論　文

幾何学的当てはめにおけるモデル選択

金谷　健一†

Model Selection for Geometric Fitting

Kenichi KANATANI†

あらまし　「幾何学的 AIC」と「幾何学的 MDL」を根本原理に戻って導出し，種々の問題点を明らかにする．そして行列のランクの推定の計算例を示す．

キーワード　幾何学的当てはめ，幾何学的 AIC，幾何学的 MDL，モデル選択, 行列のランクの推定

1. ま え が き

筆者は誤差のある画像に内在する構造を推論するための指標として「幾何学的 AIC」を提案し [8]，その応用例を示した [12], [15]．幾何学的 AIC は赤池の AIC [1], [2] にヒントを得ているが，同様によく知られている Rissanen の MDL [19], [20] を用いればどうなるのかという疑問が生じる．コンピュータビジョンの分野でも MDL を用いると称する研究はあったが，統計的推測と同じ形の問題にそのまま適用したものや [5], [16]，単に記述の短い解を得るというものが多かった [7], [14], [17], [22]．

そこで筆者らは Rissanen の MDL の対応として「幾何学的 MDL」を定義し，幾何学 AIC との違いを観察したが [10], [11]，議論にはあいまいなところがあった．本論文では幾何学的 AIC と幾何学的 MDL を根本原理に戻って導出し，種々の問題点を明らかにする．そして行列のランクの推定の計算例を示す．

2. 定　　義

2.1 幾何学的当てはめ

N 個の m 次元データ $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$ が与えられ，各 $\boldsymbol{x}_\alpha$ は真の値 $\bar{\boldsymbol{x}}_\alpha$ から期待値 $\boldsymbol{0}$，共分散行列 $V[\boldsymbol{x}_\alpha]$ の独立な正規分布に従う誤差だけずれているとする．真の値 $\bar{\boldsymbol{x}}_\alpha$ は p 次元ベクトル $\boldsymbol{u}$ でパラメータ化された r 個の拘束条件

$$F^{(k)}(\bar{\boldsymbol{x}}_\alpha, \boldsymbol{u}) = 0, \quad k = 1, ..., r \tag{1}$$

を満たすとする．データ $\{\boldsymbol{x}_\alpha\}$ の定義域 $\mathcal{X}$ を「データ空間」，ベクトル $\boldsymbol{u}$ の定義域 $\mathcal{U}$ を「パラメータ空間」，r を拘束条件の「ランク」と呼ぶ．r 個の方程式 $F^{(k)}(\boldsymbol{x}, \boldsymbol{u}) = 0$ は互いに独立で，データ空間 $\mathcal{X}$ に余次元 r の多様体 $\mathcal{S}$ を定義するとする．拘束条件 (1) は真の値 $\{\bar{\boldsymbol{x}}_\alpha\}$ が $\mathcal{S}$ 上にあることを要請している．問題は誤差のあるデータ $\{\boldsymbol{x}_\alpha\}$ からパラメータ $\boldsymbol{u}$ を推定することである．

以上は，データ $\{\boldsymbol{x}_\alpha\}$ が $\mathcal{X}$ のある多様体上に拘束され，パラメータ $\boldsymbol{u}$ も $\mathcal{U}$ のある多様体上に拘束され (例えば $\boldsymbol{x}_\alpha$ も $\boldsymbol{u}$ も単位ベクトル)，式 (1) が互いに独立でない場合に容易に拡張できる [8]．

データの共分散行列 $V[\boldsymbol{x}_\alpha]$ を次のように書く．

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha] \tag{2}$$

定数 ϵ を「ノイズレベル」，$V_0[\boldsymbol{x}_\alpha]$ を「正規化共分散行列」と呼ぶ．このように記述するのが実際的である理由は，多くの場合に誤差の絶対量は未知であるのに対して，その定性的挙動は比較的容易に，例えば画像の濃淡値から測定できるためである [13]．

もう一つの理由は，最ゆう推定において共分散行列の定数倍は解に影響しないので，正規化共分散行列 $V_0[\boldsymbol{x}_\alpha]$ のみを知れば十分だからである．実際，$V_0[\boldsymbol{x}_\alpha]$ に関する「マハラノビス距離」の 2 乗和

$$J = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)) \tag{3}$$

を拘束条件 (1) のもとで最小化すればよい．ただし，

† 岡山大学工学部情報工学科，岡山市
Department of Information Technology, Okayama University, Okayama-shi, 700-8530 Japan

ベクトル $\boldsymbol{a}$, $\boldsymbol{b}$ の内積を $(\boldsymbol{a},\boldsymbol{b})$ と記す．

誤差が小さいと仮定して式 (1) を線形化し，ラグランジュ乗数を導入して拘束条件を消去すると，上式は次のようになる [8]．

$$J=\sum_{\alpha=1}^{N}\sum_{k,l=1}^{r}W_{\alpha}^{(kl)}F^{(k)}(\boldsymbol{x}_{\alpha},\boldsymbol{u})F^{(l)}(\boldsymbol{x}_{\alpha},\boldsymbol{u}) \quad (4)$$

$W_{\alpha}^{(kl)}$ は $(\nabla_{\mathbf{x}}F_{\alpha}^{(k)},V_0[\boldsymbol{x}_{\alpha}]\nabla_{\mathbf{x}}F_{\alpha}^{(l)})$ を (kl) 要素とする $r\times r$ 行列の逆行列の (kl) 要素である（添え字 α は $\boldsymbol{x}=\boldsymbol{x}_{\alpha}$ を代入することを表す）．

2.2 漸近変数

式 (4) を最小化する解 $\hat{\boldsymbol{u}}$ の共分散行列は $\epsilon\to 0$ で $\boldsymbol{O}$ に収束するだけでなく，$O(\epsilon^4)$ の項を除いて精度の理論限界を達成する [8]．これは統計的推測において最ゆう推定量の共分散行列が観測数 $n\to\infty$ で $\boldsymbol{O}$ に収束する（「一致性」）のみならず，$O(1/n^2)$ の項を除いてクラメル・ラオの下界を達成すること（「漸近有効性」）に対応している．

一般に厳密な解析が困難な問題でも，ある変数が十分大きい，あるいは小さい場合に簡単な形が得られ，数学的構造が明らかになることが多い．そのような変数を仮に「漸近変数」と呼ぶ．

統計的推測では観測数 n が漸近変数にとられる．これはランダム誤差を繰返し観測によって克服するという思想に基づいている．そのため $n\to\infty$ で性能が急速に増大する推定方法が望まれる．なぜなら，そのような方法は同じ性能を達成するのに必要なデータ数が他の方法より少なくてすむからである．

これに対して幾何学的当てはめではノイズレベル ϵ が漸近変数にとられる．これは最小限の解像度で最大限の精度を得るという思想に基づいている [8]．そのため $\epsilon\to 0$ で性能が急速に増大する推定方法が望まれる．なぜなら，そのような方法は同じ性能を達成するのに必要な解像度が他の方法より低くてよいからである．

一般に統計的推測の $n\to\infty$ の性質が幾何学的当てはめでは $\epsilon\to 0$ で成立する [8], [9]．これは一つの画像を仮想的に繰返して観測する度に独立な誤差が入ると考え，その平均をデータとみなすと，仮想的な観測数を増やすことと誤差が減少することが等価だからである [9]．

2.3 モデルとモデル選択

統計的推測はランダム現象の解明を目的とするから，データ $\boldsymbol{x}$ はパラメータを用いて確定的な式とランダム誤差によって表現される．抽象化すると，パラメータ $\boldsymbol{\theta}$ をもつ確率密度 $P(\boldsymbol{x}|\boldsymbol{\theta})$ から発生したデータ列 $\{\boldsymbol{x}_i\}$ より $\boldsymbol{\theta}$ を推定することであり，$\boldsymbol{\theta}$ は確定的な要因を記述する係数と期待値や分散のようなランダム誤差の特性から成っている．この確率密度に複数の可能性 $P_1(\boldsymbol{x}|\boldsymbol{\theta}_1)$, $P_2(\boldsymbol{x}|\boldsymbol{\theta}_2)$, ... がある場合，各々は「（確率）モデル」と呼ばれ，どれが妥当かを判定するのが「（確率）モデル選択」である．

これに対して，データ $\boldsymbol{x}$ の満たすべき拘束条件 $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{u})=0$ のパラメータ $\boldsymbol{u}$ を推定するのが幾何学的当てはめであり，$\boldsymbol{u}$ はランダム誤差の特性を含んでいない．その拘束条件に複数の可能性 $\boldsymbol{F}_1(\boldsymbol{x},\boldsymbol{u}_1)=\boldsymbol{0}$, $\boldsymbol{F}_2(\boldsymbol{x},\boldsymbol{u}_2)=\boldsymbol{0}$, ... がある場合，各々は「（幾何学的）モデル」と呼ばれ，どれが妥当かを判定するのが「（幾何学的）モデル選択」である [8]．

3. 幾何学的 AIC

赤池の AIC [1], [2] の導出は幾何学的当てはめでは次のように対応する．

3.1 モデルのよさ

モデル (1) のもとでは $\{\boldsymbol{x}_{\alpha}\}$ は次の確率密度からの“一つ”のサンプルである（以下，確率変数を大文字で，その実現値を小文字で区別する．$|\cdot|$ は行列式を表す）．

$$P=\prod_{\alpha=1}^{N}\frac{e^{-(\boldsymbol{X}_{\alpha}-\bar{\boldsymbol{x}}_{\alpha},V[\boldsymbol{x}_{\alpha}]^{-1}(\boldsymbol{X}_{\alpha}-\bar{\boldsymbol{x}}_{\alpha}))/2}}{\sqrt{(2\pi)^m|V[\boldsymbol{x}_{\alpha}]|}} \quad (5)$$

ただし $\{\bar{\boldsymbol{x}}_{\alpha}\}$ は式 (1) で拘束されている．赤池の用いたモデルのよさの尺度はこの確率密度から真の密度 P_T までの「カルバック情報量」であり [1], [2]，次のように書ける．

$$\begin{aligned}D &= \int\cdots\int P_T\log\frac{P_T}{P}d\boldsymbol{X}_1\cdots d\boldsymbol{X}_N\\ &= E[\log P_T]-E[\log P] \quad (6)\end{aligned}$$

$E[\,\cdot\,]$ は真の確率密度 P_T に関する期待値である．D が小さいほどよいモデルとみなせるが，最後の辺の第 1 項は個々のモデルによらないので，$-E[\log P]$ が小さいほどよい．式 (2) を用いて書き直すと次のようなる．

$$\begin{aligned}&-E[\log P]\\ &= \frac{1}{2\epsilon^2}E\left[\sum_{\alpha=1}^{N}(\boldsymbol{X}_{\alpha}-\bar{\boldsymbol{x}}_{\alpha},V_0[\boldsymbol{x}_{\alpha}]^{-1}(\boldsymbol{X}_{\alpha}-\bar{\boldsymbol{x}}_{\alpha}))\right]\end{aligned}$$

$$+\frac{mN}{2}\log 2\pi\epsilon^2+\frac{1}{2}\sum_{\alpha=1}^{N}\log|V_0[\boldsymbol{x}_\alpha]| \qquad (7)$$

最後の 2 項は個々のモデルによらないので，第 1 項に $2\epsilon^2$ を掛けた次の「期待残差」が小さいほどよい (ϵ はモデルパラメータではないから，ϵ のみに依る正数を掛けてもモデル選択には影響しない) .

$$E=E\left[\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha-\bar{\boldsymbol{x}}_\alpha,V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{X}_\alpha-\bar{\boldsymbol{x}}_\alpha))\right] \quad (8)$$

3.2 期待値の評価

式 (8) 中の $E[\cdot]$ をどう評価するかが統計的推測と幾何学的当てはめで異なる .

統計的推測ではデータは (原理的には) いくらでも多く観測できる状況を想定するので，密度 $P_T(\boldsymbol{X})$ から独立な観測値 $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_n$ が得られれば，統計量 $Y(\boldsymbol{X})$ の期待値 $\int Y(\boldsymbol{X})P_T(\boldsymbol{X})d\boldsymbol{X}$ はサンプル平均 $(1/n)\sum_{i=1}^{n}Y(\boldsymbol{x}_i)$ で近似でき，$n\to\infty$ で真の値に収束する (「大数の法則」) . 赤池の AIC もこれに基づいている [1], [2] .

幾何学的当てはめでは $\{\boldsymbol{x}_\alpha\}$ は確率変数 $\{\boldsymbol{X}_\alpha\}$ の "1 回" の観測値なので，期待値をサンプル平均で置き換えることができない . しかし，幾何学的当てはめでは (原理的には) いくらでも高い解像度の装置が利用できる状況を想定するので，$\epsilon\to 0$ で一致する近似を用いればよい . 明らかに統計量 $Y(\{\boldsymbol{X}_\alpha\})$ の期待値 $\int\cdots\int Y(\{\boldsymbol{X}_\alpha\})P_T d\boldsymbol{X}_1\cdots d\boldsymbol{X}_N$ は $Y(\{\boldsymbol{x}_\alpha\})$ で近似できる . なぜなら $\epsilon\to 0$ で $P_T\to\prod_{\alpha=1}^{N}\delta(\boldsymbol{X}_\alpha-\bar{\boldsymbol{x}}_\alpha)$ であり ($\delta(\cdot)$ はデルタ関数)，$\int\cdots\int Y(\{\boldsymbol{X}_\alpha\})P_T d\boldsymbol{X}_1\cdots d\boldsymbol{X}_N$ と $Y(\{\boldsymbol{x}_\alpha\})$ は共に $Y(\{\bar{\boldsymbol{x}}_\alpha\})$ に収束するからである . したがって式 (8) は次式で近似できる .

$$J=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha,V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha)) \qquad (9)$$

3.3 偏差の除去

式 (9) には未知数 $\{\bar{\boldsymbol{x}}_\alpha\}$, $\boldsymbol{u}$ が含まれている . 式 (9) が小さいほどよいモデルであるという観点から，拘束条件 (1) のもとで式 (9) を最小にする「最ゆう推定量」$\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ を仮定するのが当然である . 素朴な考えはこれら最ゆう推定量を式 (9) に代入した「残差 (平方和)」

$$\hat{J}=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{x}}_\alpha,V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{x}}_\alpha)) \qquad (10)$$

を用いることである . しかし，これは論理的矛盾である .

式 (1) は $\{\bar{\boldsymbol{x}}_\alpha\}$, $\boldsymbol{u}$ をパラメータとするモデルのクラスを定義し，特定の値 $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ を代入して特定のモデルが得られる . **3.1** の論理からはそのよさは $E[\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha-\hat{\boldsymbol{x}}_\alpha,V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{X}_\alpha-\hat{\boldsymbol{x}}_\alpha))]$ で測るべきであり，**3.2** の論理からはその期待値は $\{\boldsymbol{X}_\alpha\}$ の代表的な実現値 $\{\boldsymbol{x}_\alpha\}$ を代入して近似できる . しかし，$\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ はその $\{\boldsymbol{x}_\alpha\}$ から計算しているので，$\{\boldsymbol{x}_\alpha\}$ は仮定したモデルと相関があり，$\{\boldsymbol{X}_\alpha\}$ の代表的な実現値とはみなせない .

実際 $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ は $\hat{J}$ を最小にするように定めているので，$\hat{J}$ は一般に $E[\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha-\hat{\boldsymbol{x}}_\alpha,V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{X}_\alpha-\hat{\boldsymbol{x}}_\alpha))]$ より小さい . この困難に対する赤池の解決法 [1], [2] をこの場合に翻訳すると次のようになる .

3.2 の論理から，仮定したモデルと無関係の実現値 $\{\boldsymbol{x}^*_\alpha\}$ (「将来のデータ」) を用いて

$$\hat{J}^*=\sum_{\alpha=1}^{N}(\boldsymbol{x}^*_\alpha-\hat{\boldsymbol{x}}_\alpha,V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}^*_\alpha-\hat{\boldsymbol{x}}_\alpha)) \qquad (11)$$

を評価する . しかし現実には $\{\boldsymbol{x}_\alpha\}$ しか得られていないから(注1)，次のように偏差を補正する .

$$\hat{J}^*=\hat{J}+b\epsilon^2 \qquad (12)$$

$\hat{J}^*$ も $\hat{J}$ も確率変数であるから b も確率変数であり，その期待値は次のように評価される [8] .

$$E^*[E[b]]=2(Nd+p)+O(\epsilon^2) \qquad (13)$$

$E[\cdot]$, $E^*[\cdot]$ はそれぞれ $\{\boldsymbol{x}_\alpha\}$, $\{\boldsymbol{x}^*_\alpha\}$ に関する期待値である . 上式より $\hat{J}^*$ の不偏推定量が第 1 近似により次のように得られる .

$$\text{G-AIC}=\hat{J}+2(Nd+p)\epsilon^2 \qquad (14)$$

ただし $d=m-r$ は拘束条件 $F^{(k)}(\boldsymbol{x},\boldsymbol{u})=0$, $k=1$, ..., r がデータ空間 $\mathcal{X}$ に定義する多様体 $\mathcal{S}$ の次元である . 式 (14) が「幾何学的 AIC」[8] である .

3.4 漸近解析と摂動解析

赤池の AIC の導出は次の性質を利用している [1], [2] .

- 最ゆう推定量は $n\to\infty$ で真の値に収束する (「大数の法則」) .

(注1): そのようなデータ $\{\boldsymbol{x}^*_\alpha\}$ が存在する場合は「クロス・バリデーション」，そのようなデータを計算機で発生させる場合は「ブートストラップ」[4] と呼ばれる .

● 最ゆう推定量は $n \to \infty$ で漸近的に正規分布に従う(「中心極限定理」).

● 標準化した正規分布に従う変数の2次形式は χ^2 分布に従い,その期待値はその自由度に等しい.

一方,式 (13) の証明は次の性質を利用している [8].

● 最ゆう推定量は $\epsilon \to 0$ で真の値に収束する.

● 誤差は正規分布と仮定しているから,線形な拘束条件のもとで最ゆう推定量は正規分布に従う.拘束条件が非線形でも ϵ が小さい極限では解の近傍での線形近似が正当化される.

● 標準化した正規分布に従う変数の2次形式は χ^2 分布に従い,その期待値はその自由度に等しい.

4. 幾何学的 MDL

Rissanen の MDL [19] ~ [21] の導出は幾何学的当てはめでは次のように対応する.

4.1 MDL 原理

Rissanen の MDL は情報理論的な符号長をモデルのよさの尺度とするものであるが,次の問題が生じる.

● 実数を扱う問題を符号化するのに無限大の符号長を要する.

● 最短符号化するための確率分布に未知数が含まれる.

● 最短符号長の厳密な評価が困難である.

Rissanen [19], [20] は,個々の(確率)モデルに依らない形で量子化して有限長で記述し,未知数の代わりに最ゆう推定量を用いた.これも量子化し,その量子化幅を全体の記述長が最短となるように選んだ(「2段階符号化」).そしてデータ長 n を漸近変数にとる漸近評価を行った.この考えをノイズレベル ϵ を漸近変数にとる幾何学的当てはめに適用する.

データ $\{\boldsymbol{x}_\alpha\}$ が確率密度 (5) から得られるなら,その定義域を量子化し,定義域と量子化幅にのみ依る定数を除いて次の長さに最短語頭符号化できる [6], [24].

$$-\log P = \frac{J}{2\epsilon^2} + \frac{mN}{2}\log 2\pi\epsilon^2 + \frac{1}{2}\sum_{\alpha=1}^{N}\log|V_0[\boldsymbol{x}_\alpha]| \qquad (15)$$

ただし log は自然対数であり,符号長を測るのに $\log_2 e$ ビットを1とする単位を用いる.式中の J は式 (3) の2乗マハラノビス距離である.

4.2 2段階符号化

式 (5) によって符号化するには式中に含まれる真の値 $\{\bar{\boldsymbol{x}}_\alpha\}$ とパラメータ $\boldsymbol{u}$ が必要である.これらに式 (15) (中の J) を最小化する最ゆう推定量 $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ を用いる.式 (15) の最後の2項は個々の(幾何学的)モデルに依らないから無視し,$\{\bar{\boldsymbol{x}}_\alpha\}$ に $\{\hat{\boldsymbol{x}}_\alpha\}$ を代入すると記述長は $\hat{J}/2\epsilon^2$ となる.ただし $\hat{J}$ は式 (10) で与えられる残差である.以下,符号長からモデルに依らない定数を無視したものを "記述長" と呼んで区別する.

最ゆう推定量 $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ も実数であるから量子化しなければならない.しかし量子化した近似値を用いると記述長 $\hat{J}/2\epsilon^2$ が増大する.そこで全体の記述長が最短になるように量子化する.そのための基礎は,式 (4) が次のように書けることである [8].

$$J = \hat{J} + \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\hat{\boldsymbol{x}}_\alpha]^-(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha)) + (\boldsymbol{u} - \hat{\boldsymbol{u}}, V_0[\hat{\boldsymbol{u}}]^{-1}(\boldsymbol{u} - \hat{\boldsymbol{u}})) + O(\epsilon^3) \qquad (16)$$

上添え字 $-$ は(ムーア・ペンローズの)一般逆行列を表し,$V_0[\hat{\boldsymbol{x}}_\alpha]$, $V_0[\hat{\boldsymbol{u}}_\alpha]$ は $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ の(事後)共分散行列であり,次のように与えられる [8].

$$V_0[\hat{\boldsymbol{x}}_\alpha] = V_0[\boldsymbol{x}_\alpha] - \sum_{k,l=1}^{r} W_\alpha^{(kl)}(V[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}}F_\alpha^{(k)})(V[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}}F_\alpha^{(k)})^\top$$

$$V_0[\hat{\boldsymbol{u}}] = \Big(\sum_{\alpha=1}^{N}\sum_{k,l=1}^{r} W_\alpha^{(kl)}(\nabla_{\mathbf{u}}F_\alpha^{(k)})(\nabla_{\mathbf{u}}F_\alpha^{(l)})^\top\Big)^{-1} \qquad (17)$$

式中の $W_\alpha^{(kl)}$ は式 (4) 中のものと同じである.

4.3 パラメータの符号化

p 次元パラメータ空間 $\mathcal{U}$ に適当な(一般には曲線)座標系 (u_i) を導入し,各座標を幅 δu_i で刻んでその格子点を指定する.$\hat{\boldsymbol{u}}$ が第 i 座標の幅が L_i の領域の内部にあるとすれば,格子の頂点数は $\prod_{i=1}^{p}(L_i/\delta u_i)$ 個あり,その一つを指定する符号長は次のようになる.

$$\log\prod_{i=1}^{p}\frac{L_i}{\delta u_i} = \log V_u - \sum_{i=1}^{p}\log\delta u_i \qquad (18)$$

ただし $V_u = \prod_{i=1}^{p} L_i$ は $\hat{\boldsymbol{u}}$ の存在する領域の体積である.上式を小さくするには刻み幅 δu_i を大きくとればよいが,$\hat{\boldsymbol{u}}$ を格子点で置き換えるために記述長 $\hat{J}/2\epsilon^2$ が増大する.その増加量は,ベクトル $\delta\boldsymbol{u} = (\delta u_i)$ を定義すると式 (16) から ϵ に関する第1近似のもとで

$(\delta\boldsymbol{u}, V_0[\hat{\boldsymbol{u}}]^{-1}\delta\boldsymbol{u})/2\epsilon^2$ である．これと式 (18) との和を δu_i で微分して 0 と置くと次式を得る．

$$\frac{1}{\epsilon^2}\Big(V_0[\hat{\boldsymbol{u}}]^{-1}\delta\boldsymbol{u}\Big)_i = \frac{1}{\delta u_i} \tag{19}$$

ただし $(\,\cdot\,)_i$ は第 i 成分を表す．パラメータ空間 $\mathcal{U}$ の座標系を $V_0[\hat{\boldsymbol{u}}]^{-1}$ が対角行列になるようにとれば次の解を得る．

$$\delta u_i = \frac{\epsilon}{\sqrt{\lambda_i}} \tag{20}$$

λ_i は $V_0[\hat{\boldsymbol{u}}]^{-1}$ の第 i 固有値である．上式より格子単位の体積が次のように書ける．

$$v_u = \prod_{i=1}^{p}\delta u_i = \frac{\epsilon^p}{\sqrt{|V_0[\hat{\boldsymbol{u}}]^{-1}|}}. \tag{21}$$

これから領域 V_u 中の格子単位の総数が次のように見積もれる．

$$N_u = \int_{V_u}\frac{d\boldsymbol{u}}{v_u} = \frac{1}{\epsilon^p}\int_{V_u}\sqrt{|V_0[\hat{\boldsymbol{u}}]^{-1}|}d\boldsymbol{u} \tag{22}$$

この内の一つを指定する符号長は次のようになる．

$$\log N_u = \log\int_{V_u}\sqrt{|V_0[\hat{\boldsymbol{u}}]^{-1}|}d\boldsymbol{u} - \frac{p}{2}\log\epsilon^2 \tag{23}$$

4.4 真の値の符号化

データ $\{\boldsymbol{x}_\alpha\}$ の定義域は m 次元データ空間 $\mathcal{X}$ であるが，$\{\hat{\boldsymbol{x}}_\alpha\}$ はその中の (前節で符号化した) $\hat{\boldsymbol{u}}$ の指定する d 次元多様体 $\hat{\mathcal{S}}$ に拘束されている．各 $\hat{\boldsymbol{x}}_\alpha$ は式 (17) のように別々の正規化共分散行列 $V_0[\hat{\boldsymbol{x}}_\alpha]$ をもつので，各 $\hat{\boldsymbol{x}}_\alpha$ に別々の曲線座標系 $(\xi_{i\alpha})$ を用い，各座標を幅 $\delta\xi_{i\alpha}$ で刻んでその格子点を指定する．

$\{\hat{\boldsymbol{x}}_\alpha\}$ が第 i 座標の幅が $l_{i\alpha}$ の領域の内部にあるとすれば，格子の頂点数は $\prod_{i=1}^{d}(l_{i\alpha}/\delta\xi_{i\alpha})$ 個あり，その一つを指定する符号長は次のようになる．

$$\sum_{i=1}^{d}\log\frac{l_{i\alpha}}{\delta\xi_{i\alpha}} = \log V_{x\alpha} - \sum_{i=1}^{d}\log\delta\xi_{i\alpha} \tag{24}$$

ただし $V_{x\alpha} = \prod_{i=1}^{d} l_{i\alpha}$ はその領域の体積である．上式を小さくするには刻み幅 $\delta\xi_{i\alpha}$ を大きくとればよいが，$\hat{\boldsymbol{x}}_\alpha$ を格子点で置き換えるために記述長 $\hat{J}/2\epsilon^2$ が増大する．$\delta\xi_{i\alpha}$, $i = 1, ..., p$ で指定される $\hat{\mathcal{S}}$ 上の変位をデータ空間 $\mathcal{X}$ から見た m 次元ベクトルを $\delta\bar{\boldsymbol{x}}_\alpha$ とすると，記述長の増加量は式 (16) から ϵ の第 1 近似において $(\delta\boldsymbol{x}_\alpha, V_0[\hat{\boldsymbol{x}}_\alpha]^-\delta\boldsymbol{x}_\alpha)/2\epsilon^2$ である．これと式 (24) との和を $\delta\xi_{i\alpha}$ で微分して 0 と置くと次式を得る．

$$\frac{1}{\epsilon^2}\Big(V_0[\hat{\boldsymbol{x}}_\alpha]^-\delta\bar{\boldsymbol{x}}_\alpha\Big)_i = \frac{1}{\delta\xi_{i\alpha}} \tag{25}$$

$V_0[\hat{\boldsymbol{x}}_\alpha]^-$ は多様体 $\hat{\mathcal{S}}$ の $\hat{\boldsymbol{x}}_\alpha$ における接空間でのみ値をもつランク d の半正値対称行列である [8]．$\hat{\mathcal{S}}$ の曲線座標系をその基底が $\hat{\boldsymbol{x}}_\alpha$ において d 本の正規直交系を成すようにとり，それに直交する $m-d$ 本の正規直交系をとってデータ空間 $\mathcal{X}$ の正規直交系とする．更に $\hat{\mathcal{S}}$ 内の d 本の正規直交系を $V_0[\hat{\boldsymbol{x}}_\alpha]^-$ が対角行列になるようにとれば，式 (25) は次の解をもつ．

$$\delta\xi_{i\alpha} = \begin{cases} \epsilon/\sqrt{\lambda_{i\alpha}} & i = 1, ..., d \\ 0 & i = d+1, ..., m \end{cases} \tag{26}$$

$\lambda_{1\alpha}$, ..., $\lambda_{d\alpha}$ は $V_0[\hat{\boldsymbol{x}}_\alpha]^-$ の d 個の正の固有値である．これから格子単位の体積が次のように書ける．

$$v_{x\alpha} = \prod_{i=1}^{d}\delta\xi_{i\alpha} = \frac{\epsilon^d}{\sqrt{|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d}} \tag{27}$$

$|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d$ は $V_0[\hat{\boldsymbol{x}}_\alpha]^-$ の正の固有値の積である．上式から領域 V_x 中の格子単位の総数が次のように見積もれる．

$$N_\alpha = \int_{V_{x\alpha}}\frac{d\boldsymbol{x}}{v_{x\alpha}} = \frac{1}{\epsilon^d}\int_{V_{x\alpha}}\sqrt{|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d}d\boldsymbol{x} \tag{28}$$

この内の一つを指定する符号長は次のようになる．

$$\log N_\alpha = \log\int_{V_{x\alpha}}\sqrt{|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d}d\boldsymbol{x} - \frac{d}{2}\log\epsilon^2 \tag{29}$$

4.5 幾何学的 MDL

式 (23), (29) から $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ の符号長の合計が次のようになる．

$$\sum_{\alpha=1}^{N}\log\int_{V_{x\alpha}}\sqrt{|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d}d\boldsymbol{x} + \log\int_{V_u}\sqrt{|V_0[\hat{\boldsymbol{u}}]^{-1}|}d\boldsymbol{u} - \frac{Nd+p}{2}\log\epsilon^2 \tag{30}$$

一方，量子化による記述長 $\hat{J}/2\epsilon^2$ の増大量は ϵ の第 1 近似において $(\delta\boldsymbol{x}_\alpha, V_0[\hat{\boldsymbol{x}}_\alpha]^-\delta\boldsymbol{x}_\alpha)/2\epsilon^2 + (\delta\boldsymbol{u}, V_0[\hat{\boldsymbol{u}}]^{-1}\delta\boldsymbol{u})/2\epsilon^2$ であり，式 (20), (26) と $V_0[\hat{\boldsymbol{x}}_\alpha]^- = \mathrm{diag}(1/\lambda_{1\alpha}, ..., 1/\lambda_{d\alpha}, 0, ..., 0)$, $V_0[\hat{\boldsymbol{u}}]^{-1}$

$= \mathrm{diag}(1/\lambda_1, ..., 1/\lambda_p)$ を代入すると，ϵ の第 1 近似において次のようになる．

$$\frac{(\delta\bar{\boldsymbol{x}}_\alpha, V_0[\hat{\boldsymbol{x}}_\alpha]^- \delta\bar{\boldsymbol{x}}_\alpha)}{2\epsilon^2} + \frac{(\delta\boldsymbol{u}, V_0[\hat{\boldsymbol{u}}]^{-1}\delta\boldsymbol{u})}{2\epsilon^2} = \frac{Nd+p}{2} \quad (31)$$

式 (20), (26) の量子化は $O(\epsilon)$ であるから上式の省略項は $o(1)$ であり，全体の記述長は次のようになる．

$$\frac{\hat{J}}{2\epsilon^2} - \frac{Nd+p}{2}\log\epsilon^2 + \sum_{\alpha=1}^{N}\log\int_{V_{x\alpha}}\sqrt{|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d}d\boldsymbol{x} + \log\int_{V_u}\sqrt{|V_0[\hat{\boldsymbol{u}}]^{-1}|}d\boldsymbol{u} + \frac{Nd+p}{2} + o(1) \quad (32)$$

ϵ はモデルパラメータではないから ϵ のみに依る正数を掛けてもモデル選択には影響しない．全体を $2\epsilon^2$ 倍し，

$$\begin{aligned}\text{G-MDL} \;=\; & \hat{J} - (Nd+p)\epsilon^2\log\epsilon^2 \\ & +2\epsilon^2\Big(\sum_{\alpha=1}^{N}\log\int_{V_{x\alpha}}\sqrt{|V_0[\hat{\boldsymbol{x}}_\alpha]^-|_d}d\boldsymbol{x} \\ & +\log\int_{V_u}\sqrt{|V_0[\hat{\boldsymbol{u}}]^{-1}|}d\boldsymbol{u}\Big) \\ & +(Nd+p)\epsilon^2 + o(\epsilon^2) \end{aligned} \quad (33)$$

を「幾何学的 MDL」と呼ぶ．

4.6 スケールの問題

式 (33) の右辺第 3 項は評価が難しい．これは式 (17) の $V_0[\hat{\boldsymbol{x}}_\alpha]$, $V_0[\hat{\boldsymbol{u}}]$ が複雑で積分が困難であるためであるが，より根本的な問題は積分できるためには領域 V_x, V_u が有界でなければならないことである．このため，データ空間 $\mathcal{X}$ やパラメータ空間 $\mathcal{U}$ が無限領域のときは，その中にデータや解がありそうな有界領域を指定しなければならない．これはパラメータに事前分布を与える「ベイズの立場」にほかならない．

そもそもモデル選択を符号長に帰着させるという思想がベイズの立場を要求している．なぜなら値が無限領域中のどこにあってもよいなら量子化しても有限長で符号化することは不可能だからである．これを避ける便法は，高次の微小量を省略したり，モデルによらない発散量を無視したりすることである．$\epsilon \to 0$ で $-\log\epsilon^2 \gg 1$ であるから式 (33) の $O(\epsilon^2)$ の項を無視すると次のようになる．

$$\text{G-MDL} = \hat{J} - (Nd+p)\epsilon^2\log\epsilon^2 \quad (34)$$

これは筆者らが最初に提案した形である [11]．この形では積分の問題が生じないが，代わりにスケールの問題が生じる．データを測る単位を例えば 10 倍すると ϵ^2 も $\hat{J}$ も 1/100 倍され，N, d, p は無次元量であるから G-MDL も 1/100 倍されるべきであるが，$\log\epsilon^2$ は $\log\epsilon^2 - \log 100$ になる．それに対して式 (33) では右辺第 2, 3 項の変化が打ち消されて不変になる．

そもそも log は無次元量にしか定義できないので，式 (34) は本来は次のように表されなければならない．

$$\text{G-MDL} = \hat{J} - (Nd+p)\epsilon^2\log\Big(\frac{\epsilon}{L}\Big)^2 \quad (35)$$

ここに L はある基準長であり，厳密には式 (33) を変形して第 3 項から定めることができるが，第 3 項の評価が困難である．そこで妥協として L を $\boldsymbol{x}_\alpha/L$ が $O(1)$ となるように選ぶ．これはデータ空間 $\mathcal{X}$ 中の体積 L^m の領域に事前分布を与えることと解釈できる．例えば $\{\boldsymbol{x}_\alpha\}$ が画素データの場合は L を画像サイズにとればよい．

誤差はデータに比べて十分小さいと仮定するから $-\log(\epsilon/L)^2 \gg 1$ である．したがって，L を $L' = \gamma L$ としても $-\log(\epsilon/L')^2 = -\log(\epsilon/L)^2 + \log\gamma^2$ において $\gamma \approx 1$ なら $\log\gamma^2 \approx 0$ であり，L はオーダーが同じであればモデル選択に影響しない．

4.7 統計的推測における MDL

前節の内容は恣意的に見えるが，データ長 n を漸近変数にとる Rissanen の MDL でも同様な問題が生じる．Rissanen の MDL は当初は次の形とされた [19]．

$$\text{MDL} = -\log\prod_{i=1}^{n}P(\boldsymbol{x}_i|\hat{\boldsymbol{\theta}}) + \frac{k}{2}\log n + O(1) \quad (36)$$

各データ $\boldsymbol{x}_i$ はパラメータ $\boldsymbol{\theta}$ をもつ確率モデル $P(\boldsymbol{x}|\boldsymbol{\theta})$ から独立に生成されるとし，$\hat{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ の最ゆう推定量である．k は $\boldsymbol{\theta}$ の次元であり，$O(1)$ は $n \to \infty$ の評価である．式 (34) は上式に対応させたものである．

しかしこの形ではデータの単位の問題が生じる．例えばデータ $\{\boldsymbol{x}_i\}$ を二つずつまとめて，$\{(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{x}_3, \boldsymbol{x}_4), ...\}$ が確率密度 $P(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) = P(\boldsymbol{x}|\boldsymbol{\theta})P(\boldsymbol{y}|\boldsymbol{\theta})$ から生成されるとすれば，問題としては同一でもデータ長が半分になり，式 (36) の右辺第 2 項が $(k/2)\log 2$ だけ小さくなる．しかし，その後 Rissanen の MDL は次の形とされる [21]．

$$\text{MDL} \;=\; -\log\prod_{i=1}^{n}P(\boldsymbol{x}_i|\hat{\boldsymbol{\theta}}) + \frac{k}{2}\log\frac{n}{2\pi}$$

$$+\log\int_{V_\theta}\sqrt{|\boldsymbol{I}(\theta)|}d\boldsymbol{\theta}+o(1) \qquad (37)$$

$\boldsymbol{I}(\theta)$ は $P(\boldsymbol{x}|\boldsymbol{\theta})$ フィッシャー情報行列である．この形ならデータの単位を変えてもフィッシャー情報行列が変化し，変化が打ち消される．しかし，パラメータの領域 V_θ が無限の場合は式 (33) と同じ問題が生じ，事前分布を仮定する等の処置が必要となる．

このように，幾何学的 MDL と Rissanen の MDL では漸近変数がデータ長 n かノイズレベル ϵ (仮想的な観測数に対応) かの違いがあっても，基本的に同じ思想であるから，一方の性質が他方の性質に対応する．

5. ノイズレベルの推定

幾何学的 AIC でも幾何学的 MDL でもその評価にノイズレベル ϵ が必要となる．これが未知の場合は推定する必要がある．ϵ は画像の解釈とは無関係に画像や処理アルゴリズムから定まる特性であるから，個々のモデルとは独立に推定しなければならない．

正しいモデルが既知なら，これは残差 $\hat{J}$ から推定できる．$\hat{J}/\epsilon^2$ は式 $F^{(k)}(\boldsymbol{x},\hat{\boldsymbol{u}})=0,\ k=1,...,r$ が定義するデータ空間 $\mathcal{X}$ 中の多様体 $\hat{\mathcal{S}}$ から各データ点までの (マハラノビス) 距離の 2 乗和である．$\hat{\mathcal{S}}$ の余次元は r であるから，$\hat{J}/\epsilon^2$ の期待値は rN となるべきであるが，$\hat{\mathcal{S}}$ をデータに当てはめているので，その自由度だけ減って $rN-p$ となる [8]．これから ϵ^2 の不偏推定量が次のように得られる．

$$\hat{\epsilon}^2=\frac{\hat{J}}{rN-p} \qquad (38)$$

正しいモデルが既知ならモデル選択が必要かという疑問が生じるが，モデル選択が有効であるのは多くの場合，“退化” の検出である．拘束条件 (1) はシーンに関する知識 (物体は剛体運動をする，等) に対応するが，それが例外的に退化した場合 (例えば運動が 0 である，等) ではそれを前提にした計算が破たんする．厳密な退化でなくても退化に近いと計算が不安定になる．このような場合，退化を自動的に検出して退化を記述するモデルに切り換えれば計算が安定化される [12], [15]．

退化とは拘束条件 (1) に新たな式 (ある量が 0 である，等) が加わることを意味する．しかし拘束条件 (1) そのものは成立している．そのような「一般モデル」は「退化モデル」が成立しようがしまいが成立するから，ϵ^2 は一般モデルの残差 $\hat{J}$ から式 (38) によって推定すればよい．

一方，統計的推定では誤差とは仮定したデータ生成機構と実際の観測データの食い違いを記述するものであるから，その分散はモデルパラメータであり，個々のモデルに基づいて推定する必要がある．

6. ランク推定への応用例

複数の物体が独立に移動する動画像中の物体数は，誤差がなければ画像データから計算したある行列のランクから計算できる [10]．誤差があるときは特異値分解し，微小な特異値を 0 とみなしてランクを推定すればよいが，しきい値の設定が困難である．

$n\times m$ 行列のランク r は m 本の列ベクトルの張る部分空間の次元であるから，これに幾何学的モデル選択が適用できる．n 次元空間の r 次元部分空間の自由度は $r(n-r)$ であるから，各要素に期待値 0，標準偏差 ϵ の正規分布に従う誤差が独立に入るとすると，幾何学的 AIC，幾何学的 MDL はそれぞれ次のようになる．

$$\begin{aligned}\text{G-AIC} &= \hat{J}_r+2r(m+n-r)\epsilon^2\\ \text{G-MDL} &= \hat{J}_r-r(m+n-r)\epsilon^2\log\Big(\frac{\epsilon}{L}\Big)^2\end{aligned} \qquad (39)$$

残差 $\hat{J}_r$ は次のように表せる．

$$\hat{J}_r=\sum_{i=r+1}^{\nu}\lambda_i^2 \qquad (40)$$

$\{\lambda_i\}$ は降順に並べた特異値である ($\nu=\min(n,m)$)．$r=1,\ 2,\ ...$ に対して式 (39) が最小になる値をランク r とすればよい．

ノイズレベル ϵ が未知の場合は，ランク r がある値 $r_{\max}$ 以下であることが既知であれば，式 (38) により次のように推定できる．

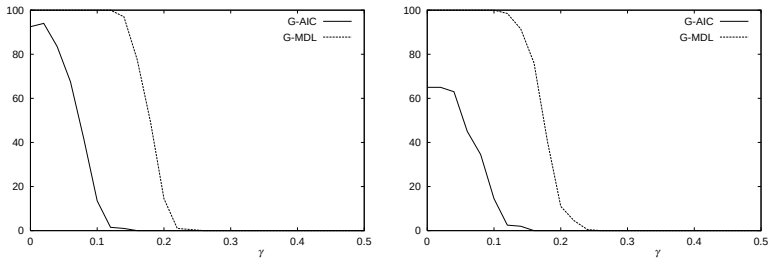
$$\hat{\epsilon}^2=\frac{\hat{J}_{r_{\max}}}{(n-r_{\max})(N-r_{\max})} \qquad (41)$$

区間 $[-1,1]$ 上の一様乱数を独立に 200 個発生させ，これを要素とする 10×20 行列を $\boldsymbol{V}\,\mathrm{diag}(\lambda_1,...,\lambda_{10})\boldsymbol{U}^\top$ と特異値分解し，次の行列 $\boldsymbol{A}$ を定義した．

$$\boldsymbol{A}=\boldsymbol{V}\,\mathrm{diag}(\lambda_1,...,\lambda_5,\gamma\lambda_5,0,...,0)\boldsymbol{U}^\top \qquad (42)$$

$\lambda_1,\ ...,\ \lambda_5$ は順に 3.81, 3.58, 3.09, 2.98, 2.75 である．

図 1 では横軸に γ をとり，$\boldsymbol{A}$ の各要素に期待値 0，



(a) (b)

図 1 幾何学的 AIC(実線) と幾何学的 MDL(破線) によるランクの推定の正解率 (%)．(a) 既知のノイズレベル，(b) 推定したノイズレベル

Fig. 1 The ratio (%) of correct estimation of matrix rank by the geometric AIC (solid line) and the geometric MDL (dashed line) using (a) the true noise level and (b) the estimated noise level.

標準偏差 0.05 の誤差を独立に加え，$r_{\max} = 6$ として 200 回の試行でランクが 5 と判定される割合を縦軸にプロットした．基準長 L は 1 とした．(a) はノイズレベル ϵ を既知とし，(b) は式 (41) から推定した場合である．

幾何学的 AIC は $\gamma = 0$ (真のランクが 5) のときでも一定の割合でランクが 6 と判定し，$\gamma > 0$ (真のランクが 6) のときは γ が非常に小さくてもランクは 6 と判定する．これに対して幾何学的 MDL は $\gamma = 0$ (真のランクが 5) のときは常にランクが 5 と判定するが，$\gamma > 0$ (真のランクが 6) のときでも γ がある程度の値までランクが 5 と判定する．これは統計的推測における "MDL の一致性" (真のモデルを選ぶ割合がデータ長 ∞ の極限で 100%に収束すること) に対応している [11]．このため幾何学的 AIC はランクを大き目に，幾何学的 MLD は小さ目に選択する傾向がある．このような幾何学的 AIC と幾何学的 MDL の対照的な挙動は他の応用でも見られる [11]．

7. むすび

AIC と MDL をめぐっては今日でも統計学者や情報理論学者の間で論争があり，またより優れるとする他の基準も数多く提案されている．しかし本論文ではどの基準の支持も不支持もせず，背景にある思想 (カルバック情報量，最小記述原理，等) の正当化もしない．本論文ではそれらが "幾何学的当てはめ" ではどう定式化されるかに限定した．

コンピュータビジョンの分野でも MDL が AIC に優れると考える研究者が多いようである．しかし，幾何学的 AIC は退化のもとでもある割合で一般モデルを選ぶのに対して，幾何学的 MDL は退化のもとでは常に退化モデルを選ぶが，退化していなくても退化モデルを選ぶ傾向が強い．これは幾何学的 MDL のほうが幾何学的 AIC よりモデルの複雑さに対するペナルティが大きいことから当然である．

先に指摘したように，幾何学的 MDL はスケールの問題が生ずる．実際問題としてはスケールを 1/10 から 10 倍程度変化させてもモデル選択には影響を与えないが，思想的にベイズの立場を要求し，これを認めるか否かは議論が別れる．これは Rissanen の MDL でも同じであり，統計学者の中にも MDL のほうがユーザが調節できるパラメータが多いからよいという意見もあり，何のためのモデル選択かという問題にかかわってくる．

特別の要求がなければ，このような微妙な問題を避けるためにも幾何学的 AIC を用いるのが簡明である．それで解決しない問題はユニバーサルな基準をいじるより，個々の問題に即して対応するのが現実的であろう．

最近，長尾ら [18] も「幾何学的 AIC」,「幾何学的 MDL」と呼ぶ基準を定義したが，これはデータ数 N を漸近変数にとるものである．このとき未知数 $\{\bar{\boldsymbol{x}}_\alpha\}$ が N とともに増加するので，推定挙動が変則的となる (このため「かく乱母数」と呼ばれる)．また符号長も急速に増大する．そこで彼らは $\{\bar{\boldsymbol{x}}_\alpha\}$ を母数未知の分布からのサンプルとみなし，その母数を推定する方式をとっている．これを一般化したものは「セミパラメトリックモデル」と呼ばれる [3]．

このような取扱いは従来の統計学の思想に沿った発展であり [3], [18]，データ数が任意に増やせる応用に

有効である．例えば時系列データから信号源の数を推定する問題は行列のランクの推定に帰着するが，それに AIC や MDL を適用する試みがある [23]．それに対して，画像処理のような最小限の解像度で最大限の精度を得ようとする応用では本論文の方法が適していると思われる．

謝辞　有益な議論を頂いたオーストラリア Monash 大学の David Suter 教授，理化学研究所の甘利俊一博士，電気通信大学の韓太舜教授，群馬大学の関庸一助教授，NEC の竹内純一氏，(株) 朋栄の松永力氏に感謝する．本研究の一部は文部科学省科学研究費基盤研究 C (2)(No. 13680432) によった。

文　献

[1] H. Akaike, “A new look at the statistical model identification,” IEEE Trans. Autom. Control, vol.16, no.6, pp.716–723, Dec. 1974.

[2] 赤池弘次, “情報量基準 AIC とは何か—その意味と将来への展望,” 数理科学, no.153, pp.5–11, March 1976.

[3] 甘利俊一, 川鍋元明, “線形関係の推定—最小 2 乗法は最良であるのか?,” 応用数理, vol.6, no.2, pp.96–109, June 1996.

[4] B. Efron and R. J. Tibshirani, An Introduction to Bootstrap, Chapman-Hall, New York, 1993.

[5] K. Bubna and C. V. Stewart, “Model selection techniques and merging rules for range data segmentation algorithms,” Comput. Vision Image Understand., vol.80, no.2, pp.215–245, 2000.

[6] 韓　太舜, 小林欣吾, 情報と符号化の数理, 培風館, 東京, 1999.

[7] 顧　海松, 浅田　稔, 白井良明, “動き情報に基づくエッジセグメントの最適分割,” 信学論 (D-II), vol.J76-D-II, no.8, pp.1544–1553, Aug. 1993.

[8] K. Kanatani, Statistical Optimization for Geometric Computation: Theory and Practice, Elsevier, Amsterdam, 1996.

[9] 金谷健一, “統計的推測と幾何学的当てはめにおけるモデル選択,” 情処研資 2000-CVIM-122-1, pp.1–8, May 2000.

[10] 金谷健一, 黒澤典義, 松永　力, モデル選択によるランク推定と複数運動の分離, 情処研資, 2001-CVIM-126-3, pp.17–24, March 2001.

[11] 金谷健一, 松永　力, “幾何学的ＭＤＬとそのメディア応用,” 情処研資 2000-CVIM-122-2, pp.9–16, May 2000.

[12] 金澤　靖, 金谷健一, “幾何学的 AIC による画像モザイク生成の安定化,” 信学論 (A), vol.J83-A, no.6, pp.686–693, June 2000.

[13] 金澤　靖, 金谷健一, “画像の特徴点に共分散行列は本当に必要か?,” 情処研資, 2001-CVIM-126-1, pp.1–8, March 2001.

[14] Y. G. Leclerc, “Constructing simple stable descriptions for image partitioning.” Int. J. Comput. Vision, vol.3, no.1, pp.73–102, 1989.

[15] 松永　力, 金谷健一, “平面パタンを用いる移動カメラの校正：最適計算，信頼性評価，および幾何学的 AIC による安定化,” 信学論 (A), vol.J83-A, no.6, pp.694–701, June 2000.

[16] B. A. Maxwell, “Segmentation and interpretation of multicolored objects with highlights,” Comput. Vision Image Understand., vol.77, no.1, pp.1–24, 2000.

[17] 宮島耕治, 武川直樹, 岡田　守, “MDL 原理に基づく正則化—不連続性に対する正則化パラメータの推定—,” 信学論 (D-II), vol.J80-D-II, no.9, pp.2369–2378, Sept. 1997.

[18] 長尾淳平, 韓　太舜, “かく乱母数を含む場合の MDL 基準の構築と空間図形モデル推定問題への応用,” 信学論 (A), vol.J83-A, no.1, pp.83–95, Jan. 2000.

[19] J. Rissanen, “Universal coding, information, prediction and estimation,” IEEE Trans. Inf. Theory, vol.30, no.4, pp.629–636, July 1984.

[20] J. Rissanen, Stochastic Complexity in Statistical Inquiry, World Scientific, Singapore, 1989.

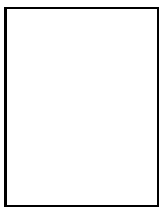
[21] J. Rissanen, “Fisher information and stochastic complexity,” IEEE Trans. Inf. Theory, vol.42, no.1, pp.40–47, January 1996.

[22] P. H. S. Torr, “An assessment of information criteria for motion model selection,” Proc. IEEE Conf. Comput. Vision Patt. Recogn., pp. 47–53, Puerto Rico, June 1997.

[23] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” IEEE Trans. Acoust., Speech & Signal Process., vol.33, no.2, pp.387–392, April 1985.

[24] 山西健司, 韓　太舜, “MDL 入門：情報理論の立場から,” 人工知能学会誌, vol.7, no.3, pp.427–434, May 1992.

（平成 13 年 1 月 15 日受付）



金谷　健一　（正員）

1972 東大・工・計数（数理工学）卒．1979 同大大学院博士課程了．工博．群馬大学工学部情報工学科教授を経て，現在，岡山大学工学部情報工学科教授．米国 Maryland 大学，デンマーク Copenhagen 大学，英国 Oxford 大学，フランス INRIA 客員研究員歴任．