution	B' $N \times H$ matrix
$E_n\{ \}$	the expectation value over all sets of n training samples	A_1 $r \times r$ matrix
$K_n(w)$	empirical Kullback information	A_2 $(H-r) \times r$ matrix
f, f_1, f_2	real analytic functions	A_3 $r \times (M-r)$ matrix
V, V_w	neighborhood of w	$A_4 = (a_{ij})$ $(H-r) \times (M-r)$ matrix
U, U_w	real analytic manifold	B_1 $r \times r$ matrix
μ	proper analytic map from U to V	B_2 $(N-r) \times r$ matrix
\mathcal{E}	subset of U	B_3 $r \times (H-r)$ matrix
$u = (u_1, \dots, u_d)$	local analytic coordinate	$B_4 = (b_{ij})$ $(N-r) \times (H-r)$ matrix
s_1, \dots, s_d	non-negative integers	Φ', Φ'' functions of C_1, C_2, C_3, A_4 and B_4
W	compact subset	$\psi'(w')$ C^∞ -function with compact support W'
g, g_1, g_2	C^∞ -functions with compact support W	E $r \times r$ unit matrix
$\zeta(z)$	zeta function	s positive integer
U_i	d -dimensional real space	$A^{(s+1)}$ $(H-r-s) \times (M-r-s)$
\mathcal{M}	a real manifold	$B^{(s+1)}$ $(N-r) \times (H-r-s)$
$(\xi_{1i}, \dots, \xi_{di})$	coordinate of U_i	\mathbf{b}_{s+1} $N-r$ vector
π	proper analytic map from \mathcal{M} to \mathbf{R}^d	$D_i(a_{ki})$ function of the entries of the matrix A_4 excluding the entries of $A^{(s+1)}$
X	subspace of \mathbf{R}^d	$\tilde{\mathbf{a}}_{s+1}$ $M-r-s-1$ vector
A	$H \times M$ matrix	\mathbf{a}_{s+1} $H-r-s-1$ vector
B	$N \times H$ matrix	$\text{Col}_1(D_i)$ first column of D_i
		D'_i $D_i = (\text{Col}_1(D_i) D'_i)$
		$\ell(s)$ function from integers to real numbers

Then the zeta function for the learning theory is defined by

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

Watanabe (1999, 2001a,b) proved that the maximum pole of $J(z)$ (as real numbers) is $-\lambda$ and its order is m , calculated by using the blowing-up process. For regular models, $\lambda = d/2$ and $m = 1$, where d is the dimension of the parameter space. Non-regular models have smaller value λ

than $d/2$, so they are effective learning machines than regular ones provided that the Bayes estimation is applied.

In Watanabe and Watanabe (2003), the upper bound of the constant λ for the reduced rank regression model was obtained. The exact value for λ has been left unknown.

In this paper, we use the inductive method to obtain the exact value λ for the reduced rank regression model, and give the asymptotic form of the stochastic complexity explicitly. Reduced rank regression estimates the conditional probability by using a reduced rank linear operator

from higher dimensional input to higher dimensional output. The aim of this model is to find small rank relation between input and output. The model is a three-layer neural network with linear hidden units (Baldi & Hornik, 1995). In order to be able to understand characters of layered neural networks, it is important to analyze the model mathematically.

The proposed method in this paper is recursive blowing-ups. By Hironaka’s Theorem (1964), it is known that the desingularization of an arbitrary polynomial can be obtained by using the blowing-up process. However, the desingularization of any polynomial in general, although it is known as a finite process, is very difficult.

It is well-known that there are many information criteria for statistical model selection of regular statistical models, for example, model selection methods AIC (Akaike, 1974), TIC (Takeuchi, 1976), HQ (Hannan & Quinn, 1979), NIC (Murata, Yoshizawa, & Amari, 1994), BIC (Schwarz, 1978), MDL (Rissanen, 1984). However, the theory of regular statistical models cannot be applied to analyzing such non-regular models. The result of this paper clarifies the asymptotic behavior of the marginal likelihood and the stochastic complexity.

In practical usage, the stochastic complexity is calculated by some numerical calculation, for example, the Markov Chain Monte Carlo method (MCMC). By the MCMC method, the estimated values of marginal likelihoods had 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 are given in this paper. This enables us to construct mathematical foundation for analyzing and developing the precision of the MCMC method.

2. Bayesian learning models

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

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

Let $X^n = (X_1, X_2, \dots, X_n)$, $X_i = (x_i, y_i)$ be arbitrary n training samples. X_i ’s are randomly selected from the true probability distribution $p(x, y|w_0)$. Then, 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

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

So 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 (the learning efficiency) as follows

$$G(n) = E_n \left\{ \int p(x, y|w_0) \log \frac{p(x, y|w_0)}{p(x, y|X^n)} dx dy \right\},$$

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

Then the average stochastic complexity (the free energy)

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

satisfies

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

where

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

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 information

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

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

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

and

$$G(n) \cong \lambda/n - (m - 1)/(n \log n), \tag{2}$$

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

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

3. Resolutions of singularities

In this section, we introduce the Hironaka’s theorem (1964) on the resolution of singularities. 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 (Watanabe, 1999, 2001a,b).

Theorem 1. (Hironaka (1964)). *Let f be a real analytic function in a neighborhood of $w = (w_1, \dots, w_d) \in \mathbf{R}^d$ 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*

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

(2) for each $u \in U$, there is a local analytic coordinate (u_1, \dots, u_d) such that $f(\mu(u)) = \pm u_1^{s_1} u_2^{s_2} \dots u_d^{s_d}$, where s_1, \dots, s_d are non-negative integers.

The above theorem is an analytic version of the Hironaka's theorem used by Atiyah (1970).

Theorem 2. (Atiyah (1970), Bernstein (1972), Björk (1979), Sato and Shintani (1974)). Let $f(w)$ be an analytic function of a variable $w \in \mathbf{R}^d$. Let $g(w)$ be a C^∞ -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 extended to a meromorphic function on the entire complex plane. Its poles are negative rational numbers.

Theorem 2 follows from Theorem 1.

Applying the Hironaka's theorem to the Kullback information $K(w)$, 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 of (1) and (2). Then the local integration on V_w of the zeta function $J(z)$ of the learning model is

$$J_w(z) = \int_{V_w} K(w)^z \psi(w) dw \\ = \int_{U_w} (u_1^{2s_1} u_2^{2s_2} \dots u_d^{2s_d})^z \psi(\mu_w(u)) |\mu_w'(u)| du.$$

Therefore, the values $J_w(z)$ can be obtained. For each $w \in \text{WK}^{-1}(0)$, 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 set of parameters W is compact, the poles and their orders of $J(z)$ are computable.

Next, we explain the construction of blowing up. There are three kinds of blowing up: 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 is a generalization of the blowing up at the point. The blowing up with respect to the coherent sheaf of ideals is a generalization of the 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 $U_i \cong \mathbf{R}^d$, $i = 1, 2, \dots, k$ ($d \geq k$) as follows.

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

Define the equivalence relation

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

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

$$\xi_{ij} = 1/\xi_{ji}, \quad \xi_{jj} = \xi_{ii}\xi_{ji}, \\ \xi_{hj} = \xi_{hi}/\xi_{ji}, \quad 1 \leq h \leq k, h \neq i, j, \\ \xi_{\ell j} = \xi_{\ell i}, \quad k + 1 \leq \ell \leq d.$$

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

Define the blowing map $\pi : \mathcal{M} \rightarrow \mathbf{R}^d$ by

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

for each $(\xi_{1i}, \dots, \xi_{di}) \in U_i$.

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

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

The blowing map satisfies

- (1) $\pi : \mathcal{M} \rightarrow \mathbf{R}^d$ is proper.
- (2) $\pi : \mathcal{M} - \pi^{-1}(X) \rightarrow \mathbf{R}^d - X$ is an isomorphism.

4. Learning curves of reduced rank regression model

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

Let

$$\{w = (A, B) \mid A \text{ is an } H \times M \text{ matrix, } B \text{ is an } N \times H \text{ matrix}\},$$

be the set of parameters.

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

Denote the input value by $x \in \mathbf{R}^M$ with a probability density function $q(x)$. Assume that all eigenvalues of the $M \times M$ matrix $\mathcal{X} = (\int x_i x_j q(x) dx)$ are positive numbers. Such a matrix is called a positive definite.

Then the output value $y \in \mathbf{R}^N$ of the reduced rank regression model is given by

$$y = BAx.$$

Consider the statistical model

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

with Gaussian noise. Let $w_0 = (A_0, B_0)$ be the true parameter. Assume that the a priori probability density function $\psi(w)$ is a C^∞ -function with compact support W , satisfying $\psi(A_0, B_0) > 0$.

We can apply Section 2, by using $p(x, y|w_0) = p(y|x, w_0) \times q(x)$ and n training samples $X^n = (X_1, X_2, \dots, X_n)$, $X_i = (x_i, y_i)$ which are randomly selected from the true probability distribution $p(x, y|w_0)$.

In Main Theorem, we give the formulas for the parameters λ and m appearing Eqs. (1) and (2).

Lemma 1. *There exist constants $c_1 > 0$ and $c_2 > 0$ such that*

$$c_1 \|BA - B_0A_0\|^2 \leq K(w) \leq c_2 \|BA - B_0A_0\|^2. \quad (3)$$

Proof. Put

$$\lambda = \max \left\{ \frac{(N + M)r - r^2 + s(N - r) + (M - r - s)(H - r - s)}{2} \mid 0 \leq s \leq \min\{M + r, H + r\} \right\}.$$

$$q(x, y) = p(y|x, (A_0, B_0))q(x).$$

Then we have the Kullback information

$$\begin{aligned} K(w) &= \int q(x, y) \log \frac{p(y|x, (A_0, B_0))}{p(y|x, w)} dx dy \\ &= \frac{1}{2} \int \| (BA - B_0A_0)x \|^2 q(x) dx. \end{aligned}$$

Let $S = BA - B_0A_0 = (s_{ij})$ and Q an orthogonal matrix such that $Q^t \mathcal{X} Q$ is diagonal.

Then, we have

$$\begin{aligned} K(w) &= \frac{1}{2} \int \|Sx\|^2 q(x) dx = \frac{1}{2} \int \sum_i \left(\sum_j s_{ij} x_j \right)^2 q(x) dx \\ &= \frac{1}{2} \sum_{i,j_1,j_2} S_{ij_1} S_{ij_2} \int x_{j_1} x_{j_2} q(x) dx = \frac{1}{2} \text{Tr}(S \mathcal{X} S^t) \\ &= \frac{1}{2} \text{Tr}(S Q Q^t \mathcal{X} Q (S Q)^t). \end{aligned}$$

Since we assume all eigenvalues of \mathcal{X} are positive numbers, there exist $c_1 > 0$ and $c_2 > 0$ such that

$$\begin{aligned} c_1 \text{Tr}(S Q (S Q)^t) &= c_1 \text{Tr}(S S^t) \leq K(w) \leq c_2 \text{Tr}(S Q (S Q)^t) \\ &= c_2 \text{Tr}(S S^t). \end{aligned}$$

Since $\text{Tr}(S S^t) = \|S\|^2$, this completes the proof. \square

Lemma 2. (Watanabe (2001c)). *Let $f(w), f_1(w), f_2(w)$ be analytic functions of $w \in \mathbf{R}^d$. Let $g(w), g_1(w), g_2(w)$ be C^∞ -functions with compact support W .*

Put

$$\zeta(z) = \int_W |f(w)|^2 g(w) dw.$$

Denote the maximum pole of $\zeta(z)$ by $-A(f, g)$.

If $|f_1| \leq |f_2|$ and $g_1 \geq g_2$ then we have $A(f_1, g_1) \leq A(f_2, g_2)$.

Lemmas 1 and 2 yield that the zeta function can be written as follows:

$$J(z) = \int_W \|BA - B_0A_0\|^{2z} \psi(w) dw.$$

Main Theorem

Let r be the rank of B_0A_0 .

$F(n)$ and $G(n)$ in Eqs. (1) and (2) are given by using the following maximum pole $-\lambda$ of $J(z)$ and its order m :

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

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

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

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

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

Case (2) Let $M + H < N + r$. Then $m = 1$ and $\lambda = (HM - Hr + Nr)/2$.

Case (3) Let $N + H < M + r$. Then $m = 1$ and $\lambda = (HN - Hr + Mr)/2$.

Case (4) Let $M + N < H + r$. Then $m = 1$ and $\lambda = MN/2$.

For practical use, the case of $M \gg H$ and $N \gg H$ are considered, so Case (4) does not occur.

This model has $MH + NH$ dimensional parameter space. Therefore, the maximum pole is $-(MH + NH)/2$ for regular models with $MH + NH$ dimensional parameter space. In other words, it does not depend on the true distribution parameter w_0 for regular models. However, non-regular models have λ depending on w_0 . So, it is difficult to construct the model selection methods for non-regular models.

The Fisher information matrix of the model is singular for each case since $\lambda < (MH + NH)/2$.

In order to prove Main Theorem, we need the following three lemmas.

Lemma 3. Let U be a neighborhood of $w_0 \in \mathbf{R}^d$. Let $T_1(w), T_2(w), T(w)$ be functions from U to $\text{Mat}(N', H')$, $\text{Mat}(N', M')$, $\text{Mat}(H', M')$, respectively.

Assume that the function $\|T(w)\|$ is bounded.

Then, there exist positive constants $\alpha > 0$ and $\beta > 0$ such that

$$\begin{aligned} \alpha (\|T_1\|^2 + \|T_2\|^2) &\leq \|T_1\|^2 + \|T_2 + T_1 T\|^2 \\ &\leq \beta (\|T_1\|^2 + \|T_2\|^2). \end{aligned}$$

Proof. Since $\|T(w)\|$ is bounded, there exists $\beta > 3$ such that

$$\begin{aligned} \|T_1\|^2 + \|T_2 + T_1T\|^2 &\leq \|T_1\|^2 + 2\|T_2\|^2 + 2\|T_1T\|^2 \\ &\leq \beta(\|T_1\|^2 + \|T_2\|^2). \end{aligned}$$

Also, there exists $\gamma > 3$ such that

$$\begin{aligned} \|T_2\|^2 &\leq 2(\|T_2 + T_1T\|^2 + \|-T_1T\|^2) \\ &\leq 2(\|T_2 + T_1T\|^2 + \gamma\|T_1\|^2), \end{aligned}$$

and hence

$$\begin{aligned} \|T_1\|^2 + \|T_2\|^2 &\leq 2\|T_2 + T_1T\|^2 + (2\gamma + 1)\|T_1\|^2 \\ &\leq (2\gamma + 1)(\|T_2 + T_1T\|^2 + \|T_1\|^2). \end{aligned}$$

Putting $\alpha = 1/(2\gamma + 1)$ completes the proof. \square

Lemma 4. Let U be a neighborhood of $w_0 \in \mathbf{R}^d$. Also let $T(w)$ be a function from U to $\text{Mat}(H', M')$.

Let P_0, Q_0 be any regular $M' \times M', H' \times H'$ matrices, respectively.

Then there exist positive constants $\alpha > 0, \beta > 0$ such that

$$\alpha\|T\|^2 \leq \|P_0TQ_0\|^2 \leq \beta\|T\|^2.$$

Proof. There exists $\beta > 0$ such that

$$\|P_0TQ_0\|^2 \leq \beta\|T\|^2.$$

Also, there exists $\gamma > 0$

$$\|T\|^2 = \|P_0^{-1}P_0TQ_0Q_0^{-1}\|^2 \leq \gamma\|P_0TQ_0\|^2.$$

The proof follows by putting $\alpha = 1/\gamma$. \square

Lemma 5. Put

$$\Phi = \|BA - B_0A_0\|^2.$$

Then there exist a function Φ' and an a priori probability density function $\psi'(w')$ such that

- (a) $\Phi' = \|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|B_4A_4\|^2$, where C_1 is an $r \times r$ matrix, C_2 is an $(N-r) \times r$ matrix, C_3 is an $r \times (M-r)$ matrix, A_4 is an $(H-r) \times (M-r)$ matrix and B_4 is an $(N-r) \times (H-r)$ matrix.
- (b) $\psi'(w')$ is a C^∞ -function with compact support W' , where $\psi'(0) > 0$ and $w' = (C_1, C_2, C_3, B_4, A_4)$.
- (c) the maximum pole of $\int_W \Phi^z \psi dw$ is equal to that of $\int_{W'} \Phi'^z \psi' dw'$.

Proof. Since the rank of $B_0 A_0$ is r , there exist regular matrices P_0, Q_0 such that

$$P_0^{-1}B_0A_0Q_0^{-1} = \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix},$$

where E is the $r \times r$ identity matrix.

Change variables from B, A to B', A' by $B' = P_0^{-1}B$ and $A' = AQ_0^{-1}$.

Then

$$\Phi = \left\| P_0 \left(B'A' - \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} \right) Q_0 \right\|^2.$$

Let

$$A' = \begin{pmatrix} A_1 & A_3 \\ A_2 & A_4 \end{pmatrix},$$

and

$$B' = \begin{pmatrix} B_1 & B_3 \\ B_2 & B_4 \end{pmatrix},$$

where

A_1 is an $r \times r$ matrix, A_3 is an $r \times (M-r)$ matrix,
 A_2 is an $(H-r) \times r$ matrix, A_4 is an $(H-r) \times (M-r)$ matrix,
 B_1 is an $r \times r$ matrix, B_3 is an $r \times (H-r)$ matrix,
 B_2 is an $(N-r) \times r$ matrix, B_4 is an $(N-r) \times (H-r)$ matrix.

Let $U_{(A',B')}$ be a sufficiently small neighborhood of any point (A',B') with

$$B'A' - \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

Since the rank $(B_1, B_3) \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ is r , we can assume A_1 is regular. Thus we can change the variables from B_1, B_2 to C_1, C_2 by $C_1 = B_1A_1 + B_3A_2 - E$ and $C_2 = B_2A_1 + B_4A_2$.

Thus,

$$B'A' - \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} C_1 & (C_1 + E - B_3A_2)A_1^{-1}A_3 + B_3A_4 \\ C_2 & (C_2 - B_4A_2)A_1^{-1}A_3 + B_4A_4 \end{pmatrix}.$$

Changing the variables from A_4 to A'_4 by $A'_4 = -A_2A_1^{-1}A_3 + A_4$ gives

$$B'A' - \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} C_1 & C_1A_1^{-1}A_3 + A_1^{-1}A_3 + B_3A'_4 \\ C_2 & C_2A_1^{-1}A_3 + B_4A'_4 \end{pmatrix}.$$

By changing the variables from A_3 to A'_3 by $A'_3 = A_1^{-1}A_3 + B_3A'_4$, we obtain

$$B'A' - \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} C_1 & C_1(A'_3 - B_3A'_4) + A'_3 \\ C_2 & C_2(A'_3 - B_3A'_4) + B_4A'_4 \end{pmatrix},$$

and

$$\Phi = \left\| P_0 \begin{pmatrix} C_1 & C_1(A'_3 - B_3A'_4) + A'_3 \\ C_2 & C_2(A'_3 - B_3A'_4) + B_4A'_4 \end{pmatrix} Q_0 \right\|^2.$$

By Lemmas 2 and 4, the maximum pole of $\int_{U_{(A',B')}} \Phi^z \psi dw$ is equal to that of

$$\int_{U_{(A',B')}} \left\| \begin{pmatrix} C_1 & C_1(A'_3 - B_3A'_4) + A'_3 \\ C_2 & C_2(A'_3 - B_3A'_4) + B_4A'_4 \end{pmatrix} \right\|^{2z} \psi dw.$$

Then Lemmas 2 and 3 yield that the maximum pole of $\int_{U_{(A',B')}} \Phi^z \psi \, dw$ is equal to that of

$$\int_{U_{(A',B')}} \left\| \begin{pmatrix} C_1 & A'_3 \\ C_2 & B_4 A'_4 \end{pmatrix} \right\|^{2z} \psi \, dw.$$

Let $C_3 = A'_3, A_4 = A'_4$ and

$$\psi'(C_1, C_2, C_3, A_4, B_4) = \psi(A, B).$$

The proof follows from the fact that the poles of the above function are same when (A', B') with

$$B'A' - \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} = 0$$

varies. \square

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

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

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

instead of

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

Proof of Main Theorem

$$\text{Let } A_4 = \begin{pmatrix} a_{11} & \cdots & a_{1,M-r} \\ a_{21} & \cdots & a_{2,M-r} \\ \vdots & & \\ a_{H-r,1} & \cdots & a_{H-r,M-r} \end{pmatrix},$$

$$B_4 = \begin{pmatrix} b_{11} & \cdots & b_{1,H-r} \\ b_{21} & \cdots & b_{2,H-r} \\ \vdots & & \\ b_{N-r,1} & \cdots & b_{N-r,H-r} \end{pmatrix}$$

Suppose that C_1, C_2 and C_3 are as in Lemma 5. Denote $C_1 = (c_{ij}^{(1)})$, $C_2 = (c_{ij}^{(2)})$ and $C_3 = (c_{ij}^{(3)})$. We need to calculate poles of the following function by using the blowing-up process together with an inductive method.

Let $\ell(j) = (N + M)r - r^2 + j(N - r) + (M - r - j) \times (H - r - j) - 1$ for $j=0, \dots, \min\{H-r, M-r\}$.

Assume

$$\Phi'' = u_{11}^2 \dots u_{ss}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \sum_{i=1}^s \|\mathbf{b}_i\|^2 + \left\| \sum_{i=1}^s \mathbf{b}_i D_i + B^{(s+1)} A^{(s+1)} \right\|^2 \right), \tag{4}$$

where

$$B^{(s+1)} = \begin{pmatrix} b_{1,s+1} & \cdots & b_{1,H-r} \\ b_{2,s+1} & \cdots & b_{2,H-r} \\ \vdots & & \\ b_{N-r,s+1} & \cdots & b_{N-r,H-r} \end{pmatrix},$$

$$A^{(s+1)} = \begin{pmatrix} a_{s+1,s+1} & \cdots & a_{s+1,M-r} \\ a_{s+2,s+1} & \cdots & a_{s+2,M-r} \\ \vdots & & \\ a_{H-r,s+1} & \cdots & a_{H-r,M-r} \end{pmatrix} \text{ and}$$

$$\mathbf{b}_i = \begin{pmatrix} b_{1i} \\ \vdots \\ b_{N-r,i} \end{pmatrix} \text{ for } i = 1, \dots, H - r.$$

$D_i(a_{kl})$ is a function, defined on the entries of the matrix, obtained from A_4 by omitting the entries of $A^{(s+1)}$. The definition of the function $D_i(a_{kl})$ will be given recursively later on in Eq. (5).

Also we inductively have poles

$$\begin{aligned} & -\frac{\ell(s) + 1}{2} \\ & = -\frac{(N + M)r - r^2 + s(N - r) + (M - r - s)(H - r - s)}{2}. \end{aligned}$$

(Basis of the induction). Construct the blowing up of Φ' along the submanifold $\{C_1 = C_2 = C_3 = A_4 = 0\}$

$$\text{Let } \begin{cases} c_{11}^{(1)} = v, \\ c_{ij}^{(1)} = v c_{ij}^{(1)}, \quad (i, j) \neq (1, 1), \\ C_2 = v C_2, \quad C_3 = v C_3, \quad A_4 = v A_4. \end{cases}$$

Then we have

$$\Phi' = v^2 \left(1 + \sum_{(i,j) \neq (1,1)} (c_{ij}^{(1)})^2 + \|C_2\|^2 + \|C_3\|^2 + \|B_4 A_4\|^2 \right).$$

Here, the Jacobian is $v^{\ell(0)}$. Therefore, we have the pole

$$-\frac{\ell(0) + 1}{2},$$

since

$$\Phi'^z \, dw' = \Phi'^z v^{\ell(0)} \, dv$$

$$\times \prod_{(i,j) \neq (1,1)} dc_{ij}^{(1)} \prod_{(i,j)} dc_{ij}^{(2)} \prod_{(i,j)} dc_{ij}^{(3)} \prod_{(i,j)} da_{ij} \prod_{(i,j)} db_{ij},$$

in this coordinate.

If we set the general case as $c_{j,i}^{(1)} = v$, $c_{j,i}^{(2)} = v$, $c_{j,i}^{(3)} = v$, we obtain the same pole.

Consider another transformation.

$$\text{Let } \begin{cases} C_1 = u_{11}C_1, & C_2 = u_{11}C_2, & C_3 = u_{11}C_3, \\ a_{11} = u_{11}, \\ a_{ij} = u_{11}a_{ij}, & (i,j) \neq (1,1). \end{cases}$$

By the symmetry of the norm function, this setting is the general case as $a_{i,j} = u_{11}$.

Then we have

$$\Phi' = u_{11}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|\mathbf{b}_1\right. \\ \left. + B^{(2)}\mathbf{a}_1\|^2 + \left\| \left(\mathbf{b}_1 B^{(2)} \right) \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A^{(2)} \end{pmatrix} \right\|^2 \right),$$

where $\tilde{\mathbf{a}}_1 = (a_{12} \dots a_{1,M-r})$ and $\mathbf{a}_1 = (a_{21} \dots a_{H-r,1})^T$. T denotes the transpose.

Put $\mathbf{b}_1 = \mathbf{b}_1 + B^{(2)}\mathbf{a}_1$. Then

$$\Phi' = u_{11}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|\mathbf{b}_1\|^2 \right. \\ \left. + \left\| \left(\mathbf{b}_1 - B^{(2)}\mathbf{a}_1 \quad B^{(2)} \right) \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A^{(2)} \end{pmatrix} \right\|^2 \right) \\ = u_{11}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|\mathbf{b}_1\|^2 \right. \\ \left. + \left\| \left(\mathbf{b}_1 \quad 0 \right) \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A^{(2)} \end{pmatrix} + B^{(2)} \left(-\mathbf{a}_1 \quad E \right) \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ A^{(2)} \end{pmatrix} \right\|^2 \right) \\ = u_{11}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|\mathbf{b}_1\|^2 \right. \\ \left. + \|\mathbf{b}_1\tilde{\mathbf{a}}_1 + B^{(2)}(-\mathbf{a}_1\tilde{\mathbf{a}}_1 + A^{(2)})\|^2 \right).$$

Let $A^{(2)} = -\mathbf{a}_1\tilde{\mathbf{a}}_1 + A^{(2)}$, then we have Eq. (4) with $s=1$:

$$\Phi' = u_{11}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \|\mathbf{b}_1\|^2 \right. \\ \left. + \|\mathbf{b}_1\tilde{\mathbf{a}}_1 + B^{(2)}A^{(2)}\|^2 \right).$$

The Jacobian of this setting is $u_{11}^{\ell(0)}$.

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

Now we apply the induction method to Eq. (4).

(Inductive step). Construct the blowing up of Φ'' in (4) along the submanifold $\{C_1 = C_2 = C_3 = \mathbf{b}_i = A^{(s+1)} = 0, 1 \leq i \leq s\}$.

$$\text{Let } \begin{cases} c_{11}^{(1)} = v, \\ c_{ij}^{(1)} = vc_{ij}^{(1)}, & (i,j) \neq (1,1), \\ \mathbf{b}_j = v\mathbf{b}_j, & 1 \leq j \leq s, & C_2 = vC_2, & C_3 = vC_3, \\ A^{(s+1)} = vA^{(s+1)}. \end{cases}$$

Substituting them into Eq. (4) gives

$$\Phi'' = u_{11}^2 \dots u_{ss}^2 v^2 \left(1 + \sum_{(i,j) \neq (1,1)} (c_{ij}^{(1)})^2 + \|C_2\|^2 + \|C_3\|^2 \right. \\ \left. + \sum_{i=1}^s \|\mathbf{b}_i\|^2 + \left\| \sum_{i=1}^s \mathbf{b}_i D_i + B^{(s+1)}A^{(s+1)} \right\|^2 \right).$$

Here, the Jacobian is $u_{11}^{\ell(0)} \dots u_{ss}^{\ell(s-1)} v^{\ell(s)}$.

Because

$$\Phi'^z dw' = \Phi''^z u_{11}^{\ell(0)} \dots u_{ss}^{\ell(s-1)} v^{\ell(s)} dw'',$$

in this new coordinate w'' , we have the poles

$$-\frac{\ell(0)+1}{2}, \dots, -\frac{\ell(s)+1}{2}.$$

If we set the general case as $c_{j,i}^{(1)} = u$, $c_{j,i}^{(2)} = u$, $c_{j,i}^{(3)} = u$, $b_{j,i} = u$, we obtain the same pole by symmetry

$$\text{Next let } \begin{cases} a_{s+1,s+1} = u_{s+1,s+1}, \\ a_{j\ell} = u_{s+1,s+1}a_{j\ell}, & s+1 \leq j \leq H-r, \\ s+1 \leq \ell \leq M-r, & (j,\ell) \neq (s+1, s+1), \\ C_1 = u_{s+1,s+1}C_1, & C_2 = u_{s+1,s+1}C_2, \\ C_3 = u_{s+1,s+1}C_3, & \mathbf{b}_i = u_{s+1,s+1}\mathbf{b}_i, \quad 1 \leq i \leq s. \end{cases}$$

We also obtain the same pole by setting $a_{j\ell} = u_{s+1,s+1}$ for any (j,ℓ) .

Substituting our new variables into Eq. (4) implies

$$\Phi'' = u_{11}^2 \dots u_{ss}^2 u_{s+1,s+1}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 \right. \\ \left. + \sum_{i=1}^s \|\mathbf{b}_i\|^2 + \left\| \sum_{i=1}^s \mathbf{b}_i D_i \right. \right. \\ \left. \left. + \left(\mathbf{b}_{s+1} \quad B^{(s+2)} \right) \begin{pmatrix} 1 & \tilde{\mathbf{a}}_{s+1} \\ \mathbf{a}_{s+1} & A^{(s+2)} \end{pmatrix} \right\|^2 \right) \\ = u_{11}^2 \dots u_{ss}^2 u_{s+1,s+1}^2 \left(\|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 \right. \\ \left. + \sum_{i=1}^s \|\mathbf{b}_i\|^2 + \left\| \sum_{i=1}^s \mathbf{b}_i D_i + \left(\mathbf{b}_{s+1} + B^{(s+2)}\mathbf{a}_{s+1} \quad 0 \right) \right. \right. \\ \left. \left. + \left(\mathbf{b}_{s+1} \quad B^{(s+2)} \right) \begin{pmatrix} 0 & \tilde{\mathbf{a}}_{s+1} \\ A^{(s+2)} \end{pmatrix} \right\|^2 \right),$$

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

Denote the first column of D_i by $\text{Col}_1(D_i)$. Let $D_i = (\text{Col}_1(D_i) \quad D_i')$.

Put $\mathbf{b}_{s+1} = \mathbf{b}_{s+1} + B^{(s+2)}\mathbf{a}_{s+1} + \sum_{i=1}^s \mathbf{b}_i \text{Col}_1(D_i)$. Then

$$\begin{aligned} & \Phi''/u_{11}^2 \cdots u_{s+1,s+1}^2 \\ &= \|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \sum_{i=1}^s \|\mathbf{b}_i\|^2 + \|\mathbf{b}_{s+1}\|^2 \\ &+ \left\| \sum_{i=1}^s \mathbf{b}_i D'_i + \left(\mathbf{b}_{s+1} - B^{(s+2)}\mathbf{a}_{s+1} \right. \right. \\ &\left. \left. - \sum_{i=1}^s \mathbf{b}_i \text{Col}_1(D_i) B^{(s+2)} \right) \begin{pmatrix} \tilde{\mathbf{a}}_{s+1} \\ A^{(s+2)} \end{pmatrix} \right\|^2 \\ &= \|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \sum_{i=1}^{s+1} \|\mathbf{b}_i\|^2 \\ &+ \left\| \sum_{i=1}^s \mathbf{b}_i D'_i + \left(\mathbf{b}_{s+1} - \sum_{i=1}^s \mathbf{b}_i \text{Col}_1(D_i) \ 0 \right) \begin{pmatrix} \tilde{\mathbf{a}}_{s+1} \\ A^{(s+2)} \end{pmatrix} \right. \\ &\left. + \left(-B^{(s+2)}\mathbf{a}_{s+1} \ B^{(s+2)} \right) \begin{pmatrix} \tilde{\mathbf{a}}_{s+1} \\ A^{(s+2)} \end{pmatrix} \right\|^2 \\ &= \|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \sum_{i=1}^{s+1} \|\mathbf{b}_i\|^2 \\ &+ \left\| \sum_{i=1}^s \mathbf{b}_i (D'_i - \text{Col}_1(D_i)\tilde{\mathbf{a}}_{s+1}) + \mathbf{b}_{s+1}\tilde{\mathbf{a}}_{s+1} \right. \\ &\left. + B^{(s+2)}(-\mathbf{a}_{s+1}, E) \begin{pmatrix} \tilde{\mathbf{a}}_{s+1} \\ A^{(s+2)} \end{pmatrix} \right\|^2 \\ &= \|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \sum_{i=1}^{s+1} \|\mathbf{b}_i\|^2 \\ &+ \left\| \sum_{i=1}^s \mathbf{b}_i (D'_i - \text{Col}_1(D_i)\tilde{\mathbf{a}}_{s+1}) + \mathbf{b}_{s+1}\tilde{\mathbf{a}}_{s+1} \right. \\ &\left. + B^{(s+2)}(-\mathbf{a}_{s+1}\tilde{\mathbf{a}}_{s+1} + A^{(s+2)}) \right\|^2. \end{aligned}$$

Now let $A^{(s+2)} = -\mathbf{a}_{s+1}\mathbf{a}_{s+1} + A^{(s+2)}$. Then,

$$\begin{aligned} & \Phi''/u_{11}^2 \cdots u_{s+1,s+1}^2 \\ &= \|C_1\|^2 + \|C_2\|^2 + \|C_3\|^2 + \sum_{i=1}^{s+1} \|\mathbf{b}_i\|^2 \\ &+ \left\| \sum_{i=1}^s \mathbf{b}_i (D'_i - \text{Col}_1(D_i)\tilde{\mathbf{a}}_{s+1}) + \mathbf{b}_{s+1}\tilde{\mathbf{a}}_{s+1} \right. \\ &\left. + B^{(s+2)}A^{(s+2)} \right\|^2. \end{aligned}$$

The Jacobian is $u_{11}^{\ell(0)} \cdots u_{s+1,s+1}^{\ell(s)}$.

Repeat this whole process by setting

$$D_i = D'_i - \text{Col}_1(D_i)\tilde{\mathbf{a}}_{s+1} (1 \leq i \leq s) \quad \text{and} \quad D_{s+1} = \tilde{\mathbf{a}}_{s+1}. \tag{5}$$

Then, s will be replaced by $s+1$ in (4) and so on. Therefore, we obtain poles

$$-\frac{\ell(s) + 1}{2},$$

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

(i) If $(M+H-N-r)/2 < 0$, then the maximum pole at $s=0$ is

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

and its order m is 1.

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

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

and its order m is 1.

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

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

and its order m is 2.

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

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

and its order m is 1.

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

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

and its order m is 1.

So Main Theorem follows.

Remark 1. Let $g(w)$ be a C^∞ -function with compact support W and $g(w_0) \neq 0$ for a fixed parameter $w_0 \in W$. Let $f_1(w), f_2(w)$ be analytic functions of $w \in W$ with $f_1(w_0) = f_2(w_0) = 0$. Assume that $0 \leq \alpha f_1 \leq f_2 \leq \beta f_1$ for any constants $\alpha > 0$ and $\beta > 0$.

By the assumption together with the Hironaka's theorem, we have a proper analytic map μ from an analytic manifold U to a neighborhood V of w_0 satisfying the followings;

- (1) $\mu : U - \mathcal{E} \rightarrow V - f_1^{-1}(0)$ is an isomorphism, where $\mathcal{E} = \mu^{-1}(f_1^{-1}(0))$.
- (2) For each $u \in U$, there is a local analytic coordinate (u_1, \dots, u_d) such that $f_1(\mu(u)) = u_1^{2s_1} u_2^{2s_2} \cdots u_d^{2s_d} f'_1$ and $f_2(\mu(u)) = u_1^{2s_1} u_2^{2s_2} \cdots u_d^{2s_d} f'_2$, where s_1, \dots, s_d are non-negative integers and $f'_1 \neq 0, f'_2 \neq 0$.

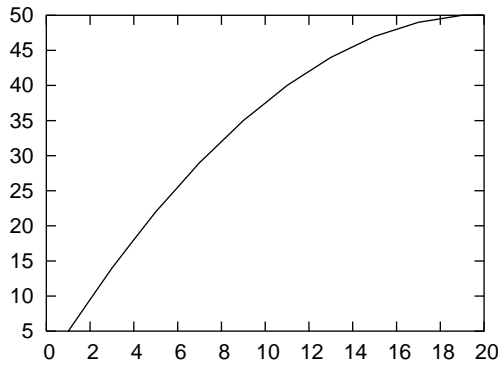


Fig. 1. The curve of λ -values in y-axis and H -values in x-axis, when $M=N=10$.

The Jacobian of μ is $J_\mu(u_1, \dots, u_d) = u_1^{m_1} \cdots u_d^{m_d} \tilde{J}_\mu(u_1, \dots, u_d)$, where m_1, \dots, m_d are non-negative integers and $\tilde{J}_\mu(0, \dots, 0) \neq 0$.

Then, $\int_W f_1(w)^z g(w) dw$ and $\int_W f_2(w)^z g(w) dw$ have poles $-\frac{m_1 + 1}{2s_1}, \dots, -\frac{m_d + 1}{2s_d}$.

Remark 2. The blowing-up computation in the proof of Main Theorem shows that $\int_{W'} \Phi^{l'z} \psi^{l'} dw^{l'}$ have always poles $-(\ell(s) + 1)/2$ for any $\psi^{l'}$ with $\psi^{l'}(A_0, B_0) > 0$. Furthermore, there is a maximum among the poles $-(\ell(s) + 1)/2$. From Remark 1, $\int_W \Phi^z \psi dw$ and $\int_{W'} \Phi^{l'z} \psi^{l'} dw^{l'}$ have poles $-(\ell(s) + 1)/2$. Note that $\int_W \Phi^z \psi dw$ and $\int_{W'} \Phi^{l'z} \psi^{l'} dw^{l'}$ have many other poles than $-(\ell(s) + 1)/2$.

5. Discussion and conclusion

In this paper, we introduce a computational method to obtain the poles of the zeta functions for the reduced rank regression model.

Note that if the rank r of $A_0 B_0$ is zero, then H, M and N can be permuted in the formula for λ in Main Theorem.

Fig. 1 shows the graphs of the maximum poles λ with λ -values in y-axis and H -values in x-axis, when $M=N=10$ and $r=0$. It is clear that the curve is not linear. If the reduced rank approximation was a regular statistical model, λ would be $(M+N)H/2$ and linear. The behaviors of λ for regular and no-regular models are so different.

In this paper, we assume that $\int (y|x)^2 q(x) dx = 0$, if and only if $y=0$, where $(y|x)$ is the inner product of y and x . If $\int (y_0|x)^2 q(x) dx = 0$ for some $y_0 \neq 0 \in \mathbf{R}^M$, then $q(x)$ is the function defined on the hypersurface $(y_0|x)=0$. Then the dimension becomes $M-1$. So the assumption is natural.

Algebraic methods can be effectively used to solve the problems in Learning theory.

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

Acknowledgements

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

References

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

Atiyah, M. F. (1970). Resolution of singularities and division of distributions. *Communications on Pure and Applied Mathematics*, 13, 145–150.

Baldi, P., & Hornik, K. (1995). Learning in linear networks: A survey. *IEEE Transactions on Neural Networks*, 6(4), 837–858.

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

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

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

Murata, N. J., Yoshizawa, S. G., & Amari, S. (1994). Network information criterion—Determining the number of hidden units for an artificial neural network model. *IEEE Transactions on Neural Networks*, 5(6), 865–872.

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

Sato, M., & Shintani, T. (1974). On zeta functions associated with prehomogeneous vector space. *Annals of Mathematics*, 100, 131–170.

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

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

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

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

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

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.

Watanabe, K., & Watanabe, S. (2003). Upper bounds of bayesian generalization errors in reduced rank regression. *IEICE Transactions*, J86-A(3), 278–287 (in Japanese).