Neural Networks 172 (2024) 106132

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Full Length Article

Consideration on the learning efficiency of multiple-layered neural networks with linear units

Miki Aoyagi

College of Science & Technology, Nihon University, 1–8–14, Surugadai, Kanda, Chiyoda-ku, Tokyo 101–8308, Japan

ARTICLE INFO

Keywords:
Resolution map
Singular learning theory
Multiple-layered neural networks with linear units
Algebraic geometry

ABSTRACT

In the last two decades, remarkable progress has been done in singular learning machine theories on the basis of algebraic geometry. These theories reveal that we need to find resolution maps of singularities for analyzing asymptotic behavior of state probability functions when the number of data increases. In particular, it is essential to construct normal crossing divisors of average log loss functions. However, there are few examples for obtaining these for singular models. In this paper, we determine the resolution map and normal crossing divisors for multiple-layered neural networks with linear units. Moreover, we have the exact values for the learning efficiency, which is so called learning coefficients. Multiple-layered neural networks with linear units are simple, however, very important models because these models give the essential information from data of input–output pairs. Moreover, these models are very close to multiple-layered neural networks with rectified linear units (ReLU). We show the learning coefficients of multiple-layered neural networks with linear units are bounded even though the number of layers goes to infinity, which means that the main term of asymptotic expansion of the free energy and generalization error of singular models are much smaller than the dimension of its parameter space.

1. Introduction

Singular learning models include hierarchical learning machines such as neural networks, reduce rank regression, Boltzmann machine, normal mixture models and so on. Even though these models have been widely used to analyze real data, theoretical analysis have not been developed sufficiently. These models have non-positive Fisher information matrices and therefore cannot be analyzed by classical theories (Watanabe, 2009).

For example, let us consider the reduce rank regression model, which is one of singular models. Let

$$\{w = (A^{(1)}, A^{(2)}) \mid \ A^{(s)} \text{ is an } H^{(s)} \times H^{(s+1)} \text{ matrix } \},$$

be the set of parameters. Assume that the input value is a vector $X \in \mathbf{R}^{H^{(3)}}$ with probability density function $r(X)$ and output value a vector $Y \in \mathbf{R}^{H^{(1)}}$ given by

$$Y = A^{(1)}A^{(2)}X + (\text{noise}),$$

with Gaussian noise. Then, the statistical learning model is obtained by

$$pl(X,Y|w) = \frac{1}{(\sqrt{2\pi})^{H^{(1)}}} \exp(-\frac{1}{2}\|Y - A^{(1)}A^{(2)}X\|^2)r(X).$$

This model has $H^{(3)}$ input units, $H^{(1)}$ output units, and $H^{(2)}$ hidden units. Let us assume that the $w^* = (A^{*(1)}, A^{*(2)})$ be the true parameter, that is, we have the true probability density function

$$r(X,Y) = \frac{1}{(\sqrt{2\pi})^{H^{(1)}}} \exp(-\frac{1}{2}\|Y - A^{*(1)}A^{*(2)}X\|^2)r(X).$$

Let $(X,Y)^n := \{(X_i,Y_i)\}_{i=1}^n$ be n training samples selected independently and identically from $r(X,Y)$.

Define the average log loss function $L(w)$ by $L(w) = -E_{X,Y}[\log pl(X,Y|w)]$ and the set of optimal parameters W_0 by

$$W_0 = \{w_0 \in W \,|\, L(w_0) = \min_{w' \in W} L(w')\} = \{(A^{(1)}, A^{(2)}) \mid A^{*(1)}A^{*(2)} = A^{(1)}A^{(2)}\}.$$

Because $r(X,Y) = pl(X,Y|(A^{*(1)}, A^{*(2)}))$, we have $r(X,Y) = pl(X,Y|w_0)$ for all $w_0 \in W_0$.

If the rank of $A^{*(1)}A^{*(2)}$ is less than $\max\{H^{(1)}, H^{(2)}, H^{(3)}\}$, then W_0 is not one point, but its dimension is positive. It is obvious that the Fisher matrix function $(\frac{\partial^2 L(w)}{\partial w_i \partial w_j})$ at $w^* = (A^{*(1)}, A^{*(2)})$ has zero eigenvalues, therefore, we cannot apply classical theories, 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), because their methods need regular conditions such as a positive definite Fisher information matrix and the unique point minimizing a log loss function.

E-mail address: aoyagi.miki@nihon-u.ac.jp.

https://doi.org/10.1016/j.neunet.2024.106132
Received 27 March 2023; Received in revised form 10 November 2023; Accepted 15 January 2024
Available online 17 January 2024

In general, let an unknown true probability density function $r(X,Y)$, and let $(X,Y)^n := \{(X_i,Y_i)\}_{i=1}^n$ be n training samples selected independently and identically from $r(X,Y)$. Consider a learning model that is written in probabilistic form as $pl(X,Y|w)$, where $w \in W \subset \mathbf{R}^d$ is a parameter. Let $\varphi(w)$ be an *a priori* probability density function on a parameter set W and $ps_n(w)$ be the *a posteriori* probability density function,

$$ps_n(w) = \frac{1}{\int_W \varphi(w)\prod_{i=1}^n pl(X_i,Y_i|w)\mathrm{d}w}\varphi(w)\prod_{i=1}^n pl(X_i,Y_i|w).$$

We then have the predictive density function in Bayesian estimation

$$pr_n(X,Y) = \int pl(X,Y|w)ps_n(w)dw.$$

Assume that we have $pl_0(X,Y) = pl(X,Y|w_0)$ for all $w_0 \in W_0$, that is, $pl_0(X,Y)$ is essentially unique for $r(X,Y)$, where

$$L(w) = -E_{X,Y}[\log pl(X,Y|w)],$$

and

$$W_0 = \{w \in W | L(w) = \min_{w' \in W} L(w')\}.$$

Let G_n be the Bayes generalization loss,

$$\begin{aligned} G_n &= -\int r(X,Y)\log pr_n(X,Y)dXdY \\ &= -\int r(X,Y)\log r(X,Y)dXdY + \int r(X,Y)\log\frac{r(X,Y)}{pr_n(X,Y)}dXdY \\ &= -\int r(X,Y)\log r(X,Y)dXdY + E_{X,Y}[\log\frac{r(X,Y)}{pr_n(X,Y)}]. \end{aligned}$$

$E_{X,Y}[\log\frac{r(X,Y)}{pr_n(X,Y)}]$ is the Kullback function, which is the distance between $r(X,Y)$ and $pr_n(X,Y)$. Watanabe (2001a, 2001b, 2010, 2018) proved the following relation,

$$E[G_n] = L(w_0) + \frac{1}{n}\lambda + o(\frac{1}{n}),$$

where λ is the learning coefficient. In the case of reduce rank regression, the value of λ was obtained (Aoyagi & Watanabe, 2005b, Theorem 2 in this paper). These values are much smaller than the dimension $H^{(1)}H^{(2)} + H^{(2)}H^{(3)}$ of parameter spaces. This fact means learning efficiencies are increased more than regular models, as a model's complexity increases.

Let

$$L_n(w) = -\frac{1}{\beta n}\log\prod_{i=1}^n pl(X_i,Y_i|w)^\beta.$$

Based on the free energy,

$$\begin{aligned} F_n(\beta) &= -\frac{1}{\beta}\log\int\prod_{i=1}^n pl(X_i,Y_i|w)^\beta\varphi(w)dw \\ &= nL_n(w_0) + \frac{\lambda}{\beta}\log(n) - \frac{\theta-1}{\beta}\log\log(n) + o_p(1) \end{aligned}$$

for inverse temperature $\beta > 0$, which was shown by Watanabe (2009), we have the model-selection methods, the "widely applicable Bayesian information criterion" (WBIC) (Watanabe, 2013) and "singular Bayesian information criterion" (sBIC Drton, 2012). sBIC uses the learning coefficients λ and its order θ very effectively with a fix point equation system of marginal likelihoods. In addition, the learning efficiency λ and its order θ of a leaning model give mathematical indicators for analyzing and developing the precision of numerical methods, such as the widely-applicable information criterion WAIC (Akaike, 1974; Watanabe, 2001a, 2001b, 2001c, 2009, 2010, 2018) and cross-validation in the Bayesian estimation approach model selection methods.

The learning efficiency λ is equal to a log canonical threshold of a Kullback function in algebraic geometry whose exact value is obtained using recursive blow-up process (Hironaka, 1964). However, it is difficult to obtain these thresholds, because we need to determine all branches of its blow-up process to obtain normal crossing divisors and Kullback functions degenerate with respect to their Newton polyhedra (Fulton, 1993). Moreover, theories for log canonical thresholds are not fruitful for the real field, compared with the complex field or algebraically closed fields (Kashiwara, 1976; Kollár, 1997; Mustata, 2002). Therefore, it is of interest in various fields, even in mathematics, to obtain these thresholds. In recent studies, we obtained exact values or bounded values of the learning coefficients for Vandermonde matrix-type singularities, which are related to the three-layered neural networks and normal mixture models, among others (Aoyagi, 2006, 2013a, 2019a, 2019b; Aoyagi & Watanabe, 2005a). We have also exact values for the restricted Boltzmann machine (Aoyagi, 2013b). Additionally, Rusakov and Geiger (2002, 2005), Zwiernik (2011) and Drton, Lin, Weihs, and Zwiernik (2017) respectively, obtained these coefficients for naive Bayesian networks, directed tree models with hidden variables, and the Gaussian latent tree and forest models.

Our purpose in this paper is to obtain λ and θ for multiple-layered neural networks with linear units.

In Section 2, we overview the definitions of resolution of singularities and log canonical thresholds, and we show our main results in Section 3. In Section 4, the main theorem's proof is obtained, and we conclude in Section 5.

2. Resolution of singularities and log canonical threshold

We denote constants by superscript $*$, for example, a^*, b^*, and w^*.

Definition 1. The log canonical threshold for an analytic function F and a C^∞ function $\varphi(w)$ with a compact support in a neighborhood U of w^*, is defined as

$$\lambda_{w^*}(F,\varphi) = \sup\{c : \int_U |F|^{-kc}\varphi(w)dw(d\overline{w})^{k-1} < \infty\},$$

where $k=1$ over the real field and $k=2$ over the complex field. The value $-\lambda_{w^*}(F,\varphi)$ is equal to the largest pole of the zeta function $\int_U |F|^{kz}\varphi(w)dw(d\overline{w})^{k-1}$ for $z \in \mathbf{C}$, respectively. Also let $\theta_{w^*}(F,\varphi)$ be its order.

If $\varphi(w^*) \neq 0$, then denote $\lambda_{w^*}(F) = \lambda_{w^*}(F,\varphi)$ and $\theta_{w^*}(F) = \theta_{w^*}(F,\varphi)$ because the log canonical threshold and its order are independent of φ.

For ideal J, generated by real analytic functions $F_1,\ldots,F_m$ in a neighborhood of w^*, define $\lambda_{w^*}(J) = \lambda_{w^*}(F_1^2+\cdots+F_m^2)$.

Here, $\lambda_{w^*}(J)$ for ideal J is well-defined by Lemma 1.

Lemma 1 (*Aoyagi, 2009, 2010; Lin, 2010*)**.** *Let $\mathcal{J} = \langle F_1,\ldots,F_n\rangle$ be the ideal generated by analytic functions $F_1,\ldots,F_n$ on a neighborhood U of $w^* \in \mathbf{R}^d$. Also, let $G_1,\ldots,G_m$ be analytic functions on U.*

We have

$$\lambda_{w^*}(G_1^2+\cdots+G_m^2) \le \lambda_{w^*}(F_1^2+\cdots+F_n^2),$$

if $G_1,\ldots,G_m \in \mathcal{J}$.

In particular, if $\mathcal{J} = \langle G_1,\ldots,G_m\rangle$, then we have

$$\lambda_{w^*}(F_1^2+\cdots+F_n^2) = \lambda_{w^*}(G_1^2+\cdots+G_m^2).$$

Definition 2. Let $C = (c_{ij})$ be a matrix. Define $\|C\|$ as the norm of C by $\|C\| = \sqrt{\sum_{i,j}|c_{ij}|^2}$. Also define $\langle C\rangle$ as the ideal generated by all elements c_{ij} of C.

Applying Hironaka's Theorem (Hironaka, 1964) to the function

$$K(w) = E_{X,Y}[\log\frac{pl_0(X,Y)}{pl(X,Y|w)}] = \int r(X,Y)\log\frac{pl_0(X,Y)}{pl(X,Y|w)}dXdY,$$

where $pl(X,Y|w)$ is a learning model with a parameter $w \in W$ and $pl_0(X,Y|w) = p(X,Y|w_0)$ for all $w_0 \in W_0$, we obtain the proper analytic map π from a manifold Q to neighborhood $V \subset W$,

$$K(\pi(u)) = u_1^{2k_1}u_2^{2k_2}\cdots u_d^{2k_d},$$

$$\pi'(u)\varphi(\pi(u)) = u_1^{h_1} u_2^{h_2} \cdots u_d^{h_d},$$

where $\varphi(w)$ is an a priori density function, $(u_1, \ldots, u_d)$ is a local analytic coordinate system on $U \subset Q$, and $k_1, \ldots, k_d, h_1, \ldots, h_d$ are non-negative integers. Then, we have

$$\int_V |K(w)|^{-c} \varphi dw = \int_Q |K(g'(u))|^{-c} \varphi(\pi(u))\pi'(u) du.$$

Therefore, the log canonical threshold, i.e., the learning coefficient is

$$\lambda = \min_U \min_{1 \le j \le d} \frac{h_j + 1}{2k_j},$$

and its order

$$\theta = \max_u \mathrm{Card}\{j : \frac{h_j + 1}{2k_j} = \lambda\},$$

where $\mathrm{Card}(S)$ denotes the cardinality of a set S.

3. Multiple-layered neural networks with linear units

Define matrices $A^{(s)}$ of size $H^{(s)} \times H^{(s+1)}$ for $s = 1, \ldots, L$,

$$A^{(s)} = (a_{ij}^{(s)}), (1 \le i \le H^{(s)}, 1 \le j \le H^{(s+1)}).$$

Let W be the set of parameters

$$W = \{w = \{A^{(s)}\}_{1 \le s \le L} \mid A^{(s)} \text{ is an } H^{(s)} \times H^{(s+1)} \text{ matrix }\}.$$

Denote the input value by $X \in \mathbf{R}^{H^{(L+1)}}$ with probability density function $r(X)$ and output value $Y \in \mathbf{R}^{H^{(1)}}$ for the multiple-layered neural network with linear units, which is given by

$$Y = \prod_{s=1}^{L} A^{(s)} X + (\text{noise}),$$

with Gaussian noise. Consider the statistical model

$$pl(Y|X, w) = \frac{1}{(\sqrt{2\pi})^{H^{(1)}}} \exp(-\frac{1}{2}\|Y - \prod_{s=1}^{L} A^{(s)} X\|^2),$$
$$pl(X, Y|w) = pl(Y|X, w) r(X).$$

The model has $H^{(L+1)}$ input units, $H^{(1)}$ output units, and $H^{(s)}$ hidden units in each hidden layer. Assume that $pl_0(X, Y)$ is essentially unique for $r(X, Y)$, and that

$$pl_0(Y|X) = \frac{1}{(\sqrt{2\pi})^{H^{(1)}}} \exp(-\frac{1}{2}\|Y - \prod_{s=1}^{L} A^{*(s)} X\|^2),$$
$$pl_0(X, Y) = pl_0(Y|X) r(X),$$

where

$$w^* = \{A^{*(s)}\}_{1 \le s \le L} \in W_0.$$

This is the over-parameterized case. Moreover, assume that the *a priori* probability density function $\varphi(w)$ is a $C^\infty-$ function with compact support W, satisfying $\varphi(w^*) > 0$. Then, the λ for the model corresponding to the log canonical threshold, denoted as $\lambda_{w^*}(\| \prod_{s=1}^{L} A^{(s)} - \prod_{s=1}^{L} A^{*(s)}\|^2)$, and its associated order, denoted as θ, are as follows.

Definition 3. Let r be the rank of $\prod_{s=1}^{L} A^{*(s)}$ and $M^{(s)} = H^{(s)} - r$ for $s = 1, \ldots, L+1$. Define $\mathcal{M} \subset \{1, \ldots, L+1\}$ such that

$$\ell = \mathrm{Card}(\mathcal{M}) - 1,$$
$$\mathcal{M} = \{S_1, \ldots, S_{\ell+1}\},$$
$$M^{(S_j)} < M^{(s)} \text{ for } S_j \in \mathcal{M} \text{ and } s \notin \mathcal{M},$$
$$\sum_{k=1}^{\ell+1} M^{(S_k)} \ge \ell M^{(s)} \text{ for } s \in \mathcal{M}$$
$$\sum_{k=1}^{\ell+1} M^{(S_k)} < \ell M^{(s)} \text{ for } s \notin \mathcal{M}.$$

Let M be the integer such that

$$M - 1 < \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell} \le M,$$

and

$$a = \sum_{k=1}^{\ell+1} M^{(S_k)} - (M-1)\ell.$$

Theorem 1. *We have*

$$\begin{aligned}\lambda &= \frac{-r^2 + r(H^{(1)} + H^{(L+1)})}{2} + \frac{a(\ell - a)}{4\ell} \\ &\quad - \frac{\ell(\ell-1)}{4}(\frac{\sum_{j=1}^{\ell+1} M^{(S_j)}}{\ell})^2 + \frac{1}{2}\sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)} \\ &= \frac{-r^2 + r(H^{(1)} + H^{(L+1)})}{2} \\ &\quad + \frac{a(\ell - a)}{4\ell} - \frac{\ell(\ell-1)}{4}(M + \frac{a-\ell}{\ell})^2 + \frac{1}{2}\sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)} \\ &= \frac{-r^2 + r(H^{(1)} + H^{(L+1)})}{2} - \frac{1}{4}(\ell - a - 1)(\ell - a) \\ &\quad - \frac{\ell(\ell-1)}{4}(M^2 + 2\frac{a-\ell}{\ell}M) + \frac{1}{2}\sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)}\end{aligned}$$

and

$$\theta = a(\ell - a) + 1.$$

Remark 1. If $\mathrm{Card}\{s \mid M^{(s)} = 0\} = 1$, then we have

$$\mathcal{M} = \{s \mid M^{(s)} = \min\{M^{(s')} \mid M^{(s')} > 0\}, \text{ or } M^{(s)} = 0\}.$$

If $\mathrm{Card}\{s \mid M^{(s)} = 0\} > 1$, then we have

$$\mathcal{M} = \{s \mid M^{(s)} = 0\}.$$

Some calculations show that for both cases, $\lambda = \frac{-r^2 + r(H^{(1)} + H^{(L+1)})}{2}$ and $\theta = 1$.

Example 1. If $M^{(1)} = M^{(2)} = \cdots = M^{(L+1)}$, then, we have $\ell = L$, the integer M such that $M - 1 < \frac{L+1}{L} M^{(1)} \le M$ and $a = (L+1)M^{(1)} - (M-1)L$.

We have

$$\lambda = \frac{-r^2 + 2rH^{(1)}}{2} + \frac{a(L-a)}{4L} + \frac{L+1}{4L}(M^{(1)})^2$$

and

$$\theta = a(L - a) + 1.$$

Fig. 1 shows that the curve of λ and θ, when $r = 0$, $M^{(1)} = 100$ and $L = 2, \ldots, 120$ in Example 1.

Theorem 2 (Aoyagi & Watanabe, 2005b). *In the paper Aoyagi and Watanabe (2005b), the learning coefficients for the reduced rank regression model which corresponds to $L = 2$ in Theorem 1 were obtained. We used the following expression. Note that $H^{(s)} = M^{(s)} + r$.*

(1) If $H^{(1)}+r \le H^{(2)}+H^{(3)}$, $H^{(2)}+r \le H^{(1)}+H^{(3)}$, $H^{(3)}+r \le H^{(1)}+H^{(2)}$ and $H^{(1)} + H^{(2)} + H^{(3)} + r$ is even, we have

- $\mathcal{M} = \{1, 2, 3\}$, $S_1 = 1, S_2 = 2, S_3 = 3$, $\ell = 2$,
- $M^{(s)} \le \frac{\sum_{j=1}^{3} M^{(S_j)}}{2}$ *for* $s = 1, 2, 3$,
- $M = \frac{\sum_{j=1}^{3} M^{(S_j)}}{2}$, $a = 2$,
- $\theta = 1$,

$$\lambda = \frac{-r^2 + r(H^{(1)} + H^{(3)})}{2} - \frac{1}{2}(\frac{\sum_{j=1}^{3} M^{(S_j)}}{2})^2 + \frac{1}{2}\sum_{1 \le i < j \le 3} M^{(S_i)} M^{(S_j)}$$

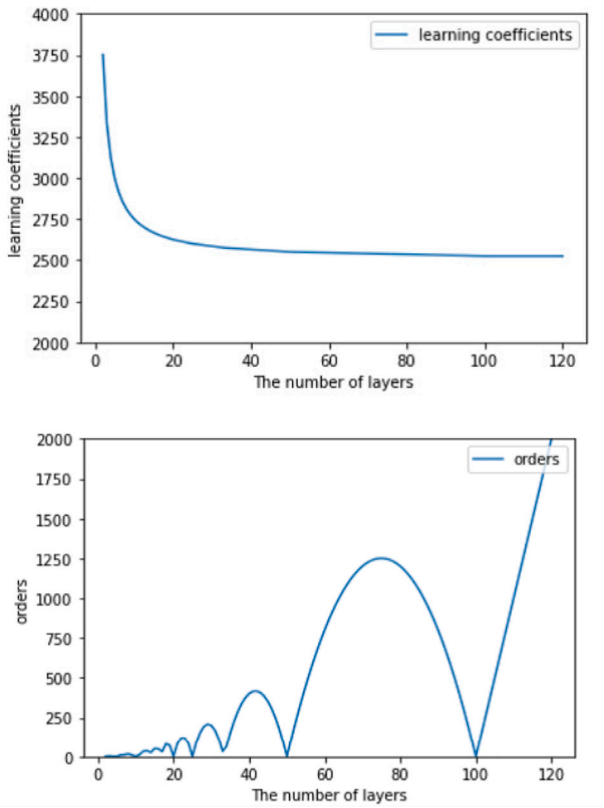

Fig. 1. The curves of λ and θ, when $r = 0$, $M^{(1)} = 100$ and $L = 2, \ldots, 120$ in Example 1.

$$= \frac{-(H^{(2)}+r)^2 - {H^{(1)}}^2 - {H^{(3)}}^2 + 2(H^{(2)}+r)H^{(1)}}{8} + \frac{2(H^{(2)}+r)H^{(3)} + 2H^{(1)}H^{(3)}}{8}.$$

(2) If $H^{(1)}+r \le H^{(2)}+H^{(3)}$, $H^{(2)}+r \le H^{(1)}+H^{(3)}$, $H^{(3)}+r \le H^{(1)}+H^{(2)}$ and $H^{(1)}+H^{(2)}+H^{(3)}+r$ is odd, we have

- $\mathcal{M} = \{1,2,3\}$, $S_1 = 1, S_2 = 2, S_3 = 3$, $\ell = 2$,
- $M^{(s)} \le \frac{\sum_{j=1}^{3} M^{(S_j)}}{2}$ *for* $s = 1,2,3$,
- $M = \frac{\sum_{j=1}^{3} M^{(S_j)}+1}{2}$, $a = 1$,
- $\theta = 2$,

$$\begin{aligned}\lambda &= \frac{-r^2 + r(H^{(1)}+H^{(3)})}{2} + \frac{1}{8} \\ &\quad -\frac{1}{2}\left(\frac{\sum_{j=1}^{3} M^{(S_j)}}{2}\right)^2 + \frac{1}{2}\sum_{1\le i<j\le 3} M^{(S_i)}M^{(S_j)} \\ &= \frac{-(H^{(2)}+r)^2 - {H^{(1)}}^2 - {H^{(3)}}^2 + 2(H^{(2)}+r)H^{(1)}}{8} \\ &\quad + \frac{2(H^{(2)}+r)H^{(3)} + 2H^{(1)}H^{(3)} + 1}{8}.\end{aligned}$$

(3) If $H^{(1)} + H^{(2)} < H^{(3)} + r$, we have

- $\mathcal{M} = \{1,2\}$, $S_1 = 1, S_2 = 2$, $\ell = 1$,
- $M^{(s)} \le \sum_{j=1}^{2} M^{(S_j)}$ *for* $s = 1,2$,
- $M^{(3)} > \sum_{j=1}^{2} M^{(S_j)}$,
- $M = \sum_{j=1}^{2} M^{(S_j)}$, $a = 1$,
- $\theta = 1$,

$$\begin{aligned}\lambda &= \frac{-r^2 + r(H^{(1)}+H^{(3)})}{2} + \frac{1}{2}M^{(1)}M^{(2)} \\ &= \frac{H^{(2)}H^{(1)} - H^{(2)}r + H^{(3)}r}{2}.\end{aligned}$$

(4) If $H^{(2)} + H^{(3)} < H^{(1)} + r$, we have

- $\mathcal{M} = \{2,3\}$, $S_1 = 2, S_2 = 3$, $\ell = 1$,
- $M^{(s)} \le \sum_{j=1}^{2} M^{(S_j)}$ *for* $s = 2,3$,
- $M^{(1)} > \sum_{j=1}^{2} M^{(S_j)}$,
- $M = \sum_{j=1}^{2} M^{(S_j)}$, $a = 1$,
- $\theta = 1$,

$$\begin{aligned}\lambda &= \frac{-r^2 + r(H^{(1)}+H^{(3)})}{2} + \frac{1}{2}M^{(2)}M^{(3)} \\ &= \frac{H^{(2)}H^{(3)} - H^{(2)}r + H^{(1)}r}{2}.\end{aligned}$$

(5) If $H^{(1)} + H^{(3)} < H^{(2)} + r$, we have

- $\mathcal{M} = \{1,3\}$, $S_1 = 1, S_2 = 3$, $\ell = 1$,
- $M^{(s)} \le \sum_{j=1}^{2} M^{(S_j)}$ *for* $s = 1,3$,
- $M^{(2)} > \sum_{j=1}^{2} M^{(S_j)}$,
- $M = \sum_{j=1}^{2} M^{(S_j)}$, $a = 1$,
- $\theta = 1$,

$$\begin{aligned}\lambda &= \frac{-r^2 + r(H^{(1)}+H^{(3)})}{2} + \frac{1}{2}M^{(1)}M^{(3)} \\ &= \frac{H^{(1)}H^{(3)}}{2}.\end{aligned}$$

4. Proof of main theorem

Define E_s be the identity matrix of size s.

Lemma 2. *Let $A = (a_{ij})$ be an $h_1 \times h_2$ matrix of variables with rank r_1. Let $r \le r_1$ and define $A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}$, where A_1 is a regular matrix of size $r \times r$.*

Then, there exist regular matrices $Q_1 = \begin{pmatrix} E_r & O \\ F_3 & E_{h_1-r} \end{pmatrix}$ and $Q_2 = \begin{pmatrix} E_r & F_2 \\ O & E_{h_2-r} \end{pmatrix}$ such that

$$Q_1AQ_2 = \begin{pmatrix} A_1 & O \\ O & C_4 \end{pmatrix},$$

with $\mathrm{rank}(C_4) = r_1 - r$.

Proof. Because A_1 is regular, let $Q_1 = \begin{pmatrix} E_r & O \\ -A_3A_1^{-1} & E_{h_1-r} \end{pmatrix}$. Then,

$$Q_1A = \begin{pmatrix} A_1 & A_2 \\ O & -A_3A_1^{-1}A_2 + A_4 \end{pmatrix}.$$

We transform the variables A_4 to $C_4 = -A_3A_1^{-1}A_2 + A_4$, to obtain

$$Q_1A = \begin{pmatrix} A_1 & A_2 \\ O & C_4 \end{pmatrix}.$$

Let $Q_2 = \begin{pmatrix} E_r & -A_1^{-1}A_2 \\ O & E_{h_2-r} \end{pmatrix}$, then we have $Q_1AQ_2 = \begin{pmatrix} A_1 & O \\ O & C_4 \end{pmatrix}$.

We transform the variables A_2 and A_3 to $F_2 = -A_1^{-1}A_2$ and $F_3 = -A_3A_1^{-1}$, respectively. Then, we have

$$Q_1 = \begin{pmatrix} E_r & O \\ F_3 & E_{h_1-r} \end{pmatrix}$$

$$Q_2 = \begin{pmatrix} E_r & F_2 \\ O & E_{h_2-r} \end{pmatrix},$$

which completes the proof. □

Lemma 3. *Let C, D, P and Q be an $h_1 \times h_2$ matrix, an $h_0 \times h_2$ matrix, an $h_0 \times h_1$ matrix and an $h_2 \times h_3$ matrix, respectively.*

Then, we have the followings.

(1) $\langle PC\rangle \subset \langle C\rangle$. $\langle CQ\rangle \subset \langle C\rangle$.

(2) $\langle PC+D\rangle+\langle C\rangle=\langle D\rangle+\langle C\rangle$.
(3) If $h_0=h_1$, $h_2=h_3$ *and* P,Q *are regular, we have*

$$\langle PC\rangle=\langle C\rangle,\quad \langle CQ\rangle=\langle C\rangle.$$

Proof. Since all elements in PC are linear combinations of those in C, we have $\langle PC\rangle\subset\langle C\rangle$. Therefore,

$$\langle PC+D\rangle+\langle C\rangle=\langle PC+D\rangle+\langle PC\rangle+\langle C\rangle=\langle D\rangle+\langle PC\rangle+\langle C\rangle=\langle D\rangle+\langle C\rangle.$$

If P is regular, we have $\langle C\rangle=\langle P^{-1}PC\rangle\subset\langle PC\rangle$. Therefore, we have $\langle PC\rangle=\langle C\rangle$. The same argument holds for Q as well. □

Let the matrix $A^{(s)}$ be in a neighborhood of $A^{*(s)}$ of rank $r^{(s)}$. In considering $\prod_{s=1}^L A^{(s)}$ in a neighborhood of a matrix of rank r, we have $r^{(s)}\geq r$. By Lemmas 1 and 3 (3), we can assume $\prod_{s=1}^L A^{*(s)}=\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}$.

Theorem 3. *There exist regular matrices* $P_1=\begin{pmatrix}E_r & O\\ F_3 & E_{H^{(1)}-r}\end{pmatrix}$ *and* $P_2=\begin{pmatrix}E_r & F_2\\ O & E_{H^{(L+1)}-r}\end{pmatrix}$ *such that*

$$P_1(\prod_{s=1}^L A^{(s)})P_2=\begin{pmatrix}C_1 & O\\ O & \prod_{s=1}^L C^{(s)}\end{pmatrix},$$

where C_1 *is in a neighborhood of* E_r, *rank*$(C^{(s)})$ *is* $r^{(s)}-r$, *and* $\prod_{s=1}^L C^{(s)}$ *is in a neighborhood of* O.

Proof. Let $A^{(1)}=\begin{pmatrix}A_1^{(1)} & A_2^{(1)}\\ A_3^{(1)} & A_4^{(1)}\end{pmatrix}$, where $A_1^{(1)}$ is a matrix of size $r\times r$. Because $\prod_{s=1}^L A^{*(s)}=\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}$, we can assume that rank$(A_1^{(1)})=r$. By Lemma 2, there exist regular matrices $Q_1^{(1)}=\begin{pmatrix}E_r & O\\ F_3^{(1)} & E_{H^{(1)}-r}\end{pmatrix}$ and $Q_2^{(1)}=\begin{pmatrix}E_r & F_2^{(1)}\\ O & E_{H^{(2)}-r}\end{pmatrix}$ such that

$$Q_1^{(1)}A^{(1)}Q_2^{(1)}=\begin{pmatrix}A_1^{(1)} & O\\ O & C_4^{(1)}\end{pmatrix}.$$

Assume that we have $Q_1'=\begin{pmatrix}E_r & O\\ F_3' & E_{H^{(1)}-r}\end{pmatrix}$ and $Q_2'=\begin{pmatrix}E_r & F_2'\\ O & E_{H^{(S+1)}-r}\end{pmatrix}$ such that

$$Q_1'(\prod_{s=1}^S A^{(s)})Q_2'=\begin{pmatrix}C_1' & O\\ O & \prod_{s=1}^S C^{(s)}\end{pmatrix}.$$

where C_1' is regular.

Let $A'^{(S+1)}=Q_2'^{-1}A^{(S+1)}=\begin{pmatrix}A_1' & A_2'\\ A_3' & A_4'\end{pmatrix}$. Then, we have

$$Q_1'(\prod_{s=1}^S A^{(s)})A^{(S+1)}=Q_1'(\prod_{s=1}^S A^{(s)})Q_2'A'^{(S+1)}=\begin{pmatrix}C_1'A_1' & C_1'A_2'\\ \prod_{s=1}^S C^{(s)}A_3' & \prod_{s=1}^S C^{(s)}A_4'\end{pmatrix}.$$

Because $Q_1'\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}=\begin{pmatrix}E_r & O\\ F_3' & O\end{pmatrix}$, we can assume $C_1'A_1'$ is regular. Let $Q_1''=\begin{pmatrix}E_r & O\\ -\prod_{s=1}^S C^{(s)}A_3'(C_1'A_1')^{-1} & E_{H^{(1)}-r}\end{pmatrix}$. Then,

$$Q_1''Q_1'(\prod_{s=1}^S A^{(s)})A^{(S+1)}=\begin{pmatrix}C_1'A_1' & C_1'A_2'\\ O & -\prod_{s=1}^S C^{(s)}A_3'A_1'^{-1}A_2'+\prod_{s=1}^S C^{(s)}A_4'\end{pmatrix}.$$

We also transform the variables A_4' to $C^{(S+1)}=-A_3'A_1'^{-1}A_2'+A_4'$, to obtain

$$Q_1''Q_1'(\prod_{s=1}^S A^{(s)})A^{(S+1)}=\begin{pmatrix}C_1'A_1' & C_1'A_2'\\ O & \prod_{s=1}^{S+1} C^{(s)}\end{pmatrix}.$$

Let $Q_2''=\begin{pmatrix}E_r & -A_1'^{-1}A_2'\\ O & E_{H^{(S+2)}-r}\end{pmatrix}$, then we have

$$Q_1''Q_1'(\prod_{s=1}^S A^{(s)})A^{(S+1)}Q_2''=\begin{pmatrix}C_1'A_1' & O\\ O & \prod_{s=1}^{S+1} C^{(s)}\end{pmatrix}.$$

We transform the variables A_2' and F_3' to $F_2''=-A_1'^{-1}A_2'$ and $F_3''=F_3'-\prod_{s=1}^S C^{(s)}A_3'(C_1'A_1')^{-1}$, respectively. Then, we have

$$Q_1''Q_1'=\begin{pmatrix}E_r & O\\ F_3'' & E_{H^{(1)}-r}\end{pmatrix}$$

$$Q_2''=\begin{pmatrix}E_r & F_2''\\ O & E_{H^{(S+2)}-r}\end{pmatrix},$$

which completes the proof by induction. □

By Theorem 3, we have

$$P_1\left(\prod_{s=1}^L A^{(s)}-\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}\right)P_2=\begin{pmatrix}C_1 & O\\ O & \prod_{s=1}^L C^{(s)}\end{pmatrix}-\begin{pmatrix}E_r & F_2\\ F_3 & F_3F_2\end{pmatrix}$$
$$=\begin{pmatrix}C_1-E_r & -F_2\\ -F_3 & \prod_{s=1}^L C^{(s)}-F_3F_2\end{pmatrix}.$$

By Lemma 3,

$$\langle\prod_{s=1}^L A^{(s)}-\prod_{s=1}^L A^{*(s)}\rangle=\langle\begin{pmatrix}C_1-E_r & -F_2\\ -F_3 & \prod_{s=1}^L C^{(s)}-F_3F_2\end{pmatrix}\rangle$$
$$=\langle C_1-E_r\rangle+\langle F_2\rangle+\langle F_3\rangle+\langle\prod_{s=1}^L C^{(s)}\rangle.$$

Since $\lambda_{w^*}\langle C_1-E_r\rangle=\frac{r^2}{2}$, $\lambda_{w^*}\langle F_2\rangle=\frac{r(H^{(L+1)}-r)}{2}$ and $\lambda_{w^*}\langle F_3\rangle=\frac{r(H^{(1)}-r)}{2}$, we have

$$\lambda_{w^*}\langle\prod_{s=1}^L A^{(s)}-\prod_{s=1}^L A^{*(s)}\rangle=\frac{r^2+r(H^{(1)}+H^{(L+1)}-2r)}{2}+\lambda_{w^*}\langle\prod_{s=1}^L C^{(s)}\rangle$$
$$=\frac{-r^2+r(H^{(1)}+H^{(L+1)})}{2}+\lambda_{w^*}\langle\prod_{s=1}^L C^{(s)}\rangle.$$

Let $M^{(s)}=H^{(s)}-r$ for $s=1,\ldots,L+1$.

Theorem 4 (*Method for Determining the Deepest Singular Point (Aoyagi, 2013a)*). *Let* $F_1(w_1,\ldots,w_d)$, …, $F_m(w_1,\ldots,w_d)$ *be homogeneous functions of* $w_1,\ldots,w_j$ $(1\leq j\leq d)$. *Also, let* φ *be a* C^∞ *function such that* $\varphi(0,\ldots,0,w^*_{j+1},\ldots,w^*_d)\geq\varphi(w^*_1,\ldots,w^*_d)$ *and* φ_w *is a homogeneous function of* $w_1,\ldots,w_j$ *in a small neighborhood of* $(0,\ldots,0,w^*_{j+1},\ldots,w^*_d)$. *Then, we have*

$$\lambda_{(0,\ldots,0,w^*_{j+1},\ldots,w^*_d)}(\langle F_1,\ldots,F_m\rangle,\varphi)\leq\lambda_{(w^*_1,\ldots,w^*_j,w^*_{j+1},\ldots,w^*_d)}(\langle F_1,\ldots,F_m\rangle,\varphi).$$

By Theorem 4 in the case where $\varphi>0$, we can set $r^{(s)}=r$ for $s=1,\ldots,L$, without loss of generality. Note that if $\varphi>0$, the degree of φ is 0.

Define for $s=1,\ldots,L$,

$$C^{(s)}=(c^{(s)}_{ij}),(1\leq i\leq M^{(s)},1\leq j\leq M^{(s+1)}),$$

and consider the log canonical threshold of $\|\prod_{s=1}^L C^{(s)}\|^2$.

We prove the main theorem by developing blow-up method along submanifolds to obtain log canonical thresholds, i.e., learning coefficients. We use a recursive blow-up process.

Define the numerical sequence by

$$M(S)=\min\{M^{(s)}\mid 1\leq s\leq S\}$$

for $S = 1, \ldots, L+1$. Let

$T_{S,k} = (t^{(1)}_{S,k}, \ldots, t^{(L)}_{S,k})$.

Definition 4. For two vectors $T = (t_1, \ldots, t_L)$ and $T' = (t'_1, \ldots, t'_L) \in \mathbf{R}^L$, denote

$T \le T'$ if $t_1 \le t'_1, \ldots, t_L \le t'_L$,

and

$T < T'$ if $t_1 \le t'_1, \ldots, t_L \le t'_L$ and $T \ne T'$.

(Inductive statement)

Let us prove the following inductive statement for $J \ge 0$ and $S \ge 1$

$$\text{(A)}\quad \langle \prod_{s=1}^{L} C^{(s)} \rangle = \langle \mathrm{diag}(b_1, \ldots, b_{M(S)}) \begin{pmatrix} E_J & O \\ O & D_J \end{pmatrix} \prod_{s=S+1}^{L} C^{(s)} \rangle,$$

where $\mathrm{diag}(b_1, \ldots, b_{M(S)})$ is a diagonal matrix of size $M(S) \times M(S)$.

Let $D_J = (d_{i,j})$ for $J+1 \le i \le M(S), J+1 \le j \le M^{(S+1)}$ be a matrix of size $(M(S) - J) \times (M^{(S+1)} - J)$ and

$$\begin{aligned} b_0 &= 1 \\ b_i &= \prod_{\{(s,k) | \tilde{t}_{s,k} = i-1\}} u_{s,k} b_{i-1}, \quad i = 1, \ldots, M(S), \\ \prod_{s=1}^{L} \prod_{i=1}^{M^{(s)}} \prod_{j=1}^{M^{(s+1)}} \mathrm{d}c^{(s)}_{ij} &= (\prod_{s=1}^{S-1} \prod_{k=1}^{M(s+1)} u_{s,k}^{M_{s,k}-1} \mathrm{d}u_{s,k})(\prod_{k=1}^{J} u_{S,k}^{M_{S,k}-1} \mathrm{d}u_{S,k}) \\ &\quad \times (\prod_{i=J+1}^{M(S)} \prod_{j=J+1}^{M^{(S+1)}} \mathrm{d}d_{ij})(\prod_{s=S+1}^{L} \prod_{i=1}^{M^{(s)}} \prod_{j=1}^{M^{(s+1)}} \mathrm{d}c^{(s)}_{ij}), \\ \tilde{t}_{s,k} &= \min\{t^{(S')}_{s,k} : 1 \le S' \le L\}, \\ &\qquad \text{for } 1 \le s \le S-1, 1 \le k \le M(s+1), \\ \tilde{t}_{S,k} &= \min\{t^{(S')}_{S,k} : 1 \le S' \le L\} = t^{(S)}_{S,k}, \text{ for } 1 \le k \le J, \end{aligned}$$

by a blow-up process.

Moreover, we have

$$(B)\quad T_{s,k} \le T_{s',k'}, \text{ or } T_{s,k} \ge T_{s',k'},$$

$$\text{for } 1 \le s, s' \le S, 1 \le k \le \begin{cases} M(s+1), & \text{if } s, s' < S, \\ J, & \text{if } s = S, s' = S. \end{cases}$$

(End of the inductive statement)

Inductively, $T_{s,k}$ and $t^{(S)}_{s,k}$ are defined, updated and used to demonstrate that $u_{s,k}$ is composed of b_i when $t^{(S)}_{s,k} \le i-1$, and λ is expressed by one of $t^{(S)}_{s,k}$'s. The relations (B) is used to obtain the value θ.

The above is obvious for $S = 1$ and $J = 0$, where the right side of Equation (A) is the same as the left side. From now on, we will consider

Case 1 : $b_{J+1} = b_{J+2} = \cdots = b_{J+J_1}, b_{J+J_1+1} \ne b_{J+J_1}$.

Case 2: $b_{J+1} = b_{J+2} = \cdots = b_{M(S)}$.

Case 1

Assume that

$$b_{J+1} = b_{J+2} = \cdots = b_{J+J_1}, b_{J+J_1+1} \ne b_{J+J_1},$$

that is, $\{\tilde{t}_{s,k} \mid \tilde{t}_{s,k} = i\} = \phi$ for $i = J+1, \ldots, J+J_1-1$.

Fix $u_{s,k}$ such that $\tilde{t}_{s,k} = J + J_1$ and $T_{s,k} \le T_{s',k'}$ for $\tilde{t}_{s',k'} = J + J_1$.

Construct the blow-up along the submanifold

$$\{d_{ij} = 0 (i = J+1, \ldots, J+J_1, j = J+1, \ldots, M^{(S+1)}), u_{s,k} = 0\},$$

Case 1 (1)

Consider instances in which

$$\begin{pmatrix} d_{J+1,J+1} & d_{J+1,J+2} & \cdots & d_{J+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d_{J+J_1,J+1} & d_{J+J_1,J+2} & \cdots & d_{J+J_1,M^{(S+1)}} \end{pmatrix}$$

$$= u_{s,k} \begin{pmatrix} d'_{J+1,J+1} & d'_{J+1,J+2} & \cdots & d'_{J+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d'_{J+J_1,J+1} & d'_{J+J_1,J+2} & \cdots & d'_{J+J_1,M^{(S+1)}} \end{pmatrix}.$$

$$\text{Let} \begin{cases} t^{(S)}_{s,k} = t^{(S+1)}_{s,k} = \cdots = t^{(L)}_{s,k} = J, \tilde{t}_{s,k} = J \\ b'_{J+1} = u_{s,k} b_{J+1} \\ \vdots \\ b'_{J+J_1} = u_{s,k} b_{J+J_1} \\ M'_{s,k} = M_{s,k} + J_1(M^{(S+1)} - J). \end{cases}$$

Because $\{\tilde{t}_{s',k'} \mid \tilde{t}_{s',k'} = i\} = \phi$ for $i = J+1, \ldots, J+J_1-1$, that is, $\tilde{t}_{s',k'} \le J$ or $\tilde{t}_{s',k'} \ge J + J_1$, and $T_{s,k} \le T_{s',k'}$ for $\tilde{t}_{s',k'} = J + J_1$, and we therefore have

$$T_{s,k} \le T_{s',k'} \text{ or } T_{s,k} \ge T_{s',k'},$$

$$\text{for } 1 \le s' \le S \text{ and } 1 \le k' \le \begin{cases} M(s+1), & \text{if } s' < S, \\ J, & \text{if } s' = S. \end{cases}.$$

Again, let $b'_i, M'_{s,k}$ be $b_i, M_{s,k}$, and we have the inductive statement with the number of elements in $\{u_{s',k'} \mid \tilde{t}_{s',k'} = J + J_1\}$ decreased by one.

Case 1 (2)

Consider instances in which

$$\begin{pmatrix} d_{J+1,J+1} & d_{J+1,J+2} & \cdots & d_{J+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d_{J+J_1,J+1} & d_{J+J_1,J+2} & \cdots & d_{J+J_1,M^{(S+1)}} \end{pmatrix}$$

$$= u_{S,J+1} \begin{pmatrix} 1 & d'_{J+1,J+2} & \cdots & d'_{J+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d'_{J+J_1,J+1} & d'_{J+J_1,J+2} & \cdots & d'_{J+J_1,M^{(S+1)}} \end{pmatrix}$$

and

$$u_{s,k} = u_{S,J+1} u'_{s,k}.$$

$$\text{Let} \begin{cases} t^{(i)}_{S,J+1} = t^{(i)}_{s,k}, (i = 1, \ldots, S-1) \\ t^{(S)}_{S,J+1} = t^{(S+1)}_{S,J+1} = \cdots = t^{(L)}_{S,J+1} = J, \tilde{t}_{S,J+1} = J \\ b'_{J+1} = u_{S,J+1} b_{J+1} \\ \vdots \\ b'_{M(S)} = u_{S,J+1} b_{M(S)} \\ M'_{S,J+1} = M_{s,k} + J_1(M^{(S+1)} - J). \end{cases}$$

Because $\{\tilde{t}_{s',k'} \mid \tilde{t}_{s',k'} = i\} = \phi$ for $i = J+1, \ldots, J+J_1-1$, that is, $\tilde{t}_{s',k'} \le J$ or $\tilde{t}_{s',k'} \ge J + J_1$, and $T_{s,k} \le T_{s',k'}$ for $\tilde{t}_{s',k'} = J + J_1$, and we therefore have

$$T_{S,J+1} \le T_{s',k'} \text{ or } T_{S,J+1} \ge T_{s',k'},$$

$$\text{for } 1 \le s' \le S \text{ and } 1 \le k' \le \begin{cases} M(s+1), & \text{if } s' < S, \\ J, & \text{if } s' = S. \end{cases}.$$

Define regular matrix Q

$$Q = \begin{pmatrix} 1 & -d'_{J+1,J+2} & -d'_{J+1,J+3} & \cdots & -d'_{J+1,M^{(S+1)}} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & & 1 \end{pmatrix}.$$

and transform the variables introducing

$$D''_J = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ d''_{J+2,J+1} & d''_{J+2,J+2} & \cdots & d''_{J+2,M^{(S+1)}} \\ d''_{J+3,J+1} & d''_{J+3,J+2} & \cdots & d''_{J+3,M^{(S+1)}} \\ \vdots & \vdots & \ddots & \vdots \\ d''_{J+J_1,J+1} & d''_{J+J_1,J+2} & \cdots & d''_{J+J_1,M^{(S+1)}} \\ d''_{J+J_1+1,J+1} & d''_{J+J_1+1,J+2} & \cdots & d''_{J+J_1+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d''_{M(S),J+1} & d''_{M(S),J+2} & \cdots & d''_{M(S),M^{(S+1)}} \end{pmatrix} = \begin{pmatrix} 1 & d'_{J+1,J+2} & \cdots & d'_{J+1,M^{(S+1)}} \\ d'_{J+2,J+1} & d'_{J+2,J+2} & \cdots & d'_{J+2,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d'_{J+J_1,J+1} & d'_{J+J_1,J+2} & \cdots & d'_{J+J_1,M^{(S+1)}} \\ d_{J+J_1+1,J+1} & d_{J+J_1+1,J+2} & \cdots & d_{J+J_1+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d_{M(S),J+1} & d_{M(S),J+2} & \cdots & d_{M(S),M^{(S+1)}} \end{pmatrix} Q.$$

Let $C_J^{(S+1)} = \begin{pmatrix} c^{(S+1)}_{J+1,1} & c^{(S+1)}_{J+1,2} & \cdots & c^{(S+1)}_{J+1,M^{(S+2)}} \\ c^{(S+1)}_{J+2,1} & c^{(S+1)}_{J+2,2} & \cdots & c^{(S+1)}_{J+2,M^{(S+2)}} \\ \vdots & \vdots & \cdots & \vdots \\ c^{(S+1)}_{M^{(S+1)},1} & c^{(S+1)}_{M^{(S+1)},2} & \cdots & c^{(S+1)}_{M^{(S+1)},M^{(S+2)}} \end{pmatrix}$ and transform the variables $C_J^{(S+1)}$ to $C_J'^{(S+1)}$ using $C_J'^{(S+1)} = Q^{-1}C_J^{(S+1)}$.

Define regular matrix P

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\frac{b'_{J+2}}{b'_{J+1}} d''_{J+2,J+1} & 1 & 0 & 0 & \cdots & 0 \\ -\frac{b'_{J+3}}{b'_{J+1}} d''_{J+3,J+1} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{b'_{M(S)}}{b'_{J+1}} d''_{M(S),J+1} & 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}.$$

and let

$$D'''_J = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & d''_{J+2,J+2} & \cdots & d''_{J+2,M^{(S+1)}} \\ 0 & d''_{J+3,J+2} & \cdots & d''_{J+3,M^{(S+1)}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & d''_{M(S),J+2} & \cdots & d''_{M(S),M^{(S+1)}} \end{pmatrix} = \begin{pmatrix} 1 & O \\ O & D_{J+1} \end{pmatrix}.$$

We then have

$$\begin{aligned} & P\mathrm{diag}(b_{J+1}, \ldots, b_{M(S)}) D_J C_J^{(S+1)} \\ &= P u_{S,J+1} \mathrm{diag}(b'_{J+1}, \ldots, b'_{M(S)}) D''_J C_J'^{(S+1)} \\ &= u_{S,J+1} P \mathrm{diag}(b'_{J+1}, \ldots, b'_{M(S)}) D''_J C_J'^{(S+1)} \\ &= u_{S,J+1} \mathrm{diag}(b'_{J+1}, \ldots, b'_{M(S)}) D'''_J C_J'^{(S+1)}. \end{aligned}$$

Again, let $b'_i, M'_{s,k}, d'''_{ij}, c'_{ij}$ be $b_i, M_{s,k}, d_{ij}, c_{ij}$.

If $J+1 \le M(S+1) = \min\{M(S), M^{(S+1)}\}$, then we have the inductive statement with J increased by one.

If $J+1 > M(S+1) = \min\{M(S), M^{(S+1)}\}$, then $D'''_J = (1, 0, \ldots, 0)$ or $D'''_J = (1, 0, \ldots, 0)^t$ with t the operation of transposition. Therefore, we have

$$\langle \prod_{s=1}^{L} C^{(s)} \rangle = \langle \mathrm{diag}(b_1, \ldots, b_J, b_{J+1}, \ldots, b_{M(S+1)}) C'^{(S+1)} \prod_{s=S+2}^{L} C^{(s)} \rangle,$$

where $C'^{(S+1)} = \begin{pmatrix} c^{(S+1)}_{1,1} & c^{(S+1)}_{1,2} & \cdots & c^{(S+1)}_{1,M^{(S+2)}} \\ c^{(S+1)}_{2,1} & c^{(S+1)}_{2,2} & \cdots & c^{(S+1)}_{2,M^{(S+2)}} \\ \vdots & \vdots & \vdots & \vdots \\ c^{(S+1)}_{J,1} & c^{(S+1)}_{J,2} & \cdots & c^{(S+1)}_{J,M^{(S+2)}} \\ c^{(S+1)}_{J+1,1} & c^{(S+1)}_{J+1,2} & \cdots & c^{(S+1)}_{J+1,M^{(S+2)}} \\ \vdots & \vdots & \vdots & \vdots \\ c^{(S+1)}_{M(S+1),1} & c^{(S+1)}_{M(S+1),2} & \cdots & c^{(S+1)}_{M(S+1),M^{(S+2)}} \end{pmatrix}$, which is the inductive statement with S increased by one.

Case 2

Assume that

$$b_{J+1} = b_{J+2} = \cdots = b_{M(S)},$$

that is, $\{\tilde{t}_{s,k} \mid \tilde{t}_{s,k} = i\} = \phi$ for $i = J+1, \ldots, M(S)-1$.

Construct the blow-up along submanifold

$$\{d_{ij} = 0, (i = J+1, \ldots, M(S), j = J+1, \ldots, M^{(S+1)})\}.$$

Consider instances in which

$$\begin{pmatrix} d_{J+1,J+1} & d_{J+1,J+2} & \cdots & d_{J+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d_{M(S),J+1} & d_{M(S),J+2} & \cdots & d_{M(S),M^{(S+1)}} \end{pmatrix} = u_{S,J+1} \begin{pmatrix} 1 & d'_{J+1,J+2} & \cdots & d'_{J+1,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d'_{M(S),J+1} & d'_{M(S),J+2} & \cdots & d'_{M(S),M^{(S+1)}} \end{pmatrix}.$$

Let $\begin{cases} t^{(i)}_{S,J+1} = M(i+1), (i = 1, \ldots, S-1) \\ t^{(S)}_{S,J+1} = t^{(S+1)}_{S,J+1} = \cdots = t^{(L)}_{S,J+1} = J, \tilde{t}_{S,J+1} = J \\ b'_{J+1} = u_{S,J+1} b_{J+1} \\ \vdots \\ b'_{M(S)} = u_{S,J+1} b_{M(S)} \\ M'_{S,J+1} = (M(S)-J)(M^{(S+1)}-J). \end{cases}$

Because $\{\tilde{t}_{s',k'} \mid \tilde{t}_{s',k'} = i\} = \phi$ for $i = J+1, \ldots, M(S)-1$, that is, $\tilde{t}_{s',k'} \le J$ or $\tilde{t}_{s',k'} = M(S)$, we have

$$T_{S,J+1} \ge T_{s',k'},$$

for $1 \le s' \le S$ and $1 \le k' \le \begin{cases} M(s+1), & \text{if } s' < S, \\ J, & \text{if } s' = S. \end{cases}$.

Define regular matrix Q

$$Q = \begin{pmatrix} 1 & -d'_{J+1,J+2} & -d'_{J+1,J+3} & \cdots & -d'_{J+1,M^{(S+1)}} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & & 1 \end{pmatrix}.$$

and transform the variables using

$$D''_J = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ d''_{J+2,J+1} & d''_{J+2,J+2} & \cdots & d''_{J+2,M^{(S+1)}} \\ d''_{J+3,J+1} & d''_{J+3,J+2} & \cdots & d''_{J+3,M^{(S+1)}} \\ \vdots & \vdots & \ddots & \vdots \\ d''_{M(S),J+1} & d''_{M(S),J+2} & \cdots & d''_{M(S),M^{(S+1)}} \end{pmatrix} = \begin{pmatrix} 1 & d'_{J+1,J+2} & \cdots & d'_{J+1,M^{(S+1)}} \\ d'_{J+2,J+1} & d'_{J+2,J+2} & \cdots & d'_{J+2,M^{(S+1)}} \\ \vdots & \vdots & \cdots & \vdots \\ d'_{M(S),J+1} & d'_{M(S),J+2} & \cdots & d'_{M(S),M^{(S+1)}} \end{pmatrix} Q.$$

Let $C_J^{(S+1)} = \begin{pmatrix} c_{J+1,1}^{(S+1)} & c_{J+1,2}^{(S+1)} & \cdots & c_{J+1,M^{(S+2)}}^{(S+1)} \\ c_{J+2,1}^{(S+1)} & c_{J+2,2}^{(S+1)} & \cdots & c_{J+2,M^{(S+2)}}^{(S+1)} \\ \vdots & \vdots & \cdots & \vdots \\ c_{M^{(S+1)},1}^{(S+1)} & c_{M^{(S+1)},2}^{(S+1)} & \cdots & c_{M^{(S+1)},M^{(S+2)}}^{(S+1)} \end{pmatrix}$ and transform the variables $C_J^{(S+1)}$ to $C_J'^{(S+1)}$ using $C_J'^{(S+1)} = Q^{-1}C_J^{(S+1)}$.

Define regular matrix P

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\frac{b'_{J+2}}{b'_{J+1}} d''_{J+2,J+1} & 1 & 0 & 0 & \cdots & 0 \\ -\frac{b'_{J+3}}{b'_{J+1}} d''_{J+3,J+1} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{b'_{M(S)}}{b'_{J+1}} d''_{M(S),J+1} & 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}.$$

and let

$$D_J''' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & d''_{J+2,J+2} & \cdots & d''_{J+2,M^{(S+1)}} \\ 0 & d''_{J+3,J+2} & \cdots & d''_{J+3,M^{(S+1)}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & d''_{M(S),J+2} & \cdots & d''_{M(S),M^{(S+1)}} \end{pmatrix} = \begin{pmatrix} 1 & O \\ O & D_{J+1} \end{pmatrix}.$$

We then have

$$\begin{aligned} & P\mathrm{diag}(b_{J+1}, \ldots, b_{M(S)}) D_J C_J^{(S+1)} \\ = \ & P u_{S,J+1} \mathrm{diag}(b'_{J+1}, \ldots, b'_{M(S)}) D_J'' C_J'^{(S+1)} \\ = \ & u_{S,J+1} P \mathrm{diag}(b'_{J+1}, \ldots, b'_{M(S)}) D_J'' C_J'^{(S+1)} \\ = \ & u_{S,J+1} \mathrm{diag}(b'_{J+1}, \ldots, b'_{M(S)}) D_J''' C_J'^{(S+1)}. \end{aligned}$$

Again, let $b'_{ij}, M'_{ij}, d'''_{ij}, c'_{ij}$ be $b_{ij}, M_{ij}, d_{ij}, c_{ij}$.

If $J+1 \le M(S+1) = \min\{M(S), M^{(S+1)}\}$, then we have the inductive statement with J increased by one.

If $J+1 > M(S+1) = \min\{M(S), M^{(S+1)}\}$, then $D_J''' = (1, 0, \ldots, 0)$ or $D_J''' = (1, 0, \ldots, 0)^t$. Therefore, we have

$$\langle \prod_{s=1}^{L} C^{(s)} \rangle = \langle \mathrm{diag}(b_1, \ldots, b_J, b'_{J+1}, \ldots, b'_{M(S+1)}) C'^{(S+1)} \prod_{s=S+2}^{L} C^{(s)} \rangle,$$

where $C'^{(S+1)} = \begin{pmatrix} c_{1,1}^{(S+1)} & c_{1,2}^{(S+1)} & \cdots & c_{1,M^{(S+2)}}^{(S+1)} \\ c_{2,1}^{(S+1)} & c_{2,2}^{(S+1)} & \cdots & c_{2,M^{(S+2)}}^{(S+1)} \\ \vdots & \vdots & \vdots & \vdots \\ c_{M(S+1),1}^{(S+1)} & c_{M(S+1),2}^{(S+1)} & \cdots & c_{M(S+1),M^{(S+2)}}^{(S+1)} \end{pmatrix}$, which is the inductive statement with S increased by one.

We finally have the inductive statement for $S = L+1$,

$$\langle \prod_{s=1}^{L} C^{(s)} \rangle = \langle \mathrm{diag}(b_1, \ldots, b_{M(L+1)}) \rangle.$$

Candidates for the log canonical threshold of $\| \prod_{s=1}^{L} C^{(s)} \|^2$ on this local coordinate are

$$\frac{1}{2} \min\{ M_{s,k}, \tilde{t}_{s,k} = 0 \}$$

and

$$M_{s,k} = (M^{(1)} - t_{s,k}^{(1)})(M^{(2)} - t_{s,k}^{(1)}) + \sum_{j=2}^{L} (t_{s,k}^{(j-1)} - t_{s,k}^{(j)})(M^{(j+1)} - t_{s,k}^{(j)}).$$

Define ℓ, H_i and S_j such that

$$H_i \in \{ t_{s,k}^{(1)}, \ldots, t_{s,k}^{(L)} \},$$

$$M_{s,k} = (M^{(S_1)} - H_1)(M^{(S_2)} - H_1) + \sum_{j=2}^{\ell} (H_{j-1} - H_j)(M^{(S_{j+1})} - H_j),$$

where $S_j < S_{j+1}$, $H_j \le H_{j-1}$, $H_j \le M(S_{j+1})$ and $H_\ell = 0$. For every $T_{s,k}$, we have such sequences (H_i) and (S_i).

Let $F_j = H_{j-1} - H_j + M^{(S_{j+1})}$ for $j = 2, \ldots, \ell$ and $F_1 = M^{(S_1)} + M^{(S_2)} - H_1$. Then, we have

$$H_j = H_{j-1} - F_j + M^{(S_{j+1})} = H_1 - \sum_{l=2}^{j} (F_l - M^{(S_{l+1})}) = -\sum_{l=1}^{j} F_l + \sum_{l=1}^{j+1} M^{(S_l)}.$$

Because $H_{j-1} - H_j \ge 0$ and $H_\ell = 0$, we have

$$F_j - M^{(S_{j+1})} \ge 0,$$

$$H_\ell = -\sum_{l=1}^{\ell} F_l + \sum_{l=1}^{\ell+1} M^{(S_l)} = 0.$$

Also, we have

$$F_\ell - M^{(S_{\ell+1})} = -\sum_{l=1}^{\ell-1} F_l + \sum_{l=1}^{\ell} M^{(S_l)} \ge 0,$$

and

$$\begin{aligned} M_{s,k} &= (F_1 - M^{(S_2)})(F_1 - M^{(S_1)}) + \sum_{j=2}^{\ell} (F_j - M^{(S_{j+1})})(\sum_{l=1}^{j} F_l - \sum_{l=1}^{j} M^{(S_l)}) \\ &= \sum_{j=1}^{\ell-1} (F_j - M^{(S_{j+1})})(\sum_{l=1}^{j} F_l - \sum_{l=1}^{j} M^{(S_l)}) + (-\sum_{j=1}^{\ell-1} F_j + \sum_{j=1}^{\ell} M^{(S_j)}) M^{(S_{\ell+1})} \\ &= \sum_{j=1}^{\ell-1} F_j^2 - (\sum_{k=1}^{\ell+1} M^{(S_k)})(\sum_{j=1}^{\ell-1} F_j) + \sum_{1 \le i < j \le \ell-1} F_i F_j + \sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)} \\ &= \sum_{j=1}^{\ell-1} (F_j - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 + \sum_{j=1}^{\ell-1} \frac{2 \sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell} F_j - (\ell - 1)(\frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 \\ &\quad - (\sum_{k=1}^{\ell+1} M^{(S_k)}) \sum_{j=1}^{\ell-1} F_j + \sum_{1 \le i < j \le \ell-1} (F_i - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})(F_j - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell}) \\ &\quad + \frac{(\ell-2) \sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell} \sum_{j=1}^{\ell-1} F_j - \frac{(\ell-1)(\ell-2)}{2} (\frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 \\ &\quad + \sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)} \\ &= \sum_{j=1}^{\ell-1} (F_j - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 + \sum_{1 \le i < j \le \ell-1} (F_i - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})(F_j - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell}) \\ &\quad - \frac{\ell(\ell-1)}{2} (\frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 + \sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)} \\ &= \frac{1}{2} \sum_{j=1}^{\ell-1} (F_j - \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 + \frac{1}{2} (\sum_{j=1}^{\ell-1} F_j - \frac{\ell-1}{\ell} \sum_{j=1}^{\ell+1} M^{(S_j)})^2 \\ &\quad - \frac{\ell(\ell-1)}{2} (\frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell})^2 + \sum_{1 \le i < j \le \ell+1} M^{(S_i)} M^{(S_j)}. \end{aligned}$$

Therefore, if $M^{(S_{k'})} > \frac{\sum_{k=1}^{\ell+1} M^{(S_k)}}{\ell}$ or $M^{(S_{k'})}(\ell - 1) > \sum_{k \ne k'} M^{(S_k)}$, we can set $F_{k'-1} = M^{(S_{k'})}$ to minimize $M_{s,k}$.

Lemma 4. *Let ℓ and $0 \le a, b \le \ell - 1$ be natural numbers and let*

$$\frac{A(b)}{\ell^2} = b(\frac{\ell - a}{\ell})^2 + (\ell - 1 - b)(-\frac{a}{\ell})^2 + (b\frac{\ell - a}{\ell} - (\ell - 1 - b)\frac{a}{\ell})^2.$$

Then,

$$\min\{ A | b = 0, \ldots, \ell - 1 \} = A(a-1) = A(a) = a\ell(\ell - a).$$

Proof. We have

$$\frac{A(b)}{\ell^2} = \frac{b(\ell - a)^2 + (\ell - 1 - b)a^2 + (b(\ell - a) - (\ell - 1 - b)a)^2}{\ell^2}.$$

and

$$\frac{\partial A(b)}{\partial b} = (\ell - a)^2 - a^2 + 2(b(\ell - a) - (\ell - 1 - b)a)\{(\ell - a) + a\} = \ell^2(1 + 2b - 2a).$$

Therefore, we have established the proof. □

Define ℓ, $\mathcal{M}$, M, S_j and a as in Definition 3.

By Lemma 4, we have

$$
\begin{aligned}
&2\lambda_O(\|\prod_{s=1}^{L} C^{(s)}\|^2)\\
&= \frac{a(\ell-a)}{2\ell} - \frac{\ell(\ell-1)}{2}\Big(\frac{\sum_{j=1}^{\ell+1} M^{(S_j)}}{\ell}\Big)^2 + \sum_{1\le i<j\le \ell+1} M^{(S_i)}M^{(S_j)}\\
&= \frac{a(\ell-a)}{2\ell} - \frac{\ell(\ell-1)}{2}\Big(M+\frac{a-\ell}{\ell}\Big)^2 + \sum_{1\le i<j\le \ell+1} M^{(S_i)}M^{(S_j)}\\
&= \frac{a(\ell-a)}{2\ell} - \frac{\ell(\ell-1)}{2}\Big(M^2+2\frac{a-\ell}{\ell}M+\frac{(a-\ell)^2}{\ell^2}\Big) + \sum_{1\le i<j\le \ell+1} M^{(S_i)}M^{(S_j)}\\
&= -\frac{1}{2}(\ell-a-1)(\ell-a) - \frac{\ell(\ell-1)}{2}\Big(M^2+2\frac{a-\ell}{\ell}M\Big) + \sum_{1\le i<j\le \ell+1} M^{(S_i)}M^{(S_j)}
\end{aligned}
$$

Let

$$
\begin{aligned}
\tilde{H}_j &= \sum_{l=1}^{j+1} M^{(S_l)} - jM, \text{ for } j=1,\dots,a,\\
\tilde{H}_j &= \sum_{l=1}^{j+1} M^{(S_l)} - aM - (j-a)(M-1), \text{ for } j=a+1,\dots,\ell,
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{H}'_j &= \sum_{l=1}^{j+1} M^{(S_l)} - j(M-1), \text{ for } j=1,\dots,\ell-a,\\
\tilde{H}'_j &= \sum_{l=1}^{j+1} M^{(S_l)} - (\ell-a)(M-1) - (j-\ell+a)M \text{ for } j=\ell-a+1,\dots,\ell.
\end{aligned}
$$

Then, we have $\tilde{H}_\ell = \tilde{H}'_\ell = 0$. Moreover, let $\tilde{T}$ and $\tilde{T}'$ be the vectors corresponding to $(\tilde{H}_j)$ and $(\tilde{H}'_j)$, respectively. That (i) $\tilde{T}$ and $\tilde{T}'$ correspond to λ, and (ii) $\tilde{T}$ has minimum one and $\tilde{T}'$ has maximum one, are obvious from Definition 4.

Note that since $M^{(S_l)} \le M-1$ if $a<\ell$, we have

$$
\begin{aligned}
\tilde{H}_j &= M(S_{j+1}) + \sum_{M^{(S_l)}\ne M(S_{j+1})} (M^{(S_l)}-M) \le M(S_{j+1}), \text{ for } j=1,\dots,a,\\
\tilde{H}_j &= M(S_{j+1}) + \sum_{M^{(S_l)}\ne M(S_{j+1})} (M^{(S_l)}-(M-1)) - a \le M(S_{j+1}),\\
&\qquad \text{for } j=a+1,\dots,\ell,
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{H}'_j &= M(S_{j+1}) + \sum_{M^{(S_l)}\ne M(S_{j+1})} (M^{(S_l)}-(M-1)) \le M(S_{j+1}),\\
&\qquad \text{for } j=1,\dots,\ell-a,\\
\tilde{H}'_j &= M(S_{j+1}) + \sum_{M^{(S_l)}\ne M(S_{j+1})} (M^{(S_l)}-(M-1)) - (j-\ell+a) \le M(S_{j+1}),\\
&\qquad \text{for } a<\ell, j=\ell-a+1,\dots,\ell,\\
\tilde{H}'_j &= M(S_{j+1}) + \sum_{M^{(S_l)}\ne M(S_{j+1})} (M^{(S_l)}-M) \le M(S_{j+1}),\\
&\qquad \text{for } a=\ell, j=1,\dots,\ell.
\end{aligned}
$$

Also we have $\tilde{H}'_1 \ge \tilde{H}'_2 \ge \cdots \ge \tilde{H}'_\ell = 0$, and $\tilde{H}'_1 \ge \tilde{H}'_2 \ge \cdots \ge \tilde{H}'_\ell = 0$.

For obtaining the values θ, we prepare the following Lemma 5.

Lemma 5. *If the vector $T_{s,k} = (t_{s,k}^{(1)},\dots,t_{s,k}^{(L)})$ satisfies*

$$\tilde{T} \le T_{s,k} \le \tilde{T}'$$

and if ℓ, (H_j), (S_j) which corresponds to $T_{s,k}$ satisfies $H_{j-1}-H_j+M^{(S_{j+1})} = t_{s,k}^{(S_j-1)} - t_{s,k}^{(S_{j+1}-1)} + M^{(S_{j+1})} = M-1$ or M, then, $T_{s,k}$ corresponds to λ.

Proof. Because $\tilde{T} \le T_{s,k} \le \tilde{T}'$ and $\tilde{H}_\ell = \tilde{H}'_\ell = 0$, we have $H_\ell = 0$. Therefore, the number of elements of $\{j : H_{j-1}-H_j+M^{(S_{j+1})} = M\}$ is a and that of $\{j : H_{j-1}-H_j+M^{(S_{j+1})} = M-1\}$ is $\ell-a$. □

Lemma 6. *We have*

$$\theta = a(\ell-a)+1.$$

Proof. Given $t_{s,k}^{(S_{j+1}-1)} = H_j$, then by Lemma 5, the number of vectors $T_{s,k}$ corresponding to λ is less than the sum of the number of the set $\cup_{j=1}^{\ell-1}\{H : \tilde{H}_j \le H \le \tilde{H}'_j\}$. We have the number of the set $\{H : \tilde{H}_j \le H \le \tilde{H}'_j\}$ as

$$
\begin{aligned}
j+1, &\quad \text{if } j \le \min\{a,\ell-a\}\\
\min\{a,\ell-a\}+1, &\quad \text{if } \min\{a,\ell-a\}+1 \le j \le \max\{a,\ell-a\}\\
\min\{a,\ell-a\}+1+\max\{a,\ell-a\}-j, &\quad \text{if } \max\{a,\ell-a\}+1 \le j \le \ell.
\end{aligned}
$$

Because J is increased by one for Case 1(2) in the proof, we have $\theta \le a(\ell-a)+1$. Finally, by using $T_{s,k}$, we construct the local coordinate for $\theta = a(\ell-a)+1$ as follows.

Let γ be $\tilde{H}'_j = M(S_{j+1})$, $(j<\gamma)$ and $\tilde{H}'_\gamma < M(S_{\gamma+1})$.

Define

$$
t_{s,k}^{(S)} = \begin{cases} M(S+1), & S<s,\\ k-1, & S=s,\\ 0, & S>s, \end{cases} \tag{1}
$$

for $s < S_2-1$,

$$
t_{s,k}^{(S)} = \begin{cases} M(S+1), & S<S_2-1,\\ \tilde{H}_{j-1}, & S_j-1 \le S < S_{j+1}-1, j<j_0\\ \tilde{H}_{j_0-1}, & S_{j_0}-1 \le S < s,\\ k-1, & S=s,\\ 0, & S>s, \end{cases} \tag{2}
$$

for $S_{j_0}-1 \le s < S_{j_0+1}-1, k < \tilde{H}_{j_0-1}+1, j_0 \ge 2$,

$$
t_{s,k}^{(S)} = \begin{cases} M(S+1), & S<s,\\ k-1, & S=s,\\ \tilde{H}'_{j_0-1}, & s<S<S_{j_0+1}-1,\\ \tilde{H}'_{j-1}, & S_j-1 \le S < S_{j+1}-1, j>j_0, \end{cases} \tag{3}
$$

for $S_{j_0}-1 \le s < S_{j_0+1}-1, k \ge \tilde{H}'_{j_0-1}+1, j_0 \ge 2, (s,k) \ne (S_{\gamma+1}-1, \tilde{H}'_\gamma+1)$,

$$
t_{s,k}^{(S)} = \begin{cases} M(S+1), & S<s,\\ \tilde{H}'_{j-1}, & S_j-1 \le S < S_{j+1}-1, \gamma+1 \le j \le \ell \end{cases} \tag{4}
$$

for $s = S_{\gamma+1}-1, k = \tilde{H}'_\gamma+1$,

$$
t_{s,k}^{(S)} = \begin{cases} M(S+1), \quad S<S_2-1,\\ \sum_{l=1}^{j} M^{(S_l)} - \max\{j-j_0+\alpha,0\}M - \min\{j-1, j_0-\alpha-1\}(M-1),\\ \qquad S_j-1 \le S < S_{j+1}-1, 2 \le j \le j_0\\ \sum_{l=1}^{j_0+1} M^{(S_l)} - \alpha M - (j_0-\alpha)(M-1),\\ \qquad S_{j_0+1}-1 \le S < S_{j_0+2}-1,\\ \sum_{l=1}^{j} M^{(S_l)} - (j-j_0-1+\alpha)M - (j_0-\alpha)(M-1),\\ \qquad S_j-1 \le S < S_{j+1}-1, j_0+2 \le j \le j_0+(a-\alpha)+1\\ \sum_{l=1}^{j} M^{(S_l)} - aM - (j-a-1)(M-1),\\ \qquad S_j-1 \le S < S_{j+1}-1, j \ge j_0+(a-\alpha)+2, \end{cases} \tag{5}
$$

for $s = S_{j_0}-1, k-1 = \sum_{l=1}^{j_0} M^{(S_l)} - \alpha M - (j_0-1-\alpha)(M-1), 0<\alpha \le j_0-1$, $k \ne \tilde{H}'_{j_0-1}+1$, and

$$t_{s,k}^{(S)} = \begin{cases} M(S+1), & S < S_2 - 1, \\ \sum_{l=1}^{j} M^{(S_l)} - \max\{j - j_0 + \alpha, 0\}M - \min\{j-1, j_0 - \alpha - 1\}(M-1) \\ \quad S_j - 1 \le S < S_{j+1} - 1, 2 \le j \le j_0 - 1 \\ \sum_{l=1}^{j_0} M^{(S_l)} - (\alpha - 1)M - (j_0 - \alpha)(M-1), \\ \quad S_{j_0} - 1 \le S < s, \\ \sum_{l=1}^{j_0} M^{(S_l)} - \alpha M - (j_0 - \alpha - 1)(M-1), \\ \quad s \le S < S_{j_0+1} - 1, \\ \sum_{l=1}^{j} M^{(S_l)} - (j - j_0 + \alpha)M - (j_0 - \alpha - 1)(M-1), \\ \quad S_j - 1 \le S < S_{j+1} - 1, j_0 + 1 \le j \le j_0 + (a - \alpha) \\ \sum_{l=1}^{j} M^{(S_l)} - aM - (j - a - 1)(M-1), \\ \quad S_j - 1 \le S < S_{j+1} - 1, j \ge j_0 + (a - \alpha) + 1, \end{cases} \tag{6}$$

for $S_{j_0} - 1 < s < S_{j_0+1} - 1, k - 1 = \sum_{l=1}^{j_0} M^{(S_l)} - \alpha M - (j_0 - 1 - \alpha)(M-1)$, $0 < \alpha \le j_0 - 1$, $k \ne \tilde{H}'_{j_0-1} + 1$.

We construct the blow-up process with $T_{s,k}$ in Eq. (5) in **Case 1 (2)**. Thus, the proof is established. □

Remark 2. To describe the proof simply, we construct a blowing-up process without the condition $T_{s,k} \le T_{s',k'}$ or $T_{s,k} \ge T_{s',k'}$. We can construct a blowing-up process with that condition; however, it results in a more complex expression.

Example 2. This is the example of T and t for $M^{(1)} = 6$, $M^{(2)} = M^{(3)} = 2$, $M^{(4)} = 3$, $M^{(5)} = M^{(6)} = 2$ in Lemma 6 with blowing up process. We have $\mathcal{M} = \{2, 3, 5, 6\}$, $M = 3$, $a = 2$, $\ell = 3$, $\lambda = \frac{3}{2}$, $\theta = 3$. Also we have $\tilde{H}_1 = 1, \tilde{H}_2 = 0, \tilde{H}_3 = 0$, and $\tilde{H}'_1 = 2, \tilde{H}_2 = 1, \tilde{H}_3 = 0$.

The blowing up process as follows.

$S = 1, J = 0$	Case 2	$T_{1,1} = (0, 0, 0, 0, 0)$
$S = 1, J = 1$	Case 2	$T_{1,2} = (1, 1, 1, 1, 1)$
$S = 2, J = 0$	Case 1(1) with $u_{1,2}$	$T_{1,2} = (1, 0, 0, 0, 0)$
$S = 2, J = 0$	Case 2	$T_{2,1} = (2, 0, 0, 0, 0)$
$S = 2, J = 1$	Case 2	$T_{2,2} = (2, 1, 1, 1, 1)$
$S = 3, J = 0$	Case 1(2) with $u_{2,2}$	$T_{3,1} = (2, 1, 0, 0, 0)$
$S = 3, J = 1$	Case 2	$T_{3,2} = (2, 2, 1, 1, 1)$
$S = 4, J = 0$	Case 1(1) with $u_{3,2}$	$T_{3,2} = (2, 2, 1, 0, 0)$
$S = 4, J = 0$	Case 1(2) with $u_{2,2}$	$T_{4,1} = (2, 1, 1, 0, 0)$
$S = 4, J = 1$	Case 2	$T_{4,2} = (2, 2, 2, 1, 1)$
$S = 5, J = 0$	Case 1(1) with $u_{4,2}$	$T_{4,2} = (2, 2, 2, 1, 0)$
$S = 5, J = 0$	Case 1(2) with $u_{2,2}$	$T_{2,2} = (2, 1, 1, 1, 0)$
$S = 5, J = 0$	Case 2	$T_{5,1} = (2, 2, 2, 2, 0)$

We have λ by $T_{4,1} = (2, 1, 1, 0, 0)$, $T_{4,2} = (2, 2, 2, 1, 0)$ and $T_{2,2} = (2, 1, 1, 1, 0)$.

5. Conclusions

We have determined the exact values for the learning coefficients of multiple-layered neural networks with linear units, which generalizes the result of paper (Aoyagi & Watanabe, 2005b). We use the inductive method and a recursive blow-up method and obtain a manifold with a resolution map of singularities.

The main theorems imply that even though model complexities are increased, generalization errors which are expressed by the learning coefficients λ are decreased when the number of layers is increased. This fact seems to be one of the reasons for multiple-layered neural networks having better efficiencies for learning. For example, if the matrix sizes are all the same $H^{(1)} \times H^{(1)}$ then

$$\lambda = \frac{-r^2 + 2rH^{(1)}}{2} + \frac{a(L-a)}{4L} + \frac{L+1}{4L}(H^{(1)} - r)^2$$
$$\to \frac{-r^2 + 2rH^{(1)}}{2} + \frac{H^{(1)} - r}{4} + \frac{1}{4}(H^{(1)} - r)^2$$

if $L \to \infty$, where r denotes the rank of the probability density function $pl_0(X, Y)$, $L + 1$ the number of layers, and $a = L(\frac{L+1}{L}(H^{(1)} - r) - \lceil \frac{L+1}{L}(H^{(1)} - r) \rceil + 1)$; here $\lceil x \rceil$ denotes the ceiling function of the real number x. The theoretical value λ shows that as the value $H^{(s)}$'s becomes larger, the generalization error increases, and as the depth of the layers increases, it decreases, when we consider over-parameterized regimes. This seems to explain the reason why double descent (Nakkiran, Kaplun, Bansal, Yang, Barak, & Sutskever, 2020) occurs in machine learning. On the other hand, its order $\theta = a(L-a) + 1 \to \infty$ as $L \to \infty$, since we have $M = H^{(1)} - r + 1$ and $a = H^{(1)} - r$ when $L > H^{(1)} - r$. Note that if $L \le H^{(1)} - r$ then the curve of θ oscillates. Recently, these theoretical values of the learning coefficients have been used effectively in numerical experiments such as information criteria, the Markov chain Monte Carlo (Nagata & Watanabe, 2008a, 2008b), and model selection methods. We are also preparing numerical experiments to compare with the theory and will consider these applications in the future.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research was funded by Grants-in-Aid for Scientific Research - KAKENHI, Japan - under grant number 18K11479. We would like to express our gratitude to the reviewers for their insightful discussions, which greatly contributed to the enhancement of our paper. We also extend our appreciation to Richard Haase, PhD, from Edanz (https://jp.edanz.com/ac), for editing a draft of this manuscript.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
Aoyagi, M. (2006). The zeta function of learning theory and generalization error of three layered neural perceptron. *RIMS Kokyuroku, Recent Topics on Real and Complex Singularities*, *1501*, 153–167.
Aoyagi, M. (2009). Log canonical threshold of vandermonde matrix type singularities and generalization error of a three layered neural network. *International Journal of Pure and Applied Mathematics*, *52*(2), 177–204.
Aoyagi, M. (2010). A Bayesian learning coefficient of generalization error and vandermonde matrix-type singularities. *Communications in Statistics. Theory and Methods*, *39*(15), 2667–2687.
Aoyagi, M. (2013a). Consideration on singularities in learning theory and the learning coefficient. *Entropy*, *15*(9), 3714–3733.
Aoyagi, M. (2013b). Learning coefficient in Bayesian estimation of restricted Boltzmann machine. *Journal of Algebraic Statistics*, *4*(1), 30–57.
Aoyagi, M. (2019a). Learning coefficient of vandermonde matrix-type singularities in model selection. *Entropy (Information Theory, Probability and Statistics)*, *21*(6–561), 1–12.
Aoyagi, M. (2019b). Learning coefficients and information criteria. *Frontiers in Artificial Intelligence and Applications*, 351–362.
Aoyagi, M., & Watanabe, S. (2005a). Resolution of singularities and the generalization error with Bayesian estimation for layered neural network. *IEICE Trans. J88-D-II*, *10*, 2112–2124.
Aoyagi, M., & Watanabe, S. (2005b). Stochastic complexities of reduced rank regression in Bayesian estimation. *Neural Networks*, *18*, 924–933.
Drton, M. (2012). Conference lecture: Bayesian information criterion for singular models. Algebraic Statistics 2012 in the Alleghenies at the Pennsylvania State University, http://jasonmorton.com/aspsu2012/.
Drton, M., Lin, S., Weihs, L., & Zwiernik, P. (2017). Marginal likelihood and model selection for Gaussian latent tree and forest models. *Bernoulli*, *23*(2), 1202–1232.
Fulton, W. (1993). *Introduction to toric varieties, annals of mathematics studies*. Princeton, NJ, USA: Princeton University Press.

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 Math*, *79*, 109–326.

Kashiwara, M. (1976). B-functions and holonomic systems. *Inventions Mathematics*, *38*, 33–53.

Kollár, J. (1997). Singularities of pairs. *Algebraic geometry-Santa Cruz 1995, Proc. Symp. Pure Math., American Mathematical Society, Providence, RI*, *62*, 221–287.

Lin, S. (2010). Asymptotic approximation of marginal likelihood integrals. (preprint).

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.

Mustata, M. (2002). Singularities of pairs via jet schemes. *Journal of the American Mathematical Society*, *15*, 599–615.

Nagata, K., & Watanabe, S. (2008a). Exchange Monte Carlo sampling from Bayesian posterior for singular learning machines. *IEEE Transactions on Neural Networks*, *19*(7), 1253–1266.

Nagata, K., & Watanabe, S. (2008b). Asymptotic behavior of exchange ratio in exchange Monte Carlo method. *International Journal of Neural Networks*, *21*(7), 980–988.

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2020). Deep double descent: Where bigger models and more data hurt. ICLR2020, https://arxiv.org/pdf/1912.02292.pdf.

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

Rusakov, D., & Geiger, D. (2002). Asymptotic model selection for naive Bayesian networks. *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, 438–445.

Rusakov, D., & Geiger, D. (2005). Asymptotic model selection for naive Bayesian networks. *Journal of Machine Learning Research*, *6*, 1–35.

Schwarz, G. (1978). Estimating the dimension of a model. *The 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.

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, S. (2009). *Algebraic geometry and statistical learning theory, vol. 25*. New York, USA: Cambridge University Press.

Watanabe, S. (2010). Equations of states in singular statistical estimation. *Neural Networks*, *23*(1), 20–34.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, (14), 867–897.

Watanabe, S. (2018). *Mathematical theory of Bayesian statistics*. New York, NY, USA.

Zwiernik, P. (2011). An asymptotic behavior of the marginal likelihood for general Markov models. *Journal of Machine Learning Research*, *12*, 3283–3310.