Cramer-Rao Lower Bounds for Curve Fitting

Kenichi Kanatani

Department of Computer Science, Gunma University, Kiryu, Gunma 376, Japan
kanatani@cs.gunma-u.ac.jp

Abstract

We point out that the derivation of the Cramer-Rao lower bound for estimating a circular arc center and its radius by Y. T. Chan and S. M. Thomas in *Graphical Models and Image Processing*, **57**-6, 1995, 527–532, has some problems although the final result is correct. Examining the mathematical structure of the problem carefully, we first correct their mistakes and then present a suitable formulation for the problem. We show that the result can be extended to more general problems including line and conic fitting.

1 Introduction

The purpose of this paper is to point out that the derivation of the Cramer-Rao lower bound for estimating a circular arc center and its radius by Chan and Thomas [1] has some problems although the final result is correct. Examining the mathematical structure of the problem carefully, we first correct their mistakes and then present a suitable formulation to the problem. We show that the result can be extended to more general problems including line and conic fitting.

The problem Chan and Thomas [1] considered is to estimate the center and radius of a circle by observing data points $\{(x_\alpha, y_\alpha)\}$, $\alpha = 1, ..., N$, which are generally deviated from the circle to be found by noise. Formally, the problem can be stated as follows[1]:

Problem 1 *Estimate c_x, c_y, and R such that*

$$(\bar{x}_\alpha - c_x)^2 + (\bar{y}_\alpha - c_y)^2 = R^2 \tag{1}$$

for some values $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ by observing their noisy data $\{(x_\alpha, y_\alpha)\}$.

Evidently, we need some knowledge as to how the noisy data $\{(x_\alpha, y_\alpha)\}$ are related to their true values $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$. Let us write[2]

$$x_\alpha = \bar{x}_\alpha + \Delta x_\alpha, \quad y_\alpha = \bar{y}_\alpha + \Delta y_\alpha. \tag{2}$$

[1]Chan and Thomas [1] wrote $(\bar{x}, \bar{y})$ for the center of the circle, $(\tilde{x}_i, \tilde{y}_i)$ for the ith data point, and (x_i, y_i) for its true position. In order to prevent confusion in the subsequent sections, we write (c_x, c_y), (x_α, y_α), and $(\bar{x}_\alpha, \bar{y}_\alpha)$ for them, respectively. Also, we use α, β, ... to label the data, and i, j, k, ... for the indices of vectors and matrices.

[2]Chan and Thomas [1] used symbols ε_i and e_i for the noise terms, but here we use Δx_α and Δy_α for the convenience of subsequent discussions.

Chan and Thomas [1] assumed that Δx_α and Δy_α, $\alpha = 1, ..., N$, were independent Gaussian random variables of mean 0 and variance σ^2.

Chan and Thomas [1] asserted that $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ were *functions* of c_x, c_y, and R. This is clearly wrong; the αth position $(\bar{x}_\alpha, \bar{y}_\alpha)$ can be anywhere as long as it is on the circle $(x - c_x)^2 + (y - c_y)^2 = R^2$.

2 Regression vs. gometric fitting

It is not difficult to understand why Chan and Thomas [1] made an apparently wrong assertion: the traditional Cramer-Rao lower bound can be derived only for problems for which that assumption applies. To see this, let us generalize the problem a little. Let $\boldsymbol{x}_\alpha = (x_{\alpha(1)}, ..., x_{\alpha(m)})^\top$, $\alpha = 1, ..., N$, be the data, which are deviated from their true values $\bar{\boldsymbol{x}}_\alpha = (\bar{x}_{\alpha(1)}, ..., \bar{x}_{\alpha(m)})^\top$, $\alpha = 1, ..., N$, by noise, and let $\boldsymbol{u} = (u_1, ..., u_n)$ be the parameter to be estimated (the superscript $\top$ denotes transpose). In Problem 1, we have $m = 2$ and $n = 3$:

$$\boldsymbol{x}_\alpha = \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix}, \quad \bar{\boldsymbol{x}}_\alpha = \begin{pmatrix} \bar{x}_\alpha \\ \bar{y}_\alpha \end{pmatrix}, \quad \boldsymbol{u} = \begin{pmatrix} c_x \\ c_y \\ R \end{pmatrix}. \tag{3}$$

We now distinguish the following two problems:

Problem 2 *Given a vector-valued function $\boldsymbol{F}(t, \boldsymbol{u})$ and known constants $\{t_\alpha\}$, estimate the parameter $\boldsymbol{u}$ that satisfies*

$$\bar{\boldsymbol{x}}_\alpha = \boldsymbol{F}(t_\alpha, \boldsymbol{u}) \tag{4}$$

for some values $\{\bar{\boldsymbol{x}}_\alpha\}$ by observing their noisy data $\{\boldsymbol{x}_\alpha\}$.

Problem 3 *Given a function $F(\boldsymbol{x}, \boldsymbol{u})$, estimate the parameter $\boldsymbol{u}$ that satisfies*

$$F(\bar{\boldsymbol{x}}_\alpha, \boldsymbol{u}) = 0 \tag{5}$$

for some values $\{\bar{\boldsymbol{x}}_\alpha\}$ by observing their noisy data $\{\boldsymbol{x}_\alpha\}$.

In Problem 2, the data $\{\boldsymbol{x}_\alpha\}$ can be viewed as noisy observations of a "time"-varying quantity $\boldsymbol{F}(t, \boldsymbol{u})$, say economic growth, parameterized by $\boldsymbol{u}$ and sampled at discrete time $t = t_1, t_2, ..., t_N$. Known as linear/nonlinear *regression* (depending on whether $\boldsymbol{F}(t, \boldsymbol{u})$ is linear or nonlinear in $\boldsymbol{u}$), this type of estimation is a central theme in traditional statistics. Estimation of this type also plays an important role in physics; it is often referred to as the *inverse problem* [3].

Eq. (4) can also be expressed in the form

$$\boldsymbol{x}_\alpha = \boldsymbol{F}(t_\alpha, \boldsymbol{u}) + \boldsymbol{\epsilon}_\alpha, \tag{6}$$

where $\boldsymbol{\epsilon}_\alpha$ is the observation noise in the datum $\boldsymbol{x}_\alpha$; Eq. (6) is called the "model" of the regression. It is for this class of problems that the Cramer-Rao lower bound is established. And within this framework, various variants of the Cramer-Rao lower bound have been studied (e.g., the case in which the parameter $\boldsymbol{u}$ has some constraints [2]).

In contrast, Problem 3 is a *constraint satisfaction problem*, because the equation $F(\boldsymbol{x}, \boldsymbol{u}) = 0$ is usually called the *constraint*; in computer vision, such constraints play a

central role in reconstructing 2D/3D shapes from image data. Geometrically, the problem can be viewed as *fitting* a curve/surface $F(\boldsymbol{x}, \boldsymbol{u}) = 0$ to data points $\{\boldsymbol{x}_\alpha\}$ in m dimensions. Let us call such a problem *geometric fitting*. The accuracy bound for this class of problems was studied by Kanatani [7], who called it the "Cramer-Rao lower bound".

Clearly, the problem Chan and Thomas [1] posed (i.e., Problem 1) belongs to geometric fitting (i.e., Problem 3), in which case the traditional Cramer-Rao lower bound cannot be derived. So, they regarded the problem as regression (i.e., Problem 2) by introducing *artificial angle variables* $\{\phi_\alpha\}$: the data are supposed to be observed at given "angular time" $t = \phi_1, \phi_2, ..., \phi_N$. Chan and Thomas [1] then made a strange assertion (the symbols are altered into our convention):

> The three independent variables are c_x, c_y, and R and $\bar{x}_\alpha$, $\bar{y}_\alpha$ are functions of these three dependent variables. In fact, we can choose an angle variable ϕ_α and then show that ϕ_α is a function of the three independent variables c_x, c_y and R.

This is supposed to be the core of the paper of Chan and Thomas [1], but evidently this is not correct. In fact, angle ϕ_α can be *any* value; it cannot be uniquely determined as a function of c_x, c_y and R.

Perhaps, the correct assertion should be:

> We have $2N + 3$ unknowns: c_x, c_y, R, $(\bar{x}_\alpha, \bar{y}_\alpha)$, $\alpha = 1, ..., N$. But for each α, $(\bar{x}_\alpha, \bar{y}_\alpha)$ are constrained by Eq. (1). So, if we introduce angle variables $\{\phi_\alpha\}$ and *assume they are known constants*, we can express $(\bar{x}_\alpha, \bar{y}_\alpha)$ in terms of three unknowns c_x, c_y, and R.

Thus introducing artificial variables and additional assumptions, one can convert Problem 1 into the form of Problem 2 and apply the traditional theory of the Cramer-Rao lower bound. In the final results, the constants $\{\phi_\alpha\}$ are expressed in terms of the unknowns c_x, c_y, R, and $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$. This is what Chan and Thomas [1] should have supposedly done.

In this paper, we show that an accuracy bound can be derived directly for Problem 1 by regarding c_x, c_y, R, and $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ as unknowns without introducing any artificial variables and additional assumptions. We also show that the result thus obtained is identical to that obtained with artificial variables and additional assumptions.

3 Unbiasedness of regression

The principle that underlies the derivation of the traditional Cramer-Rao lower bound for Problem 2 is as follows [7]. Let $\hat{\boldsymbol{u}}$ be an estimator, i.e., a function of the data $\{\boldsymbol{x}_\alpha\}$: $\hat{\boldsymbol{u}} = \hat{\boldsymbol{u}}(\{\boldsymbol{x}_\alpha\})$. Suppose this is an *unbiased* estimator of $\boldsymbol{u}$ in the sense that

$$E[\hat{\boldsymbol{u}}] = \boldsymbol{u}, \tag{7}$$

where $E[\,\cdot\,]$ denotes expectation. Let $p_\alpha(\boldsymbol{x}_\alpha; \bar{\boldsymbol{x}}_\alpha)$ be the probability density of datum $\boldsymbol{x}_\alpha$. For example, if the noise is subject to an identical and independent Gaussian distribution with mean 0 and variance σ^2 in all components, we have $p_\alpha(\boldsymbol{x}_\alpha; \bar{\boldsymbol{x}}_\alpha) = p(\boldsymbol{x}_\alpha; \bar{\boldsymbol{x}}_\alpha)$, $\alpha = 1$, ..., N, for

$$p(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \frac{1}{(\sqrt{2\pi\sigma^2})^m} e^{-\sum_{i=1}^m (x_i - \bar{x}_i)^2 / 2\sigma^2}. \tag{8}$$

The unbiasedness condition (7) can be written as

$$\int \cdots \int \hat{\boldsymbol{u}}(\{\boldsymbol{x}_\beta\}) \prod_{\alpha=1}^{N} p_\alpha(\boldsymbol{x}_\alpha; \bar{\boldsymbol{x}}_\alpha) d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_N = \boldsymbol{u}, \tag{9}$$

where $d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_N$ denotes $\prod_{\alpha=1}^{N} \prod_{i=1}^{m} dx_{\alpha(i)}$ (by $x_{\alpha(i)}$, we denote the ith component of vector $\boldsymbol{x}_\alpha$). This unbiasedness condition should hold *irrespective of the value of* $\boldsymbol{u}$. This means that if $\boldsymbol{u}$ is perturbed into $\boldsymbol{u} + \delta\boldsymbol{u}$ in Eq. (9), the first variation (i.e., the variation obtained by omitting second and higher order terms in $\delta\boldsymbol{u}$) should be equal on both sides. Since $\bar{\boldsymbol{x}}_\alpha$ is a function of $\boldsymbol{u}$ in the form of Eq. (4), its first variation $\delta\bar{\boldsymbol{x}}_\alpha = (\delta\bar{x}_{\alpha(i)})$ has the form

$$\delta\bar{x}_{\alpha(i)} = \sum_{k=1}^{n} \frac{\partial \bar{x}_{\alpha(i)}}{\partial u_k} \delta u_k. \tag{10}$$

Hence, the first variation of the left-hand sides of Eq. (9) becomes as follows (note the identity $\partial p_\alpha / \partial(\cdot) = p_\alpha \partial \log p_\alpha / \partial(\cdot)$):

$$\int \cdots \int \hat{\boldsymbol{u}}(\{\boldsymbol{x}_\gamma\}) \sum_{\alpha=1}^{N} \sum_{i=1}^{m} \sum_{k=1}^{n} \frac{\partial \log p_\alpha}{\partial \bar{x}_{\alpha(i)}} \frac{\partial \bar{x}_{\alpha(i)}}{\partial u_k} \delta u_k \prod_{\beta=1}^{N} p_\beta d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_N. \tag{11}$$

This should be equal to the first variation of the right-hand side of Eq. (9). Hence, we obtain the fundamental equality

$$\sum_{\alpha=1}^{N} \sum_{i=1}^{m} E[\hat{u}_k \frac{\partial \log p_\alpha}{\partial \bar{x}_{\alpha(i)}}] \frac{\partial \bar{x}_{\alpha(i)}}{\partial u_l} = \delta_{kl}, \tag{12}$$

where δ_{kl} is the Kronecker delta. For the probability density in the form of Eq. (8), we have

$$E[\frac{\partial \log p_\alpha}{\partial \bar{x}_{\alpha(i)}} \frac{\partial \log p_\beta}{\partial \bar{x}_{\beta(j)}}] = \frac{1}{\sigma^2} \delta_{\alpha\beta} \delta_{ij}. \tag{13}$$

From these, the Cramer-Rao lower bound for the probability density (8) can be derived in the form $\boldsymbol{J}^{-1}$, where $\boldsymbol{J} = (J_{kl})$ is the $n \times n$ Fisher information matrix given by

$$J_{kl} = \frac{1}{\sigma^2} \sum_{\alpha=1}^{N} \sum_{i=1}^{m} \frac{\partial \bar{x}_{\alpha(i)}}{\partial u_k} \frac{\partial \bar{x}_{\alpha(i)}}{\partial u_l}. \tag{14}$$

The Cramer-Rao lower bound for a general probability density can be derived similarly.

If we attempt to apply Eq. (14) to Problem 1 directly, we are forced to evaluate the partial derivatives $\partial\bar{x}_\alpha/\partial c_x$, $\partial\bar{x}_\alpha/\partial c_y$, $\partial\bar{x}_\alpha/\partial R$, $\partial\bar{y}_\alpha/\partial c_x$, $\partial\bar{y}_\alpha/\partial c_y$, and $\partial\bar{x}_\alpha/\partial R$. But this is impossible *unless* $(\bar{x}_\alpha, \bar{y}_\alpha)$ *are functions of* c_x, c_y, *and* R.

4 Unbiasedness of geometric fitting

Consider Problem 3. Let $\hat{\boldsymbol{u}}$ be an unbiased estimator of $\boldsymbol{u}$ in the sense of Eq. (7). Since $\bar{\boldsymbol{x}}_\alpha$ is no longer a function of $\boldsymbol{u}$, we do not have an expression like Eq. (10). But note that Eq. (9) should hold for any $\{\bar{\boldsymbol{x}}_\alpha\}$ and any $\boldsymbol{u}$ *subject to the constraint* $F(\bar{\boldsymbol{x}}_\alpha, \boldsymbol{u}) = 0$. Hence, if $\bar{\boldsymbol{x}}_\alpha$ and $\boldsymbol{u}$ are perturbed in the form

$$\bar{\boldsymbol{x}}_\alpha \to \bar{\boldsymbol{x}}_\alpha + \delta\bar{\boldsymbol{x}}_\alpha, \quad \boldsymbol{u} \to \boldsymbol{u} + \delta\boldsymbol{u}, \tag{15}$$

the first variations on both sides of Eq. (9) should be equal for any variations $\delta\bar{\boldsymbol{x}}_\alpha$ and $\delta\boldsymbol{u}$ *subject to*

$$\sum_{i=1}^{m} \frac{\partial F}{\partial \bar{x}_{\alpha(i)}} \delta \bar{x}_{\alpha(i)} + \sum_{k=1}^{n} \frac{\partial F}{\partial u_k} \delta u_k = 0. \tag{16}$$

Eq. (12) is now replaced by

$$\sum_{\alpha=1}^{N} \sum_{i=1}^{m} E[\hat{u}_k \frac{\partial \log p_\alpha}{\partial \bar{x}_{\alpha(i)}}] \delta \bar{x}_{\alpha(i)} = \delta u_k. \tag{17}$$

Combining Eqs. (16), (17), and (13), we can easily derive a lower bound for the probability density (8). Instead of pursuing the derivation for this specific problem, it is much more convenient to obtain a more general result, as we now show.

5 Generalization

Let $\{\boldsymbol{a}_\alpha\}$, $\alpha = 1, ..., N$, be N instances of an m-dimensional variable $\boldsymbol{a}$ constrained to be an m'-dimensional *manifold* $\mathcal{A} \subset \mathcal{R}^m$ ($\mathcal{R}^m$ is the m-dimensional Cartesian space); we call $\mathcal{A}$ the *data space*. We assume that each $\boldsymbol{a}_\alpha$ is a perturbed datum

$$\boldsymbol{a}_\alpha = \bar{\boldsymbol{a}}_\alpha + \Delta\boldsymbol{a}_\alpha, \tag{18}$$

where $\Delta\boldsymbol{a}_\alpha$ is a random variable subject to a (not necessarily Gaussian) probability density[3] p_α parameterized by the unknown true value $\bar{\boldsymbol{a}}_\alpha$.

Suppose the true values $\{\bar{\boldsymbol{a}}_\alpha\}$ are constrained by the following equations:

$$F^{(k)}(\bar{\boldsymbol{a}}_\alpha, \boldsymbol{u}) = 0, \quad k = 1, ..., L. \tag{19}$$

Here, $\boldsymbol{u}$ is an n-dimensional vector that parameterizes the constraint; $F^{(k)}(\boldsymbol{a}, \boldsymbol{u})$ is a smooth function of variables $\boldsymbol{a}$ and $\boldsymbol{u}$. We assume that the domain of $\boldsymbol{u}$ is not necessarily the entire n-dimensional space $\mathcal{R}^n$ but an n'-dimensional manifold $\mathcal{U} \subset \mathcal{R}^n$; we call $\mathcal{U}$ the *parameter space*. Our task is formally stated as follows:

Problem 4 *Estimate the parameter* $\boldsymbol{u}$ *that satisfies Eq. (19) for some values* $\{\bar{\boldsymbol{a}}_\alpha\}$ *by observing their noisy data* $\{\boldsymbol{a}_\alpha\}$.

We assume that the L constraint equations are not necessarily independent and the essential number of independent equations is r ($\leq L$). In other words, we assume[4] that Eq. (19) defines a manifold $\mathcal{S}$ of *codimension* r in the data space $\mathcal{A}$ parameterized by $\boldsymbol{u}$; we call $\mathcal{S}$ the (*geometric*) *model*. Problem 4 has the following geometric interpretation:

Problem 5 *Estimate the parameter* $\boldsymbol{u}$ *in such a ways that the parameterized manifold* $\mathcal{S}$ *passes through* N *points* $\{\bar{\boldsymbol{a}}_\alpha\}$ *in* $\mathcal{A}$ *by observing their noisy positions* $\{\boldsymbol{a}_\alpha\}$.

[3]Here, we are assuming that the noise is small as compared with the "curvature" of the data space $\mathcal{A}$ and that the domain of p_α can be identified with the tangent space $T_{\bar{\boldsymbol{a}}_\alpha}(\mathcal{A})$ to $\mathcal{A}$ at $\bar{\boldsymbol{a}}_\alpha$ [7].

[4]To be precise, we are assuming that each of the L constraint equations defines a manifold of "codimension 1" (a "hypersurface") in such a way that any two of such manifolds intersect "transversally" [7].

Following the logic described in the preceding section, we can obtain a lower bound as follows. First, we define the following quantities:

- The *score* is defined as the m-dimensional vector

$$\boldsymbol{l}_\alpha = \nabla_{\bar{\boldsymbol{a}}_\alpha} \log p_\alpha, \tag{20}$$

  where $\nabla_{(\cdot)}$ denotes differentiation with respect to each component of $(\cdot)$ (e.g., the above equation becomes $l_{\alpha(k)} = \partial \log p_\alpha(\boldsymbol{a}_\alpha; \bar{\boldsymbol{a}}_\alpha)/\partial \bar{a}_{\alpha(k)}$ in elements).

- The *Fisher information matrix* is define by

$$\boldsymbol{J}_\alpha = E[\boldsymbol{l}_\alpha \boldsymbol{l}_\alpha^\top]. \tag{21}$$

- The *weight matrix* $\bar{\boldsymbol{W}}_\alpha = (\bar{W}_\alpha^{(kl)})$ is defined by

$$(\bar{W}_\alpha^{(kl)}) = \left((\nabla_{\boldsymbol{a}} \bar{F}_\alpha^{(k)}, \boldsymbol{J}_\alpha^- \nabla_{\boldsymbol{a}} \bar{F}_\alpha^{(l)}) \right)^- , \tag{22}$$

  where $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ and the superscript $-$ denotes the (Moore-Penrose)generalized inverse. The above equation means that $\bar{W}_\alpha^{(kl)}$ is the (kl) element of the generalized inverse of the matrix whose (kl) element is $(\nabla_{\boldsymbol{a}} \bar{F}_\alpha^{(k)}, \boldsymbol{J}_\alpha^- \nabla_{\boldsymbol{a}} \bar{F}_\alpha^{(l)})$. The bar and the subscript α of F mean that the value is evaluated at $\boldsymbol{a} = \bar{\boldsymbol{a}}_\alpha$ (e.g., the ith element of vector $\nabla_{\boldsymbol{a}} \bar{F}_\alpha^{(k)}$ is $\partial F^{(k)}(\boldsymbol{a}, \boldsymbol{u})/\partial a_i|_{\boldsymbol{a}=\bar{\boldsymbol{a}}_\alpha}$).

- The *moment matrix* $\boldsymbol{M}$ is defined by

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} \bar{W}_\alpha^{(kl)} \left(\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} \nabla_{\boldsymbol{u}} \bar{F}_\alpha^{(k)} \right) \left(\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} \nabla_{\boldsymbol{u}} \bar{F}_\alpha^{(l)} \right)^\top , \tag{23}$$

  where $\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}}$ is the *projection matrix* onto the tangent space $T_{\boldsymbol{u}}(\mathcal{U})$ to $\mathcal{U}$ (note that the ith element of vector $\nabla_{\boldsymbol{u}} \bar{F}_\alpha^{(k)}$ is $\partial F^{(k)}(\boldsymbol{a}, \boldsymbol{u})/\partial u_i|_{\boldsymbol{a}=\bar{\boldsymbol{a}}_\alpha}$).

Let $\hat{\boldsymbol{u}}$ be an unbiased estimator of $\boldsymbol{u}$ in the sense that[5]

$$\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} E[\hat{\boldsymbol{u}} - \boldsymbol{u}] = \boldsymbol{0}. \tag{24}$$

Its covariance matrix is defined by

$$V[\hat{\boldsymbol{u}}] = E[\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} (\hat{\boldsymbol{u}} - \boldsymbol{u})(\hat{\boldsymbol{u}} - \boldsymbol{u})^\top \boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}}]. \tag{25}$$

Then, we can obtain the following inequality [7]:

$$V[\hat{\boldsymbol{u}}] \succ \boldsymbol{M}^-. \tag{26}$$

Inequality $\boldsymbol{A} \succ \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is a positive semi-definite symmetric matrix (hence element-wise inequality holds for the diagonal elements by the Sylvester theorem).

[5]The usual definition (7) cannot be used here: if the parameter space $\mathcal{U}$ is "curved", then $E[\hat{\boldsymbol{u}}] \notin \mathcal{U}$ although $\hat{\boldsymbol{u}} \in \mathcal{U}$. Eqs. (24) and (25) means we are identifying the domain of $\boldsymbol{u}$ with the tangent space to $\mathcal{U}$ by assuming that noise very small [7].

It can be proved that this inequality is tight. In fact, the *maximum likelihood estimator* obtained by minimizing

$$J[\boldsymbol{u}] = \sum_{\alpha=1}^{N} W_{\alpha}^{(kl)} F_{\alpha}^{(k)} F_{\alpha}^{(l)} \tag{27}$$

can attain the bound in the first order (i.e., ignoring higher order terms in the noise and the curvature of the model) if the probability density p_α belongs to a class called the *exponential family* [7], among which the Gaussian distribution is a typical example. In Eq. (27), $F_{\alpha}^{(k)}$ is the abbreviation of $F^{(k)}(\boldsymbol{a}_\alpha; \boldsymbol{u})$; $W_{\alpha}^{(kl)}$ is the quantity obtained by replacing the true values $\{\bar{\boldsymbol{a}}_\alpha\}$ by the data $\{\boldsymbol{a}_\alpha\}$ in the definitions of $\bar{W}_{\alpha}^{(kl)}$ and by projecting it into a matrix of rank r (the smallest r eigenvalues are replaced by 0 in the canonical form); this operation is necessary for preventing possible occurrence of numerical instability [7].

6 Circle fitting

Let us go back to Problem 1 Chan and Thomas [1] posed. Suppose the noise is Gaussian; it may be anisotropic and different from point to point. The covariance matrix of the αth point is defined by

$$E[\begin{pmatrix} (x_\alpha - \bar{x}_\alpha)^2 & (x_\alpha - \bar{x}_\alpha)(y_\alpha - \bar{y}_\alpha) \\ (y_\alpha - \bar{y}_\alpha)(x_\alpha - \bar{x}_\alpha) & (y_\alpha - \bar{y}_\alpha)^2 \end{pmatrix}] = \begin{pmatrix} \sigma_{\alpha(x)}^2 & \gamma_{\alpha(xy)} \\ \gamma_{\alpha(xy)} & \sigma_{\alpha(y)}^2 \end{pmatrix}. \tag{28}$$

Let $\hat{x}$, $\hat{y}$, and $\hat{R}$ be unbiased estimators of c_x, c_y, and R, respectively. Their covariance matrix $V[\hat{x}, \hat{y}, \hat{R}]$ is defined by

$$V[\hat{x}, \hat{y}, \hat{R}] = E[\begin{pmatrix} (\hat{x} - c_x)^2 & (\hat{x} - c_x)(\hat{y} - c_y) & (\hat{x} - c_x)(\hat{R} - R) \\ (\hat{y} - c_y)(\hat{x} - c_x) & (\hat{y} - c_y)^2 & (\hat{y} - c_y)(\hat{R} - R) \\ (\hat{R} - R)(\hat{x} - c_x) & (\hat{R} - R)(\hat{y} - c_y) & (\hat{R} - R)^2 \end{pmatrix}]. \tag{29}$$

The lower bound has the form

$$V[\hat{x}, \hat{y}, \hat{R}] \succ \begin{pmatrix} \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{x}_\alpha - c_x)^2 & \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{x}_\alpha - c_x)(\bar{y}_\alpha - c_y) & \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{x}_\alpha - c_x) R \\ \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{y}_\alpha - c_y)(\bar{x}_\alpha - c_x) & \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{y}_\alpha - c_y)^2 & \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{y}_\alpha - c_y) R \\ \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{x}_\alpha - c_x) R & \sum_{\alpha=1}^N \bar{W}_\alpha (\bar{y}_\alpha - c_y) R & \sum_{\alpha=1}^N \bar{W}_\alpha R^2 \end{pmatrix}^{-1}, \tag{30}$$

where

$$\bar{W}_\alpha = \frac{1}{\sigma_{\alpha(x)}^2 (\bar{x}_\alpha - c_x)^2 + 2\gamma_{\alpha(xy)} (\bar{x}_\alpha - c_x)(\bar{y}_\alpha - c_y) + \sigma_{\alpha(y)}^2 (\bar{y}_\alpha - c_y)^2}. \tag{31}$$

The lower bound can be attained in the first order by the maximum likelihood estimator obtained by minimizing

$$J[c_x, c_y, R] = \sum_{\alpha=1}^{N} W_\alpha \left((x_\alpha - c_x)^2 + (y_\alpha - c_y)^2 - R^2 \right)^2, \tag{32}$$

where W_α is the value of $\bar{W}_\alpha$ obtained by replacing $\bar{x}_\alpha$ and $\bar{y}_\alpha$ by x_α and y_α, respectively, in Eq. (31), i.e.,

$$W_\alpha = \frac{1}{\sigma^2_{\alpha(x)}(x_\alpha - c_x)^2 + 2\gamma_{\alpha(xy)}(x_\alpha - c_x)(y_\alpha - c_y) + \sigma^2_{\alpha(y)}(y_\alpha - c_y)^2}. \tag{33}$$

If the noise is isotropic and identical for all the points in particular, as assumed by Chan and Thomas [1], we have $\sigma^2_{\alpha(x)} = \sigma^2_{\alpha(y)} = \sigma^2$ and $\gamma_{\alpha(xy)} = 0$. Then, Eq. (31) is

$$\bar{W}_\alpha = \frac{1}{\sigma^2\left((\bar{x}_\alpha - c_x)^2 + (\bar{y}_\alpha - c_y)^2\right)} = \frac{1}{\sigma^2 R^2}. \tag{34}$$

Hence, the right-hand side of Eq. (30) reduces to

$$\sigma^2 R^2 \begin{pmatrix} \sum_{\alpha=1}^N (\bar{x}_\alpha - c_x)^2 & \sum_{\alpha=1}^N (\bar{x}_\alpha - c_x)(\bar{y}_\alpha - c_y) & \sum_{\alpha=1}^N (\bar{x}_\alpha - c_x)R \\ \sum_{\alpha=1}^N (\bar{y}_\alpha - c_y)(\bar{x}_\alpha - c_x) & \sum_{\alpha=1}^N (\bar{y}_\alpha - c_y)^2 & \sum_{\alpha=1}^N (\bar{y}_\alpha - c_y)R \\ \sum_{\alpha=1}^N (\bar{x}_\alpha - c_x)R & \sum_{\alpha=1}^N (\bar{y}_\alpha - c_y)R & NR^2 \end{pmatrix}^{-1}. \tag{35}$$

Although the appearance is slightly different, this result is identical to that given by Chan and Thomas [1]. Thus, we have shown that one can reach the same conclusion as Chan and Thomas [1] without introducing artificial variables and additional assumptions.

The lower bound can be attained in the first order by minimizing

$$\tilde{J}[c_x, c_y, R] = \frac{1}{R^2}\sum_{\alpha=1}^N \left((x_\alpha - c_x)^2 + (y_\alpha - c_y)^2 - R^2\right)^2. \tag{36}$$

Note that the weight $1/R^2$ is necessary; otherwise the circle tends to have a smaller radius, i.e., *statistical bias* [7] occurs.

7 Conic fitting

The above result can immediately be extended to fitting a conic of the form

$$Ax^2 + 2Bxy + Cy^2 + 2(Dx + Ey) + F = 0. \tag{37}$$

For notational convenience, we represent a point (x, y) by a three-dimensional vector

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{38}$$

Its covariance matrix is a singular matrix of rank 2:

$$V[\boldsymbol{x}] = E[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^\top] = \begin{pmatrix} \sigma^2_{\alpha(x)} & \gamma_{\alpha(xy)} & 0 \\ \gamma_{\alpha(xy)} & \sigma^2_{\alpha(y)} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{39}$$

Consider the 3×3 matrix

$$\boldsymbol{Q} = \begin{pmatrix} A & B & D \\ B & C & E \\ D & E & F \end{pmatrix}, \tag{40}$$

and let $\hat{\boldsymbol{Q}}$ be an unbiased estimator of it. We write

$$\hat{\boldsymbol{Q}} = \boldsymbol{Q} + \Delta\boldsymbol{Q}, \tag{41}$$

and define its *covariance tensor* by

$$\mathcal{V}[\hat{\boldsymbol{Q}}] = E[\Delta\boldsymbol{Q} \otimes \Delta\boldsymbol{Q}], \tag{42}$$

where the symbol $\otimes$ denotes tensor product (the tensor product $\boldsymbol{A} \otimes \boldsymbol{B}$ of matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$ is a quantity with four indices; its $(ijkl)$ element is $A_{ij}B_{kl}$). Then, we obtain the following inequality [10]:

$$\mathcal{V}[\hat{\boldsymbol{Q}}] \succ \left(\sum_{\alpha=1}^{N} \frac{\mathcal{P}(\bar{\boldsymbol{x}}_\alpha \otimes \bar{\boldsymbol{x}}_\alpha) \otimes \mathcal{P}(\bar{\boldsymbol{x}}_\alpha \otimes \bar{\boldsymbol{x}}_\alpha)}{4(\bar{\boldsymbol{x}}_\alpha, \boldsymbol{Q}V[\boldsymbol{x}_\alpha]\boldsymbol{Q}\bar{\boldsymbol{x}}_\alpha)} \right)^{-}. \tag{43}$$

Here, the inner product of matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$ is defined by $(\boldsymbol{A}; \boldsymbol{B}) = \sum_{i,j=1}^{3} A_{ij}B_{ij}$. The symbol $\mathcal{P} = (P_{ijkl})$ denotes the *projection tensor* defined by

$$P_{ijkl} = \delta_{ik}\delta_{jl} - Q_{ij}Q_{kl}. \tag{44}$$

The lower bound can be attained in the first order by the maximum likelihood estimator obtained by minimizing

$$J[\boldsymbol{Q}] = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{x}_\alpha, \boldsymbol{Q}\boldsymbol{x}_\alpha)^2}{4(\boldsymbol{x}_\alpha, \boldsymbol{Q}V[\boldsymbol{x}_\alpha]\boldsymbol{Q}\boldsymbol{x}_\alpha)}. \tag{45}$$

Minimization can be efficiently performed by a procedure called *renormalization* [10]. This procedure can easily be adapted to circle fitting by simply reducing the number of parameters. The statistical bias that occurs if a naive least-square method is applied to conic fitting has also been studied experimentally [10].

8 Line fitting

In order to see how our theory works, let us finally consider the simplest problem: we are to fit a line

$$Ax + By + C = 0 \tag{46}$$

to $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$. For analytical convenience, we use the three-dimensional representation (37). Let

$$\boldsymbol{n} = \begin{pmatrix} A \\ B \\ C \end{pmatrix}. \tag{47}$$

Since the scale of the vector $\boldsymbol{n}$ is indeterminate, we adopt normalization $\|\boldsymbol{n}\| = 1$, where $\|\cdot\|$ denotes the norm of a vector.

Let $\hat{\boldsymbol{n}}$ be an unbiased estimator of $\boldsymbol{n}$ in the sense that

$$E[\boldsymbol{P}_{\boldsymbol{n}}(\hat{\boldsymbol{n}} - \boldsymbol{n})] = \boldsymbol{0}, \tag{48}$$

where $\boldsymbol{P_n}$ $(= \boldsymbol{I} - \boldsymbol{n}\boldsymbol{n}^\top)$ is the projection matrix onto the plane orthogonal to $\boldsymbol{n}$. We define the covariance matrix of $\hat{\boldsymbol{n}}$ by

$$V[\hat{\boldsymbol{n}}] = E[\boldsymbol{P_n}(\hat{\boldsymbol{n}} - \boldsymbol{n})(\hat{\boldsymbol{n}} - \boldsymbol{n})^\top \boldsymbol{P_n}]. \tag{49}$$

The inequality has the following form [9]:

$$V[\hat{\boldsymbol{n}}] \succ \left(\sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{x}}_\alpha \bar{\boldsymbol{x}}_\alpha^\top}{(\boldsymbol{n}, V[\boldsymbol{x}_\alpha]\boldsymbol{n})} \right)^- . \tag{50}$$

Equality holds in the first order by the maximum likelihood estimator obtained by minimizing

$$J[\boldsymbol{n}] = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{n}, \boldsymbol{x}_\alpha)^2}{(\boldsymbol{n}, V[\boldsymbol{x}_\alpha]\boldsymbol{n})}. \tag{51}$$

If the noise is isotropic and identical for all points in particular, the covariance matrix (39) reduces to

$$V[\boldsymbol{x}_\alpha] = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{52}$$

Then, the right-hand side of Eq. (50) becomes

$$\frac{\sigma^2}{1+d^2} \begin{pmatrix} \sum_{\alpha=1}^{N} \bar{x}_\alpha^2 & \sum_{\alpha=1}^{N} \bar{x}_\alpha \bar{y}_\alpha & \sum_{\alpha=1}^{N} \bar{x}_\alpha \\ \sum_{\alpha=1}^{N} \bar{y}_\alpha \bar{x}_\alpha & \sum_{\alpha=1}^{N} \bar{y}_\alpha^2 & \sum_{\alpha=1}^{N} \bar{y}_\alpha \\ \sum_{\alpha=1}^{N} \bar{x}_\alpha & \sum_{\alpha=1}^{N} \bar{y}_\alpha & N \end{pmatrix}^- , \tag{53}$$

where d is the distance of the line from the image origin. In this case, minimizing Eq. (51) is equivalent to what is known as the (*total*) *least-squares method*, i.e., minimizing the sum of the squared distances from the data points to the line to be fitted. It can be shown by experiments that Eq. (53) is a good measure of the reliability of the line fitted to edge pixels by least squares [9] and can be used to derive a reliability measure for edges in many applications such as camera calibration [4, 5, 8]. The renormalization technique can also be used to compute an optimal fit and its reliability simultaneously [9].

9 Visualization of reliability

In their paper, Chan and Thomas [1] computed the effect of arc length on the bound. Since their bound is identical to (35), the same conclusion can be reached by our formulation. However, since Eq. (35) is a special case of a more general result, we can say much more than Chan and Thomas [1] for not only circles but also general curves.

Asymptotic expression

If the data points are edge pixels constituting a segment, they are very dense. Hence, we can approximate the summation over data points by line integral over that segment along the curve; the result depends solely on the *geometry* of that segment. Such asymptotic expressions are very useful for estimating the reliability of fitting *a priori* (i.e., before any data points are observed) [4, 5, 8].

Primary deviation pair

Let $F(\boldsymbol{x}, \boldsymbol{u}) = 0$ be the curve to be fitted. Since the bound has the form of the covariance matrix $V[\hat{\boldsymbol{u}}]$ of the estimator $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$, it defines a probability distribution in the parameter space if the distribution is approximated as Gaussian. Then, we can find the "most likely orientation" of deviation in that parameter space. This is given as the unit eigenvector $\boldsymbol{v}$ of the covariance matrix $V[\hat{\boldsymbol{u}}]$, and the corresponding eigenvalue λ is the variance in that orientation. This is similar to the "modal analysis" of vibration with many degrees of freedom familiar in physics.

It follows that if $\hat{\boldsymbol{u}}$ is the estimated parameter, the values

$$\hat{\boldsymbol{u}}^+ = \hat{\boldsymbol{u}} + \sqrt{\lambda}\boldsymbol{v}, \quad \hat{\boldsymbol{u}}^- = \hat{\boldsymbol{u}} - \sqrt{\lambda}\boldsymbol{v} \tag{54}$$

are the most typical values. Accordingly, we obtain two curves

$$F(\boldsymbol{x}, \hat{\boldsymbol{u}}^+) = 0, \quad F(\boldsymbol{x}, \hat{\boldsymbol{u}}^-) = 0, \tag{55}$$

which we call the *primary deviation pair* of the fitted curve $F(\boldsymbol{x}, \hat{\boldsymbol{u}}) = 0$. We can visualize how the decrease of the arc length pushes the primary deviation pair away from the fitted conic on both sides [9, 10].

10 Concluding remarks

We have pointed out that the derivation of the Cramer-Rao lower bound for circle fitting by Chan and Thomas [1] has some problems. Examining the mathematical structure of the problem carefully, we corrected their mistakes and presented a suitable formulation for the problem. We have shown that the result can be extended to more general problems including line and conic fitting.

The theoretical framework described here also applies to more abstract problems. For example, 3-D structure estimation from motion images can also be viewed as geometric fitting; 3-D structure can be reconstructed by "fitting" a manifold (defined by what is known as the *epipolar equation*) to noisy data points in high dimensions [7]. The general formulation of the lower bound and the maximum likelihood estimation presented here yields a statistically optimal reconstruction algorithm and a tight reliability evaluation scheme [6]. The same principle can be applied to optical flow detection [11] and 3-D reconstruction from optical flow [12].

Thus, the Cramer-Rao lower bound as formulated here plays an essential role in many applications of geometric estimation from noisy data.

This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research C(2) (No. 09680352).

References

[1] Y. T. Chan and S. M. Thomas, Cramer-Rao lower bound for estimation of a circular arc center and its radius, *Graphical Models and Image Processing*, **57**-6, 1995, 527–532.

[2] J. D. Gorman and A. O. Hero, Lower bounds for parameter estimation with constraints, *IEEE Trans. Inform. Theory*, **26**-6, 1990, 1285–1301.

[3] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences. 1. Auflage*, Friedr. Vieweg & Sohn, Wiesbaden, Germany, 1993.

[4] K. Kanatani, Statistical analysis of focal-length calibration using vanishing points, *IEEE Trans. Robotics Automation*, **8**-6, 1992, 767–775.

[5] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, U.K., 1993.

[6] K. Kanatani, Renormalization for motion analysis: Statistically optimal algorithm, *IEICE Trans. Inf. & Sys.*, **E77-D**-11, 1994, 1233–1239.

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

[8] K. Kanatani and T. Maruyama, Optimal grid pattern for focal length calibration, *Advanced Robotics*, **10**-1, 1996, 81–103.

[9] Y. Kanazawa and K. Kanatani, Optimal line fitting and reliability evaluation, *IEICE Trans. Inf. & Sys.*, **E79-D**-9, 1996, 1317–1322.

[10] Y. Kanazawa and K. Kanatani, Optimal conic fitting and reliability evaluation, *IEICE Trans. Inf. & Sys.*, **E79-D**-9, 1996, 1323–1328.

[11] N. Ohta, Optical flow detection using a general noise model, *IEICE Trans. Inf. & Sys.*, **E79-D**-7, 1996, 951–957.

[12] N. Ohta and K. Kanatani, Optimal structure-from-motion algorithm for optical flow, *IEICE Trans. Inf. & Sys.*, **E78-D**-12, 1995, 1559–1566.