# [チュートリアル講演] 特異点解消と学習理論への応用

青柳　美輝†　　渡辺　澄夫†

† 上智大学理工学部　〒102 − 8554 東京都千代田区紀尾井町 7–1
† 東京工業大学精密工学研究所,　〒226–8503 神奈川県横浜市緑区長津田 4259
E-mail: †miki-a@sophia.ac.jp, ††swatanab@pi.titech.ac.jp

**あらまし**　はじめに代数多様体上の代数的部分集合や特異点などの定義を行い，特異点解消を具体的に求めるための方法である blow up 操作について解説する．この応用としては，カルバック情報量と事前分布より定まる 1 変数複素関数である学習モデルのゼーター関数の極を求めることが出来る．論文 [1]〜[3] において，この極は，今まで解析困難であった非正則学習モデルの場合でも，ベイズ推測に関する学習曲線の漸近展開の情報を与えることが示されている．
**キーワード**　統計的推定，階層モデル，非正則モデル，ベイズ推定，特異点解消

# Resolution of singularities and its application to learning theory

Miki AOYAGI† and Sumio WATANABE†

† Faculty of Science and Technology, Sophia University
7–1, Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan
† Precision and Intelligence Laboratory, Tokyo Institute of Technology,
4259 Nagatsuda, Midori-ku, Yokohama, 226–8503 Japan
E-mail: †miki-a@sophia.ac.jp, ††swatanab@pi.titech.ac.jp

**Abstract**　Our purpose in this paper is to give an introduction to algebraic geometry and especially construction of the blowing-up. This construction is the main tool in the resolution of singularities of an algebraic variety. By using the blowing-up process, it can be calculated the poles of a zeta function which is integral of the Kullback distance and a certain priori probability density function. Hierarchical learning machines such as layered neural networks and Gaussian mixtures are non-regular learning machines. It was proved that the largest pole of the zeta function gives asymptotically the stochastic complexity of non-regular learning machine (see [1]〜[3])
**Key words**　stochastic complexity, layered neural networks, non-regular learning machines, Bayesian estimate, resolution of singularities

## 1. まえがき

ニューラルネットワークのように，入力からも出力からも直接には働きを定められていない隠れた部分を持つ学習モデルは，パラメータが特定不能になり，フィッシャー計量が縮退するという特徴をもつために，従来の統計理論だけでは解析不可能であったが，この場合に関して，次のように特異点解消により構成された方法論と数理的に緊密な関係を持つことが明らかになっている [1]〜[3]．サンプル数が $n$ のときの学習モデルの確率的複雑さ $F(n)$ はある有理数 $\lambda$ と自然数 $m$ が存在して

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

となる．ここで $O(1)$ は $n$ の有界な関数である．ベイズ推測における汎化誤差（真の分布からベイズ推測された分布までのカルバック情報量の平均値）$G(n)$ は，確率的複雑さの差分 $F(n+1) - F(n)$ に等しいものであるが，漸近展開可能であれば，上記の定数 $\lambda$ および $m$ を用いて

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

となる．この $\lambda$ および $m$ が，カルバック情報量と事前分布より定まる 1 変数複素関数である学習モデルのゼーター関数より，決まり，特異点解消の方法を用いて計算される量である．正則な統計モデルでは，$\lambda =$ パラメータ数$/2$, $m = 1$ である．すなわち複雑な構造を持つ学習モデルのベイズ推測に関する学習曲線は，たくさんの特異点が局所的に引き起こす全ての学習曲線の下側にあり，この結果，正則な統計モデルよりも，常に優れた学習が実現されている．

この論文では，まず代数幾何学における特異点解消の定理を

紹介し，応用として，多層神経回路網の汎化誤差の漸近展開を考察する．

## 2. 特異点解消

以下の説明は, [4]〜[8] を主に参照している．

### 2.1 定義および特異点解消定理

最初に，代数幾何でよく使用される用語の定義を行う．

以下，$k$ は標数 0 の体 (例えば実数体，複素数体) であるとする．$k[x] = k[x_1, x_2, \cdots, x_n]$ を体 $k$ を係数とする $x_1, \cdots, x_n$ の多項式全体の集合とする．$f_1, \cdots, f_l \in k[x]$ に対して，

$$V(f_1, \cdots, f_l) = \{a = (a_1, \cdots, a_n) \in k^n | f_\alpha(a) = 0, 1 \leqq \alpha \leqq l\}$$

を $f_1, \cdots, f_l$ が定める代数的集合という．同様に多項式環 $k[x]$ のイデアル $J$ に対して

$$V(J) = \{b \in k^n | J \text{ の任意の元 } g \text{ に対して } g(b) = 0\}$$

をイデアル $J$ が定める代数的集合またはアフィン代数的集合と呼ぶ．ここでイデアル $J$ とは，$k[x]$ の部分集合で，ある $f_1, \cdots, f_l$ が存在して，

$$J = \{g(x) = \sum_{\alpha=1}^{l} g_\alpha(x) f_\alpha(x) | g_\alpha(x) \in k[x]\}$$

と書けるもののことである．従って，$V(J) = V(f_1, \cdots, f_l)$ である．代数的集合の全体 $V(J)$ は閉集合の公理を満たし，こうして定義された $k^n$ の位相をザリスキー位相という．以下，慣例に従って，$k^n$ を $\mathbf{A}_k^n$ とかく．

代数的集合 $V$ は，$V = V_1 \cup V_2$, $V \neq V_1$, $V \neq V_2$ と 2 つの代数的集合 $V_1$, $V_2$ の和集合になるときに可約と呼ばれる．可約でないとき既約という．

イデアル $J$ が，$f_1 f_2 \in J$ ならば，$f_1 \in J$, $f_2 \in J$ を満たすとき，素イデアルという．一方，$I(V) = \{f \in k[x] |$ 全ての $a \in V$ に対して $f(a) = 0\}$ とおけば，$J \subset I(V(J))$ が成り立つ．代数的集合 $V(J)$ が既約であるための必要十分条件は $I(V(J))$ が素イデアルであることであり，この時，$V(J)$ をアフィン代数多様体という．

代数的集合 $V$ に対して，そのイデアルを $I(V)$ とする．$k[V] = k[x]/I(V)$ は $V$ 上の関数の環とみなすことが出来る．これを $V$ の座標環という．もし，$I(V)$ が素イデアルならば剰余環 $k[V]$ は整域 ($fg = 0$, $f, g \in k[V]$ ならば $f = 0$ または $g = 0$) となる．

写像 $\phi : V \to W$ が多項式写像であるということを多項式 $f_1, \cdots, f_m \in k[x]$ によって

$$\phi(a) = (f_1(a), \cdots, f_m(a)), a \in V$$

と表せるもので定義する．多項式関数の概念の一般化である．

アフィン代数多様体 $V \subset \mathbf{A}_k^n$ の座標環 $k[V]$ は整域になるが，その商体

$$k(V) = \{f = g/h | g, h \in k[V]\}$$

を $V$ の関数体といい，この要素を $V$ 上の有理関数とよぶ．$f$ は $h(a) \neq 0$ となる点 $a$ でのみ値が定義されている．

$a \in V$ に対して，環 $\mathcal{O}_a(V) = \{f \in k(V) |$ ある $g, h \in k[V]$ が存在して，$h(a) \neq 0$ かつ $f = g/h\}$ を $a$ における $V$ の局所環といい，その元を $a$ で正則という．$f$ の定義域を $\mathrm{dom} f = \{a \in V | a$ で $f$ は正則 $\}$ と書く．

有理写像 $f : V - - \to \mathbf{A}_k^n$ とは，部分集合上で定義された写像で，

$$f = (f_1, \cdots, f_n), f_1, \cdots, f_n \in k(V)$$

で書けるものである．

2 つのアフィン代数多様体 $V \subset \mathbf{A}_k^n$, $W \subset \mathbf{A}_k^m$ に対して，$V$ から $W$ への有理写像 $f : V - - \to W$ とは，有理写像 $f : V - - \to \mathbf{A}_k^n$ で，$f(\mathrm{dom} f) \subset W$ を満たすもののことである．

次に有理写像 $f = (f_1, \cdots, f_m)$, $g = (g_1, \cdots, g_n)$ の合成写像 $g \circ f$ を自然に定義したいが，合成写像は $\mathrm{dom} f \cap f^{-1}(\mathrm{dom} g)$ 上で定義されるべき写像になる．しかし，この集合が空になることがある．例えば，$f(x) = (1/x, 1/x^2)$, $g(y_1, y_2) = 1/(y_1^2 - y_2)$ であれば，$g \circ f(x) = 1/(1/x^2 - 1/x^2)$ となり，意味が無くなる．

そこで，次のような概念を導入する．有理写像 $f : V - - \to W$ に対して，$f(\mathrm{dom} f)$ がザリスキー位相に関して $W$ で稠密である時，$f$ は支配的であるという．このときはいつでも合成写像 $g \circ f$ が定義できる．

アフィン代数多様体 $V \subset \mathbf{A}_k^n$, $W \subset \mathbf{A}_k^m$ と開部分集合 $U_1 \subset V$, $U_2 \subset W$ に対して，有理写像 $f : V - - \to W$ が $U_1 \subset \mathrm{dom} f$，$f(U_1) \subset U_2$ を満たすとき，$f : U_1 \to U_2$ は正則写像という．特に，正則写像 $f : V \to W$ は多項式写像と一致する．正則写像 $f : U_1 \to U_2$ が逆写像も正則写像であるとき正則同型（写像）とよぶ．正則同型 $f : U_1 \to U_2$ が存在するとき $U_1$ と $U_2$ は双正則同値とよぶ．

有理写像 $f : V - - \to W$ が双有理写像であるとは，$f$ が有理逆写像をもつことである．双有理写像 $f : V - - \to W$ が存在するとき $V$ と $W$ は双有理同値 とよぶ．また $f$ は支配的になる．

次に射影空間の定義を行う．$W = k^n - \{(0, 0 \cdots, 0)\}$ に対して，同値関係 $\sim$ を

$$(a_0, a_1, \cdots, a_n) \sim (b_0, b_1, \cdots, b_n)$$

$$\Longleftrightarrow \quad (a_0, a_1, \cdots, a_n) = \alpha(b_0, b_1, \cdots, b_n)$$
$$\text{を満足する元 } \alpha \in k - \{0\} \text{ が存在する}$$

で定義する．このとき，$\mathbf{P}_k^n = W/\sim$ を射影空間と呼ぶ．

この空間は，$U_i = \mathbf{A}_k^n$, $i = 1, \cdots, n$ を次のように貼り合わせたものである．

$\mathbf{P}_k^n$ の元を代表元を用いて $(a_0 : a_1 : \cdots : a_n)$ で表すと，写像

$$\nu_i : U_i \to \mathbf{P}_k^n \ ; \ (a_{i0}, \cdots, a_{ii-1}, a_{ii+1}, \cdots, a_{in})$$
$$\mapsto (a_{i0}, \cdots, a_{ii-1}, 1, a_{ii+1}, \cdots, a_{in})$$

は，$\{(a_0 : a_1 : \cdots : a_n) \in \mathbf{P}_k^n | a_i \neq 0\}$ への全単射写像である．

一方，各 $U_i = \{(a_{i0}, \cdots, a_{ii-1}, a_{ii+1}, \cdots, a_{in}) \in \mathbf{A}_k^n\}$ 上で $a_{ij} \neq 0 (i < j)$ をみたす開集合を $U_{ij}$ と表せば，

$$U_{ij} \to U_{ji} \quad ; \quad (a_{i0}, \cdots, a_{ii-1}, a_{ii+1}, \cdots, a_{in})$$
$$\mapsto (a_{i0}/a_{ij}, \cdots, a_{ii-1}/a_{ij}, 1/a_{ij}, a_{ii+1}/a_{ij}, \cdots, a_{ij-1}/a_{ij}, a_{ij+1}/a_{ij}, \cdots, a_{in}/a_{ij})$$

は同型を与える．この同型により $\{U_i, \nu_i\}$ を貼り合わせると，$\mathbf{P}_k^n$ と同型になる．$\mathbf{P}_k^n$ の元を $a = (a_0 : a_1 : \cdots : a_n)$ と書くときこれを $a$ の斉次座標という．

次に斉次イデアルを定義する．射影空間においては，代数的集合などが，斉次イデアルを用いて定義される．

$d$ 次斉次多項式であるとは，

$$f = \sum_{i_0+\cdots+i_n=d} x_0^{i_0} x_1^{i_1} \cdots x_n^{i_n}$$

と書けるものである．イデアル $I \subset k[x_0 \cdots, x_n]$ が斉次イデアルであるとは任意の $f \in I$ に対して，$f$ を斉次多項式に分解して $f = f_0 + \cdots + f_N$, ($f_i$ は $i$ 次斉次多項式) としたとき，$f_i$ が $I$ に属することである．これは，$I$ が有限個の斉次多項式で生成されることと同値である．

$I \subset k[x_0 \cdots, x_n]$ を斉次イデアルとする．$\mathbf{P}_k^n$ のイデアル $I$ が定める代数的集合を

$$V(I) = \{b \in k^{n+1} | I \text{ の任意の元 } g \text{ に対して } g(b) = 0\}$$

とする．これらの集合によって定義される $\mathbf{P}_k^n$ の位相をザリスキー位相という．また，$\mathbf{A}_k^n$ と同様に $I(V) = \{f \in k[x_1, \cdots, x_n] |$ 全ての $a \in V$ に対して $f(a) = 0, f$ は斉次 $\}$ が定義される．

$V \subset \mathbf{P}_k^n$ を既約な代数的集合，即ち，射影代数多様体とし，そのイデアルを $I(V)$ とする．

$V$ 上には定数以外の多項式は存在しないので，有理関数から定義する．

$V$ 上の有理関数とは，同じ次数の斉次多項式 $g, h \in k[x]$ の商によって与えられる関数 $f : V - - \to k$ のことである．有理関数全体のなす集合，有理関数体は

$$k(V) = \{g/h | g, h \in k[x] \text{ は斉次かつ同じ次数で } h \notin I(V)\}/\sim$$

と表せる．但し，同値関係 $\sim$ は

$$g/h \sim g'/h' \iff h'g - g'h \in I(V).$$

アフィンの場合と同様に，$\mathcal{O}_a(V) = \{f \in k(V) |$ ある $g, h \in k[V]$ が存在して, $h(a) \neq 0$ かつ $f = g/h\}$ を $a$ における $V$ の局所環といい，その元を $a$ で正則という．$f$ の定義域を $\mathrm{dom} f = \{a \in V | a$ で $f$ は正則 $\}$ と書く．

有理写像 $f : V - - \to \mathbf{A}_k^n$ とは，

$$f = (f_1, \cdots, f_n), f_1, \cdots, f_n \in k(V)$$

で表されるものであり，同様に，有理写像 $f : V - - \to \mathbf{P}_k^n$ とは，

$$f = (f_0 : f_1 : \cdots : f_n), f_0, f_1, \cdots, f_n \in k(V)$$

で表されるものである．

射影代数多様体 $V \subset \mathbf{P}_k^n$, $W \subset \mathbf{P}_k^m$ と開部分集合 $U_1 \subset V$, $U_2 \subset W$ に対して，有理写像 $f = (f_0 : \cdots : f_m) : V - - \to W$ が，次の 3 つの条件 (1)$U_1$ 上 $f_i$ は正則であり，(2) 少なくとも一つの $i$ に対して $f_i \neq 0$，また，(3)$f(U_1) \subset U_2$ を満たすとき，$f : U_1 \to U_2$ は正則写像という．

同様に，$\mathbf{A}_k^n \times \mathbf{P}_k^m$ にも代数的集合 $V(I)$ などが自然に定義される．この論文では，これらをまとめて代数多様体(注1) と呼ぶことにする．

代数多様体 $V, W$ に対して，実は，$V$ と $W$ が双正則同値ならば $k[V] \cong k[W]$ が成り立ち，$V$ と $W$ が双有理同値ならば $k(V) \cong k(W)$ が成り立つ．

一般に代数多様体 $V(J) = V(f_1, \cdots, f_\ell)$ 上では，ある整数 $n$ が存在して，

$$\mathrm{rank} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x), & \frac{\partial f_1}{\partial x_2}(x), & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & & \vdots \\ \frac{\partial f_\ell}{\partial x_1}(x), & \frac{\partial f_\ell}{\partial x_2}(x), & \cdots & \frac{\partial f_\ell}{\partial x_n}(x) \end{pmatrix} \leqq \ell - n$$

かつ，稠密開集合上では等号が成立する．等号が成立する点全体を滑らかな点または非特異点という．また，そうでない点を特異点という．この $n$ は $V$ の次元と呼ばれる．

代数多様体の射 $f : Y \to X$ が射影的であるとは，埋め込み $i : Y \hookrightarrow X \times \mathbf{P}^n$ が存在して，$f = p_1 \circ i$ と分解することである．但し，$p_1 : X \times \mathbf{P}^n \to X$ は第 1 成分への射影．

与えられた代数多様体 $X$ に対して，滑らかな代数多様体 $Y$ からの双有理で射影的な射 $f : Y \to X$ を特異点解消という．

### 2. 2　Blow up

特異点解消の手段である blow up を紹介する．最初に，$\mathbf{A}_k^n$ の原点 $0 = (0, \cdots, 0)$ を中心とする blow up を定義する．$Y$ を $\mathbf{A}_k^n \times \mathbf{P}_k^{n-1}$ の部分集合で

$$Y = \{(0, \cdots, 0)\} \times \mathbf{P}_k^{n-1} \cup \{(a_1, \cdots, a_n) \times (a_1 : \cdots : a_n) | (a_1, \cdots, a_n) \in \mathbf{A}_k^n - (0)\}$$

とおく．$\mathbf{A}_k^n \times \mathbf{P}_k^{n-1}$ の座標を $(x_1, \cdots, x_n) \times (t_1 : \cdots : t_n)$ であらわす．この時，$Y$ の定義イデアル $I(Y)$ は $x_i t_j - x_j t_i$ たちで生成される斉次イデアルである．

$$I(Y) = (x_i t_j - x_j t_i) \quad \subset k[x_1, \cdots, x_n, t_1, \cdots, t_n]$$

従って，$Y$ は代数多様体であり，この時，$Y$ から $\mathbf{A}_k^n$ への第 1 成分への射影 $\mu = p_1$ を blow up と呼ぶ．明らかに，$\mu$ は双有理で射影的な射である．

Y の開集合 $U_i = \{t_i \neq 0\} \cap Y$ に対して，写像

$$Y_i (\cong \mathbf{A}_k^n) \ni (\xi_{1i}, \cdots, \xi_{ni})$$
$$\mapsto (\xi_{ii}\xi_{1i}, \cdots, \xi_{ii}\xi_{i-1i}, \xi_{ii}, \xi_{ii}\xi_{i+1i}, \cdots, \xi_{ii}\xi_{ni})$$
$$\times (\xi_{1i} : \cdots : \xi_{i-1i} : 1 : \xi_{i+1i} : \cdots : \xi_{ni}) \in U_i$$

は $\mathbf{A}_k^n$ から $U_i$ への同型を与える．即ち，$Y$ は $Y_i$ を貼り合せたもので構成される．

(注1)：一般に代数多様体はアフィン多様体や射影多様体などよりも広い概念である．

代数多様体 $Z \subset \mathbf{A}_k^n$ に対しては，$\mu^{-1}(Z-\{0\})(\cong Z-\{0\})$ の $Y$ での閉包 $Z'$ を $Z$ の固有変換といい，$\mu$ を $Z'$ に制限したものを blow up という．

次にイデアルによる blow up を定義する．イデアル $J \subset k[x_1,\cdots,x_n]$ を $f_1,\cdots,f_m$ で生成されるものとする．

$Y$ を $\mathbf{A}_k^n \times \mathbf{P}_k^{n-1}$ の部分集合で

$$Y = V(J) \times \mathbf{P}_k^{n-1} \cup$$

$$\{a \times (f_1(a):\cdots:f_m(a)) | a \in \mathbf{A}_k^n - V(J)\}$$

とおく．この時，$Y$ の定義イデアル $I(Y)$ は $f_it_j - f_jt_i$ たちで生成される $k[x_1,\cdots,x_n,t_1,\cdots,t_n]$ のイデアルで $t$ に関しては斉次のものである．従って，$Y$ は代数多様体であり，$Y$ から $\mathbf{A}_k^n$ への第 1 成分への射影 $\mu = p_1$ をイデアル $J$ による blow up と呼ぶ．これは $J$ の生成元 $f_1,\cdots,f_m$ の取り方によらず同型を除いて一意的に定まる．

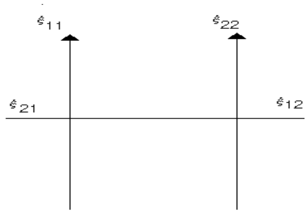
代数多様体 $Z \subset \mathbf{A}_k^n$ に対しては，$\mu^{-1}(Z-V(J))(\cong Z-V(J))$ の $Y$ での閉包 $Z'$ を $Z$ の固有変換といい，$\mu$ を $Z'$ に制限したものを blow up という．

［定理 1］（広中の定理） 代数多様体は，適当な blow up 列の合成により特異点を解消できる．

### 2. 3　特異点解消の例

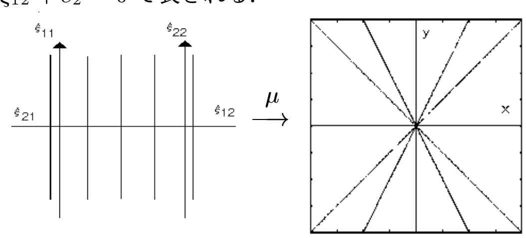
この節では，簡単な例を用いて特異点解消を説明する．

［例 1］ $f(x,y) = c_1x + c_2y$，($c_1, c_2$ は定数) で定義される代数多様体 $V(f)$ を考える．特異点はないが blow up によってどのように変化するかを調べる．$\mathbf{A}_k^2$ を原点 0 で blow up すれば，上記の記号をそのまま用いれば，$Y$ は $Y_1$，$Y_2$ を $\xi_{21} \neq 0$, $\xi_{12} \neq 0$ において，$\xi_{21} = 1/\xi_{12}$，$\xi_{11} = \xi_{22}\xi_{12}$，$\xi_{11}\xi_{21} = \xi_{22}$ で貼り合わせたものである．図では抽象的にしばしば次のようなものが用いられる．




この図は，$Y$ が 2 組の座標で表されることを示している．

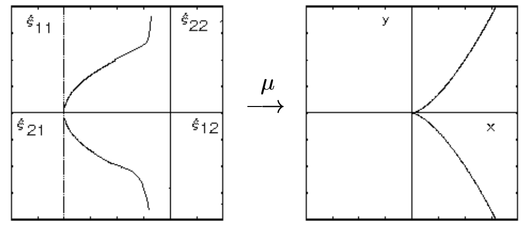
$Y_1$ 上では，$f(\xi_{11}, \xi_{11}\xi_{21}) = \xi_{11}(c_1 + c_2\xi_{21}) = 0$ となり，固有変換は，$c_1 + c_2\xi_{21} = 0$ で表される．$Y_2$ 上では $f(\xi_{12}\xi_{22}, \xi_{22}) = \xi_{22}(c_1\xi_{12} + c_2) = 0$ となり，固有変換は，$c_1\xi_{12} + c_2 = 0$ で表される．




［例 2］ $f(x,y) = x^n - y^2$, $n \geqq 3$ で定義される代数多様体 $V(f)$ を考える．特異点は原点 0 である．$\mathbf{A}_k^2$ を原点 0 で blow up すれば，$Y_1$ 上では $f(\xi_{11}, \xi_{11}\xi_{21}) = \xi_{11}^2(\xi_{11}^{n-2} - \xi_{21}^2) = 0$ となり，固有変換は，$\xi_{11}^{n-2} - \xi_{21}^2 = 0$ で表される．

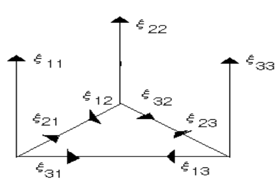
$Y_2$ 上では $f(\xi_{12}\xi_{22}, \xi_{22}) = \xi_{22}^2(\xi_{22}^{n-2}\xi_{12}^n - 1)$ となり，固有変換は，$\xi_{22}^{n-2}\xi_{12}^n - 1 = 0$ で表される．$\xi_{22}^{n-2}\xi_{12}^n - 1 = 0$ は非特異になる．$g_1 = \xi_{11}^{n-2} - \xi_{21}^2 = 0$ は同様に原点に特異点を持ち，次数が $n$ から $n-2$ に変化しているので，帰納的に定義される blow up の列により特異点が解消されることがわかる．

$n = 3$ の時の図は次のようになる．




2 次元の 1 つの多項式で定義される超曲面 $V(f) = \{f(x,y) = 0\}$ の場合は，適当な座標変換と特異点で点による blow up を繰り返し用いることによって特異点を解消できる．3 次元以上はイデアルによる blow up が必要になってくる．

［例 3］ $f(x,y) = x^4 + y^4 + z^2$ で定義される代数多様体 $V(f)$ を考える．特異点は原点 0 である．$\mathbf{A}_k^3$ を原点 0 で blow up すれば，上記の記号をそのまま用いれば，$Y$ は，$Y_1$，$Y_2$，$Y_3$ を次のように貼り合わせたものである．$Y_1$，$Y_2$ を $\xi_{21} \neq 0, \xi_{12} \neq 0$ において，$\xi_{21} = 1/\xi_{12}$，$\xi_{11} = \xi_{22}\xi_{12}$，$\xi_{11}\xi_{21} = \xi_{22}$ で貼り合わせ，$Y_1$，$Y_3$ を $\xi_{31} \neq 0$, $\xi_{13} \neq 0$ において，$\xi_{13} = 1/\xi_{31}$，$\xi_{33} = \xi_{31}\xi_{11}$，$\xi_{33}\xi_{13} = \xi_{11}$ で貼り合わせ，$Y_2$，$Y_3$ を $\xi_{32} \neq 0$, $\xi_{23} \neq 0$ において，$\xi_{32} = 1/\xi_{23}$，$\xi_{22} = \xi_{33}\xi_{23}$，$\xi_{22}\xi_{32} = \xi_{33}$，で貼り合わせる．図では抽象的にしばしば次のようなものが用いられる．




この図は 2 次元のときと同様，3 組の座標で $Y$ が成り立っていることを示すものである．

$\mathbf{A}_k^2$ を原点 0 で blow up すれば，$Y_1$ 上では $f(\xi_{11}, \xi_{11}\xi_{21}, \xi_{11}\xi_{31}) = \xi_{11}^2(\xi_{11}^2 + \xi_{11}^2\xi_{21}^4 + \xi_{31}^2) = 0$, $Y_2$ 上では $f(\xi_{22}\xi_{12}, \xi_{22}, \xi_{22}\xi_{32}) = \xi_{22}^2(\xi_{12}^4\xi_{22}^2 + \xi_{22}^2 + \xi_{32}^2) = 0$, $Y_3$ 上では $f(\xi_{33}\xi_{13}, \xi_{33}\xi_{23}, \xi_{33}) = \xi_{33}^2(\xi_{33}^2\xi_{13}^4 + \xi_{33}^2\xi_{23}^4 + 1) = 0$ で表される．$Y_3$ 上の固有変換 $\xi_{33}^2\xi_{13}^4 + \xi_{33}^2\xi_{23}^4 + 1 = 0$ は非特異になる．$Y_1$ 上の固有変換 $\xi_{11}^2 + \xi_{11}^2\xi_{21}^4 + \xi_{31}^2 = 0$ は $\xi_{11} = \xi_{31} = 0$ で $Y_2$ 上の固有変換 $\xi_{12}^4\xi_{22}^2 + \xi_{22}^2 + \xi_{32}^2 = 0$ は $\xi_{12} = \xi_{32} = 0$ で特異点を持つ．従って，有限回の点の blow up によっては決して解消は得られない．特異点解消はイデアルでの blow up によって得られる．

$\xi_{11}^2 + \xi_{11}^2\xi_{21}^4 + \xi_{31}^2 = 0$ を $x_1^2 + x_1^2y_1^4 + z_1^2 = 0$ と書き直しておく．$\mathbf{A}_k^3$ をイデアル $I = (x_1, z_1)$ で blow up する．$Y_1 \cong \mathbf{A}_k^3$ の座標を $(\xi'_{11}, \xi'_{21}, \xi'_{31})$，$Y_2 \cong \mathbf{A}_k^3$ の座標を $(\xi'_{12}, \xi'_{22}, \xi'_{32})$ とおくと，イデアル $I$ で $\mathbf{A}_k^3$ を blow up した集合 $Y$ は $Y_1$，$Y_2$ を

$\xi'_{31} \neq 0$, $\xi'_{12} \neq 0$, において，$\xi'_{31} = 1/\xi'_{12}$, $\xi'_{11} = \xi'_{32}\xi'_{12} = x_1$, $\xi'_{21} = \xi'_{22} = y_1$ $\xi'_{11}\xi'_{31} = \xi'_{32} = z_1$ で貼り合わせたものと同型になる．

$Y_1$ 上では $x_1^2 + x_1^2 y_1^4 + z_1^2 = \xi'^2_{11}(1 + \xi'^2_{21} + \xi'^2_{31}) = 0$ となり $Y_2$ 上では $x_1^2 + x_1^2 y_1^4 + z_1^2 = \xi'^2_{32}(\xi'^2_{12} + \xi'^2_{12}\xi'^2_{22} + 1) = 0$ で表される．固有変換 $1 + \xi'^2_{21} + \xi'^2_{31}$ と $\xi'^2_{12} + \xi'^2_{12}\xi'^2_{22} + 1$ はともに非特異になり特異点が解消されている．

$\xi_{12}^4\xi_{22}^2 + \xi_{22}^2 + \xi_{32}^2 = 0$ も同様に $I = (\xi_{22}, \xi_{32})$ で特異点解消できる．

従って，$V(f)$ はこれらの blow up の列で特異点解消できる．

Artin の代数化定理を用いれば，孤立特異点（特異点 $x$ のある近傍が存在して，その近傍の特異点は $x$ のみ）であるとき，十分小さい近傍に制限してみると代数多様体の孤立特異点とみなしてよい．例えば，$y^2 - x^2 = 0$ および $\sin(y-x)\sin(y+x) = 0$ は原点の近傍において同型であるとみなすことが出来る．

特異点解消定理は次のように述べることも出来る．

$k^n$ の解析的な関数 $f$ と特異点 $x$ に対して，その点の近傍 $W$ と，ある多様体 $U$ と $U$ から $W$ への固有（コンパクト集合の逆像もコンパクト）な解析関数 $\mu$ が存在して，任意の $u \in U$ について適当な局所座標 $(u_1, \cdots, u_n)$ が存在して，$f(\mu(u)) = u_1^{s_1} u_2^{s_2} \cdots u_n^{s_n}$ が成り立つ．ここで $s_1, \cdots, s_n$ は非負の整数．

## 3. 学習理論への応用

ここでは，渡辺澄夫によって得られている特異点解消の学習理論への応用について説明し，また 3 層ニューラルネットワークに適用して得られた結果について概略を述べる．

### 3.1 学習理論

入出力の空間を $x \in \mathbf{R}^N$，パラメータの空間を $w \in W \subset \mathbf{R}^d$ とする．学習モデル $p(x|w)$ とその事前分布 $\psi(w)$ が与えられているものとし，その真の分布を $p(x|w_0)$ と仮定する．

カルバック距離 $K(w)$ を $K(w) = \int p(x|w_0) \log \frac{p(x|w_0)}{p(x|w)} dx$ とおく．

$p(x|w_0)$ に従う独立なサンプルを $X^n = (X_1, X_2, ..., X_n)$ とする．与えられたサンプル $X^n$ と事前確率 $\psi(w)$ とから，事後確率 $p(w|X^n)$ を $p(w|X^n) = \frac{1}{Z_n}\psi(w)\Pi_{i=1}^n p(X_i|w)$ によって定める（$Z_n$ は正規化定数）．これよりベイズ推測を $p(x|X^n) = \int p(x|w)p(w|X^n)dw$ と定義する．汎化誤差 $G(n)$ をサンプル $X^n$ に関する平均 $E_n\{\cdot\}$ を用いて，

$$G(n) = E_n\{\int p(x|w_0) \log \frac{p(x|w_0)}{p(x|X_n)} dx\}$$

としておく．

確率的複雑さを

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

とすれば，

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

が成り立つ．

学習モデルのゼータ関数を

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

とおけば，[1]〜[3] より，この関数 $J(z)$ の最も原点に近い極を $\lambda$，その位数を $m$ とすると，

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

が成り立つ．ここで，$O(1)$ は $n$ の有界な関数である．

### 3.2 多層神経回路網の学習曲線

ここでは，入力ユニット 1 個，中間ユニット $p$ 個，出力ユニット 1 個の 3 層パーセプトロンの学習曲線を考察する．

$x$ を入力，$y$ を出力，パラメータを $w = \{a_m, b_m | m = 1, \cdots, p\}$ とし，

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

にとおく．このとき学習モデルは

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

で書ける．事前分布 $\psi(x)$ は $[-1, 1]$ 上の一様分布，真の分布は $w_0 = 0$ すなわち全ての $a_m, b_m$ は 0 の場合を考える．

このときカルバック情報量は

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

になるが，$\lambda$, $m$ は, [9] により，次の関数

$$\int_{0\ \text{の近傍}} \{\sum_{n=1}^{p} (\sum_{m=1}^{p} a_m b_m^{2n-1})^2\}^z \Pi_{m=1}^p \mathrm{d}a_m \mathrm{d}b_m$$

の原点に最も近い極と位数を求めればよい事がわかっている．また, [9] では次のような評価が得られている．

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

この節では，

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

の blow up を考察し，更に厳密な $\lambda$ の評価を紹介する．

### 3.3 blow up の計算

$$\begin{cases} b_1 = v_1 \\ b_m = v_1 b_m \ \ m = 2, \cdots, p \end{cases}$$ で変換すれば，

$$G = \{\sum_{n=1}^{p} v_1^{4n-2} (a_1 + \sum_{m=2}^{p} a_m b_m^{2n-1})^2\}^z v_1^{p-1}$$
$$\mathrm{d}a_1 \mathrm{d}v_1 \Pi_{m=2}^p \mathrm{d}a_m \mathrm{d}b_m$$

$d_1 = a_1 - \sum_{m=2}^p a_m b_m$ とおくと，

$$G=\{v_1^2(d_1^2+\sum_{n=2}^{p}v_1^{4n-4}$$
$$(d_1-\sum_{m=2}^{p}a_mb_m+\sum_{m=2}^{p}a_mb_m^{2n-1})^2)\}^z$$
$$v_1^{p-1}\mathrm{d}d_1\mathrm{d}v_1\Pi_{m=2}^{p}\mathrm{d}a_m\mathrm{d}b_m$$

ここで，次のような関数を定義しておく．

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

特にこの関数は

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

を満たしている．

$i\geqq 2$ に対して

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

とおき，$f$ を用いて，$G$ を表せば，

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

即ち，$d_1=c_2=\cdots=c_p=0$ での blow up を考察すればよい．

(1) から $\left\{\begin{array}{l}d_1=u_1\\v_1=u_1v_1\end{array}\right.$ で変換すれば，

$$\begin{aligned}G\ =\ &\{u_1^4v_1^2(1+\sum_{n=2}^{p}u_1^{4n-4}v_1^{4n-6}(u_1+f_{n,2}(b_2)c_2+\cdots\\&+f_{n,i}(b_i)c_i+\cdots+f_{n,n}(b_{n-1})c_n)^2)\}^z\\&u_1^pv_1^{p-1}\mathrm{d}u_1\mathrm{d}v_1\Pi_{m=2}^{p}\mathrm{d}a_m\mathrm{d}b_m\end{aligned}$$

従って，$-\frac{p}{2},-\frac{p+1}{4}$ の極が出てくる．

(1) を $\left\{\begin{array}{l}d_1=v_1d_1\\v_1=v_1\end{array}\right.$ で変換すれば，

$$\begin{aligned}G\ =\ &\{v_1^4(d_1^2+\sum_{n=2}^{p}v_1^{4n-6}(d_1v_1+f_{n,2}(b_2)c_2\\&+\cdots+f_{n,i}(b_i)c_i+\cdots+f_{n,n}(b_{n-1})c_n)^2)\}^z\\&v_1^p\mathrm{d}d_1\mathrm{d}v_1\Pi_{m=2}^{p}\mathrm{d}a_m\mathrm{d}b_m\end{aligned}\tag{2}$$

(2) を $\left\{\begin{array}{l}d_1=u_1\\v_1=u_1v_1\end{array}\right.$ で変換すれば，

$$\begin{aligned}G\ =\ &\{u_1^6v_1^4(1+\sum_{n=2}^{p}v_1^{4n-8}(u_1v_1+f_{n,2}(b_2)c_2+\cdots\\&+f_{n,i}(b_i)c_i+\cdots+f_{n,n}(b_{n-1})c_n)^2)\}^z\\&u_1^{p+1}v_1^p\mathrm{d}d_1\mathrm{d}v_1\Pi_{m=2}^{p}\mathrm{d}a_m\mathrm{d}b_m\end{aligned}$$

従って $-\frac{p+2}{6},-\frac{p+1}{4}$ の極が出てくる．

(2) を $\left\{\begin{array}{l}d_1=d_1v_1\\v_1=v_1\end{array}\right.$ で変換すれば，

$$\begin{aligned}G\ =\ &\{v_1^6(d_1^2+\sum_{n=2}^{p}v_1^{4n-8}(d_1v_1^2+f_{n,2}(b_2)c_2+\cdots\\&+f_{n,i}(b_i)c_i+\cdots+f_{n,n}(b_{n-1})c_n)^2)\}^z\\&v_1^{p+1}\mathrm{d}d_1\mathrm{d}v_1\Pi_{m=2}^{p}\mathrm{d}a_m\mathrm{d}b_m\end{aligned}\tag{3}$$

この論文では詳細は述べないが，このような方法により，$i^2\leqq p$ となる最大の $i$ について

$$-\frac{p+i(i-1)+2i}{4i+2}=-\frac{p+i^2+i}{4i+2}$$

が原点に近い極として出てくる．

## 4. ま と め

特異点解消を代数幾何学的観点から説明を行い，特異点解消の手段である blow up 操作について具体的に考察した．また，応用として，3 層ニューラルネットワークの学習モデルのゼータ関数の極を求める計算方法を示し，blow up により学習曲線の漸化式を得られることを示した．これは，代数幾何学の方法が，学習理論という極めて具体的な世界の複雑な問題を解決できるという意味での強さを持っていることを表している．今後は，学習モデルのゼータ関数の極の性質を代数幾何学観点から考察をおこないたい．

### 文 献


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

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

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

[4] M. リード，若林 功訳，“初等代数幾何講義,” 岩波書店，東京，1991.

[5] 石井 志保子，“特異点入門,” Springer，東京，1997.

[6] 川又 雄二郎，“共立講座 21 世紀の数学，代数多様体論,” 共立出版，東京，1997.

[7] 上野 健爾，“岩波講座　現代数学の基礎，代数幾何,” 岩波書店，東京，1997.

[8] 堀川穎二，“複素代数幾何学入門,” 岩波書店，東京，1990.

[9] Watanabe, S.,“ On the generalization error by a layered statistical model with Bayesian estimation, ”IEICE Trans., J81-A, pp. 1442-1452, 1998, (English version : Elect. and Comm. in Japan., John Wiley and Sons, Vol.83, No.6, pp.95-106, 2000.