

幾何学的モデルの選択基準について

金谷 健一[†]

群馬大学工学部情報工学科

本論文では、ロボットが画像やセンサデータから3次元環境のモデルを構築するときの幾何学的モデルの選択の基準として、AICを変形したものを導入する。3次元環境のモデル化の問題は通常の統計学と異なり、局外母数がデータ数だけ存在し、モデルが多様体によって指定されることから、その複雑さはパラメータ数(自由度)のみで評価されるのではなく、その多様体の次元や余次元などの座標系の変換に不変な量が本質的な役割を果たすことを示す。また、問題の幾何学的な性質から、推定したいのが誤差の性質ではなく幾何学的対象であるため、誤差分布に関する知識がなければAICのみではモデルの比較ができない。これを克服する方法として、モデルのよさを相対的に評価する比較検定の方式を提案する。このとき恣意的な有意水準は必要ない。これらを単純な平面上の点データに対する例題で説明する。

キーワード: AIC、最尤推定、仮説検定、幾何学的モデル、3次元環境理解、モデル選択

Selection Criterion for Geometric Models

Kenichi Kanatani

Department of Computer Science
Gunma University, Kiryu, Gunma 376, Japan

In this paper, we modify the *AIC criterion* so that it can be used for selecting a geometric model for robotics applications, such as building a geometric model of the environment from image and sensor data. Such geometric problems are very different from traditional statistical problems, to which AIC has been widely applied, because as many *nuisance parameters* as the number of the data appear and the model is defined as a *manifold*. As a result, the complexity of the model is evaluated by not only its degree of freedom but also such invariant quantities as the *dimension* and *codimension* of the manifold. Since what we want to estimate is a geometric object, not the statistical characteristics of the noise, the use of the AIC criterion requires the knowledge of the noise distribution. In order to avoid this, we present a *comparison test* for evaluating relative goodness of one model to another without using any arbitrarily set threshold such as the significance level. Our theory is illustrated by analyzing point data in two dimensions.

Key words: AIC, maximum likelihood estimation, testing of hypothesis, geometric model, 3-D environment understanding, model selection

謝辞: 本研究の発端となった討論をして頂いた東京大学工学部の甘利俊一教授に感謝します。また、本研究に参加して頂いた群馬大学工学部の太田直哉助手、群馬工業高専の金沢靖助手、および筆者の研究室の大学院諸君にも感謝します。本研究の一部は文部省科学研究費一般研究B(No. 07458067)および大川情報通信基金からの助成による。

[†] 376 桐生市天神町1-5-1 群馬大学工学部情報工学科
Tel: (0277)30-1841, Fax: (0277)30-1801
E-mail: kanatani@cs.gunma-u.ac.jp

1. はじめに

ロボットが自律的に作業するには画像やセンサデータからの3次元環境理解、すなわち3次元環境モデルの構築が不可欠である。その根源的な問題は画像やセンサデータには誤差があることである。これを解決する基本的な方法は、得られたデータに2次元および3次元物体のモデルを最適に当てはめることである。物体のモデルは直線、円、平面、円筒、球などのプリミティブから構成される。したがって、

1. モデルを選択する、
2. モデルを最適に当てはめる、

の2段階が必要である。これはより抽象的なレベルにも当てはまる。例えば動画像から対象の3次元構造を復元する動画像の3次元解析は、動画像データにそれが3次元空間を移動する物体の投影像であるというモデルを最適に当てはめる操作である。この問題は3次元復元とも呼ばれ、最適な復元の研究に数理的な方法が駆使されてきた[3, 4, 5, 6, 12, 15]。

一方、モデル選択として提案されているのは、経験的な方法による人工知能的な推論規則がほとんどである。これはロボットの作業空間では候補モデルが限定されていて、簡単な特徴抽出のみから容易に判定できると期待されるためであるが、十分に研究が進んでいるとはいえない。今後は、これに対しても数理的な方法論が確立されることが望まれる。

筆者は画像理解やロボティクスへの応用に適合する統計学を再構成する努力を行ない、特に第2段階の3次元復元、すなわち指定されたモデルの最適当てはめ問題を統計的推論として定式化する研究を行なった[6, 10]。本論文は第1段階のモデル選択に対しても、これに統計学としての基礎づけを与えようとするものである。

モデル選択の基準としてはAIC[1]がさまざまに用いられている¹。これは最尤推定量の将来のデータに対する予測能力でモデルのよさを測ろうとする立場である[2, 13]。通常AICは

$$\begin{aligned} & (\text{データとモデルのくい違いの程度}) \\ & + (\text{モデルのパラメータ数の影響}) \end{aligned}$$

と説明され、これが小さいほどよい、したがってデータが同程度に説明できる限り、パラメータ数の少ないモデルを選ぶと解釈されている。しかし幾何学的モデルの選択にはこの解釈は当たらない。

【例1】3次元空間中の点データ集合に

1. 平面を当てはめる、
2. 直線を当てはめる、

という二つのモデルを比較するとする(図1)。最小二乗法²で当てはめた平面または直線からの残差をくい

¹最近ではMLD基準[14]に関心が高まっているが本論文では触れない。

²誤差の分布が各点で独立、同一、等方な正規分布であれば、最

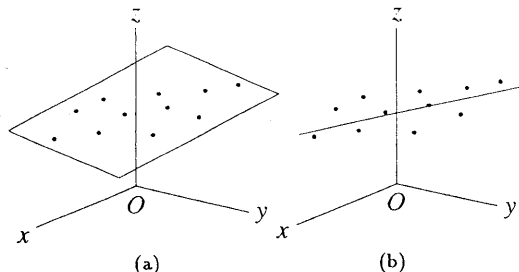


図1: 空間中の点データに(a)平面を当てはめる(b)直線を当てはめる

違いの尺度とすると、当然平面を当てはめたほうが残差が常に少ない。一方、平面は3個のパラメータで指定できるが、直線は4個のパラメータを必要とするから、パラメータに関しても平面のほうが有利である。するとどんなデータでも(例え同一直線上にあって)平面を当てはめるべきであろうか?

このパラドックスは、幾何学的モデルが多様体で記述されること、および局外母数の役割を無視したためである。正しく解釈すると、モデルの多様体を指定するパラメータ数(自由度)は本質的な意味をもたず、その次元と余次元³が重要な役割を果たす。本論文ではこのことを単純な平面上の点データに対する例題で説明し、恣意的な有意水準を用いないでモデルのよさを相対的に評価する比較検定の方式を提案する。

2. 幾何学的推定問題

n 次元空間 R^n に N 個の点 r_1, \dots, r_N が与えられているとする。各点 r_α はその真の位置 \bar{r}_α から誤差 Δr_α だけずれたものとし、

$$r_\alpha = \bar{r}_\alpha + \Delta r_\alpha \quad (1)$$

と置く。誤差 Δr_α は各 α について独立で、期待値0、共分散行列 $V[r_\alpha]$ の正規分布に従う確率変数の実現値であるとする。各共分散行列 $V[r_\alpha]$ は正則行列⁴であり、かつ既知であるとする。一方、真の位置 \bar{r}_α 、 $\alpha = 1, \dots, N$ は n 次元空間 R^n のある多様体 S の上にあるとする。次の問題を考える。

【問題1】データ $\{r_\alpha\}$ から真の位置 $\{\bar{r}_\alpha\}$ を含む多様体 S を推定せよ。

この問題を解くためにモデルを導入する。 n 次元空間中の d 個のパラメータで指定される m 次元多様体小二乗解は最尤推定量であり、その残差は定数差を除いて最大対数尤度の符号を変えたものに一致する。

³その多様体の存在する空間の次元とその多様体の次元との差。

⁴データの定義域が n 次元空間のある多様体上に拘束されていれば、その共分散行列はその多様体の接空間を値域とする特異行列となる。例えばデータが単位ベクトルであれば、これは半径1の球面上に拘束され、その共分散行列はそのベクトルに直交する方向のみに値域をもつ。このような場合も一般逆行列や射影子を用いれば同様な議論ができるが、本論文では説明の簡単のため正則な共分散行列の場合のみを考える。

S を (幾何学的) モデルと呼び、 $m, r (= n - m), d$ をそれぞれそのモデルの次元、余次元、自由度と呼ぶ。モデル S を一つ選ぶと、問題 1 は次の問題となる。

【問題 2】 与えられたモデルに対して、真の位置 $\{\hat{r}_\alpha\}$ 、および $\hat{r}_\alpha \in S$ となる多様体 S をデータ $\{r_\alpha\}$ から最適に推定せよ。

最適性の基準として最尤推定を採用する。 $\{\hat{r}_\alpha\}, S$ の推定量を $\{\hat{r}_\alpha\}, \hat{S}$ とするときの尤度は、仮定より次のように与えられる⁵。

$$L(\{\hat{r}_\alpha\}, \hat{S}) = \prod_{\alpha=1}^N \frac{e^{-(r_\alpha - \hat{r}_\alpha, V[r_\alpha]^{-1}(r_\alpha - \hat{r}_\alpha))/2}}{\sqrt{(2\pi)^n |V[r_\alpha]|}} \quad (2)$$

したがって、問題 2 は次のように書き直される。

【問題 3】 尤度 $L(\{\hat{r}_\alpha\}, \hat{S})$ を制約条件 $\hat{r}_\alpha \in \hat{S}$ のもとで最大にする位置 $\{\hat{r}_\alpha\}$ および多様体 \hat{S} を求めよ。

このようにして最尤推定量 $\{\hat{r}_\alpha\}, \hat{S}$ が得られたとする。これを指定するパラメータ数を考える。まず多様体 \hat{S} の指定に d 個必要である。多様体 \hat{S} 上に曲線座標系を定義すれば、その上の一点は m 個の座標値で指定できる。したがって、 $\{\hat{r}_\alpha\}$ の指定に mN 個必要である。ゆえにモデル S の AIC は次のように書ける。

$$AIC(S) = -2 \log L(\{\hat{r}_\alpha\}, \hat{S}) + 2(mN + d) \quad (3)$$

3. 幾何学的 AIC の定義

画像理解やロボティクスへの応用を考えると、共分散行列 $V[r_\alpha]$ を次のように分解した形で表現するのが都合が良い。

$$V[r_\alpha] = \epsilon^2 V_0[r_\alpha], \quad \alpha = 1, \dots, N \quad (4)$$

定数 ϵ は誤差の大きさの尺度であり、ノイズレベルと呼ぶ。そして $V_0[r_\alpha]$ を正規化共分散行列と呼ぶ [10]。例えば 3 次元空間で各座標軸方向に標準偏差 ϵ の誤差が発生する場合は、その ϵ をノイズレベルとすると、正規化共分散行列 $V_0[r_\alpha]$ は単位行列 I である。残差平方和を次のように定義する。

$$S(\hat{S}) = \sum_{\alpha=1}^N (r_\alpha - \hat{r}_\alpha, V_0[r_\alpha]^{-1}(r_\alpha - \hat{r}_\alpha)) \quad (5)$$

すると式 (3) は次のように書き直せる。

$$AIC(S) = \frac{S(\hat{S})}{\epsilon^2} + \sum_{\alpha=1}^N \log |V_0[r_\alpha]| + nN \log 2\pi\epsilon^2 + 2(mN + d) \quad (6)$$

ノイズレベル ϵ 、正規化共分散行列 $V_0[r_\alpha]$ 、データの次元 n および個数 N は固定されているので、モデル

⁵ 本論文ではベクトル a, b の内積を (a, b) で表す。

S を測る尺度としては右辺第 2, 第 3 項を除外してもよい。さらに全体を定数倍してもよい。そこで、モデル S の幾何学的 AIC を次のように定義する。

$$AIC_0(S) = S(\hat{S}) + 2(mN + d)\epsilon^2 \quad (7)$$

統計学との対応を考えると、真の位置 $\{\hat{r}_\alpha\}$ がいわゆる局外母数 (攪乱母数) に当たり、データの個数 N だけ存在する。一方、多様体 S を指定する d 個のパラメータが構造母数に当たるが、これはデータに確率密度関数に陽には含まれていない。含まれているのは局外母数 $\{\hat{r}_\alpha\}$ のみであり、構造母数は間接的に局外母数 $\{\hat{r}_\alpha\}$ の相互の関係を記述している。画像理解やロボティクスで現われる推定問題のほとんどすべてがこの形である。式 (7) から次のことがわかる。

1. データ数 N が大きいとき、モデル S の自由度 d は幾何学的 AIC にほとんど影響を及ぼさない。
2. 当てはめる多様体の次元 m で本質的な役割を果たす。
3. データ数が幾何学的 AIC の第 2 項に現われている。

ところで AIC は元来データ数無限大の漸近的解析から得られるものであり [2, 13]、ここで考えるようなデータ数だけ局外母数が存在する場合にも妥当か疑問に思える。しかし、データ数を無限に近づけなくても、誤差の大きさを 0 に近づける極限の摂動解析を行えば同様な結論が得られる [11]。

式 (7) の右辺の残差平方和 $S(\hat{S})$ を計算するにはモデル S の具体的な表現を与える必要がある。例えばモデル S は次の $r (= n - m)$ 個の独立な方程式⁶ で指定されるとする。

$$F^{(k)}(r, u) = 0, \quad k = 1, \dots, r \quad (8)$$

u は d 次元ベクトルであり⁷、最尤推定量 \hat{S} が $u = \hat{u}$ のとき得られるとする。このとき、次の表現が得られる [7, 8, 9]。

$$S(\hat{S}) = \sum_{\alpha=1}^N \sum_{k,l=1}^r \hat{W}_\alpha^{(kl)} F^{(k)}(r, \hat{u}) F^{(l)}(r, \hat{u}) \quad (9)$$

$$\left(\hat{W}_\alpha^{(kl)} \right) = \left((\nabla_r \hat{F}_\alpha^{(k)}, V_0[r_\alpha] \nabla_r \hat{F}_\alpha^{(l)}) \right)^{-1} \quad (10)$$

式 (10) は $(\nabla_r F^{(k)}(\hat{r}_\alpha, \hat{u}), V[r_\alpha] \nabla_r F^{(l)}(\hat{r}_\alpha, \hat{u}))$ を (kl) 要素とする行列の逆行列の (kl) 要素が $\hat{W}_\alpha^{(kl)}$ に等しいことを表す略記である。また $\nabla_r \hat{F}_\alpha^{(k)}$ は

⁶ その他いくつか、病的な例外を除く条件が必要となる。画像理解やロボティクスへの応用では独立とは限らない、すなわち冗長な方程式を含んだ表現を考える必要がある。その場合も同様に行ける (詳細省略)。

⁷ 画像理解やロボティクスへの応用ではパラメータ空間が d 次元多様体であって d より大きい次元のベクトルで指定される場合も考える必要がある。その場合も同様に行ける (詳細省略)。

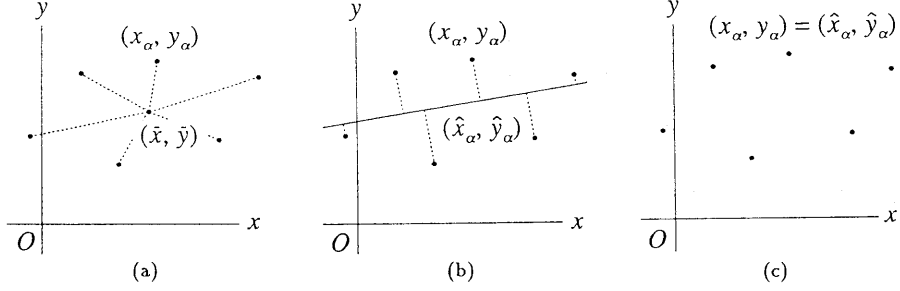


図 2: (a) 一点モデル S_0 (b) 直線モデル S_1 (c) 平面モデル S_2

$\nabla_{\mathbf{r}} F_{\alpha}^{(k)}(\hat{\mathbf{r}}_{\alpha}, \hat{\mathbf{u}})$ の略記であり、 $\nabla_{\mathbf{r}}(\cdot)$ は $\partial(\cdot)/\partial r_1, \dots, \partial(\cdot)/\partial r_n$ を要素とする列ベクトルである。

4. 平面上のデータ点

平面上に N 個のデータ (x_{α}, y_{α}) , $\alpha = 1, \dots, N$ が与えられているとする。誤差は各 x, y 成分に独立に平均 0、標準偏差 ϵ の正規分布に従うとする。 (x_{α}, y_{α}) の真の位置 $(\hat{x}_{\alpha}, \hat{y}_{\alpha})$ の与えられたモデルのもとでの最尤推定量を (\hat{x}, \hat{y}) とすると、残差平方和は次のように表せる。

$$S(\hat{S}) = \sum_{\alpha=1}^N ((x_{\alpha} - \hat{x}_{\alpha})^2 + (y_{\alpha} - \hat{y}_{\alpha})^2) \quad (11)$$

4.1 一点モデル

真の位置はある一点であるというモデルを S_0 とする。すなわち N 点 $\{(x_{\alpha}, y_{\alpha})\}$ はある共通の一点 (\hat{x}, \hat{y}) から独立な誤差によってずれたものとする (図 2(a))。このモデルの次元は 0、余次元は 2、自由度は 2 であり、幾何学的 AIC は次のようになる。

$$AIC_0(S_0) = S(\hat{S}_0) + 4\epsilon^2 \quad (12)$$

このモデルを限定して、真の位置は既知の点、例えば原点 $(0, 0)$ であるモデルを S'_0 とすると、次元は 0、余次元は 2、自由度は 0 であり、幾何学的 AIC は次のようになる。

$$AIC_0(S'_0) = S(\hat{S}'_0) \quad (13)$$

点の (\hat{x}, \hat{y}) の最尤推定量は明らかにサンプル平均

$$\hat{x} = \frac{1}{N} \sum_{\alpha=1}^N x_{\alpha}, \quad \hat{y} = \frac{1}{N} \sum_{\alpha=1}^N y_{\alpha} \quad (14)$$

に一致する。これは各 α について真の位置 $(\hat{x}_{\alpha}, \hat{y}_{\alpha})$ の最尤推定量でもあるから、残差平方和はそれぞれ

$$S(\hat{S}) = \sum_{\alpha=1}^N (x_{\alpha}^2 + y_{\alpha}^2) - N(\hat{x}^2 + \hat{y}^2) \quad (15)$$

$$S(\hat{S}') = \sum_{\alpha=1}^N (x_{\alpha}^2 + y_{\alpha}^2) \quad (16)$$

と書ける。式 (12), (13) より、 $AIC_0(S'_0) < AIC_0(S_0)$ となるのは

$$\sqrt{\hat{x}^2 + \hat{y}^2} < \frac{2\epsilon}{\sqrt{N}} \quad (17)$$

のときである。したがって、サンプル平均が原点から距離 $2\epsilon/\sqrt{N}$ 以下のとき、真の位置は原点であると判定される。

4.2 直線モデル

真の位置はある直線上にあるというモデルを S_1 とする (図 2(b))。このモデルの次元は 1、余次元は 1、自由度は 2 であり、幾何学的 AIC は次のようになる。

$$AIC_0(S_1) = S(\hat{S}_1) + 2(N+2)\epsilon^2 \quad (18)$$

このモデルを限定して、直線が原点を通るモデルを S'_1 とすれば、次元は 1、余次元は 1、自由度は 1 であり、幾何学的 AIC は次のようになる。

$$AIC_0(S'_1) = S(\hat{S}'_1) + 2(N+1)\epsilon^2 \quad (19)$$

$AIC_0(S'_1) < AIC_0(S_1)$ となるのは

$$S(\hat{S}'_1) - S(\hat{S}_1) < 2\epsilon^2 \quad (20)$$

のときであるから、このとき直線は原点を通ると判定される。

点の位置をベクトル $\mathbf{r}_{\alpha} = (x_{\alpha}, y_{\alpha})^T$ で表し、モデル S_1 の直線を $(\mathbf{n}, \mathbf{r}) = d$ と表す。 \mathbf{n} は直線の単位法線ベクトルであり、 d は原点からの距離 (\mathbf{n} の方向に正) である。サンプル平均 $\hat{\mathbf{r}}$ とモーメント行列 \mathbf{M} を

$$\hat{\mathbf{r}} = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{r}_{\alpha}, \quad \mathbf{M} = \sum_{\alpha=1}^N (\mathbf{r}_{\alpha} - \hat{\mathbf{r}})(\mathbf{r}_{\alpha} - \hat{\mathbf{r}})^T \quad (21)$$

と定義すると、 \mathbf{n} の最尤推定量は行列 \mathbf{M} の最小固有値に対する単位固有ベクトル $\hat{\mathbf{n}}$ であり、距離 d の最尤推定量は $\hat{d} = (\hat{\mathbf{n}}, \hat{\mathbf{r}})$ で与えられる [10]。各 α について、 $\hat{\mathbf{r}}_{\alpha}$ の最尤推定量 $\hat{\mathbf{r}}_{\alpha}$ はデータ点 \mathbf{r}_{α} から直線 $(\hat{\mathbf{n}}, \mathbf{r}) = \hat{d}$ に下ろした垂線の足である。したがって、残差平方和が次のように表される。

$$S(\hat{S}_1) = \sum_{\alpha=1}^N (\hat{\mathbf{n}}, \mathbf{r}_{\alpha} - \hat{\mathbf{r}})^2 \quad (22)$$

一方、モデル S'_1 の直線はその単位法線ベクトル \mathbf{n} を用いて $(\mathbf{n}, \mathbf{r}) = 0$ と表せる。 \mathbf{n} の最尤推定量はモーメント行列

$$M' = \sum_{\alpha=1}^N \mathbf{r}_\alpha \mathbf{r}_\alpha^\top \quad (23)$$

の最小固有値に対する単位固有ベクトル $\hat{\mathbf{n}}'$ であり、残差平方和は次のように表せる。

$$S(\hat{S}'_1) = \sum_{\alpha=1}^N (\hat{\mathbf{n}}', \mathbf{r}_\alpha)^2 \quad (24)$$

4.3 平面モデル

真の位置は xy 平面上にある、すなわちまったく未知であるというモデルを S_2 とする (図 2(c))。このモデルの次元は 2、余次元は 0、自由度は 0 である。各データ点の真の位置の最尤推定量は明らかにデータ点自身であり、残差平方和は 0 である。したがって幾何学的 AIC は次のようになる。

$$AIC_0(S_2) = 4N\epsilon^2 \quad (25)$$

式 (18), (25) より、 $AIC_0(S_1) < AIC_0(S_2)$ となるのは

$$S(\hat{S}_1) < 2(N-2)\epsilon^2 \quad (26)$$

であるから、このときの N 個の点が同一直上にあるというモデルが支持される。2 個の点 ($N=2$) のときは常に直線モデルが棄却される。

5. ノイズレベルの推定とモデルの強さ

前節での例題でもわかるように、幾何学的 AIC によるモデル選択では誤差の共分散行列に関する知識が必要である。しかし、現実の画像やセンサから得られたデータの精度をあらかじめ正確に推定することは実際上困難である。一方、誤差の傾向 (一樣であるとか等方であるとか、ある方向に出やすいなど) を定性的に予想するの可能であることが多い。そこで各データの共分散行列 $V[\mathbf{r}_\alpha]$ が定数倍を除いて既知であると仮定する。すなわち、正規化共分散行列 $V_0[\mathbf{r}_\alpha]$ は既知であるが、ノイズレベル ϵ は未知であるとする。

モデル S が真であれば、ノイズレベル ϵ が次のように推定できる。モデル S が真のとき、その最尤推定量を \hat{S} とすると、 $S(\hat{S})/\epsilon^2$ が自由度 $rN-d$ の χ^2 に従うことから [7, 8, 9]

$$\epsilon_S^2 = \frac{S(\hat{S})}{rN-d} \quad (27)$$

が ϵ^2 の不偏推定量になっている ($r = n - m$ はモデルの余次元)。その分散は次のようになる。

$$V[\epsilon_S^2] = \frac{\epsilon^4}{rN-d} \quad (28)$$

したがって $rN-d$ が大きければ、式 (27) は ϵ^2 のよい近似を与える期待できる。一方 $rN-d$ が小さいと残差平方和 $S(\hat{S})$ に含まれる誤差の情報が少ない。これ次のように解釈できる。

誤差はあらゆる方向に発生するが、多様体 S の「面内」に誤差が発生しても推定量は不変であり、残差平方和 $S(\hat{S})$ は変化せず、「面外」に発生して初めて $S(\hat{S})$ が増加する。これに r 次元あり、合計 rN の自由度がある。しかし、 \hat{S} は残差を最小にするように推定したものであり、 d 個のパラメータを調節して d 自由度に相当する誤差を 0 にしているから⁸、結局 $rN-d$ 自由度が残る。いい換えれば、 $rN-d$ が大きいほど強い制約を課すことになり、当てはめは難しくなるとともに、モデルが真であれば誤差が多く自由度をもち、誤差の特徴がとらえやすくなるのである。以上のことから

$$p(S) = rN - d \quad (29)$$

をモデル S の強さと定義する。 p が小さいほど「弱い」モデルである。モデル S' がモデル S より強いことを次のように書く。

$$S' \succ S \quad (30)$$

【例 2】前節の例ではモデル $S'_0, S_0, S'_1, S_1, S_2$ の強さはそれぞれ $2N, 2N-2, N-1, N-2, 0$ であり、 N が大きいときは $S'_0 \succ S_0 \succ S'_1 \succ S_1 \succ S_2$ である。すなわち真の位置は原点であるというモデル S'_0 が最も強く、真の位置に何の制約もないというモデル S_2 が最も弱い。

6. 幾何学的モデルの比較検定

式 (27) によってノイズレベル ϵ を推定しても、これを式 (7) の ϵ に代入するのは意味がない。なぜなら、モデルを知るには誤差を知らなければならないし、誤差を知るにはモデルを知らなければならないし、両方が未知なら何も判定できないからである。しかし、個々のモデルの評価はできなくても、二つのモデルの比較は可能である。いま二つのモデル S と S' が与えられたとする。それぞれの次元、余次元、自由度を m, r, d および m', r', d' とし、 $S' \succ S$ とする。そして S は弱いモデルであって、一般に正しいと考えられるものとする。一方 S' はモデル S にさらに制約を加えて得られた強いモデルであるとし、それが成立するかどうかを調べたいとする。このとき弱いモデル S をより強いモデル S' に置き換えてもよいかの検定を次のように行なう。ただし $p(S) > 0$ とする。

⁸ 例えは平面上の直線の上の点は直線に沿う 1 自由度と直線から離れる 1 自由度をもつから、2 点では 2 自由度の残差が生じるはずであるが、その 2 点を通る直線を当てはめると残差が 0 になる。これは直線のもつ 2 自由度を調節してその 2 点のもつ 2 自由度の誤差を 0 にしたためである。

モデル S のもとではノイズレベル ϵ が式 (27) によって推定できる。これを用いると、モデル S, S' の幾何学的 AIC が次のように評価できる。

$$AIC_0(S) = S(\hat{S}) + \frac{2(mN + d)}{rN - d} S(\hat{S}) \quad (31)$$

$$AIC_0(S') = S(\hat{S}') + \frac{2(m'N + d')}{rN - d} S(\hat{S}') \quad (32)$$

もし $AIC_0(S') < AIC_0(S)$ であれば、モデル S をより強いモデル S' に置き換えることに意味があると考えられる。この条件は次のように書ける。

$$\frac{S(\hat{S}')}{S(\hat{S})} < 1 + \frac{2(m - m')N + 2(d - d')}{rN - d} \quad (33)$$

これは $S(\hat{S}') < S(\hat{S})$ より弱い条件である。モデル S, S' のもとでの式 (27) による ϵ^2 の不偏推定量をそれぞれ $\hat{\epsilon}_S^2, \hat{\epsilon}_{S'}^2$ とすると、式 (33) は次のようにも書ける。

$$\frac{\hat{\epsilon}_{S'}^2}{\hat{\epsilon}_S^2} < \frac{(r + 2(m - m'))N + d - 2d'}{r'N - d'} \quad (34)$$

この比較検定の意味は統計学の仮説検定の否定的な意味とは逆であり、肯定的意味である。仮説検定においては強い仮説 (帰無仮説) が真だとすると、弱い仮説 (対立仮説) で説明できることが説明できているとはいえない、したがって強い仮説を採用する積極的な理由がないという意味である。ここでは逆に、弱いモデルが真だとすると、それで説明できることは強いモデルでも説明できる、すなわち強いモデルを採用する積極的な理由があるということを意味する。このとき、ノイズの強さを知る必要がないだけでなく、通常に仮説検定にみられる恣意的な有意水準⁹を必要としない。

【例 3】 5 節のモデル S_0, S_1 を比較する。これは、点の真の位置が同一直線上にあると思われるとき、それらが同一点かどうかを検定する問題である。式 (33) から、次の条件が成立すればモデル S_0 に取り替えることができる。

$$\frac{S(\hat{S}_0)}{S(\hat{S}_1)} < 3 + \frac{1}{N - 2} \quad (35)$$

すなわち、一点を当てはめた残差平方和が直線当てはめの残差平方和の約 3 倍以内であれば、一点とみなしたほうが予測能力の意味ですぐれていると考えられる。

【例 4】 例 1 の二つのモデルを比較する。空間中の点の真の位置が平面上にあるモデル S_{Π} と、それらが同一直線上にあるモデル S_L を考える。式 (33) から、モデル S_{Π} が正しいとき、次の条件が成立すればモデル S_L に取り替えることができる。

$$\frac{S(\hat{S}_L)}{S(\hat{S}_{\Pi})} < 3 + \frac{4}{N - 2} \quad (36)$$

⁹ いわゆる神聖な 5%。

すなわち、直線を当てはめた残差平方和が平面当てはめの残差平方和の約 3 倍以内であれば、一直線とみなしたほうが予測能力の意味ですぐれていると考えられる。

7. まとめ

本論文ではロボットが画像やセンサデータから 3 次元環境のモデルを構築するときの幾何学的モデルの選択の基準として幾何学的 AIC を導入した。そして、通常の統計学の取り扱いと異なり、局外母数がデータの数だけ存在し、モデルが多様体によって指定されることから、モデルの複雑さがパラメータ数 (自由度) のみで評価されるのではなく、その多様体の次元や余次元という座標系の変換に不変な量が本質的な役割を果たすことを示した。また、恣意的な有意水準を用いないでモデルのよさを相対的に評価する比較検定的方式を提案した。これらを単純な平面上の点データに対する例題で説明した。

参考文献

- [1] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automation Control*, 19 (1974), 176-723.
- [2] 赤池弘次, 情報量基準 AIC とは何か, 数理科学, 153 (1976), 716-723.
- [3] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993.
- [4] K. Kanatani, *Group-Theoretical Methods in Image Understanding*, Springer, Berlin, F.R.G., 1990.
- [5] 金谷健一, 「画像理解 — 3次元認識の数理 —」, 森北出版, 1990.
- [6] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford Univ. Press, Oxford, 1993.
- [7] 金谷健一, パラメトリック当てはめの精度の理論的限界, 情報処理学会研究報告, 94-CV-91 (1994-11).
- [8] 金谷健一, パラメトリック当てはめの最尤推定, 情報処理学会研究報告, 94-CV-91 (1994-11).
- [9] 金谷健一, 幾何学的補正問題の精度の理論限界と最適計算法, 情報処理学会研究報告, 95-CV-94 (1995-5).
- [10] 金谷健一, 「空間データの数理 — 3次元コンピューティングに向けて —」, 朝倉書店, 1995.
- [11] 金谷健一, 幾何学的モデル選択の情報量基準, 電子情報通信学会技術研究報告, IT95 (1995-7).
- [12] S. Maybank, *Theory of Reconstruction from Image Motion*, Springer, Berlin, 1993.
- [13] 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.
- [14] 山西健司, 韓大舜, MDL 入門: 情報理論の立場から, 人工知能学会誌, 7-3 (1992), 427-434.
- [15] J. Weng, T. S. Huang and N. Ahuja, *Motion and Structure from Image Sequences*, Springer, Berlin, 1993.