PAPER

Ellipse Fitting with Hyperaccuracy

Kenichi KANATANI[†a)], ***Member***

SUMMARY For fitting an ellipse to a point sequence, ML (maximum likelihood) has been regarded as having the highest accuracy. In this paper, we demonstrate the existence of a "hyperaccurate" method which outperforms ML. This is made possible by error analysis of ML followed by subtraction of high-order bias terms. Since ML nearly achieves the theoretical accuracy bound (the KCR lower bound), the resulting improvement is very small. Nevertheless, our analysis has theoretical significance, illuminating the relationship between ML and the KCR lower bound.

key words: *ellipse fitting, maximum likelihood estimation, KCR lower bound, error analysis, hyperaccuracy correction*

1. Introduction

Circular objects in a 3-D scene are projected onto ellipses on a 2-D image, and their 3-D positions can be computed from their projections [13]. For this reason, fitting an ellipse to a point sequence is one of the first steps of various vision applications, and numerous papers have been written on this subject. They are classified into two categories:

1. How can we judge whether a sequence of edge points entirely consists of points on an ellipse or it contains other points ("outliers")?
2. How can we fit the equation of an ellipse to a sequence of points as accurately as possible?

For the first task, many algorithms have been proposed in the past [6], [10], [11], [19], [30]–[32], [36], [38]. An abundance of literature exist on the second task, too. Most of the proposed methods are based on heuristics combining voting and least squares (LS) [2], [3], [12], [21], [28], [29], [33]–[35], but there are also theoretical treatments, mainly by statisticians, treating the problem as statistical estimation [22], [23], [25], [37]. However, their major concern is the consistency and efficiency of the estimator in the asymptotic limit as the number of points increases.

A contrasting approach was presented by Kanatani [15], who generalized ellipse fitting into an abstract framework, which he called "geometric fitting". Having actual image processing in mind, he pursued fitting schemes whose accuracy rapidly increases as the noise level decreases for a fixed number of points. He asserted that such methods can tolerate larger image processing uncertainty for a desired accuracy level [18].

Manuscript received February 16, 2006.
Manuscript revised May 29, 2006.

[†]The author is with the Department of Computer Science, Okayama University, Okayama-shi, 700-8530 Japan.
a) E-mail: kanatani@suri.it.okayama-u.ac.jp

DOI: 10.1093/ietisy/e88-d.10.2653

Table 1 Accuracy of ellipse fitting methods.

low accuracy	LS, Taubin's method, etc.
high accuracy	ML (renormalization, HEIV, FNS)
hyperaccuracy	studied in this paper

In his framework, a lower bound on the covariance matrix of the estimator is obtained [15], [17]. Chernov and Lesort [4] called it the "KCR (Kanatani-Cramer-Rao) lower bound" and showed that it can be derived under a weaker assumption.

It can be shown that maximum likelihood (ML) achieves that bound except for higher order terms in the noise level [4], [15], [18]. The solution can be computed by iterative schemes such as FNS [5] and HEIV [24]. Kanatani's renormalization* [14], [15], [20] also has accuracy nearly equivalent to ML [18].

Let us call, for convenience, those methods whose accuracy is comparable to ML *high accuracy* methods and other methods (e.g., LS and Taubin' method [35]) *low accuracy* methods. In contrast, we say methods that outperform ML, if they exist, have *hyperaccuracy* (Table 1).

In the past, ML has been regarded as the most accurate method for ellipse fitting. In other domains, such as structure from motion [8], [26], [27] and fundamental matrix computation [37], methods purported to be better than ML have been shown to exist, but they are effective only *asymptotically* as the number of data points grows. Their basic strategy is to estimate the *asymptotic distribution* of the data points and make use of that knowledge. Such approaches are known as *semi-parametric models* [1].

In contrast, our approach is to do error analysis of ML and subtract high-order bias terms. Bias removal can be done by numerical resampling called *bootstrap* [7]. Cabrera and Meer [3] applied this technique to ellipse fitting, but the computational burden is too heavy to be effective. Here, we derive an *analytical* correction formula, which can be applied to any *finite* (typically very small) number of data.

The goal of this paper is to show the following:

1. There *does* exist a hyperaccurate method.
2. The accuracy gain is very small, because ML already nearly achieves the KCR lower bound.

The second fact implies that we cannot expect dramatic accuracy improvement in practice. Nevertheless, our

*The program is available from the following site: http://www.suri.it.okayama-u.ac.jp analysis has theoretical significance, illuminating the relationship between ML and the KCR lower bound.

Our analysis is not limited to ellipse fitting; it is directly applied to general quadratic curves and surfaces, in fact to any algebraic curves and surfaces in general spaces. However, we restrict ourselves to ellipses, partly because this is the most important problem in practical vision applications and partly because the effect of our method is easy to visualize for ellipses.

2. KCR Lower Bound for Ellipse Fitting

We want to fit an ellipse to N points $\{(x_\alpha, y_\alpha)\}$, $\alpha = 1, ..., N$. An ellipse is represented by

$$Ax^2+2Bxy+Cy^2+2f_0(Dx+Ey)+Ff_0^2 = 0, \quad (1)$$

where f_0 is an arbitrary scaling constant†. If we define

$$\boldsymbol{u} = \begin{pmatrix} A & B & C & D & E & F \end{pmatrix}^\top,$$
$$\boldsymbol{\xi} = \begin{pmatrix} x^2 & 2xy & y^2 & 2f_0x & 2f_0y & f_0^2 \end{pmatrix}^\top, \quad (2)$$

Equation (1) is written as

$$(\boldsymbol{u}, \boldsymbol{\xi}) = 0. \quad (3)$$

Throughout this paper, we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$. Since the magnitude of the vector $\boldsymbol{u}$ is indeterminate, we adopt normalization $\|\boldsymbol{u}\| = 1$. Geometrically, Eq. (3) describes a hyperplane in the 6-dimensional space $\mathcal{R}^6$ of the variable vector $\boldsymbol{\xi}$. The N points $\{(x_\alpha, y_\alpha)\}$, $\alpha = 1, ..., N$, can be regarded as points in $\mathcal{R}^6$ via the embedding $\boldsymbol{\xi} : \mathcal{R}^2 \to \mathcal{R}^6$ defined by the second of Eqs. (2). Thus, ellipse fitting is converted to hyperplane fitting in $\mathcal{R}^6$.

Equation (1) describes not necessarily an ellipse but also a parabola, a hyperbola, and their degeneracy (e.g., two lines), generically called a *conic*. For this reason, fitting a curve in the form of Eq. (1) is often called *conic fitting* [13]. Even if the points $\{(x_\alpha, y_\alpha)\}$ are sampled from an ellipse, the fitted equation may represent a hyperbola or other curves in the presence of large noise, and a technique for preventing that has been proposed [9]. Here, however, we do not impose any such constraints, assuming that noise is sufficiently small. We also assume that outliers have already been removed in the preceding image processing stage.

Suppose each point (x_α, y_α) is perturbed from its true position $(\bar{x}_\alpha, \bar{y}_\alpha)$ by Gaussian noise of mean 0 and standard deviation σ in each component independently. Then, the covariance matrix of $\boldsymbol{\xi}_\alpha$ has the form $4\sigma^2 V_0[\boldsymbol{\xi}_\alpha]$, where $V_0[\boldsymbol{\xi}_\alpha]$, which we call the *normalized covariance matrix*, has the following form after omitting higher order terms††in σ:

$$\begin{pmatrix} \bar{x}_\alpha^2 & \bar{x}_\alpha\bar{y}_\alpha & 0 & f_0\bar{x}_\alpha & 0 & 0 \\ \bar{x}_\alpha\bar{y}_\alpha & \bar{x}_\alpha^2+\bar{y}_\alpha^2 & \bar{x}_\alpha\bar{y}_\alpha & f_0\bar{y}_\alpha & f_0\bar{x}_\alpha & 0 \\ 0 & \bar{x}_\alpha\bar{y}_\alpha & \bar{y}_\alpha^2 & 0 & f_0\bar{y}_\alpha & 0 \\ f_0\bar{x}_\alpha & f_0\bar{y}_\alpha & 0 & f_0^2 & 0 & 0 \\ 0 & f_0\bar{x}_\alpha & f_0\bar{y}_\alpha & 0 & f_0^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (4)$$

Since $\boldsymbol{\xi}_\alpha$ has only 2 degrees of freedom (i.e., x_α and y_α), the matrix $V_0[\boldsymbol{\xi}_\alpha]$ has rank 2.

Let $\hat{\boldsymbol{u}}$ be an estimator of $\boldsymbol{u}$ obtained by some means. Its accuracy is measured by the following covariance matrix:

$$V[\hat{\boldsymbol{u}}] = E[(\boldsymbol{P}_{\mathbf{u}}\hat{\boldsymbol{u}})(\boldsymbol{P}_{\mathbf{u}}\hat{\boldsymbol{u}})^\top]. \quad (5)$$

Here, $E[\,\cdot\,]$ denotes expectation with respect to the noise in the data $\{(x_\alpha, y_\alpha)\}$, and $\boldsymbol{P}_{\mathbf{u}}$ is the projection matrix ($\boldsymbol{I}$ denotes the unit matrix)

$$\boldsymbol{P}_{\mathbf{u}} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^\top, \quad (6)$$

which projects $\hat{\boldsymbol{u}}$ onto the hyperplane orthogonal to $\boldsymbol{u}$. Since the parameter vector $\boldsymbol{u}$ is normalized to unit norm, its domain is the unit sphere $\mathcal{S}^5$ in $\mathcal{R}^6$. Following the approach of Kanatani [15], we focus on the asymptotic limit of small noise and identify the domain of the errors with the tangent hyperplane to $\mathcal{S}^5$ at $\boldsymbol{u}$. So, the error is evaluated after projecting it onto that hyperplane. Hence, the covariance matrix $V[\hat{\boldsymbol{u}}]$ is a singular matrix of rank 5.

In this setting, Kanatani [15], [18] proved that if $\boldsymbol{\xi}_\alpha$ is regarded as an independent Gaussian random variable of mean $\bar{\boldsymbol{\xi}}_\alpha$ and covariance matrix $V[\boldsymbol{\xi}_\alpha]$, the following inequality holds for an arbitrary unbiased estimator $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$:

$$V[\hat{\boldsymbol{u}}] \succ \Big(\sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big)^-. \quad (7)$$

Here, $\succ$ means that the left-hand side minus the right is positive semidefinite, and the superscript $-$ denotes the generalized inverse (of rank 5).

Chernov and Lesort [4] called the right-hand side of Eq. (7) the *KCR* (*Kanatani-Cramer-Rao*) *lower bound* and showed that it holds except for $O(\sigma^4)$ even if $\hat{\boldsymbol{u}}$ is not unbiased; it is sufficient that $\hat{\boldsymbol{u}}$ is "consistent" in the sense that $\hat{\boldsymbol{u}} \to \boldsymbol{u}$ as $\sigma \to 0$.

3. Maximum Likelihood Estimation

The best known method for solving the above problem is the *least squares* (*LS*), also known as *algebraic distance minimization*, minimizing

$$J_{\mathrm{LS}} = \sum_{\alpha=1}^{N} (\boldsymbol{u}, \boldsymbol{\xi}_\alpha)^2, \quad (8)$$

which is written as $J_{\mathrm{LS}} = (\boldsymbol{u}, \boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u})$ if we define

$$\boldsymbol{M}_{\mathrm{LS}} = \sum_{\alpha=1}^{N} \boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top. \quad (9)$$

The solution $\hat{\boldsymbol{u}}_{\mathrm{LS}}$ is the unit eigenvector of $\boldsymbol{M}_{\mathrm{LS}}$ for

†One can set $f_0 = 1$ unless the data have too large magnitudes, in which case a large value of f_0 would stabilize numerical computation.

††We confirmed by experiment that inclusion of the omitted higher order terms has no noticeable effects in our numerical results shown later.

the smallest eigenvalue. However, the LS solution $\boldsymbol{u}_{\rm LS}$ is known to have large statistical bias [15].

If $\boldsymbol{\xi}_\alpha$ is regarded as an independent Gaussian random variable of mean $\bar{\boldsymbol{\xi}}_\alpha$ and covariance matrix $V[\boldsymbol{\xi}_\alpha]$, *maximum likelihood* (*ML*) is to minimize the sum of the square Mahalanobis distances of the data points $\boldsymbol{\xi}_\alpha$ to the hyperplane to be fitted, minimizing

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

subject to the constraint $(\boldsymbol{u}, \bar{\boldsymbol{\xi}}_\alpha) = 0$, $\alpha = 1$, ..., N. We can use $V_0[\boldsymbol{\xi}_\alpha]$ instead of the full covariance matrix $4\sigma^2 V_0[\boldsymbol{\xi}_\alpha]$, because the solution is unchanged if $V_0[\boldsymbol{\xi}_\alpha]$ is multiplied by a positive constant. Introducing Lagrange multipliers for the constraint $(\boldsymbol{u}, \bar{\boldsymbol{\xi}}_\alpha) = 0$, we can reduce the problem to unconstrained minimization of the following function [5], [15], [24]:

$$J = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, \boldsymbol{\xi}_\alpha)^2}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \tag{11}$$

By differentiation with respect to $\boldsymbol{u}$, we have

$$\nabla_{\mathbf{u}} J = \sum_{\alpha=1}^{N} \frac{2(\boldsymbol{\xi}_\alpha, \boldsymbol{u})\boldsymbol{\xi}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} - \sum_{\alpha=1}^{N} \frac{2(\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \tag{12}$$

The ML estimator $\hat{\boldsymbol{u}}$ is obtained by solving $\nabla_{\mathbf{u}} J = \mathbf{0}$, or

$$\boldsymbol{M}\boldsymbol{u} = \boldsymbol{L}\boldsymbol{u}, \tag{13}$$

where we define

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} \frac{\boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}, \quad \boldsymbol{L} = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \tag{14}$$

The FNS of Chojnacki et al. [5] solves Eq. (13) by iteratively computing eigenvalue problems; the HEIV of Leedan and Meer [24] iteratively computes generalized eigenvalue problems. Kanatani's renormalization [15] also solves Eq. (13) with nearly the same accuracy [18].

4. Error Analysis of ML

Substituting $\boldsymbol{\xi}_\alpha = \bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha$ in the matrix $\boldsymbol{M}$ in Eqs. (14), we obtain

$$\boldsymbol{M} = \bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_2\boldsymbol{M}, \tag{15}$$

$$\begin{aligned} \Delta_1\boldsymbol{M} &= \sum_{\alpha=1}^{N} \frac{\Delta\boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}, \\ \Delta_2\boldsymbol{M} &= \sum_{\alpha=1}^{N} \frac{\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}, \end{aligned} \tag{16}$$

where $\bar{\boldsymbol{M}}$ is the value of the matrix $\boldsymbol{M}$ defined by the true values $\{\bar{\boldsymbol{\xi}}_\alpha\}$ of $\{\boldsymbol{\xi}_\alpha\}$. The matrix $\boldsymbol{L}$ in Eqs. (14) is written as

$$\begin{aligned} \boldsymbol{L} &= \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\ &= \sum_{\alpha=1}^{N} \frac{(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} = \Delta_2\boldsymbol{L}, \end{aligned} \tag{17}$$

so $\boldsymbol{L}$ is a second order quantity.

Letting $\boldsymbol{u}$ be the noise-free value of the solution, we expand the ML estimator $\hat{\boldsymbol{u}}$ in the form

$$\hat{\boldsymbol{u}} = \boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots, \tag{18}$$

where $\Delta_k\boldsymbol{u}$ denotes terms which contain kth powers of the components of $\Delta\boldsymbol{\xi}_\alpha$ having a magnitude of $O(\sigma^k)$. Substituting Eq. (18) into Eq. (13), we obtain

$$\begin{aligned} &(\bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M} + \Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M} + \cdots) \\ &(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots) \\ &= (\Delta_2\boldsymbol{L} + \Delta_2^*\boldsymbol{L} + \cdots)(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots), \end{aligned} \tag{19}$$

where $\Delta_1^*\boldsymbol{M}$, $\Delta_2^*\boldsymbol{M}$, and $\Delta_2^*\boldsymbol{L}$ are, respectively, the perturbation terms arising from replacing $\boldsymbol{u}$ by $\hat{\boldsymbol{u}}$ in $\boldsymbol{M}$, $\Delta_1\boldsymbol{M}$, and $\boldsymbol{L}$. They are given by

$$\Delta_1^*\boldsymbol{M} = -2\sum_{\alpha=1}^{N} \frac{(\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}, \tag{20}$$

$$\begin{aligned} \Delta_2^*\boldsymbol{M} &= -2\sum_{\alpha=1}^{N} \frac{(\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\Delta\boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top)}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\ &\quad + \sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} O(\sigma^2), \\ \Delta_2^*\boldsymbol{L} &= \sum_{\alpha=1}^{N} \frac{(\bar{\xi}_\alpha, \Delta_1\boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\ &\quad + 2\sum_{\alpha=1}^{N} \frac{(\bar{\xi}_\alpha, \Delta_1\boldsymbol{u})(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u}) V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \end{aligned} \tag{21}$$

Equating terms of $O(1)$, $O(\sigma)$, and $O(\sigma^2)$ on both sides of Eq. (19), we obtain

$$\Delta_1\boldsymbol{u} = -\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u} \tag{22}$$

$$\begin{aligned} \Delta_2\boldsymbol{u}^\perp = &-\bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{M}\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u} \\ &+ \bar{\boldsymbol{M}}^-\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u} - \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{M}\boldsymbol{u} \\ &+ \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{L}\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\boldsymbol{u}, \end{aligned} \tag{23}$$

where the superscript $\perp$ means the component orthogonal to $\boldsymbol{u}$; the component parallel to $\boldsymbol{u}$ is irrelevant, since the solution is normalized to a unit vector (Fig. 1). The derivation of Eqs. (23) is given in the Appendix.

The first order error $\Delta\boldsymbol{u}_1$ yields the variation corresponding to the KCR lower bound. In fact, we have

$$\begin{aligned} E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top] &= E[\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}\boldsymbol{u}^\top\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-] \\ &= E\Big[\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{\Delta\boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\boldsymbol{u}\boldsymbol{u}^\top \end{aligned}$$

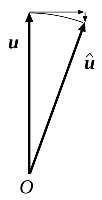


Fig. 1 The orthogonal and the parallel components of the error in $\hat{\boldsymbol{u}}$.

$$\sum_{\beta=1}^{N}\frac{\Delta\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top+\bar{\boldsymbol{\xi}}_\beta\Delta\boldsymbol{\xi}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\bar{\boldsymbol{M}}^-\Big]$$

$$=\bar{\boldsymbol{M}}^-\sum_{\alpha,\beta=1}^{N}\frac{(\boldsymbol{u},E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\bar{\boldsymbol{M}}^-$$

$$=\bar{\boldsymbol{M}}^-\sum_{\alpha=1}^{N}\frac{4\sigma^2\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\bar{\boldsymbol{M}}^-$$

$$=4\sigma^2\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^-=4\sigma^2\bar{\boldsymbol{M}}^-, \tag{24}$$

where we have used the identity† $E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]=4\sigma^2\delta_{\alpha\beta}V_0[\boldsymbol{\xi}_\alpha]$, a consequence of the noise independence.

From the definition of $\bar{\boldsymbol{M}}$ and $V_0[\boldsymbol{\xi}_\alpha]$, we can see that Eq. (24) coincides the KCR lower bound (the right-hand side of Eq. (7)). We now examine the effect of the second order error $\Delta_2\boldsymbol{u}$.

5. Bias Evaluation for ML

Since $E[\Delta\boldsymbol{\xi}_\alpha]=\boldsymbol{0}$, we have $E[\Delta_1\boldsymbol{M}]=\boldsymbol{O}$. Hence, the first order error $\Delta_1\boldsymbol{u}$ is "unbiased". So, we evaluate the bias of the second order error $\Delta_2\boldsymbol{u}$. The expectation of $\Delta_2\boldsymbol{M}$ is

$$E[\Delta_2\boldsymbol{M}]=\sum_{\alpha=1}^{N}\frac{E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}=\sum_{\alpha=1}^{N}\frac{4\sigma^2V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}$$
$$=4\sigma^2\boldsymbol{N}, \tag{25}$$

where we define

$$\boldsymbol{N}=\sum_{\alpha=1}^{N}\frac{V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \tag{26}$$

The expectation of $\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}$ is

$$E[\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}]$$
$$=E\Big[\bar{\boldsymbol{M}}^-\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top+\Delta\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\bar{\boldsymbol{M}}^-$$
$$\sum_{\beta=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\beta\Delta\boldsymbol{\xi}_\beta^\top+\Delta\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\boldsymbol{u}\Big]$$
$$=\sum_{\alpha,\beta=1}^{N}\frac{1}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}$$
$$\Big(\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha(\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta,E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})$$
$$+\bar{\boldsymbol{M}}^-E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u}(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta)\Big)$$
$$=4\sigma^2\sum_{\alpha=1}^{N}\frac{1}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}\Big((\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha,$$
$$V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha+(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{M}}^-V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}\Big). \tag{27}$$

The expectation of $\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}$ is

$$E[\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}]$$
$$=-2\sum_{\alpha=1}^{N}\frac{E[(\Delta_1\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u})]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$=-2\sum_{\alpha=1}^{N}\frac{1}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}E[(\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})$$
$$(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha]$$
$$=-2\sum_{\alpha=1}^{N}\frac{1}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}(\bar{\boldsymbol{M}}^-V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},$$
$$E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha, \tag{28}$$

where $E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]$ is evaluated as follows:

$$E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]$$
$$=E\Big[\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha(\Delta\boldsymbol{\xi}_\alpha,\boldsymbol{u})}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\sum_{\beta=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\beta^\top(\Delta\boldsymbol{\xi}_\beta,\boldsymbol{u})}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\Big]$$
$$=\sum_{\alpha,\beta=1}^{N}\frac{(\boldsymbol{u},E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top],\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}$$
$$=4\sigma^2\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}=4\sigma^2\bar{\boldsymbol{M}}. \tag{29}$$

Thus, $E[\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}]$ is finally

$$E[\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}]$$
$$=8\sigma^2\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^-V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$=8\sigma^2\sum_{\alpha=1}^{N}\frac{(V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$=-8\sigma^2\sum_{\alpha=1}^{N}\frac{(V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \tag{30}$$

The expectation of $\Delta_2^*\boldsymbol{M}\boldsymbol{u}$ is

$$E[\Delta_2^*\boldsymbol{M}\boldsymbol{u}]$$
$$=-2\sum_{\alpha=1}^{N}\frac{E[(\Delta_1\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha(\Delta\boldsymbol{\xi}_\alpha,\boldsymbol{u})]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$=2\sum_{\alpha=1}^{N}\frac{E[(\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\Delta\boldsymbol{\xi}_\alpha,\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$

†The symbol $\delta_{\alpha\beta}$ is the Kronecker delta, taking on 1 for $\alpha=\beta$ and 0 otherwise.

$$= 2\sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{M}}^{-}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}, E[\Delta_1 \boldsymbol{M}\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top \boldsymbol{u}])\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}, \quad (31)$$

where $E[\Delta_1 \boldsymbol{M}\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top \boldsymbol{u}]$ is evaluated as follows:

$$E[\Delta_1 \boldsymbol{M}\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top \boldsymbol{u}] = E\Big[\sum_{\beta=1}^{N} \frac{(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta \Delta\boldsymbol{\xi}_\alpha^\top \boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\Big]$$
$$= \sum_{\beta=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\beta(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\beta \Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}$$
$$= 4\sigma^2 \frac{\bar{\boldsymbol{\xi}}_\alpha(\boldsymbol{u}, V_0[\Delta\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = 4\sigma^2 \bar{\boldsymbol{\xi}}_\alpha. \quad (32)$$

Thus, $E[\Delta_2^* \boldsymbol{M}\boldsymbol{u}]$ is finally

$$E[\Delta_2^* \boldsymbol{M}\boldsymbol{u}] = 8\sigma^2 \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{M}}^{-}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}, \bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$= 8\sigma^2 \sum_{\alpha=1}^{N} \frac{(V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}, \bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \quad (33)$$

The expectation of $\Delta_2 \boldsymbol{L}$ is

$$E[\Delta_2 \boldsymbol{L}] = E\Big[\sum_{\alpha=1}^{N} \frac{(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}\Big]$$
$$= \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u})V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$= 4\sigma^2 \sum_{\alpha=1}^{N} \frac{V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = 4\sigma^2 \boldsymbol{N}. \quad (34)$$

The expectation of $\bar{\boldsymbol{M}}^{-}\Delta_2^* \boldsymbol{L}\boldsymbol{u}$ is

$$E[\bar{\boldsymbol{M}}^{-}\Delta_2^* \boldsymbol{L}\boldsymbol{u}]$$
$$= \bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha^\top E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top]\bar{\boldsymbol{\xi}}_\alpha V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$+2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha^\top E[\Delta_1\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}, \quad (35)$$

where $E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top]$ is evaluated as follows:

$$E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top]$$
$$= E[(\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u})(\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]$$
$$= \bar{\boldsymbol{M}}^{-}E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]\bar{\boldsymbol{M}}^{-}$$
$$= \bar{\boldsymbol{M}}^{-}E[\sum_{\alpha=1}^{N}\frac{(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\sum_{\beta=1}^{N}\frac{(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}]\bar{\boldsymbol{M}}^{-}$$
$$= \bar{\boldsymbol{M}}^{-}\sum_{\alpha,\beta=1}^{N}\frac{(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\bar{\boldsymbol{M}}^{-}$$
$$= 4\sigma^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}\bar{\boldsymbol{M}}^{-}$$
$$= 4\sigma^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\bar{\boldsymbol{M}}^{-} = 4\sigma^2\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^{-}$$
$$= 4\sigma^2\bar{\boldsymbol{M}}^{-}. \quad (36)$$

On the other hand, $E[\Delta_1\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u}$ is

$$E[\Delta_1\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u}$$
$$= -E[\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u}\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u}$$
$$= -\bar{\boldsymbol{M}}^{-}E\Big[\sum_{\beta=1}^{N}\frac{(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta\Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\Big]\boldsymbol{u}$$
$$= -\bar{\boldsymbol{M}}^{-}\sum_{\beta=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\beta(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\beta\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}$$
$$= -4\sigma^2\bar{\boldsymbol{M}}^{-}\frac{\bar{\boldsymbol{\xi}}_\alpha(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = -4\sigma^2\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha. \quad (37)$$

Thus, $E[\bar{\boldsymbol{M}}^{-}\Delta_2^*\boldsymbol{L}\boldsymbol{u}]$ is finally

$$E[\bar{\boldsymbol{M}}^{-}\Delta_2^*\boldsymbol{L}\boldsymbol{u}]$$
$$= 4\sigma^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$-8\sigma^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$
$$= -4\sigma^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \quad (38)$$

From Eqs. (25)$\sim$(38), the bias of the second order error $\Delta_2\boldsymbol{u}^\perp$ in Eq. (23) is

$$E[\Delta_2\boldsymbol{u}^\perp] = 4\sigma^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \quad (39)$$

6. Hyperaccuracy Correction

The above analysis implies that we can obtain a hyperaccurate estimator by subtracting an estimate of the bias $E[\Delta_2\boldsymbol{u}^\perp]$ from the ML estimator $\hat{\boldsymbol{u}}$ in the form

$$\tilde{\boldsymbol{u}} = N[\hat{\boldsymbol{u}} - \Delta_c\boldsymbol{u}], \quad (40)$$

where $N[\cdot]$ denotes normalization into a unit vector. The correction term $\Delta_c\boldsymbol{u}$ is given by

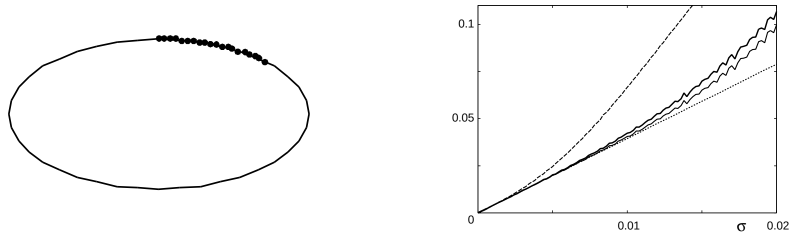
$$\Delta_c\hat{\boldsymbol{u}} = 4\hat{\sigma}^2\boldsymbol{M}^{-}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{M}^{-}\bar{\boldsymbol{\xi}}_\alpha, V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})\bar{\boldsymbol{\xi}}_\alpha}{(\hat{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})^2}, \quad (41)$$

where $\boldsymbol{u}$ and $\bar{\boldsymbol{M}}$ in Eq. (39) are replaced by the ML estimator $\hat{\boldsymbol{u}}$ and the matrix $\boldsymbol{M}$ defined from $\{\boldsymbol{\xi}_\alpha\}$, respectively. The variance σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{(\hat{\boldsymbol{u}}, \boldsymbol{M}\hat{\boldsymbol{u}})}{4(N-5)}. \quad (42)$$

7. Experiments

We defined $N = 20$ points $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ on the ellipse shown in Fig. 2(a) with equal intervals. From them, we



(a) (b)

Fig. 2 (a) 20 points on an ellipse. (b) Noise level vs. RMS error: LS (broken line), ML (thick solid line), hyperaccuracy correction (thin solid line), KCR lower bound (dotted line).

generated data points $\{(x_\alpha, y_\alpha)\}$ by adding Gaussian noise of mean 0 and standard deviation σ to the x and y coordinates independently. Then, we fitted an ellipse by different methods. For computing ML, we used the FNS of Chojnacki et al. [5].

Figure 2(b) plots for different σ the fitting error evaluated by the following RMS (root mean square) error over 10,000 independent trials:

$$D = \sqrt{\frac{1}{10000}\sum_{a=1}^{10000} \|\boldsymbol{P}_{\mathbf{u}}\hat{\boldsymbol{u}}^{(a)}\|^2}. \tag{43}$$

Here, $\hat{\boldsymbol{u}}^{(a)}$ is the ath value of $\hat{\boldsymbol{u}}$. The thick solid line is for ML; the thin solid line is the result of our hyperaccurate correction. For comparison, we also plot the LS solution $\hat{\boldsymbol{u}}_{\rm LS}$ by the broken line. The dotted line is the RMS error derived from Eq. (7):

$$D_{\rm KCR} = 2\sigma\sqrt{\mathrm{tr}\Big(\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big)^-}. \tag{44}$$

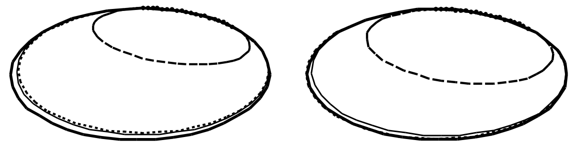
Intuitively, $[0, D]$ and $[0, D_{\rm KCR}]$ indicate the empirical and theoretical "one-sigma" ranges of the error fluctuations in all directions in $\mathcal{R}^6$ averaged.

As we can see from Fig. 2(b), LS has very low accuracy, while ML is very accurate; it almost achieves the KCR lower bound when the noise is small. As the noise increases, however, a small gap appears between the RMS error and the KCR lower bound. After adding the hyperaccurate correction, the RMS error approaches closer to the KCR lower bound†.

Figure 3(a) shows one instance of ellipse fitting (σ = 0.015). The dotted line shows the true ellipse; the broken line is for LS; the thick solid line is for ML; the thin solid line is for our hyperaccurate correction. We can see that the fitted ellipse is closer to the true shape after the correction. Figure 3(b) is another instance (σ = 0.015). In this case, the ellipse given by ML is already very accurate, and it slightly deviates from the true shape after the correction.

Thus, the accuracy sometimes improves and sometimes deteriorates. Overall, however, the cases of improvement are the majority; on average we observe

†The hyperaccuracy correction of ellipse fitting was first presented in [16], but the term $\Delta_2^*\boldsymbol{L}$ was not taken into account.



(a) (b)

Fig. 3 Two instances of ellipse fitting: LS (broken line), ML (thick solid line), hyperaccuracy correction (thin solid line), true ellipse (dotted line).

slight improvement as shown in Fig. 2(b). Closely examining many examples, we have observed that the accuracy drop occurs almost always when the ML fitted ellipse falls inside the true shape. However, the majority of the fitted ellipses are outside the true shape. Thus, the correction is effective on average.

We infer that ML is likely to produce ellipses outside the true shape because it is parameterized in the form of Eq. (1). In fact, if the major or minor axis of the ellipse is a, the coefficient of x^2 or y^2 is proportional to $1/a^2$. If $1/a^2$ is "unbiased", then a is biased to be larger than the true value, as can be easily seen from the shape of the graph of $y = 1/x^2$.

8. Conclusions

We have demonstrated for the first time the existence of "hyperaccurate" ellipse fitting which outperforms ML. This is made possible by error analysis of ML followed by subtraction of high-order bias terms. However, ML nearly achieves the KCR lower bound, meaning that even if the bias is eliminated, the solution still fluctuates with the magnitude corresponding to the KCR lower bound, which is theoretically impossible to reduce. Thus, the accuracy gain by our method is almost unnoticeable in practice, compared to which removing outliers and stabilizing the computation have far more practical significance. Nevertheless, our analysis has theoretical significance, illuminating the relationship between ML and the KCR lower bound.

Acknowledgments

The author thanks Nikolai Chernov of the University of Alabama, U.S.A., for helpful discussions. He also thanks Junpei Yamada of Okayama University, Japan, and Yasuyuki Sugaya of Toyohashi University of Technology, Japan, for doing numerical experiments. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C (No. 17500112).

References

[1] S. Amari and M. Kawanabe, "Information geometry of estimating functions in semiparametric statistical models," Bernoulli, vol.3, no.1, pp.29–54, March 1997.
[2] F.J. Bookstein, "Fitting conic sections to scattered data," Comput. Graph. Image Process., vol.9, no.1, pp.56–71, 1979.
[3] J. Cabrera and P. Meer, "Unbiased estimation of ellipses by bootstrapping," IEEE Trans. Pattern Anal. Mach. Intell., vol.18, no.7, pp.752–756, July 1996.
[4] N. Chernov and C. Lesort, "Statistical efficiency of curve fitting algorithms," Comput. Stat. Data Anal., vol.47, no.4, pp.713–728, Nov. 2004.
[5] W. Chojnacki, M.J. Brooks, A. van den Hengel, and D. Gawley, "On the fitting of surfaces to data with covariances," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.11, pp.1294–1303, Nov. 2000.
[6] D.B. Cooper and N. Yalabik, "On the computational cost of approximating and recognizing nose-perturbed straight lines and quadratic arcs in the plane," IEEE Trans. Comput., vol.C-25, no.10, pp.1020–1032, Oct. 1976.
[7] B. Efron and R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, 1994.
[8] T. Endoh, T. Toriu, and N. Tagawa, "A superior estimator to the maximum likelihood estimator on 3-D motion estimation from noisy optical flow," IEICE Trans. Inf. & Syst., vol.E77-D, no.11, pp.1240–1246, Nov. 1994.
[9] A. Fitzgibbon, M. Pilu, and R.B. Fisher, "Direct least square fitting of ellipses," IEEE Trans. Pattern Anal. Mach. Intell., vol.21, no.5, pp.476–480, May 1999.
[10] J. Illingworth and J. Kittler, "The adaptive Hough transform," IEEE Trans. Pattern Anal. Mach. Intell., vol.9, no.5, pp.690–698, Sept. 1999.
[11] D. Ioannou, W. Huda, and A.F. Laine, "Circle recognition through a 2D Hough transform and radius histogramming," Image Vis. Comput., vol.17, no.1, pp.15–26, Jan. 1999.
[12] S. H. Joseph, "Unbiased least squares fitting of circular arcs," CVGIP: Models Image Process., vol.56, no.5, pp.424–432, Sept. 1994.
[13] K. Kanatani, Geometric Computation for Machine Vision, Oxford University Press, Oxford, U.K., 1993.
[14] K. Kanatani, "Statistical bias of conic fitting and renormalization," IEEE Trans. Pattern Anal. Mach. Intell., vol.16, no.3, pp.320–326, March 1994.
[15] K. Kanatani, Statistical Optimization for Geometric Computation: Theory and Practice, Elsevier Science, Amsterdam, The Netherlands, 1996, reprinted by Dover, New York, 2005.
[16] K. Kanatani, Ellipse fitting with hyperaccuracy, Proc. 9th Euro Conf. Comput. Vision, vol.1, pp.484–495, Graz, Austria, May 2006.
[17] K. Kanatani, "Cramer-Rao lower bounds for curve fitting," Graphical Models Image Process., vol.60, no.2, pp.93–99, March 1998.
[18] K. Kanatani, "Further improving geometric fitting," Proc. 5th Int. Conf. 3-D Digital Imaging and Modeling, pp. 2–13, Ottawa, Ontario, Canada, June 2005.
[19] K. Kanatani and N. Ohta, "Automatic detection of circular objects by ellipse growing," Int. J. Image Graphics, vol.4, no.1, pp.35–50, Jan. 2004.
[20] Y. Kanazawa and K. Kanatani, "Optimal conic fitting and reliability evaluation," IEICE Trans. Inf. & Syst., vol.E79-D, no.9, pp.1323–1328, Sept. 1996.
[21] D. Keren, D. Cooper, and J. Subrahmonia, "Describing complicated objects by implicit polynomials," IEEE Trans. Pattern Anal. Mach. Intell., vol.16, no.1, pp.38–53, Jan. 1994.
[22] A. Kukush, I. Markovsky, and S. Van Huffel, "Consistent estimation in an implicit quadratic measurement error model," Comput. Stat. Data Anal., vol.47, no.1, pp.123–147, Aug. 2004.
[23] A. Kukush and E. O. Maschlce, "The efficiency of adjusted least squares in the linear functional relationship," J. Multivariate Anal., vol.87, no.2, pp.261–274, Oct. 2003.
[24] Y. Leedan and P. Meer, "Heteroscedastic regression in computer vision: Problems with bilinear constraint," Int. J. Comput. Vis., vol.37, no.2, pp.127–150, June 2000.
[25] M. Mühlich and R. Mester, "Unbiased errors-in-variables estimation using generalized eigensystem analysis," Proc. 2nd Workshop Statistical Methods in Video Processing, pp.38–49, Prague, Czech, May 2004.
[26] N. Ohta, "Motion parameter estimation from optical flow without nuisance parameters," 3rd Int. Workshop on Statistical and Computational Theory of Vision, Nice, France, Oct. 2003, http://www.stat.ucla.edu/~sczhu/Workshops/SCTV2003.html
[27] T. Okatani and K. Deguchi, "Toward a statistically optimal method for estimating geometric relations from noisy data: Cases of linear relations," Proc. IEEE Conf. Comput. Vision Pattern Recog., vol.1, pp.432–439, Madison, WI, U.S.A., June 2003.
[28] J. Porrill, "Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter," Image Vis. Comput., vol.8, no.1, pp.37–41, Feb. 1990.
[29] V. Pratt, "Direct least-squares fitting of algebraic surfaces," Comput. Grap., vol.21, no.4, pp.145–152, July 1987.
[30] G. Roth and M.D. Levine, "Extracting geometric primitives," CVGIP: Image Understand., vol.58, no.1, pp.1–22, July 1993.
[31] G. Roth and M.D. Levine, "Geometric primitive extraction using a genetic algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol.16, no.9, pp.901–905, Sept. 1994.
[32] P.L. Rosin and G.A.W. West, "Nonparametric segmentation of curves into various representations," IEEE Trans. Pattern Anal. Mach. Intell., vol.17, no.12, pp.1140–1153, Dec. 1995.
[33] R. Safee-Rad, I. Tchoukanov, B. Benhabib, and K.C. Smith, "Accurate parameter estimation of quadratic curves from gray-level images," CVGIP: Image Understand., vol.54, no.2, pp.259–274, Sept. 1991.
[34] P.D. Sampson, "Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithm," Comput. Graph. Image Process., vol.18, no. 1, pp.97–108, Jan. 1982.
[35] G. Taubin, "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.13, no.11, pp.1115–1138, Nov. 1991.
[36] S. Tsuji and F. Matsumoto, "Detection of ellipses by a modified Hough transform," IEEE Trans. Comput., vol.C-27, no.8, pp.777–781, Aug. 1978.
[37] M. Werman and D. Keren, "A Bayesian method for filtering parametric and nonparametric models to noisy data," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.5, pp.528–534, May 2001.
[38] H.K. Yuen, J. Illingworth, and J. Kittler, "Detecting partially occluded ellipses using the Hough transform," Image Vis. Comput., vol.7, no.1, pp.31–37, Feb. 1989.

Appendix: Perturbation Analysis of ML

Equating terms of $O(\sigma)$ on both sides of Eq. (19), we obtain

$$\bar{\boldsymbol{M}}\Delta_1\boldsymbol{u} + \Delta_1\boldsymbol{M}\boldsymbol{u} + \Delta_1^*\boldsymbol{M}\boldsymbol{u} = \boldsymbol{0}. \tag{45}$$

From Eq. (20), we see that

$$\Delta_1^*\boldsymbol{M}\boldsymbol{u}$$
$$= -2\sum_{\alpha=1}^{N} \frac{((\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}) + O(\sigma^2))\boldsymbol{\xi}_\alpha(\boldsymbol{\xi}_\alpha, \boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$$

$$= \mathbf{0}. \tag{46}$$

Hence, multiplication of $\bar{\boldsymbol{M}}^-$ on both sides of Eq. (45) from left yields

$$\boldsymbol{P}_{\mathbf{u}}\Delta_1\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u} = \mathbf{0}, \tag{47}$$

where $\boldsymbol{P}_{\mathbf{u}}$ is the projection matrix defined in Eq. (6). Note that $\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}} = \boldsymbol{P}_{\mathbf{u}}$. Since $\boldsymbol{u}$ is constrained to be a unit vector, the error $\Delta\boldsymbol{u}$ should be orthogonal to $\boldsymbol{u}$ to a first approximation. Hence, $\boldsymbol{P}_{\mathbf{u}}\Delta_1\boldsymbol{u} = \Delta_1\boldsymbol{u}$, from which we obtain Eq. (22).

Equating terms of $O(\sigma^2)$ on both sides of Eq. (19), we obtain

$$\begin{aligned}&\bar{\boldsymbol{M}}\Delta_2\boldsymbol{u} + \Delta_1\boldsymbol{M}\Delta_1\boldsymbol{u} + \Delta_1^*\boldsymbol{M}\Delta_1\boldsymbol{u}\\ &\quad+\Delta_2\boldsymbol{M}\boldsymbol{u} + \Delta_2^*\boldsymbol{M}\boldsymbol{u} = \Delta_2\boldsymbol{L}\boldsymbol{u} + \Delta_2^*\boldsymbol{L}\boldsymbol{u}.\end{aligned} \tag{48}$$

Multiplication of $\bar{\boldsymbol{M}}^-$ on both sides of Eq. (48) from left yields

$$\begin{aligned}&\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\Delta_2\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\Delta_1\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_1^*\boldsymbol{M}\Delta_1\boldsymbol{u}\\ &\quad+\bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{M}\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{M}\boldsymbol{u} = \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{L}\boldsymbol{u}\\ &\quad+\bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\boldsymbol{u}.\end{aligned} \tag{49}$$

Noting that $\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\Delta_2\boldsymbol{u} = \boldsymbol{P}\mathbf{u}\Delta_2\boldsymbol{u} = \Delta_2\boldsymbol{u}^\perp$, and substituting Eq. (22) for $\Delta_1\boldsymbol{u}$, we obtain Eq. (23).

Kenichi Kanatani received his B.E., M.S., and Ph.D. in applied mathematics from the University of Tokyo in 1972, 1974 and 1979, respectively. After serving as Professor of computer science at Gunma University, Gunma, Japan, he is currently Professor of computer science at Okayama University, Okayama, Japan. He is the author of many books on computer vision and received many awards including the best paper awards from IPSJ (1987) and IEICE (2005). He is an IEEE Fellow.