ヒューマンインタフェース　95－1
コンピュータビジョンと　129－1
イメージメディア
(2001. 9. 13)

幾何学的AICと幾何学的MDL再考

金谷 健一
岡山大学工学部情報工学科

「幾何学的AIC」と「幾何学的MDL」を根本原理に戻って導出し，その基礎を明確にする．そのために，ノイズレベルϵを漸近変数にとる「幾何学的当てはめ」と観測数nが漸近変数となる「統計的推測」と対比させ，この対比のもとで赤池のAICとRissanenのMDLに対応するものとして幾何学的AICと幾何学的MDLをとらえ，種々の問題点を明らかにする．

キーワード: 統計的推測, 幾何学的当てはめ, 幾何学的AIC, 幾何学的MDL, モデル選択, 退化の検出

Geometric AIC and Geometric MDL Revisited

Kenichi Kanatani
Department of Information Technology, Okayama University, Okayama 700-8530 Japan

In an attempt to give a rigorous foundation to the "geometric AIC" and the "geometric MDL", we closely examine their derivations by going back to their first principles. To this end, we contrast "geometric fitting", for which the noise level ϵ is taken as the asymptotic variable, with "statistical inference", for which the number n of observation is taken as the asymptotic variable, and regard the geometric AIC and the geometric MDL as the counterparts of Akaike's AIC and Rissanen's MDL. We discuss many problems that emerge from our analysis.

Key words: statistical inference, geometric fitting, geometric AIC, geometric MDL, model selection, degeneracy detection

1. 序論

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赤池のAIC [1, 2, 23] の考え方を用いて「幾何学的AIC」を提案し [8]，さまざまな応用例を示した [9, 13, 14, 16, 19, 24]．

さらに赤池のAICと幾何学的AICの関係を明らかにするために，問題を「統計的推測」と「幾何学的当てはめ」の対比と位置づけた [10]．前者は観測数nの増加に対する挙動を解析し，少ないデータ数でも高い精度を得ようとするのに対して，後者はノイズレベルϵの減少に対する挙動を解析し，低い解像度でも高い精度を得ようとする．

この考えに基いてRissanen [20, 21] のMDLの幾何学的当てはめにおける対応として「幾何学的MDL」を定義し，いくつかの問題に適用して幾何学AICとの違いを観察した [11, 12, 17]．

しかしこれまでの議論にはあいまいなところがあり，種々の疑問 (特にスケールに関する問題) が提起された．本論文では幾何学的AICと幾何学的MDLをより厳密に定義し，その基礎をより明確にすることによってそれらの疑問に答えるとともに，浮かび上がる諸問題を検討する．

†700-8530 岡山市津島中3－1－1
岡山大学工学部情報工学科 Tel/Fax: (086)251-8173
E-mail: kanatani@suri.it.okayama-u.ac.jp

2. 定義

まず，幾何学的AICと幾何学的MDLの基礎となる幾何学的当てはめと幾何学的モデル選択を定義する．

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$\mathbf{0}$，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$\{\boldsymbol{x}_\alpha\}$がデータ空間$\mathcal{X}$のある多様体上に拘束され，パラメータ$\boldsymbol{u}$もパラメータ空間$\mathcal{U}$のある多様体上に拘束される場合 (例えば$\boldsymbol{x}_\alpha$も$\boldsymbol{u}$も単位ベクトルの場合) に容易に拡張できる [8]．射影行列を導入し，逆行列を (ムーア・ペンローズの) 一

般逆行列に置きかえれば，以下の議論はそのまま成立する．

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 [15]．

もう一つの理由は，最尤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，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]．

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

$W_\alpha^{(kl)}$ は $(\nabla_{\boldsymbol{x}} F_\alpha^{(k)}, V_0[\boldsymbol{x}_\alpha] \nabla_{\boldsymbol{x}} F_\alpha^{(l)})$ を (kl) 要素とする $r \times r$ 行列の逆行列の (kl) 要素である (F の添え字 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 $\hat{\boldsymbol{u}}$ の共分散行列はノイズレベル $\epsilon \to 0$ 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尤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，また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．これは現象に加わるランダム誤差を**繰り返し観測によって克服しよう**という思想に基いている．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よう**という思想に基いている [7, 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もよいからである．

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 [7, 8, 10]．これは，一つの画像を仮想的に繰り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 [10]．

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 $p_1(\boldsymbol{x}|\boldsymbol{\theta}_1), p_2(\boldsymbol{x}|\boldsymbol{\theta}_2), \ldots$ 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 $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0}$ 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]．

確率的機構は既知の正規化共分散行列 $V_0[\boldsymbol{x}_\alpha]$ をもつ正規分布と仮定され，選択の余地がなく，選択すべきはデータの従属関係のみである．誤差は画像や処理アルゴリズムの特性であり，画像の意味やシーンの満たす拘束条件とは無関係である．

3. 幾何学的 AIC

赤池の AIC[1, 2, 23] 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) のもとではデータ $\{\boldsymbol{x}_\alpha\}$ 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)．

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

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 $p_T(\{\boldsymbol{X}_\alpha\})$ までの**カルバック情報量**であり [1, 2, 23]，次のように書ける．

$$D = \int\cdots\int p_T(\{\boldsymbol{X}_\alpha\})\log\frac{p_T(\{\boldsymbol{X}_\alpha\})}{p(\{\boldsymbol{X}_\alpha\})}d\boldsymbol{X}_1\cdots d\boldsymbol{X}_N = E[\log p_T(\{\boldsymbol{X}_\alpha\})] - E[\log p(\{\boldsymbol{X}_\alpha\})] \tag{6}$$

$E[\cdot]$ は真の確率密度 $p_T(\{\boldsymbol{X}_\alpha\})$ 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，$-E[\log p(\{\boldsymbol{X}_\alpha\})]$ 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．

$$-E[\log p(\{\boldsymbol{X}_\alpha\})] = \frac{1}{2\epsilon^2}E[\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{X}_\alpha - \bar{\boldsymbol{x}}_\alpha))] + \frac{mN}{2}\log 2\pi\epsilon^2 + \frac{1}{2}\sum_{\alpha=1}^{N}\log|V_0[\boldsymbol{x}_\alpha]| \tag{7}$$

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 のみに依存する正の量を掛けてもモデル選択には影響しない)．

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

これはよく知られた (正規分布の誤差のもとでの) **最小二乗基準**にほかならない．

3.2 期待値の評価

式 (8) はモデル選択基準としては使えない．なぜなら，期待値 $E[\cdot]$ は真の密度に関してとらなければならないが，それが未知だからである．**これにどう対処するかが統計的推測と幾何学的当てはめでは異なる．**

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 $p_T(\boldsymbol{X})$ から多数の独立な観測値 $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ 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**)．これはモンテカルロ法による積分評価の原理でもあり，赤池の AIC もこれに基いている [1, 2, 23]．

一方，幾何学的当てはめではデータ $\{\boldsymbol{x}_\alpha\}$ 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か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 $\int\cdots\int Y(\{\boldsymbol{X}_\alpha\})p_T(\{\boldsymbol{X}_\alpha\})d\boldsymbol{X}_1\cdots d\boldsymbol{X}_N$ は $Y(\{\boldsymbol{x}_\alpha\})$ で近似できる ($1/N$ がないことに注意)．なぜなら $\epsilon\to 0$ で $p_T(\{\boldsymbol{X}_\alpha\})\to\prod_{\alpha=1}^{N}\delta(\boldsymbol{X}_\alpha - \bar{\boldsymbol{x}}_\alpha)$ であり ($\delta(\cdot)$ はデルタ関数)，$\int\cdots\int Y(\{\boldsymbol{X}_\alpha\})p_T(\{\boldsymbol{X}_\alpha\})d\boldsymbol{X}_1\cdots d\boldsymbol{X}_N$ 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．

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

3.3 偏差の除去

式 (9) はまだモデル選択基準としては使えない．それは未知数 $\{\bar{\boldsymbol{x}}_\alpha\}$, $\boldsymbol{u}$ が含まれているからである．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**最尤推定量** $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ を仮定するのが当然である．素朴な考えはこれら最尤推定量を式 (9) に代入した**残差 (平方和)**

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

を用いることである．しかし，これは論理的に矛盾である．

式 (1) は特定のモデルではなく，$\{\bar{\boldsymbol{x}}_\alpha\}$, $\boldsymbol{u}$ をパラメータとするモデルの**クラス**を定義している．そして，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 節の論理から，そのモデルのよさは $E[\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{X}_\alpha - \hat{\boldsymbol{x}}_\alpha))]$ で測るべきであり，3.2 節の論理からその期待値は $\{\boldsymbol{X}_\alpha\}$ の代表的な実現値 $\{\boldsymbol{x}_\alpha\}$ を代入して近似できる．しかし，**$\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ はその実現値 $\{\boldsymbol{x}_\alpha\}$ 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 $E[\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{X}_\alpha - \hat{\boldsymbol{x}}_\alpha))]$ よりも小さい．これは赤池の遭遇した困難と同一である [1, 2, 23]．赤池の解決法をこの場合に翻訳すると次のようになる．

3.2 節の論理から，仮定したモデルとは無関係の実現値 $\{\boldsymbol{x}_\alpha^*\}$ を用いて

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

を評価する．$\{\boldsymbol{x}_\alpha^*\}$ は仮想的に別の実験を行った場合に得られるであろうデータであり，**将来のデータ**と呼ぶ．しかし現実には現在のデータ $\{\boldsymbol{x}_\alpha\}$ しか得ら

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

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

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

$\hat{\boldsymbol{u}}$ を符号化するため 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．

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

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．

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

したがって領域 V_u 中の格子単位の総数が次のように書ける．

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

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

$\{\hat{\boldsymbol{x}}_\alpha\}$ を符号化するためにその定義域を量子化する．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．$\hat{\boldsymbol{x}}_\alpha$ は式 (17) のように別々の正規化共分散行列 $V_0[\hat{\boldsymbol{x}}_\alpha]$ を持つので，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．各 $\hat{\boldsymbol{x}}_\alpha$ は式 (17) のように別々の正規化共分散行列 $V_0[\hat{\boldsymbol{x}}_\alpha]$ を持つので，

$\{\hat{\boldsymbol{x}}_\alpha\}$ 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．さらに $\hat{\mathcal{S}}$ 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 $V_0[\hat{\boldsymbol{x}}_\alpha]^-$ を $\hat{\mathcal{S}}$ の接空間に制限した行列式 (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書ける．

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

この内の一つを指定する符号長は次のようになる．

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

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

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

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

$\hat{\boldsymbol{u}}$ を符号化するため 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．

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

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．

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

したがって領域 V_u 中の格子単位の総数が次のように書ける．

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

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

$\{\hat{\boldsymbol{x}}_\alpha\}$ を符号化するためにその定義域を量子化する．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．$\hat{\boldsymbol{x}}_\alpha$ は式 (17) のように別々の正規化共分散行列 $V_0[\hat{\boldsymbol{x}}_\alpha]$ を持つので，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．各 $\hat{\boldsymbol{x}}_\alpha$ は式 (17) のように別々の正規化共分散行列 $V_0[\hat{\boldsymbol{x}}_\alpha]$ を持つので，

$\{\hat{\boldsymbol{x}}_\alpha\}$ 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．さらに $\hat{\mathcal{S}}$ 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 $V_0[\hat{\boldsymbol{x}}_\alpha]^-$ を $\hat{\mathcal{S}}$ の接空間に制限した行列式 (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書ける．

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

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) から最尤推定量 $\{\hat{\boldsymbol{x}}_\alpha\}$, $\hat{\boldsymbol{u}}$ の符号長の合計は次のようになる.

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

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.

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

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.

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

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,

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

を幾何学的 MDL と呼ぶ.

4.6 スケールの問題

式 (33) をモデル選択基準とするのは実際的とはいえない. まず右辺第 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. そのため, 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 MDL の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有限長で符号化することは不可能だからである. このようなベイズの立場を認めるかどうかは議論が別れるところである.

これをあいまいにして避ける便法は, 問題の生じる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.

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

これは筆者らが最初に提案した形である [12, 17]. 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ことが指摘された. 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.

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

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な便法に見えるが, 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 [20].

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

各データ $\boldsymbol{x}_i$ はパラメータ $\boldsymbol{\theta}$ を持つ確率密度 (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

尤推定量である．k は $\boldsymbol{\theta}$ の次元であり，$O(1)$ はデータ長 n を漸近変数にとる $n \to \infty$ の評価である．筆者らの初期の幾何学的 MDL(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 $\{\boldsymbol{x}_i\}$ を二つづつまとめて，$\{(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{x}_3, \boldsymbol{x}_4), ...\}$ 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．これが Rissanen の MDL に対する一つの反論ともなっていた [3]．しかし，その後 Rissanen の MDL は次の形とされる [22]．

$$\mathrm{MDL} = -\log \prod_{i=1}^{n} p(\boldsymbol{x}_i|\hat{\boldsymbol{\theta}}) + \frac{k}{2}\log\frac{n}{2\pi} + \log\int_{V_\theta}\sqrt{|\boldsymbol{I}(\boldsymbol{\theta})|}d\boldsymbol{\theta} + o(1) \qquad (37)$$

$\boldsymbol{I}(\boldsymbol{\theta})$ は $p(\boldsymbol{x}_i|\boldsymbol{\theta})$ 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．

もし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 $\mathcal{X}$ 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．

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

この式によって正しく推定できることは多くのシミュレーションで確認されている．

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綻する．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，モデル選択によって退化を自動的に検出し，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 [9, 13, 14, 16, 19, 24]．

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，個々のモデルに基いて推定する必要がある．これが統計的推測と幾何学的当てはめの基本的な相違の一つである．

6. まとめ

本論文の目的は，筆者が以前に提案した「幾何学的 AIC」と「幾何学的 MDL」に対して提起された種々の疑問に答えるために，根本原理に立ち戻って導出を詳細に点検し，その基礎を明確にすることである．そのために，ノイズレベル ϵ を漸近変数にとる「幾何学的当てはめ」と観測数 n が漸近変数となる「統計的推測」と対比させ，この対比のもとで赤池の AIC と Rissanen の MDL に対応するものとして幾何学的 AIC と幾何学的 MDL を厳密に導出した．

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**．本論文ではそれらを既にあるものとし，それらが**幾何学的当てはめにおいてはどう定式化されるか**に限定したものである．

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り，筆者が幾何学的 AIC を提案して以来，MDL を (使えるように適切に定式化した上で) 用いるべきではないかという反論が絶えず，これが幾何学的 AIC に対する一種の批判となっていた．

しかし幾何学的 AIC と幾何学的 MDL を具体的な問題に適用して比較したところ [11, 12, 17]，両者は対照的な性質をもつものの，幾何学的 MDL が幾何

学的 AIC より優れるという結果は得られなかった．

一般に幾何学的 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 (これは統計的推測のおける **MDL の一致性**に対応する)，退化していなくても退化モデルを選ぶ傾向が強い．一般に幾何学的 AIC はデータの変動をより忠実に反映し，幾何学的 MDL はデータの変動を無視しても単純なモデルを選ぶ傾向がある．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，これはデータ数を漸近変数にとるものである．データ数を漸近変数にとれば未知数 $\{\bar{x}_\alpha\}$ 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**撹乱母数**と呼ばれる)．また符号長も急速に増大する．そこで彼らは $\{\bar{x}_\alpha\}$ 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 [4]．

このような取り扱いは従来の統計学の思想に沿った自然な発展であり [4, 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 [25]．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論が適していると思われる．

謝辞: 本研究に対する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。

参考文献

[1] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automatic Control*, vol.16, no.6, pp.716–723, 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] 赤池弘次, AIC と MDL と BIC, オペレーションズ・リサーチ, vol.41, no.7, pp.375–378, 1996.

[4] 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, 1996.

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

[6] 韓太舜, 小林欣吾, 情報と符号化の数理, 培風館, 1999.

[7] 金谷健一, 空間データの数理, 朝倉書店, 1995.

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

[9] 金谷健一, 自己評価を伴うアクティブビジョン, 日本ロボット学会誌, vol.15, no.2, pp.268–274, 1997.

[10] 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, 情報処理学会研究報告 2000-CVIM-122-1, pp.1–8, May 2000.

[11] 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, 情報処理学会研究報告 2001-CVIM-126-3, pp.17–24, March 2001.

[12] 金谷健一, 松永 力, 幾何学的 MDL とそのメディア応用, 情報処理学会研究報告 2000-CVIM-122-2, pp.9–16, May 2000.

[13] Y. Kanazawa and K. Kanatani, Infinity and planarity test for stereo vision, *IEICE Trans. Inf. & Syst.*, vol.E80-D, no.8, pp.774–779, 1997.

[14] 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, 電子情報通信学会論文誌 A, vol.J83-A, no.6, pp.686–693, 2000.

[15] 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？, 情報処理学会研究報告 2001-CVIM-126-1, pp.1–8, March 2001.

[16] 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, 電子情報通信学会論文誌 A, vol.J83-A, no.6, pp.694–701, 2000.

[17] C. Matsunaga and K. Kanatani, Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by model selection, *Proc. 6th Euro. Conf. Comput. Vision*, June–July, 2000, Dublin, Ireland, vol.2, pp.595–609.

[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, 2000.

[19] N. Ohta and K. Kanatani, Moving object detection from optical flow without empirical thresholds, *IEICE Trans. Inf. & Syst.*, vol.E81-D, no.2, pp.243–245, 1998.

[20] J. Rissanen, Universal coding, information, prediction and estimation, *IEEE Trans. Inform. Theory*, vol.30, no.4, pp.629–636, 1984.

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

[22] J. Rissanen, Fisher information and stochastic complexity, *IEEE Trans. Inform. Theory*, vol.42, no.1, pp.40–47, 1996.

[23] 坂本慶行, 石黒真木夫, 北川源四郎, 情報量統計学, 共立出版, 1983.

[24] Iman Triono, N. Ohta and K. Kanatani, Automatic recognition of regular figures by geometric AIC, *IEICE Trans. Inf. & Syst.*, vol.E81-D, no.2, pp.246–248, 1998.

[25] 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, 1985.

[26] 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, 1992.