Memoirs of the Faculty of Engineering, Okayama University, Vol. 42, pp. 10–17, January 2008

Geometric BIC

Kenichi KANATANI*
Department of Computer Science, Okayama University
Okayama 700-8530 Japan

(Received November 14, 2007)

The author introduced the "geometric AIC" and the "geometric MDL" as model selection criteria for geometric fitting problems. These correspond to Akaike's "AIC" and Rissanen's "BIC", respectively, well known in the statistical estimation framework. Another criterion well known is Schwarz' "BIC", but its counterpart for geometric fitting has been unknown. This paper introduces the corresponding criterion, which we call the "geometric BIC", and shows that it is of the same form as the geometric MDL. We present the underlying logical reasoning of Bayesian estimation.

1. Introduction

The basic principle of computer vision is to assume a certain structure, or a *model*, in the observed scene, such as certain objects being there, and to do inference by extracting characteristics of the assumed structure from observed images, estimating such properties of the scene as categories, numbers, sizes, shapes, and the positions and orientations. However, we sometimes do not know what the model should be. In such a case, selecting an appropriate model from multiple candidates is called *model selection*.

If the model has a form of standard statistical estimation such as regression, various types of *model selection criteria* have been proposed. The best known are Akaike's *AIC* (*Akaike Information Criterion*) [1], Schwarz' *BIC* (*Bayesian Information Criterion*) [18], and Rissanen's *MDL* (*Minimum Description Length*) [17].

However, geometric inference for computer vision, typically structure from motion, does not have the standard form of statistical estimation [3, 9, 10]. For this, the author introduced the *geometric AIC* [3, 7] and *geometric MDL* [9], which correspond to Akaike's AIC and Rissanen's MDL in the traditional statistical estimation framework.

The main motivation of traditional statistical estimation is to do precise inference using a large but limited number of data, while the main goal of geometric inference is to do precise but robust estimation that can tolerate noise [3, 9, 10]. This is a sort of "dual" relationship. Hence, while the AIC and the MDL are derived from asymptotic analysis with respect to the number N of data, the geometric AIC and the geometric MDL are derived from perturbation analysis with respect to the noise level ε [3, 9, 10].

Then, a question arises. What corresponds to Schwarz' BIC? The BIC is also derived from asymptotic analysis with respect to the number N of data. What criterion results if we do perturbation analysis with respect to the noise level ε?

It has already been conjectured [9] that because the BIC and the MDL have the same form up to higher order terms in $1/\sqrt{N}$, the geometric AIC and the "geometric BIC" should have the same form up to higher order terms in ε. However, the concrete form has not been shown.

In this paper, we present a rigorous derivation of the geometric BIC based on Schwarz' BIC principle and confirm that it indeed has the same form as the geometric MDL. This illuminates the Bayesian logic of model selection for geometric estimation.

First, we briefly summarize Akaike's AIC, Schwarz' BIC, and Rissanen's MDL in Section 2 and the geometric AIC and the geometric MDL in Section 3. Then, we describe the mathematical framework of geometric fitting in Section 4. Section 5 is the main part of this paper, where we derive the geometric BIC. We discuss its applications in Section 6 and conclude in Section 7.

2. AIC, BIC, and MDL

We first give a brief review of the AIC, the BIC, and the MDL.

*E-mail kanatani@suri.it.okayama-u.ac.jp

2.1 Statistical Inference and Model Selection

A probability density $p(\boldsymbol{x}|\boldsymbol{\theta})$ parameterized by unknown $\boldsymbol{\theta}$ is called a (*statistical*) *model*. The goal of *statistical estimation* is to estimate $\boldsymbol{\theta}$ from multiple data $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ assumed to be independently sampled from $p(\boldsymbol{x}|\boldsymbol{\theta})$. *Maximum likelihood* (*ML*) is to find the value of $\boldsymbol{\theta}$ that maximizes the likelihood $\prod_{\alpha=1}^{N} p(\boldsymbol{x}_\alpha|\boldsymbol{\theta})$. When we have multiple candidate models[1] $p_1(\boldsymbol{x}|\boldsymbol{\theta})$, ..., $p_M(\boldsymbol{x}|\boldsymbol{\theta})$, (*statistical*) *model selection* is to find the most appropriate one from among them.

2.2 Statistical Model Selection Criteria

The AIC, the BIC, and the MDL have the forms

$$\text{AIC} = -2\sum_{\alpha=1}^{N} \log p(\boldsymbol{x}_\alpha|\hat{\boldsymbol{\theta}}) + 2k, \tag{1}$$

$$\text{BIC} = -\sum_{\alpha=1}^{N} \log p(\boldsymbol{x}_\alpha|\hat{\boldsymbol{\theta}}) + \frac{k}{2}\log N, \tag{2}$$

$$\text{MDL} = -\sum_{\alpha=1}^{N} \log p(\boldsymbol{x}_\alpha|\hat{\boldsymbol{\theta}}) + \frac{k}{2}\log N, \tag{3}$$

where k is the degree of freedom of the model (= the dimension of $\boldsymbol{\theta}$) and $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta}$ obtained by assuming that model. These criteria are computed for each candidate model, and the one that has the smallest value is adopted as the most appropriate.

AIC. Akaike's AIC principle is to choose the model that is the closest to the true model measured in the *Kullback-Leibler* (*KL*) *distance* (or *divergence*) [1]. Since the true model is unknown, it is approximated by $p(\boldsymbol{x}|\hat{\boldsymbol{\theta}})$, plugging in the ML estimator $\hat{\boldsymbol{\theta}}$. This is justified when the number N of data is large (*consistency* of ML). Since the KL distance is defined by expectation involving the true model, it is approximated by summation over the data, which is justified when N is large. However, if we use the *same* data for computing the ML estimator $\hat{\boldsymbol{\theta}}$ and approximating the expectation, mutual correction gives rise to statistical bias. Akaike [1] estimated the bias by doing asymptotic expansion, assuming that N is large, and omitting high order terms in $1/\sqrt{N}$. Subtracting it from the estimate of the KL distance, he obtained his AIC in the form of Eq. (1), excluding model-independent terms.

BIC. Schwarz' BIC principle is to assume an a prior probability of the model, evaluate the a posteriori probability using the Bayes rule, and choose the model that has the largest value of it. Schwarz [18] assumed equal prior probabilities for the candidate models and analyzed asymptotic expansion of the (logarithmic) posterior, noting that the distribution of $\boldsymbol{\theta}$ concentrates on a small neighborhood of the ML estimator $\hat{\boldsymbol{\theta}}$ when N is large. Omitting higher order terms in $1/\sqrt{N}$ and excluding model-independent terms, he obtained his BIC in the form of Eq. (2) independent of the a priori probability of $\boldsymbol{\theta}$.

MDL. Rissanen's MDL principle is to choose the model that gives the shortest description when it is optimally encoded along with the data [17]. According to information theory, the data $\{\boldsymbol{x}_\alpha\}$ are optimally encoded using its occurrence probability $p(\boldsymbol{x}|\boldsymbol{\theta})$, but since the true value $\boldsymbol{\theta}$ is unknown, the ML estimator $\hat{\boldsymbol{\theta}}$ is substituted for it. Since the data $\{\boldsymbol{x}_\alpha\}$ and the ML estimator $\hat{\boldsymbol{\theta}}$ are both real numbers, which require an infinitely long description length, they are quantized into discrete values, and the quantization width is determined so that the resulting code length is the shortest. As the model (i.e., $\hat{\boldsymbol{\theta}}$) is better approximated, the code length of the data $\{\boldsymbol{x}_\alpha\}$ becomes shorter and approaches the information theoretical limit. However, the description of the model requires a larger code length for that. Rissanen [17] evaluated their optimal balance, analyzed its asymptotic expansion, omitting higher order terms in $1/\sqrt{N}$, and obtained his MDL in the form of Eq. (3), excluding model-independent terms[2].

3. Geometric AIC and Geometric MDL

Here, we give a brief review of the geometric AIC and the geometric MDL.

3.1 Geometric Model Selection

Given N data $\{\boldsymbol{x}_\alpha\}$, *geometric fitting* is the problem of estimating the law that constrains their true values $\{\bar{\boldsymbol{x}}_\alpha\}$ in the form of an "implicit" equation

$$\boldsymbol{F}(\boldsymbol{x};\boldsymbol{u}) = \boldsymbol{0}, \tag{4}$$

parameterized by unknown $\boldsymbol{u}$. Equation (4) is called the (*geometric*) *model*.

Many computer vision problems fall in this category. The model in the form of Eq. (4) may describe curves or shapes in the image or relationships among multiple images such as the epipolar constraint. By estimating the parameter $\boldsymbol{u}$ (e.g., coefficients of equations, and the fundamental matrix for two images) so that Eq. (4) fits the data $\{\boldsymbol{x}_\alpha\}$ well, we can infer the structure of the scene or its motion [2].

[1] The same symbol $\boldsymbol{\theta}$ is used for the convenience of description, but it may have a different dimension from model to model.

[2] Eq. (3) is a crude approximation. A more detailed form involves integration involving the Fisher information matrix $\boldsymbol{I}(\boldsymbol{\theta})$ [17].

When we have multiple candidate models[3] $\boldsymbol{F}_1(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0}$, ..., $\boldsymbol{F}_M(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0}$, (*geometric*) *model selection* is to find the most appropriate one from among them.

3.2 Geometric Model Selection Criteria

The author introduced the following geometric AIC and the geometric MDL [3, 7, 9]:

$$\text{G-AIC} = \hat{J} + 2(Nd + p)\varepsilon^2, \tag{5}$$

$$\text{G-MDL} = \hat{J} - (Nd + p)\varepsilon^2 \log \varepsilon^2. \tag{6}$$

Here, $\hat{J}$ is the residual (the sum of squares of the Mahalanobis distances) of the fitted model from the data $\{\boldsymbol{x}_\alpha\}$, d is the dimension of the manifold defined by the model, p is the degree of freedom of the model (= the dimension of $\boldsymbol{u}$), and ε is the noise level. Their precise definitions are given later.

Geometric AIC. The geometric AIC is derived from Akaike's AIC principle, minimizing the KL distance of the candidate model from the true model. Since the true model is unknown, we replace the true values $\{\bar{\boldsymbol{x}}_\alpha\}$ and the unknown $\boldsymbol{u}$ by their ML estimators $\{\hat{\boldsymbol{x}}_\alpha\}$ and $\hat{\boldsymbol{u}}$, respectively, evaluating the resulting bias, and subtracting it. The only difference from the AIC is that while the AIC is based on the asymptotic expansion in $1/\sqrt{N}$, the geometric AIC is obtained by perturbation expansion in the noise level ε. The integration for evaluating the KL distance is approximated by summation over data. This is justified because the true values $\{\bar{\boldsymbol{x}}_\alpha\}$ are very close to their ML estimators $\{\hat{\boldsymbol{x}}_\alpha\}$ when ε is small. Omitting higher order terms in ε and excluding model-independent terms, we obtain the geometric AIC in the form of Eq. (5) [3, 7].

Geometric MDL. The geometric MDL is derived from Rissanen's MDL principle, minimizing the description length of both the data and the model when optimally encoded. The data $\{\boldsymbol{x}_\alpha\}$ and the ML estimators $\{\hat{\boldsymbol{x}}_\alpha\}$ and $\hat{\boldsymbol{u}}$ are quantized, and the quantization width is determined so that the resulting code length is minimized. The only difference from the MDL is that while the MDL is based on the the asymptotic expansion in $1/\sqrt{N}$, the geometric MDL is obtained by perturbation expansion in the noise level ε. Omitting higher order terms in ε and excluding model-independent terms, we obtain the geometric MDL in the form of Eq. (3) [9].

One problem is that Eq. (6) involves logarithm of ε, which has the dimension of length (in pixels). This anomaly is caused by our crude order comparison: the scale factor to divide ε for canceling the dimensionality is separated due to the additivity of the logarithm and discarded[4], because it increases less rapidly than $O(\log \varepsilon)$ as $\varepsilon \approx 0$. This anomaly could be compensated for if higher order terms in ε were included, but that would cause much complication. A realistic compromise is to introduce a typical reference length L, such as the image size, and replace $\log \varepsilon^2$ by $\log(\varepsilon/L)^2$. No practical problem arises by that [9].

4. Geometric Fitting

We now describe a mathematical framework in which the geometric BIC is derived.

4.1 Geometric Models

Let $\{\boldsymbol{x}_\alpha\}$, $\alpha = 1$, ..., N, be m-dimensional vector data[5], which are assumed to be purturbed from their true values $\{\bar{\boldsymbol{x}}_\alpha\}$ by independent Gaussian noise of mean $\boldsymbol{0}$ and covariance matrix

$$V[\boldsymbol{x}_\alpha] = \varepsilon^2 V_0[\boldsymbol{x}_\alpha], \tag{7}$$

where ε, which we call the *noise level*, is a noise magnitude independent of the data, and $V_0[\boldsymbol{x}_\alpha]$, which we call the *normalized covariance matrix*, is a matrix that depends only on the true value $\bar{\boldsymbol{x}}_\alpha$ but not on noise. Thus, ε is a statistical quantity, while $V_0[\boldsymbol{x}_\alpha]$ is a geometric quantity.

Let Eq. (4) be r-dimensional equation, and write componentwise as

$$F^{(k)}(\boldsymbol{x}; \boldsymbol{u}) = 0, \qquad k = 1, ..., r. \tag{8}$$

These r equations define a manifold (an algebraic variety) $\mathcal{S}$ parameterized by $\boldsymbol{u}$ in the m-dimensional space of the variable $\boldsymbol{x}$, which we call the *data space* and denote by $\mathcal{X}$. If the r equations in Eq. (8) are algebraically independent[6], the dimension of $\mathcal{S}$ is $d = m - r$. Geometric fitting is regarded as the problem of adjusting $\boldsymbol{u}$ so that the manifold $\mathcal{S}$ passes by the obseved data $\{\boldsymbol{x}_\alpha\}$ as closely as possible in the data space $\mathcal{X}$.

[3] As before, the same symbol $\boldsymbol{u}$ is used for convenience, but it may have a different dimension from model to model.

[4] This problem also arises to Rissanen's MDL: if we combine multiple data, e.g., a consecutive pair, into one, the apparent number N of data decreases, so the MDL changes its value. This effect is compensated for by higher order terms in $1/\sqrt{N}$ involving the Fisher information matrix $\boldsymbol{I}(\theta)$. See footnote 2).

[5] The following argument holds if each $\boldsymbol{x}_\alpha$ is constrained to have a smaller degree m' ($< m$) of freedom, e.g., being a unit vector. We only need to introduce degenerate covariance matrices, pseudoinverse, and projection on to the constrained space [3]. Here, for similicity, we assume that no such intrinsic constraints exist.

[6] The following argument holds if the r equations in Eq. (8) has redundancy with only r' ($< r$) being independent. We only need to introduce pseudoinverse and projection operation operators [3]. Here, for simplicity, we assume that the r equations are independent.

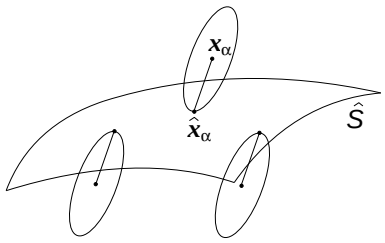

Figure 1: Fitting a manifold $\hat{\mathcal{S}}$ closest to $\boldsymbol{x}_\alpha$ measured in the Mahalanobis distance. The point $\hat{\boldsymbol{x}}_\alpha$ on it closest to $\boldsymbol{x}_\alpha$ in the Mahalanobis distance is its ML estimator. The ellipsoids represent equal probability surfaces $(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha)) = \text{constant}$.

4.2 Maximum Likelihood

From the above assumptions, the probability density of the data $\{\boldsymbol{x}_\alpha\}$ given their true values $\{\bar{\boldsymbol{x}}_\alpha\}$ and the parameter $\boldsymbol{u}$ is

$$p(\{\boldsymbol{x}_\alpha\}|\{\bar{\boldsymbol{x}}_\alpha\}, \boldsymbol{u}) = \frac{e^{-J/2\varepsilon^2}}{\sqrt{(2\pi)^{Nm}\varepsilon^{2m}|V_0[\boldsymbol{x}_\alpha]|^N}}, \quad (9)$$

where we define

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

Throughout this paper, we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$.

Equation (9) is the likelihood function if it is regarded as a function of $\{\bar{\boldsymbol{x}}_\alpha\}$ and $\boldsymbol{u}$ given the data $\{\boldsymbol{x}_\alpha\}$. The values $\{\bar{\boldsymbol{x}}_\alpha\}$ and $\boldsymbol{u}$ that maximize Eq. (9) are their ML estimators. They are the minimizer of the function J in Eq. (10) subject to

$$F^{(k)}(\bar{\boldsymbol{x}}_\alpha; \boldsymbol{u}) = 0, \quad k = 1, ..., r, \quad \alpha = 1, ..., N. \quad (11)$$

Let $\hat{J}$ be the resulting minimum value of J.

Geometrically, ML is to adjust $\boldsymbol{u}$ to fit a manifold $\hat{\mathcal{S}}$ closest to points $\boldsymbol{x}_\alpha$ in $\mathcal{X}$ measured in the sum of the square Mahalanobis distances in the form of Eq. (10) (Fig. 1). The resulting value $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$ is the ML estimator of $\boldsymbol{u}$, and the points $\hat{\boldsymbol{x}}_\alpha$ on $\hat{\mathcal{S}}$ closest to $\boldsymbol{x}_\alpha$ measured in Eq. (10) are their ML estimators

4.3 Two-Stage Estimation

We compute ML in two stages. First, we fix $\boldsymbol{u}$ and minimize Eq. (10) with respect to $\{\bar{\boldsymbol{x}}_\alpha\}$ subject to Eq. (11). Let $\tilde{J}(\boldsymbol{u})$ be the resulting minimum of Eq. (10). Next, we minimize

$$\tilde{J}(\boldsymbol{u}) = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u}), V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u}))), \quad (12)$$

with respect to $\boldsymbol{u}$; we no longer need to consider Eq. (11), which is identically satisfied by $\{\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})\}$. The value $\hat{\boldsymbol{u}}$ that minimizes Eq. (12) is the ML estimator of $\boldsymbol{u}$. The corresponding $\{\tilde{\boldsymbol{x}}_\alpha(\hat{\boldsymbol{u}})\}$ are the ML estimators of $\{\bar{\boldsymbol{x}}_\alpha\}$, and $\tilde{J}(\hat{\boldsymbol{u}})$ equals the minimum $\hat{J}$ of J.

4.4 A Posteriori Covariance Matrices

Since $\{\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})\}$ identically satisfy Eq. (11), they are constrained to be in the d-dimensional manifold $\mathcal{S}$ in the data space $\mathcal{X}$. Hence, although the original data $\{\boldsymbol{x}_\alpha\}$ have m degrees of freedom in $\mathcal{X}$, the points $\{\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})\}$ have only d degrees of freedom. They are projections of $\{\boldsymbol{x}_\alpha\}$ onto $\mathcal{S}$ defined by the Mahalanobis distance minimization. The normalized covariance matrix $V_0[\tilde{\boldsymbol{x}}_\alpha]$ of $\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})$ is the associated projection of $\boldsymbol{x}_\alpha(\boldsymbol{u})$ onto the tangent space $T_{\tilde{\mathbf{x}}_\alpha}(\mathcal{S})$ to $\mathcal{S}$ at $\tilde{\boldsymbol{x}}_\alpha$ and has the following form [3, 7]:

$$V_0[\tilde{\boldsymbol{x}}_\alpha] = V_0[\boldsymbol{x}_\alpha] - \sum_{k,l=1}^{r} W_\alpha^{(kl)} (V_0[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}} F_\alpha^{(k)})(V_0[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}} F_\alpha^{(l)})^\top. \quad (13)$$

Here, $\nabla_{\mathbf{x}} F^{(k)}$ denotes gradient of $F^{(k)}$ in Eq. (8) with respect to $\boldsymbol{x}$. The subscript α of $F^{(k)}$ means its evaluation at $\boldsymbol{x} = \boldsymbol{x}_\alpha$, and $W_\alpha^{(kl)}$ is the (kl) element of the inverse of the $r \times r$ matrix whose (kl) element is $(\nabla_{\mathbf{x}} F_\alpha^{(k)}, V_0[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}} F_\alpha^{(l)})$: we symbolically write this as

$$\left(W_\alpha^{(kl)}\right) = \left((\nabla_{\mathbf{x}} F_\alpha^{(k)}, V_0[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}} F_\alpha^{(l)})\right)^{-1}. \quad (14)$$

Note that $V_0[\tilde{\boldsymbol{x}}_\alpha]$ in Eq. (13) is an $m \times m$ matrix but has rank d $(< m)$, because it is the projection of $V_0[\boldsymbol{x}_\alpha]$ onto the d-dimensional tangent space $T_{\tilde{\mathbf{x}}_\alpha}(\mathcal{S})$ to $\mathcal{S}$.

On the other hand, the posterior covariance matrix of the ML estimator $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$ is evaluated as follows [3, 7]:

$$V[\hat{\boldsymbol{u}}] = \varepsilon^2 \hat{\boldsymbol{M}}^{-1} + O(\varepsilon^4), \quad (15)$$

$$\hat{\boldsymbol{M}} = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{r} \hat{W}_\alpha^{(kl)} \nabla_{\mathbf{u}} \hat{F}_\alpha^{(k)} \nabla_{\mathbf{u}} \hat{F}_\alpha^{(l)\top}. \quad (16)$$

Here, $\nabla_{\mathbf{u}} F^{(k)}$ denotes gradient of $F^{(k)}$ in Eq. (8) with respect to $\boldsymbol{u}$, and the subscript α of $F^{(k)}$ means evaluation at $\boldsymbol{x} = \boldsymbol{x}_\alpha$. The hats on $W_\alpha^{(kl)}$ and $F_\alpha^{(k)}$ mean substitution of $\hat{\boldsymbol{u}}$ for $\boldsymbol{u}$.

If we replace $\boldsymbol{x}_\alpha$ and $\hat{\boldsymbol{u}}$ in the expression of $\hat{\boldsymbol{M}}$ in Eq. (16) by their true values $\bar{\boldsymbol{x}}_\alpha$ and $\boldsymbol{u}$, respectively, the first term on the right-hand side of Eq. (15) gives the *KCR lower bound* on the covariance matrix of any unbiased estimator of $\boldsymbol{u}$ [3, 8, 10].

5. Geometric BIC

The above mathematical framework is the same as the one used to derive the geometric AIC and the

geometric MDL [3, 7, 9]. We now derived the geometric BIC in the same framework. The following is the originality of this paper.

5.1 A Priori and A Posteriori Probabilities

Suppose we have M models $\mathcal{M}_1$, ..., $\mathcal{M}_M$. Let $p(\mathcal{M}_i)$ be the a priori probability for the model $\mathcal{M}_i$, $p(\boldsymbol{u}|\mathcal{M}_i)$ the a priori probability density for its parameter $\boldsymbol{u}$, and $p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}, \mathcal{M}_i)$ the a priori probability density[7] of the true values $\{\bar{\boldsymbol{x}}_\alpha\}$ given $\boldsymbol{u}$.

The likelihood $p(\{\boldsymbol{x}_\alpha\}|\{\bar{\boldsymbol{x}}_\alpha\}, \boldsymbol{u}, \mathcal{M}_i)$ of the data $\{\boldsymbol{x}_\alpha\}$ given the parameter $\boldsymbol{u}$ and the true values $\{\bar{\boldsymbol{x}}_\alpha\}$ for model $\mathcal{M}_i$ is given by Eq. (9). According to the Bayes rule, the a posteriori probability $p(\mathcal{M}_i|\{\boldsymbol{x}_\alpha\})$ of model $\mathcal{M}_i$ given the data $\{\boldsymbol{x}_\alpha\}$ is given by

$$p(\mathcal{M}_i|\{\boldsymbol{x}_\alpha\}) = \iint\cdots\int p(\{\boldsymbol{x}_\alpha\}|\{\bar{\boldsymbol{x}}_\alpha\}, \boldsymbol{u}, \mathcal{M}_i) \\ p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}, \mathcal{M}_i)p(\boldsymbol{u}|\mathcal{M}_i)d\bar{\boldsymbol{x}}_1\cdots d\bar{\boldsymbol{x}}_N d\boldsymbol{u}p(\mathcal{M}_i) \\ \Big/ \sum_{i=1}^{M} p(\{\boldsymbol{x}_\alpha\}, \mathcal{M}_i), \tag{17}$$

where $p(\{\boldsymbol{x}_\alpha\}, \mathcal{M}_i)$ in the denominator is the expression in the numerator. We assume each model has the same a priori probability and choose the model that maximizes Eq. (17). Since the denominator does not depend on individual models, we choose the model that maximizes

$$L = \int e^{-J} p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})p(\boldsymbol{u})d\bar{\boldsymbol{x}}^N d\boldsymbol{u}, \tag{18}$$

where and hereafter we omit $\mathcal{M}_i$ and denote $\int\int\cdots\int d\bar{\boldsymbol{x}}_1\cdots d\bar{\boldsymbol{x}}_N d\boldsymbol{u}$ by $\int d\bar{\boldsymbol{x}}^N d\boldsymbol{u}$ to avoid notational clutter.

5.2 Expansion around ML Estimators

In order to simplify the notation, we introduce the following inner product and norm associated with the Mahalanobis distance:

$$(\boldsymbol{a}, \boldsymbol{b})_\alpha \equiv (\boldsymbol{a}, V_0[\boldsymbol{x}_\alpha]^{-1}\boldsymbol{b}), \quad \|\boldsymbol{a}\|_\alpha \equiv \sqrt{(\boldsymbol{a}, \boldsymbol{a})_\alpha}. \tag{19}$$

Then, Eq. (10) is written as

$$\begin{aligned} J &= \sum_{\alpha=1}^{N} \|\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha\|_\alpha^2 \\ &= \sum_{\alpha=1}^{N} \|(\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha) + (\tilde{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\alpha)\|_\alpha^2 \\ &= \sum_{\alpha=1}^{N} \|\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha\|_\alpha^2 + \sum_{\alpha=1}^{N} \|\tilde{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\alpha\|_\alpha^2 + \cdots, \end{aligned} \tag{20}$$

[7] Strictly, we need the subscript i for the parameter $\boldsymbol{u}$ and the functions $p(\,\cdot\,)$ and $p(\,\cdot\,|\cdots)$, because they are different from model to model. However, we omit the subscript i to avoid notational complications.

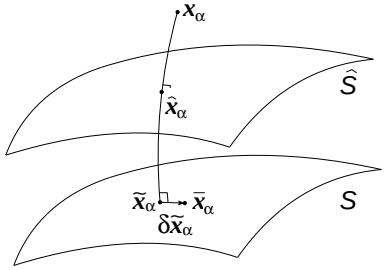

Figure 2: Measured in the Mahalanobis distance, $\tilde{\boldsymbol{x}}_\alpha$ and $\hat{\boldsymbol{x}}_\alpha$ are the closest points in the true manifold $\mathcal{S}$ and the fitted manifold $\hat{\mathcal{S}}$, respectively, from the data point $\boldsymbol{x}_\alpha$. The true position $\bar{\boldsymbol{x}}_\alpha$ is in $\mathcal{S}$.

where $\tilde{\boldsymbol{x}}_\alpha$ is a shorthand of $\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})$, and "$\cdots$" denotes higher order terms in ε. The reasoning behind Eq. (20) is as follows. Since $\tilde{\boldsymbol{x}}_\alpha$ is, by definition, the point in $\mathcal{S}$ "closest" to $\boldsymbol{x}_\alpha$ measured in the norm $\|\cdot\|_\alpha$, the displacement $\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha$ is "orthogonal" to $\mathcal{S}$ with respect to the inner product $(\,\cdot\,,\,\cdot\,)_\alpha$. Since the ML estimator $\tilde{\boldsymbol{x}}_\alpha \in \mathcal{S}$ is in the $O(\varepsilon)$ neighborhood of its true position $\bar{\boldsymbol{x}}_\alpha \in \mathcal{S}$ for small noise level ε, the deviation $\tilde{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\alpha$ is in higher order contact with $\mathcal{S}$ at $\tilde{\boldsymbol{x}}_\alpha$. Hence, Eq. (20) holds[8] (Fig. 2).

Consider the first term $\sum_{\alpha=1}^{N} \|\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha\|_\alpha^2$ ($= \sum_{\alpha=1}^{N} \|\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})\|_\alpha^2$) in the last line of Eq. (20). Letting $\boldsymbol{u} = \hat{\boldsymbol{u}} + \delta\boldsymbol{u}$, we expand the first term in $\delta\boldsymbol{u}$ around the ML estimator $\hat{\boldsymbol{u}}$. Since by definition $\hat{\boldsymbol{u}}$ minimizes this term, the first order term in $\delta\boldsymbol{u}$ vanishes. From Eq. (15), we obtain

$$\sum_{\alpha=1}^{N} \|\boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha\|_\alpha^2 = \hat{J} + (\delta\boldsymbol{u}, \hat{\boldsymbol{M}}\delta\boldsymbol{u}) + \cdots, \tag{21}$$

where "$\cdots$" denotes higher order terms in ε. This is seen as follows. Equation (15) implies that the a posteriori probability density of $\boldsymbol{u}$ should be proportional to $e^{-(\delta\mathbf{u}, V[\hat{\mathbf{u}}]^{-1}\delta\mathbf{u})} = e^{-(\delta\mathbf{u}, \hat{\mathbf{M}}\delta\mathbf{u})/2\varepsilon^2}$ except for higher order terms in ε. Hence, the likelihood in Eq. (10) should have the same expansion form.

Similarly, the second term $\|\tilde{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\alpha\|_\alpha^2$ in the last line of Eq. (20) is written from Eq. (13) in the form

$$\|\tilde{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\alpha\|_\alpha^2 = (\delta\boldsymbol{x}_\alpha, V_0[\tilde{\boldsymbol{x}}_\alpha]^{-}\delta\boldsymbol{x}_\alpha) + \cdots, \tag{22}$$

where we put $\delta\boldsymbol{x}_\alpha = \tilde{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\alpha$, and $V_0[\tilde{\boldsymbol{x}}_\alpha]^{-}$ is the pseudoinverse[9] of $V_0[\tilde{\boldsymbol{x}}_\alpha]$.

[8] This argument could be precisely formalized by rigorous order analysis, but what we need later is only the leading terms. So, higher order details are omitted.

[9] The matrix $V_0[\tilde{\boldsymbol{x}}_\alpha]$ in Eq. (13) is singular and has rank d. Its domain is the d-dimensional tangent space $T_{\tilde{\boldsymbol{x}}_\alpha}(\mathcal{S})$ to $\mathcal{S}$, whose orthogonal complement is the null space of $V_0[\tilde{\boldsymbol{x}}_\alpha]$; no deviations are allowed in it. The pseudoinverse $V_0[\tilde{\boldsymbol{x}}_\alpha]^{-}$ means the inverse operation within $T_{\tilde{\boldsymbol{x}}_\alpha}(\mathcal{S})$, preserving the same null space.

Thus, Eq. (20) has the following expansion:

$$J = \hat{J} + (\delta\boldsymbol{u}, \hat{\boldsymbol{M}}\delta\boldsymbol{u}) + \sum_{\alpha=1}^{N} (\delta\tilde{\boldsymbol{x}}_\alpha, V_0[\tilde{\boldsymbol{x}}_\alpha]^- \delta\tilde{\boldsymbol{x}}_\alpha) + ... \tag{23}$$

The reasoning we invoked above is essentially the same as that Kanatani [3, 7, 9] used to derive his geometric AIC and geometric MDL [3, 7, 9].

5.3 Expansion of A Posteriori Probability

Substituting Eq. (23) and omitting higher order terms in ε, we can write Eq. (18) as follows.

$$\begin{aligned} L &= e^{-\hat{J}/2\varepsilon^2} \int e^{-(\delta\mathbf{u}, \hat{\mathbf{M}}\delta\mathbf{u})/2\varepsilon^2} \\ &\quad \Big(\int e^{-\sum_{\alpha=1}^{N} (\delta\tilde{\mathbf{x}}_\alpha, V_0[\tilde{\mathbf{x}}_\alpha]^- \delta\tilde{\mathbf{x}}_\alpha)/2\varepsilon^2} \\ &\quad p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}) d\bar{\boldsymbol{x}}^N \Big) p(\boldsymbol{u}) d\boldsymbol{u} \\ &= e^{-\hat{J}/2\varepsilon^2} \int e^{-(\mathbf{u}-\hat{\mathbf{u}}, \hat{\mathbf{M}}(\mathbf{u}-\hat{\mathbf{u}}))/2\varepsilon^2} \\ &\quad \prod_{\alpha=1}^{N} \Big(\int e^{-(\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha, V_0[\tilde{\mathbf{x}}_\alpha]^- (\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha))/2\varepsilon^2} \\ &\quad p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}) d\bar{\boldsymbol{x}}_\alpha \Big) p(\boldsymbol{u}) d\boldsymbol{u}. \end{aligned} \tag{24}$$

The expression $e^{-(\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha, V_0[\tilde{\mathbf{x}}_\alpha]^- (\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha))/2\varepsilon^2}$ in $\bar{\boldsymbol{x}}_\alpha$ takes values close to 1 only in the $O(\varepsilon)$ neighborhood of the ML estimator $\tilde{\boldsymbol{x}}_\alpha$ and exponentially decays to 0 outside it. The a priori probability $p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})$ represents the state of our knowledge about true position of $\boldsymbol{x}_\alpha$ given $\boldsymbol{u}$, e.g., that the feature point we seek may be detected around this region in the image if the scene has the structure specified by $\boldsymbol{u}$. Hence, we may assume that $p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})$ varies smoothly around the ML estimator $\hat{\boldsymbol{x}}_\alpha$ unless specific evidence for otherwise exists. Thus, if we expand $p(\{\bar{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})$ around $\hat{\boldsymbol{x}}_\alpha$ into $p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}) + (\nabla_{\mathbf{x}} p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}), \bar{\boldsymbol{x}}_\alpha - \hat{\boldsymbol{x}}_\alpha) + \cdots$, we can ignore the second and higher order terms. Since the first order term is an odd function around $\hat{\boldsymbol{x}}_\alpha$, integration of it after multiplication by $e^{-(\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha, V_0[\tilde{\mathbf{x}}_\alpha]^- (\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha))/2\varepsilon^2}$ around $\hat{\boldsymbol{x}}_\alpha$ vanishes. The integration of the 0th order term $p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})$ is evaluated from the normalization relation of the Gaussian distribution in the form

$$\begin{aligned} &\int e^{-(\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha, V_0[\tilde{\mathbf{x}}_\alpha]^- (\bar{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\alpha))/2\varepsilon^2} p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}) d\bar{\boldsymbol{x}}_\alpha \\ &\qquad = \sqrt{(2\pi)^d \varepsilon^{2d} |V_0[\tilde{\boldsymbol{x}}_\alpha]|_+} p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}), \end{aligned} \tag{25}$$

where $|V_0[\tilde{\boldsymbol{x}}_\alpha]|_+$ denotes the product of positive eigenvalues of $|V_0[\tilde{\boldsymbol{x}}_\alpha]|$, i.e., its determinant restricted to its domain $T_{\tilde{\mathbf{x}}_\alpha}(\mathcal{S})$.

Next, we compute the product $\prod_{\alpha=1}^{N}$ of Eq. (25), multiply it by $e^{-(\delta\mathbf{u}, \hat{\mathbf{M}}\delta\mathbf{u})/2\varepsilon^2} p(\boldsymbol{u})$, and integrate the resulting expression with respect to $\boldsymbol{u}$. Again, the expression $e^{-(\mathbf{u}-\hat{\mathbf{u}}, \hat{\mathbf{M}}(\mathbf{u}-\hat{\mathbf{u}}))/2\varepsilon^2}$ in $\boldsymbol{u}$ has values close to 1 only in the $O(\varepsilon)$ neighborhood of the ML estimator $\hat{\boldsymbol{u}}$ and exponentially decays to 0 outside it. Also, $|V_0[\tilde{\boldsymbol{x}}_\alpha]|_+$, $p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})$, and $p(\boldsymbol{u})$ can be regarded as smooth functions of $\boldsymbol{u}$ around $\hat{\boldsymbol{u}}$, so their second and higher order expansion terms can be ignored. The first order terms are odd functions of $\boldsymbol{u}$ around $\hat{\boldsymbol{u}}$, so integration of them after multiplication by $e^{-(\delta\mathbf{u}, \hat{\mathbf{M}}\delta\mathbf{u})/2\varepsilon^2}$ vanishes. Hence, we only need to integrate the 0th order terms of $|V_0[\tilde{\boldsymbol{x}}_\alpha]|_+$, $p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u})$, and $p(\boldsymbol{u})$ multiplied by $e^{-(\delta\mathbf{u}, \hat{\mathbf{M}}\delta\mathbf{u})/2\varepsilon^2}$. Using the normalization relation of the Gaussian distribution, we obtain

$$\begin{aligned} &\int e^{-(\delta\mathbf{u}, \hat{\mathbf{M}}\delta\mathbf{u})/2\varepsilon^2} \prod_{\alpha=1}^{N} \sqrt{(2\pi)^d \varepsilon^{2d} |V_0[\tilde{\boldsymbol{x}}_\alpha]|_+} p(\{\hat{\boldsymbol{x}}_\alpha\}|\boldsymbol{u}) \\ &\quad \times p(\boldsymbol{u}) d\boldsymbol{u} = \sqrt{(2\pi)^p \varepsilon^{2p} |\hat{\boldsymbol{M}}|_+^{-p}} \\ &\quad \times \prod_{\alpha=1}^{N} \sqrt{(2\pi)^d \varepsilon^{2d} |V_0[\hat{\boldsymbol{x}}_\alpha]|_+} p(\{\hat{\boldsymbol{x}}_\alpha\}|\hat{\boldsymbol{u}}) + \cdots, \end{aligned} \tag{26}$$

where $|V_0[\hat{\boldsymbol{x}}_\alpha]|_+$ denotes the value of $|V_0[\tilde{\boldsymbol{x}}_\alpha]|_+$ ($= |V_0[\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{u})]|_+$) evaluated at $\boldsymbol{u} = \hat{\boldsymbol{u}}$.

The reasoning we invoked above is essentially the same as that Schwartz [18] used to derive his BIC.

5.4 Geometric BIC

Thus, Eq. (24) is evaluated except for higher order terms in ε in the following form:

$$\begin{aligned} L = e^{-\hat{J}/2\varepsilon^2} &\sqrt{(2\pi)^p \varepsilon^{2p} |\hat{\boldsymbol{M}}|_+^{-p}} p(\hat{\boldsymbol{u}}) \\ &\prod_{\alpha=1}^{N} \sqrt{(2\pi)^d \varepsilon^{2d} |V_0[\hat{\boldsymbol{x}}_\alpha]|_+} p(\{\hat{\boldsymbol{x}}_\alpha\}|\hat{\boldsymbol{u}}). \end{aligned} \tag{27}$$

Its logarithm takes the form

$$\begin{aligned} \log L = &-\frac{\hat{J}}{2\varepsilon^2} + \frac{p}{2}\log 2\pi + \frac{p}{2}\log\varepsilon^2 - \frac{p}{2}\log|\hat{\boldsymbol{M}}|_+ \\ &+ \log p(\hat{\boldsymbol{u}}) + \frac{Nd}{2}\log 2\pi + \frac{Nd}{2}\log\varepsilon^2 \\ &+ \frac{1}{2}\sum_{\alpha=1}^{N} \log|V_0[\hat{\boldsymbol{x}}_\alpha]|_+ + \sum_{\alpha=1}^{N} \log p(\{\hat{\boldsymbol{x}}_\alpha\}|\hat{\boldsymbol{u}}). \end{aligned} \tag{28}$$

The model that has the largest value of $\log L$ can be regarded as the most appropriate. Multiplying Eq. (28) by $-2\varepsilon^2$, we obtain

$$\begin{aligned} &\hat{J} - (Nd+p)\varepsilon^2 \log\varepsilon^2 \\ &+\varepsilon^2 \Big(p\log|\hat{\boldsymbol{M}}|_+ - \sum_{\alpha=1}^{N} \log|V_0[\hat{\boldsymbol{x}}_\alpha]|_+ \\ &-(Nd+p)\log 2\pi - 2\log p(\hat{\boldsymbol{u}}) \\ &-2\sum_{\alpha=1}^{N} \log p(\{\hat{\boldsymbol{x}}_\alpha\}|\hat{\boldsymbol{u}}) \Big). \end{aligned} \tag{29}$$

The last term $\varepsilon^2(\cdots)$ approaches 0 more quickly than the preceding term of $O(\varepsilon^2 \log \varepsilon^2)$ as $\varepsilon \to 0$. Omitting the last term, we obtain the *geometric BIC*

$$\text{G-BIC} = \hat{J} - (Nd + p)\varepsilon^2 \log \varepsilon^2, \tag{30}$$

which has the same form as the geometric MDL in Eq. (6). The situation corresponds to the fact that Rissanen's MDL has the same form as Schwartz' BIC as far as the leading terms are concerned, in spite of the fact that they are derived by quite different reasonings: one from information theory, the other from the Bayes rule.

Equation (30) has the same apparent anomaly as Eq. (6) in that logarithm is taken for ε^2, which has the dimension of square length. This is caused because we omit all terms of $O(\varepsilon^2)$. As in the case of the geometric MDL, this can be remedied by introducing a characteristic reference length L, such as the image size, and replacing $\log \varepsilon^2$ by $\log(\varepsilon/L)^2$, which has little effect on model selection in practice.

6. Applications

The geometric AIC and the geometric MDL have been used for model selection of various problems for computer vision, including fitting lines, curves, planes, and surfaces to 2-D and 3-D points [9], reliability evaluation of 3-D computation using a moving camera [5], detecting symmetry of 2-D shapes [4], segmenting a curve into line segments [6], inferring object shapes by stereo vision [12], moving object detection from optical flow [16], camera motion estimation for virtual studio systems [15], correspondence detection between images [14], automatic regularity enforcement on 2-D figures [20], automatic image mosaicing [13], and multibody motion segmentation [11, 19].

Almost all these applications are for *degeneracy detection*. For particular parameter values, the model degenerates and has a lower degree of freedom, or the manifold it defines has a lower dimension. For example, curves and surfaces degenerate into lines and planes if some of the coefficients vanish. Depending on the parameter values, rigid motions may degenerate into pure rotations, and projective transformations into affine transformations. If such degeneracy occurs, the computation based on a nondegenerate model may fail. For example, 3-D reconstruction fails if the assumed rigid camera motion degenerates into a pure rotation. In such a case, one needs to switch to the computation based on the degenerate model. To this, geometric model selection is called for: because the nondegenerate model always has a smaller residual than the degenerate model, the models cannot be compared by the residual alone.

For such applications, the following has been known [9].

- The geometric AIC tends to select a model that is faithful to the data. It almost always judges a nondegenerate model to be nondegenerate but sometimes judges a degenerate model to be nondegenerate.
- The geometric MDL prefers the simplicity of the model to the faithfulness to the data. It almost always judges a degenerate model to be degenerate but very often judges a nondegenerate model to be degenerate.

Hence, the choice between the geometric AIC and the geometric MDL should be based on which the user gives preference, detecting degeneracy or nondegeneracy.

Since the geometric BIC introduced in this paper has the same form as the geometric MDL, no essentially new results are obtained by it. However, we obtain a new "interpretation" that the use of the geometric MDL can be regarded also as the use of the geometric BIC, affirmatively resolving the suspicion if the geometric BIC to be defined would have the same form as the geometric MDL.

7. Conclusions

This paper has answered the question as to what Schwarz' BIC in the traditional statistical estimation framework corresponds to in the geometric fitting framework in the same way Akaike's AIC and Rissanen's MDL corresponds to the geometric AIC and the geometric MDL, respectively.

We first described the difference between the traditional statistical estimation framework and the geometric fitting framework, pointing out that the asymptotic analysis as the number N of data goes to ∞ in the former corresponds to the perturbation analysis as the noise level ε goes to 0 in the latter. Then, we introduced the Bayesian logic for geometric model selection and derived the geometric BIC using it. We found that it has the same form as the geometric MDL. We also discussed applications of geometric model selection.

References

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automation and Control*, vol. 26, no. 6, pp. 716–723, 1974.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.

[3] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, the Netherlands, 1996; Reprinted, Dover, New York, NY, U.S.A., 2005.

[4] K. Kanatani, "Comments on 'Symmetry as a Continuous Feature'," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 246–247, March 1997.

[5] K. Kanatani, "Self-evaluation for active vision by the geometric information criterion," *Proc. 7th Int'l Conf. Computer Analysis of Images and Patterns*, Kiel, Germany, pp. 247–254, September 1997.

[6] K. Kanatani, "Comments on 'Nonparametric Segmentation of Curves into Various Representations'," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1391–1392, December 1997.

[7] K. Kanatani, "Geometric information criterion for model selection," *Int'l J. Computer Vision*, vol. 26, no. 3, pp. 171–189, February/March 1998.

[8] K. Kanatani, "Cramer-Rao lower bounds for curve fitting," *Graphical Models and Image Processing*, vol. 60, no. 2, pp. 93–99, March 1998.

[9] K. Kanatani, "Uncertainty modeling and model selection for geometric inference," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1307–1319, October 2004.

[10] K. Kanatani, "Statistical optimization for geometric fitting: Theoretical accuracy analysis and high order error analysis," *Int'l J. Computer Vision*, to appear.

[11] K. Kanatani, "Motion segmentation by subspace separation: Model selection and reliability evaluation," *Int'l J. Image and Graphics*, vol. 2, no. 2, pp. 179–197, April 2002.

[12] Y. Kanazawa and K. Kanatani, "Infinity and planarity test for stereo vision," *IEICE Trans. Information and Systems*, vol. E80-D, no. 8, pp. 774–779, August 1997.

[13] Y. Kanazawa and K. Kanatani, "Stabilizing image mosaicing by model selection," M. Pollefeys, L. Van Gool, A. Zisserman, and A. Fitzgibbon Eds. *3D Structure from Images–SMILE 2000*, Springer, Berlin, 2001, pp. 35–51.

[14] Y. Kanazawa and K. Kanatani "Robust image matching preserving global consistency," *Proc. 6th Asian Conf. Computer Vision*, Jeju, Korea, vol. 2, pp. 1128–1133, January 2004.

[15] C. Matsunaga and K. Kanatani, "Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by model selection," *Proc. 7th Euro. Conf. Computer Vision*, July 2000, Dublin, Ireland, vol. 2, pp. 595–609.

[16] N. Ohta and K. Kanatani, "Moving object detection from optical flow without empirical thresholds," *IEICE Trans. Information and Systems*, vol. E81-D, no. 2, pp. 243–245, February 1998.

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

[18] G. Schwarz, "Estimating the dimension of a model,") *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, March 1978.

[19] Y. Sugaya and K. Kanatani, "Multi-stage optimization for multi-body motion segmentation," *IEICE Trans. Information and Systems*, vol. E87-D, no. 7, pp. 1935–1942, July 2004.

[20] I. Triono, N. Ohta and K. Kanatani, "Automatic recognition of regular figures by geometric AIC," *IEICE Trans. Information and Systems*, vol. E81-D, no. 2, pp. 246–248, February 1998.