Int J Comput Vis (2008) 80: 167–188
DOI 10.1007/s 11263-007-0098-0

Statistical Optimization for Geometric Fitting: Theoretical Accuracy Bound and High Order Error Analysis

Kenichi Kanatani

Receieved: 19 November 2006 / Accepted: 20 September / published online: 27 November 2007

Abstract A rigorous accuracy analysis is given to various techniques for estimating parameters of geometric models from noisy data. First, it is pointed out that parameter estimation for vision applications is very different in nature from traditional statistical analysis and hence a different mathematical framework is necessary. After a general framework is formulated, typical numerical techniques are selected, and their accuracy is evaluated up to high order terms. As a byproduct, our analysis leads to a "hyperaccurate" method that outperforms existing methods.

Keywords: geometric fitting, parameter estimation, error analysis, hyperaccuracy.

1. Introduction

Modeling the geometric structure of images in a parametric form and estimating the parameters from observations are the first steps of many computer vision applications such as 3-D reconstruction and virtual reality generation. The purpose of this paper is to present a theoretical foundation for rigorous accuracy analysis that can lead to improved estimation techniques.

This may sound simple, because parameter estimation in the presence of noise is the main theme of statistics, so all one needs to do seems simply use the established results of statistics. In Sect. 2, we first point out that *this is not so* because parameter estimation for typical computer vision applications is very different in nature from traditional statistical analysis. We then present a mathematical framework that suits geometric computations frequently encountered in computer vision applications.

In Sect. 3, we first formulate a general framework and then focus on problems for which the model equation is linear in the parameters. In Sect. 4, we select well known estimation techniques and analyze their accuracy up to high order terms. This reveals why some methods known to be superior/inferior are really so in theoretical terms. As a byproduct, our analysis leads to a "hyperaccurate" method that outperforms existing methods. In Sect. 5, we conclude.

K. Kanatani
Department of Computer Science, Okayama University, Okayama 700-8530, Japan
e-mail: kanatani@suri.it.okayama-u.ac.jp

2. Geometric Fitting

2.1 Definition

We call the class of problems to be discussed in this paper *geometric fitting*: we fit a parameterized geometric model (a curve, a surface, or a relationship in high dimensions) expressed as an *implicit* equation in the form

$$F(\boldsymbol{x};\boldsymbol{u}) = 0, \tag{1}$$

to N data $\boldsymbol{x}_\alpha$, $\alpha = 1, ..., N$, typically points in an image or point correspondences over multiple images (Kanatani 1996). The function $F(\boldsymbol{x};\boldsymbol{u})$, which may be a vector if the model is defined by multiple equations, is parameterized by $\boldsymbol{u}$. Each $\boldsymbol{x}_\alpha$ is assumed to be perturbed by independent noise from its true value $\bar{\boldsymbol{x}}_\alpha$ which strictly satisfies (1). From the parameter $\boldsymbol{u}$ of the fitted equation, one can discern the underlying geometric structure. A large class of computer vision problems fall into this category (Kanatani 1996).

Though one can speak of noise and parameter estimation, the fact that this problem does *not* straightforwardly fit the traditional framework of statistics has not been widely recognized. The following are typical distinctions of geometric fitting as compared with the traditional parameter estimation problem: smallskip

- Unlike traditional statistics, there is no *explicit* model which explains observables in terms of deterministic mechanisms and random noise. All descriptions are *implicit*.
- No inputs or outputs exist. No such concepts exist as *causes* and *effects*, or *ordinates* and *abscissas*.
- The underlying data space is usually homogeneous and isotropic with no inherent coordinate system. Hence, the estimation process should be *invariant* to changes of the coordinate system with respect to which the data are described.
- Usually, the data are geometrically constrained to be on predetermined curves, surfaces, and hypersurfaces (e.g., unit vectors or matrices of determinant 0). The parameters to be estimated may also be similarly constrained. Hence, the *Gaussian distribution*, the most fundamental noise modeling, does *not* exist in its strict sense in such constrained spaces.

We first discuss why the traditional approach does not suit our intended applications.

2.2 Reduction to Statistical Estimation

It appears that the problem can be rewritten in the traditional form. The "observable" is the set of data $\boldsymbol{x}_\alpha$, which can be rearranged into a high dimensional vector $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^\top & \boldsymbol{x}_2^\top & \cdots & \boldsymbol{x}_N^\top \end{pmatrix}^\top$. Let $\boldsymbol{\varepsilon}_\alpha$ be noise in $\boldsymbol{x}_\alpha$, and define the vector $\boldsymbol{E} = \begin{pmatrix} \boldsymbol{\varepsilon}_1^\top & \boldsymbol{\varepsilon}_2^\top & \cdots & \boldsymbol{\varepsilon}_N^\top \end{pmatrix}^\top$. Let $\bar{\boldsymbol{X}}$ be the true value of $\boldsymbol{X}$. The *statistical model* in the usual sense is

$$\boldsymbol{X} = \bar{\boldsymbol{X}} + \boldsymbol{E}. \tag{2}$$

The unknown $\bar{\boldsymbol{X}}$ needs to be estimated. Let $p(\boldsymbol{E})$ be the probability density of the noise vector $\boldsymbol{E}$. Our task is to estimate $\bar{\boldsymbol{X}}$ from $\boldsymbol{X}$, which we regard as sampled from $p(\boldsymbol{X} - \bar{\boldsymbol{X}})$. The difficulty is that the parameter $\boldsymbol{u}$, which we really want to estimate, is not contained in this model. Instead, the parameter $\boldsymbol{u}$ implicitly constrains the mutual relationships among the components of $\bar{\boldsymbol{X}}$.

In order to make the implicit constraint explicit, one needs to introducing a new parameter $\boldsymbol{t}$ to solve (1) for $\boldsymbol{u}$ in the parametric form

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{t}; \boldsymbol{u}). \tag{3}$$

For example, if we want to fit a circle $(x-a)^2 + (y-b)^2 = r^2$, we rewrite it as $x = a + r\cos\theta$, $y = b + r\sin\theta$ by introducing the directional angle θ. However, this type of parametric representation is usually very difficult to obtain.

Suppose such a parametric representation does exist. Substituting $\bar{\boldsymbol{x}}_1 = \boldsymbol{x}(\boldsymbol{t}_1, \boldsymbol{u})$, $\bar{\boldsymbol{x}}_2 = \boldsymbol{x}(\boldsymbol{t}_2, \boldsymbol{u})$, ..., $\bar{\boldsymbol{x}}_N = \boldsymbol{x}(\boldsymbol{t}_N, \boldsymbol{u})$, (2) now has the form

$$\boldsymbol{X} = \bar{\boldsymbol{X}}(\boldsymbol{t}_1, ..., \boldsymbol{t}_N; \boldsymbol{u}) + \boldsymbol{E}. \tag{4}$$

Our task is to estimate the parameters $\boldsymbol{t}_1$,..., $\boldsymbol{t}_N$ and $\boldsymbol{u}$ from $\boldsymbol{X}$.

2.3 Neyman-Scott Problem

Although the problem looks like a standard form, there is a big difference: we observe *only one* observable $\boldsymbol{X}$ for a "particular" set of parameters $\boldsymbol{t}_1$,..., $\boldsymbol{t}_N$ and $\boldsymbol{u}$. Namely, $\boldsymbol{X}$ is a *single* sample from $p(\boldsymbol{X} - \bar{\boldsymbol{X}}(\boldsymbol{t}_1, ..., \boldsymbol{t}_N; \boldsymbol{u}))$.

The tenet of statistical estimation is to observe *repeated* samples from a distribution, or an *ensemble*, and infer its unknown parameters. Naturally, estimation becomes more accurate as more samples are drawn, thanks to the "law of large numbers". Here, however, only one sample $\boldsymbol{X}$ is available.

What happens if we increase the data? If we observe another datum $\boldsymbol{x}_{N+1}$, the observable $\boldsymbol{X}$ becomes a yet higher dimensional vector, and (4) becomes a yet higher dimensional equation, which has an additional unknown $\boldsymbol{t}_{N+1}$. This means that the resulting observable $\boldsymbol{X}$ is not "another" sample of the same distribution; it is *one* sample from a *new* distribution with a new set of parameters $\boldsymbol{t}_1$,..., $\boldsymbol{t}_{N+1}$ and $\boldsymbol{u}$. However large the number of data is, the number of observables is always 1.

This (seeming) anomaly was first pointed out by Neyman and Scott (1948). Since then, this problem has been referred to as the *Neyman-Scott problem*. Even for a single observation, maximum likelihood (ML) is possible. However, Neyman and Scott (1948) pointed out that the estimated parameters do not necessarily converge to their true values as $N \to \infty$, indicating the (seeming) lack of "consistency", which is a characteristic of ML.

This is natural because increasing the number of data does not mean increasing the number of *samples* from a distribution having *particular parameters*. Though $\boldsymbol{u}$ may be unchanged as N increases, we have as many parameters $\boldsymbol{t}_1$,..., $\boldsymbol{t}_N$ as the increased number of data. Due to this (seeming) anomaly, these are called *nuisance parameters*, whereas $\boldsymbol{u}$ is called the *structural parameter* or the *parameter of interest*.

2.4 Semiparametric Models

In spite of many attempts in the past, this anomaly has never been resolved, because it does not make sense to regard what is not standard statistical estimation as standard statistical estimation. It has been realized that the

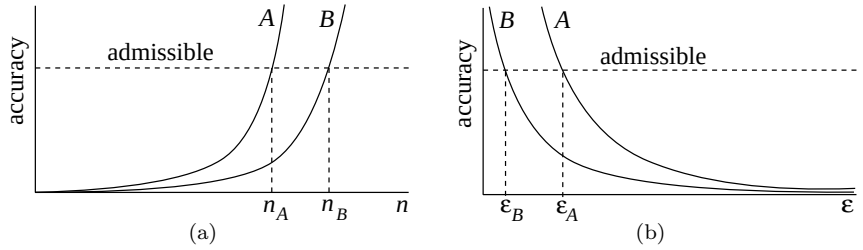


Figure 1: (a) For the standard statistical estimation, it is desired that the accuracy increases rapidly as $n \to \infty$ for the number n of observations, because admissible accuracy can be reached with a smaller number of observations. (b) For geometric fitting, it is desired that the accuracy increases rapidly as $\varepsilon \to 0$ for the noise level ε, because larger data uncertainty can be tolerated for admissible accuracy

only way to fit the problem in the standard framework is to regard $\boldsymbol{t}_1$,..., $\boldsymbol{t}_N$ not as parameters but as *data* sampled from a *fixed* probability density $q(\boldsymbol{t};\boldsymbol{v})$ with some unknown parameters $\boldsymbol{v}$ called *hyperparameters*.

The problem is now interpreted as follows. Given $\boldsymbol{u}$ and $\boldsymbol{v}$, the values $\boldsymbol{t}_1$,..., $\boldsymbol{t}_N$ are randomly drawn from $q(\boldsymbol{t};\boldsymbol{v})$. Then, (3) defines the true values $\bar{\boldsymbol{x}}_1$, ..., $\bar{\boldsymbol{x}}_N$, to which random noise drawn from $p(\boldsymbol{E})$ is added. The task is to estimate both $\boldsymbol{u}$ and $\boldsymbol{v}$ by observing $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$. For a given parametric density $q(\boldsymbol{t};\boldsymbol{v})$, statisticians call this interpretation the *structural model*, while (4) the *functional model*.

In practice, however, how can we give the density $q(\boldsymbol{t};\boldsymbol{v})$ by merely looking at a *single* set of data $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$? To cope with this difficulty, a new approach has emerged: we introduce a density $q(\boldsymbol{t};\boldsymbol{v})$ whose form is not completely specified. Such a model is said to be *semiparametric* (Amari and Kawanabe 1997; Bickel et al. 1994).

The standard procedure goes like this. We first estimate the density $q(\boldsymbol{t};\boldsymbol{v})$ (the most difficult part), then *marginalize* the model over $q(\boldsymbol{t};\boldsymbol{v})$, integrating out all $\boldsymbol{t}_1$, ..., $\boldsymbol{t}_N$ (not analytically easy), and finally search for an optimal value of $\boldsymbol{u}$. Now that the problem is reduced to repeated sampling from a fixed distribution, the consistency as $N \to \infty$ is guaranteed under mild conditions.

This approach has been adopted in some computer vision problems where a large number of data are available. Ohta (2003) showed that the semiparametric model yields a better result for 3-D interpretation of a dense optical flow field, and Okatani and Deguchi (2003) demonstrated that for estimating 3-D shape and motion from a point cloud seen in multiple images, the semiparametric model can result in higher accuracy. In both cases, however, the procedure is very complicated, and the superior performance is obtained only when the number of data is extremely large.

2.5 Dual Approach

A natural question arises: why do we need to rewrite (1) in a parametric form by introducing the new parameter $\boldsymbol{t}$? If (1) has a simple form, e.g., a polynomial, why do we need to convert it to a complicated (generally non-algebraic[1]) form, if the conversion is possible at all? Why cannot we do estimation using (1) as is?

This might be answered as follows. Statisticians try to fit the problem in the standard framework because they are motivated to analyze *asymptotic* behavior of estimation as the number n of observations increases. In particular, the "consistency", i.e., the property that the computed estimates converge to their true values as $n \to \infty$, together with the speed of convergence measured in $O((1/\sqrt{n})^k)$, is their major concern.

This concern originates from the fact that an estimation method whose accuracy increases rapidly as $n \to \infty$ can attain admissible accuracy with a fewer number of observations (Fig. 1(a)). Such a method is desirable because most statistical applications are done in the presence of large noise (e.g., agriculture, medicine, economics, psychology, and census surveys), and hence one needs a large number of repeated observations to compensate for the noise, which entails a considerable cost in real situations.

To this, Kanatani (1996, 2004d) countered, saying that the purpose of many computer vision applications is to estimate the underlying geometric structure as accurately as possible in the presence of small noise. In fact, the uncertainty introduced by image processing operations is usually around a few pixels or subpixels. He asserted that in such domains, it is more reasonable to evaluate the performance in the limit $\varepsilon \to 0$ for the noise level ε, because a method whose accuracy increases rapidly as $\varepsilon \to 0$ can tolerate larger uncertainty for admissible accuracy (Fig. 1(b)).

If our our interest is in the limit $\varepsilon \to 0$, we can build a mathematical theory of estimation directly from (1). Indeed, this is what has implicitly been done by many computer vision researchers for years without worrying about orthodox theories in the statistical literature.

[1]It is known that a polynomial (or algebraic) equation does not have an algebraically parametric representation unless its "genus" is 0 (Clebsch theorem).

Table 1: Duality between traditional statistical estimation and geometric fitting (Kanatani 2004)

Statistical estimation	Geometric fitting
data generating mechanism	geometric constraints
$\boldsymbol{x} \sim p(\boldsymbol{x}; \boldsymbol{\theta})$	$F(\boldsymbol{x}; \boldsymbol{u}) = 0$
CR lower bound	KCR lower bound
$V_{\rm CR}[\hat{\theta}] = O(1/n)$	$V_{\rm KCR}[\hat{\boldsymbol{u}}] = O(\varepsilon^2)$
ML is optimal in the	ML is optimal in the
limit $n \to \infty$	limit $\varepsilon \to 0$
Akaike's AIC	geometric AIC
AIC $= \cdots + O(1/n)$	G-AIC $= \cdots + O(\varepsilon^4)$
Rissanen's MDL	geometric MDL
MDL $= \cdots + O(1)$	G-MDL $= \cdots + O(\varepsilon^2)$

2.6 Duality of interpretation

Kanatani (1996, 2004) pushed this idea further and showed that resulting mathematical consequences have corresponding traditional results in a *dual* form, e.g., the KCR lower bound (Chernov and Lesort 2004; Kanatani 1998) corresponds to the traditional Cramer-Rao (CR) lower bound, and the geometric AIC and the geometric MDL correspond, respectively, to Akaike's AIC (Akaike 1974) and Rissannen's MDL (Rissanen 1989) (Table 1).

The correspondence is dual in the sense that small noise expansions have the form $\cdots + O(\varepsilon^k)$ for geometric fitting, to which correspond traditional asymptotic expansions in the form $\cdots + O(1/\sqrt{n^k})$. Kanatani (1996, 2004) explained this, invoking the following thought experiment.

For geometric fitting, the image data may not be exact due to the uncertainty of image processing operations, but *they always have the same value however many times we observe them*, so the number n of observations is always 1, as pointed out earlier. Suppose, hypothetically, they change their values each time we observe them as if in quantum mechanics. Then, we would obtain n different values for n observations. If we take their sample mean, its standard deviation is $1/\sqrt{n}$ times that of individual observations. This means that repeating hypothetical observations n times effectively reduces the noise level ε to $\varepsilon/\sqrt{n}$. Thus, the behavior of estimation for $\varepsilon \to 0$ is mathematically equivalent to the asymptotic behavior for $n \to \infty$ of the number n of hypothetical observations (not the number N of "data").

In the following, we adopt this approach and analyze the accuracy of existing estimation techniques around $\varepsilon \approx 0$.

3. Parameter Estimation and Accuracy

3.1 Noise Description and Estimators

Our goal is to obtain a good estimate of the parameter $\boldsymbol{u}$ from observed data $\boldsymbol{x}_\alpha$. To do mathematical analysis, however, there is a serious obstacle arising from the fact that the data $\boldsymbol{x}_\alpha$ and the parameter $\boldsymbol{u}$ may be constrained; they may be unit vectors or matrices of determinant 0, for instance. How can we define noise in the data and errors of the parameters? Evidently, direct vector calculus is not suitable. For example, if a unit vector is perturbed isotropically, the perturbed values are distributed over a unit sphere, but their average is "inside" the sphere.

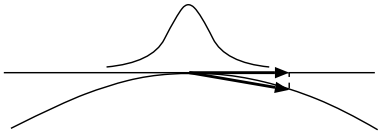


Figure 2: The displacement of a constrained variable is projected onto the tangent space, with which we identify the noise domain

A more serious problem is that noise distributions *cannot* be Gaussian, because Gaussian distributions with infinitely long tails can exist only in a Euclidean space. Since Gaussian distributions are the most fundamental of all distributions, how can we do mathematical analysis without it?

Several mathematical formulations have been proposed for probability distributions in a non-Euclidean space based on theories of Lie groups and invariant measures (e.g., Begelfor and Werman (2005) and Pennec et al. (2006)), but the results are very much complicated.

Fortunately, such complications are not necessary in our formulation, because we are focusing only on small noise effects in the dual framework. We can simply assume that noise concentrates on a small region around the true value and regard it as occurring in the *tangent space* at that point. Within this tangent space, the noise distribution can be regarded as Gaussian; the discrepancy at the tail part is of higher order terms. Accordingly, we define the *covariance matrix* of $\boldsymbol{x}_\alpha$ by

$$V[\boldsymbol{x}_\alpha] = E[\Big(\mathcal{P}_{\bar{\boldsymbol{x}}_\alpha}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)\Big)\Big(\mathcal{P}_{\bar{\boldsymbol{x}}_\alpha}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)\Big)^\top], \qquad (5)$$

where $E[\,\cdot\,]$ denotes expectation over the noise distribution, and $\mathcal{P}_{\bar{\boldsymbol{x}}_\alpha}$ denotes projection onto the tangent space to the domain $\mathcal{X}$ of the data at $\bar{\boldsymbol{x}}_\alpha$ (Fig. 2).

The geometric fitting problem in the form of (1) is solved if a procedure is given for computing an estimate $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$ in terms of observed data $\boldsymbol{x}_\alpha$, which defines a function

$$\hat{\boldsymbol{u}} = \hat{\boldsymbol{u}}(\boldsymbol{x}_1, ..., \boldsymbol{x}_N), \qquad (6)$$

called an *estimator* of $\boldsymbol{u}$. A natural requirement is that the true value should be obtained in the absence of noise:

$$\lim_{\varepsilon \to 0} \hat{\boldsymbol{u}} = \boldsymbol{u}. \qquad (7)$$

Here, ε is the noise level, and $\boldsymbol{u}$ the true parameter value. Chernov and Lesort (2004) called this condition *consistency* in the dual framework. In this paper, we consider

only consistent estimators in this sense. Confirming consistency is usually a trivial matter.

If $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ are random variables, so is $\hat{\boldsymbol{u}}$ as a function of them. Hence, we can measure its accuracy by its covariance matrix. Here again, the parameter $\boldsymbol{u}$ may be constrained and its domain $\mathcal{U}$ may not be Euclidean. So, we identify the error of $\hat{\boldsymbol{u}}$ as belonging to the tangent space to $\mathcal{U}$ at the true value $\boldsymbol{u}$. Namely, we define the covariance matrix $V[\hat{\boldsymbol{u}}]$ of $\hat{\boldsymbol{u}}$ by

$$V[\hat{\boldsymbol{u}}] = E[\Big(\mathcal{P}_{\mathbf{u}}(\hat{\boldsymbol{u}} - \boldsymbol{u})\Big)\Big(\mathcal{P}_{\mathbf{u}}(\hat{\boldsymbol{u}} - \boldsymbol{u})\Big)^{\top}], \tag{8}$$

where $\mathcal{P}_{\mathbf{u}}$ denotes projection onto the tangent space of the domain $\mathcal{U}$ at $\boldsymbol{u}$.

3.2 KCR Lower Bound

Kanatani (1996, 2005) proved that if each datum $\boldsymbol{x}_\alpha$ is an independent Gaussian random variable in the abovementioned sense with mean $\bar{\boldsymbol{x}}_\alpha$ and covariance matrix $V[\boldsymbol{x}_\alpha]$, the following inequality holds for an arbitrary unbiased estimator $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$ (see Appendix 1 for the proof):

$$V[\hat{\boldsymbol{u}}] \succ \left(\sum_{\alpha=1}^{N} \frac{(\mathcal{P}_{\mathbf{u}}\nabla_{\mathbf{u}}\bar{F}_\alpha)(\mathcal{P}_{\mathbf{u}}\nabla_{\mathbf{u}}\bar{F}_\alpha)^{\top}}{(\nabla_{\mathbf{x}}\bar{F}_\alpha, V[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}}\bar{F}_\alpha)}\right)^{-}. \tag{9}$$

Here, $\succ$ means that the left-hand side minus the right is positive semidefinite, and the superscript $-$ denotes pseudoinverse. The symbols $\nabla_{\mathbf{x}}\bar{F}_\alpha$ and $\nabla_{\mathbf{u}}\bar{F}_\alpha$ denote the gradient of the function $F(\boldsymbol{x};\boldsymbol{u})$ in (1) with respect to $\boldsymbol{x}$ and $\boldsymbol{u}$, respectively, evaluated at $\boldsymbol{x} = \bar{\boldsymbol{x}}_\alpha$. Throughout this paper, we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$.

Chernov and Lesort (2004) called the right-hand side of (9) the *KCR* (*Kanatani-Cramer-Rao*) *lower bound* and showed that it holds except for $O(\varepsilon^4)$ even if $\hat{\boldsymbol{u}}$ is not unbiased; it is sufficient that $\hat{\boldsymbol{u}}$ is "consistent" in the sense of (7).

If we worked in the traditional domain of statistics, we would obtain the corresponding *CR* (*Cramer-Rao*) *lower bound*. The statistical model is given by (4) with likelihood function $p(\boldsymbol{X} - \bar{\boldsymbol{X}}(\boldsymbol{t}_1, ..., \boldsymbol{t}_N; \boldsymbol{u}))$. So, the CR bound can be obtained by following the standard procedure described in the statistical literature.

To be specific, we first evaluate second order derivatives of $\log p(\boldsymbol{X} - \bar{\boldsymbol{X}}(\boldsymbol{t}_1, ..., \boldsymbol{t}_N; \boldsymbol{u}))$ with respect to both $\boldsymbol{t}_1$, ..., $\boldsymbol{t}_N$ and $\boldsymbol{u}$ (or multiply the first order derivatives each other) and define an $(mN + p) \times (mN + p)$ matrix, where m and p are the dimensions of the vectors $\boldsymbol{t}_\alpha$ and the vector $\boldsymbol{u}$, respectively. We then take expectation of this matrix with respect to the density $p(\boldsymbol{X} - \bar{\boldsymbol{X}}(\boldsymbol{t}_1, ..., \boldsymbol{t}_N; \boldsymbol{u}))$. The resulting matrix is called the *Fisher information matrix*. Then, we invert it and discard the nuisance parameters $\boldsymbol{t}_1$, ..., $\boldsymbol{t}_N$ by taking out only the $p \times p$ diagonal block corresponding to $\boldsymbol{u}$, resulting in the CR lower bound on $\boldsymbol{u}$ alone, which turns out to be the same as (9)

In most cases, however, this derivation process is almost intractable due to the difficulty of analytically inverting a matrix of a very large size. In contrast, the KCR lower bound in the form of (9) directly gives a bound on $\boldsymbol{u}$ alone, without involving any "nuisance parameters". This is one of the most significant advantages of working in the dual framework (Kanatani 1996, 2005).

3.3 Minimization Schemes

As we have repeatedly pointed out, our goal is to obtain a *good estimator* $\hat{\boldsymbol{u}}$ of the parameter $\boldsymbol{u}$ from observed data $\boldsymbol{x}_\alpha$. An estimator $\hat{\boldsymbol{u}}$ is *good* if it is always very close to the true value $\boldsymbol{u}$, namely if its covariance matrix $V[\hat{\boldsymbol{u}}]$ is small; an estimator is *optimal* if its covariance matrix agrees with the KCR lower bound in the sense that no better estimator can exist.

Thus, the goal is to *minimize the covariance matrix* $V[\hat{\boldsymbol{u}}]$ by defining a *good function* $\hat{\boldsymbol{u}} = \hat{\boldsymbol{u}}(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ of the data; it is *not* to minimize some cost function. It is a practical strategy to *implicitly* define an estimator through minimization of a particular cost function, but we should bear in mind that *this is not always necessary*, as we will see later.

The most widely used is what is called *least-squares* (*LS*) (and by other names such as *algebraic distance minimization*), minimizing

$$J = \sum_{\alpha=1}^{N} F(\boldsymbol{x}_\alpha; \boldsymbol{u})^2. \tag{10}$$

Another popular scheme is to minimizes what is known as *geometric distance* (and as other names such as *Sampson error*)

$$J = \sum_{\alpha=1}^{N} \frac{F(\boldsymbol{x}_\alpha; \boldsymbol{u})^2}{\|\nabla_{\mathbf{x}} F_\alpha\|^2}. \tag{11}$$

Many other minimization schemes have been proposed in the past. All of them are designed so as to make $F(\boldsymbol{x}_\alpha; \boldsymbol{u})$ *approximately* 0 for all α and at the same time let the solution $\hat{\boldsymbol{u}}$ have desirable properties (Bookstein 1979; Sampson 1982; Taubin 1991). To this, Kanatani (1996) viewed the problem as *statistical estimation* for estimating the true data values $\bar{\boldsymbol{x}}_\alpha$ that *strictly* satisfy the constraint

$$F(\bar{\boldsymbol{x}}_\alpha; \boldsymbol{u}) = 0, \qquad \alpha = 1, ..., N, \tag{12}$$

using the knowledge of the data covariance matrices $V[\boldsymbol{x}_\alpha]$.

If we assume that the noise in each $\boldsymbol{x}_\alpha$ is independent Gaussian (in the tangent space) with mean $\boldsymbol{0}$ and covariance matrix $V[\boldsymbol{x}_\alpha]$, the likelihood of observing $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ is

$$C\prod_{\alpha=1}^{N} e^{-(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha, V[\boldsymbol{x}_\alpha]^-(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha))/2}, \tag{13}$$

where C is a normalization constant. The true values $\bar{\boldsymbol{x}}_1$, ..., $\bar{\boldsymbol{x}}_N$ are constrained by (12). *Maximum likelihood* (*ML*) maximizes (13), which is equivalent to minimizing the (square) *Mahalanobis distance*

$$J=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha, V[\boldsymbol{x}_\alpha]^-(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha)), \tag{14}$$

subject to (12).

The constraint of (12) can be eliminated by introducing Lagrange multipliers and ignoring higher order terms in the noise level, which can be justified in our dual framework. The resulting form is (see Appendix 2 for the derivation)

$$J=\sum_{\alpha=1}^{N}\frac{F(\boldsymbol{x}_\alpha;\boldsymbol{u})^2}{(\nabla_{\mathbf{x}}F_\alpha, V[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}}F_\alpha)}. \tag{15}$$

It can be shown that the covariance matrix $V[\hat{\boldsymbol{u}}]$ of the resulting estimator $\hat{\boldsymbol{u}}$ achieves the KCR lower bound except for $O(\varepsilon^4)$ (Chernov and Lesort 2004; Kanatani 1996, 2005) (see Appendix 3 for the proof).

3.4 Linearized Constraint Optimization

In the rest of this paper, we concentrate on a special subclass of geometric fitting problems in which (1) reduces to the linear form

$$(\boldsymbol{\xi}(\boldsymbol{x}),\boldsymbol{u})=0, \tag{16}$$

by changing variables $\boldsymbol{\xi}=\boldsymbol{\xi}(\boldsymbol{x})$. If the data $\boldsymbol{x}_\alpha$ are m-dimensional vectors and the unknown parameter $\boldsymbol{u}$ is a p-dimensional vector, the mapping $\boldsymbol{\xi}(\,\cdot\,)$ is a (generally nonlinear) embedding from $\mathcal{R}^m$ to $\mathcal{R}^p$. In order to remove scale indeterminacy, we normalize $\boldsymbol{u}$ to $\|\boldsymbol{u}\|=1$.

The KCR lower bound ($\equiv$ the right-hand side of (9)) now has the form

$$V_{\mathrm{KCR}}[\hat{\boldsymbol{u}}]=\Big(\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u},V[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big)^-, \tag{17}$$

where we write $\bar{\boldsymbol{\xi}}_\alpha=\boldsymbol{\xi}(\bar{\boldsymbol{x}}_\alpha)$. The covariance matrix $V[\boldsymbol{\xi}_\alpha]$ of $\boldsymbol{\xi}_\alpha=\boldsymbol{\xi}(\boldsymbol{x}_\alpha)$ is given, except for higher order terms in the noise level, in the form

$$V[\boldsymbol{\xi}_\alpha]=\nabla_{\mathbf{x}}\bar{\boldsymbol{\xi}}_\alpha^\top V[\boldsymbol{x}_\alpha]\nabla_{\mathbf{x}}\bar{\boldsymbol{\xi}}_\alpha, \tag{18}$$

where $\nabla_{\mathbf{x}}\bar{\boldsymbol{\xi}}_\alpha$ is the $m\times p$ Jacobian matrix

$$\nabla_{\mathbf{x}}\boldsymbol{\xi}=\begin{pmatrix}\partial\xi_1/\partial x_1 & \cdots & \partial\xi_p/\partial x_1\\ \vdots & \ddots & \vdots\\ \partial\xi_1/\partial x_m & \cdots & \partial\xi_p/\partial x_m\end{pmatrix}. \tag{19}$$

evaluated at $\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha$. Note that in (17) we do not need the projection operator for the normalization constraint $\|\boldsymbol{u}\|=1$, because $\bar{\boldsymbol{\xi}}_\alpha$ is orthogonal to $\boldsymbol{u}$ due to (16); for the moment, we assume that no other internal constraints exist.

This subclass of geometric fitting problems covers a wide range of computer vision applications. The following are typical examples:

Example 1 Suppose we want to fit a quadratic curve (circle, ellipse, parabola, hyperbola, or their degeneracy) to N points (x_α,y_α) in the plane. The constraint has the form

$$Ax^2+2Bxy+Cy^2+2(Dx+Ey)+F=0. \tag{20}$$

If we define

$$\boldsymbol{\xi}(x,y)=(x^2\ \ 2xy\ \ y^2\ \ 2x\ \ 2y\ \ 1)^\top,$$

$$\boldsymbol{u}=(A\ \ B\ \ C\ \ D\ \ E\ \ F)^\top, \tag{21}$$

(20) is linearized in the form of (16). If independent Gaussian noise of mean 0 and standard deviation σ is added to each coordinates of (x_α,y_α), the covariance matrix $V[\boldsymbol{\xi}_\alpha]$ of the transformed $\boldsymbol{\xi}_\alpha$ has the form

$$V[\boldsymbol{\xi}_\alpha]=4\sigma^2\begin{pmatrix}\bar{x}_\alpha^2 & \bar{x}_\alpha\bar{y}_\alpha & 0 & \bar{x}_\alpha & 0 & 0\\ \bar{x}_\alpha\bar{y}_\alpha & \bar{x}_\alpha^2+\bar{y}_\alpha^2 & \bar{x}_\alpha\bar{y}_\alpha & \bar{y}_\alpha & \bar{x}_\alpha & 0\\ 0 & \bar{x}_\alpha\bar{y}_\alpha & \bar{y}_\alpha^2 & 0 & \bar{y}_\alpha & 0\\ \bar{x}_\alpha & \bar{y}_\alpha & 0 & 1 & 0 & 0\\ 0 & \bar{x}_\alpha & \bar{y}_\alpha & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}, \tag{22}$$

except for $O(\sigma^4)$, where $(\bar{x}_\alpha,\bar{y}_\alpha)$ is the true position of (x_α,y_α). □

Example 2 Suppose we have N corresponding points in two images of the same scene viewed from different positions. If point (x,y) in the first image corresponds to (x',y') in the second, they should satisfy the following *epipolar equation* (Hartley and Zisserman 2000):

$$\left(\begin{pmatrix}x\\y\\1\end{pmatrix},\boldsymbol{F}\begin{pmatrix}x'\\y'\\1\end{pmatrix}\right)=0. \tag{23}$$

Here, $\boldsymbol{F}$ is a matrix of rank 2, called the *fundamental matrix*, that depends only on the intrinsic parameters of

the two cameras that took the two images and their relative 3-D positions, but not on the scene and the location of the identified points (Hartley and Zisserman 2000). If we define

$$\boldsymbol{\xi}(x,y,x',y') = (xx'\ xy'\ x\ yx'\ yy'\ y\ x'\ y'\ 1)^\top,$$
$$\boldsymbol{u} = (F_{11}\ F_{12}\ F_{13}\ F_{21}\ F_{22}\ F_{23}\ F_{31}\ F_{32}\ F_{33})^\top, \quad (24)$$

(23) is linearized in the form of (16). If independent Gaussian noise of mean 0 and standard deviation σ is added to each coordinates of the corresponding points (x_α, y_α) and (x'_α, y'_α), the covariance matrix $V[\boldsymbol{\xi}_\alpha]$ of the transformed $\boldsymbol{\xi}_\alpha$ has the form

$$V[\boldsymbol{\xi}_\alpha] = \sigma^2 \begin{pmatrix} \bar{x}_\alpha^2 + \bar{x}_\alpha'^2 & \bar{x}'_\alpha\bar{y}'_\alpha & \bar{x}'_\alpha & \bar{x}_\alpha\bar{y}_\alpha \\ \bar{x}'_\alpha\bar{y}'_\alpha & \bar{x}_\alpha^2 + \bar{y}_\alpha'^2 & \bar{y}'_\alpha & 0 \\ \bar{x}'_\alpha & \bar{y}'_\alpha & 1 & 0 \\ \bar{x}_\alpha\bar{y}_\alpha & 0 & 0 & \bar{y}_\alpha^2 + \bar{x}_\alpha'^2 \\ 0 & \bar{x}_\alpha\bar{y}_\alpha & 0 & \bar{x}'_\alpha\bar{y}'_\alpha \\ 0 & 0 & 0 & \bar{x}'_\alpha \\ \bar{x}_\alpha & 0 & 0 & \bar{y}_\alpha \\ 0 & \bar{x}_\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\begin{pmatrix} 0 & 0 & \bar{x}_\alpha & 0 & 0 \\ \bar{x}_\alpha\bar{y}_\alpha & 0 & 0 & \bar{x}_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \bar{x}'_\alpha\bar{y}'_\alpha & \bar{x}'_\alpha & \bar{y}_\alpha & 0 & 0 \\ \bar{y}_\alpha^2 + \bar{y}_\alpha'^2 & \bar{y}'_\alpha & 0 & \bar{y}_\alpha & 0 \\ \bar{y}'_\alpha & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \bar{y}_\alpha & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (25)$$

except for $O(\sigma^4)$, where $(\bar{x}_\alpha, \bar{y}_\alpha)$ and $(\bar{x}'_\alpha, \bar{y}'_\alpha)$, are the true positions of (x_α, y_α) and (x'_α, y'_α), respectively. The fundamental matrix has, aside from scale normalization, the constraint that its determinant is 0. If we take this constraint into consideration, the KCR lower bound of (17) involves the corresponding projection operation (Kanatani and Ohta 2003). □

As we can see from (22) and (25), the covariance matrix $V[\boldsymbol{\xi}_\alpha]$ is usually factored into the form

$$V[\boldsymbol{\xi}_\alpha] = \varepsilon^2 V_0[\boldsymbol{\xi}_\alpha], \quad (26)$$

where ε is a constant that characterizes the noise and $V_0[\boldsymbol{\xi}_\alpha]$ is a matrix that depends only on the true data values. Hereafter, we assume this form and define ε to be the *noise level*; we call $V_0[\boldsymbol{\xi}_\alpha]$ the *normalized covariance matrix*.

The true values in (22) and (25) can be approximated by their observed values in actual computations, because the covariance matrix is a "property" of individual data, not a "function" of its location. The covariance matrix describes its location uncertainty determined by the image processing operator (e.g., the Harris operator) that detected it and the characteristics of the image region surrounding it (e.g., homogeneous or textured). Hence, the covariance matrix is treated as a given constant in the optimization process.

4. Accuracy of Parameter Estimation

We now give a rigorous accuracy analysis of typical estimation techniques up to high order error terms. The following analysis has evolved from the author's earlier studies of ellipse fitting (Kanatani 2006). The methods used and the results derived there also have meaning in a general situation.

4.1 Least Squares (LS)

For the linearized constraint of (16), minimization of (10) reduces to minimization of

$$J = \sum_{\alpha=1}^N (\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 = \sum_{\alpha=1}^N (\boldsymbol{u}, \boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top\boldsymbol{u}) = (\boldsymbol{u}, \boldsymbol{M}_{\rm LS}\boldsymbol{u}), \quad (27)$$

where

$$\boldsymbol{M}_{\rm LS} \equiv \sum_{\alpha=1}^N \boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top. \quad (28)$$

The quadratic form $(\boldsymbol{u}, \boldsymbol{M}_{\rm LS}\boldsymbol{u})$ is minimized by the unit eigenvector for the smallest eigenvalue of $\boldsymbol{M}_{\rm LS}$.

To do error analysis, we write

$$\boldsymbol{M}_{\rm LS}\hat{\boldsymbol{u}} = \lambda\hat{\boldsymbol{u}}, \quad (29)$$

into which we substitute $\boldsymbol{\xi}_\alpha = \bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha$ and $\hat{\boldsymbol{u}} = \boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots$, where Δ_k denotes perturbations corresponding to the kth orders in $\Delta\boldsymbol{\xi}_\alpha$. Then, we have

$$(\bar{\boldsymbol{M}}_{\rm LS} + \Delta_1\boldsymbol{M}_{\rm LS} + \Delta_2\boldsymbol{M}_{\rm LS})(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots)$$
$$= (\Delta_1\lambda + \Delta_2\lambda + \cdots)(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots), \quad (30)$$

where $\bar{\boldsymbol{M}}_{\rm LS}$ is the value of $\boldsymbol{M}_{\rm LS}$ obtained by replacing $\boldsymbol{\xi}_\alpha$ in (29) by their true values $\bar{\boldsymbol{\xi}}_\alpha$, and

$$\Delta_1\boldsymbol{M}_{\rm LS} = \sum_{\alpha=1}^N (\bar{\boldsymbol{\xi}}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top + \Delta\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top),$$

$$\Delta_2\boldsymbol{M}_{\rm LS} = \sum_{\alpha=1}^N \Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top. \quad (31)$$

We also expand the eigenvalue λ in (29) into $\Delta_1\lambda + \Delta_2\lambda + \cdots$. Since $\lambda = 0$ in the absence of noise, its 0th order term does not exist.

Equating first and second order terms on both sides of (30), we obtain

$$\bar{\boldsymbol{M}}_{\mathrm{LS}}\Delta_1\boldsymbol{u} + \Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u} = \Delta_1\lambda\boldsymbol{u}, \tag{32}$$

$$\begin{aligned}&\bar{\boldsymbol{M}}_{\mathrm{LS}}\Delta_2\boldsymbol{u} + \Delta_1\boldsymbol{M}_{\mathrm{LS}}\Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}\\ &=\Delta_1\lambda\Delta_1\boldsymbol{u} + \Delta_2\lambda\boldsymbol{u}.\end{aligned} \tag{33}$$

Computing the inner product with $\boldsymbol{u}$ on both sides of (32) and noting that $(\boldsymbol{u}, \bar{\boldsymbol{M}}_{\mathrm{LS}}\Delta_1\boldsymbol{u})$ and $(\boldsymbol{u}, \Delta_1\bar{\boldsymbol{M}}_{\mathrm{LS}}\boldsymbol{u})$ identically vanish due to $(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{u}) = 0$, we see that $\Delta_1\lambda = 0$. Multiplying $\bar{\boldsymbol{M}}_{\mathrm{LS}}^-$ on both sides of (32) and noting that $\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{M}}_{\mathrm{LS}} = \boldsymbol{P}_{\boldsymbol{u}}$ ($\equiv \boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^\top$, the projection matrix onto the hyperplane orthogonal to $\boldsymbol{u}$) and $\Delta_1\boldsymbol{u}$ is orthogonal to $\boldsymbol{u}$ to a first approximation (because $\|\boldsymbol{u}\| = 1$), we conclude that

$$\Delta_1\boldsymbol{u} = -\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}. \tag{34}$$

Evidently, $E[\Delta_1\boldsymbol{u}] = \boldsymbol{0}$. Its covariance matrix is

$$\begin{aligned}V[\Delta_1\boldsymbol{u}] &= E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top]\\ &= \bar{\boldsymbol{M}}_{\mathrm{LS}}^- E[(\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u})(\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u})^\top]\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\\ &= \bar{\boldsymbol{M}}_{\mathrm{LS}}^- E\Big[\sum_{\alpha=1}^N(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha \sum_{\beta=1}^N(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta^\top\Big]\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\\ &= \bar{\boldsymbol{M}}_{\mathrm{LS}}^- \sum_{\alpha\beta=1}^N(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\\ &= \varepsilon^2\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Big(\sum_{\alpha=1}^N(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big)\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\\ &= \varepsilon^2\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{M}}'_{\mathrm{LS}}\bar{\boldsymbol{M}}_{\mathrm{LS}}^-,\end{aligned} \tag{35}$$

where we define

$$\bar{\boldsymbol{M}}'_{\mathrm{LS}} \equiv \sum_{\alpha=1}^N(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top, \tag{36}$$

and use the identity $E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top] = \varepsilon^2\delta_{\alpha\beta}V_0[\boldsymbol{\xi}_\alpha]$ implied by our assumption about the noise ($\delta_{\alpha\beta}$ is the Kronecker delta, taking 1 for $\alpha = \beta$ and 0 otherwise).

Multiplying $\bar{\boldsymbol{M}}_{\mathrm{LS}}^-$ on both sides of (33) and solving for $\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{M}}_{\mathrm{LS}}\Delta_2\boldsymbol{u}$ ($\equiv \boldsymbol{P}_{\boldsymbol{u}}\Delta_2\boldsymbol{u}$), we obtain

$$\begin{aligned}\Delta_2\boldsymbol{u}^\perp &= -\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_1\boldsymbol{M}_{\mathrm{LS}}\Delta_1\boldsymbol{u} - \bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_2\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}\\ &= \bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_1\boldsymbol{M}_{\mathrm{LS}}\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u} - \bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_2\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u},\end{aligned} \tag{37}$$

where $\Delta_2\boldsymbol{u}^\perp$ ($\equiv \boldsymbol{P}_{\boldsymbol{u}}\Delta_2\boldsymbol{u}$) is the component of $\Delta_2\boldsymbol{u}$ orthogonal to $\boldsymbol{u}$. The parallel component $\Delta_2\boldsymbol{u}^\parallel$ can also be computed, but it is not important, since it arises solely for enforcing the normalization constraint $\|\hat{\boldsymbol{u}}\| = 1$ (Fig. 3). Thus, we can measure the accuracy only by examining the orthogonal component, as discussed in Sect. 3.1.

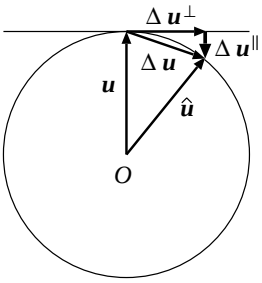


Figure 3: The orthogonal error component $\Delta\boldsymbol{u}^\perp$ and the parallel error component $\Delta\boldsymbol{u}^\parallel$ of an estimate $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$. The accuracy can be measured by the orthogonal component $\Delta\boldsymbol{u}^\perp$

If we note that

$$\begin{aligned}&E[\Delta_1\boldsymbol{M}_{\mathrm{LS}}\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}]\\ &= E\Big[\sum_{\alpha=1}^N(\bar{\boldsymbol{\xi}}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top + \Delta\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top)\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\sum_{\beta=1}^N(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta\Big]\\ &= \sum_{\alpha,\beta=1}^N(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\beta\Delta\boldsymbol{\xi}_\alpha^\top]\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{\xi}}_\beta)\bar{\boldsymbol{\xi}}_\alpha\\ &\quad + \sum_{\alpha,\beta=1}^N(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{\xi}}_\beta)E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u}\\ &= \varepsilon^2\sum_{\alpha=1}^N(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha\\ &\quad + \varepsilon^2\sum_{\alpha=1}^N(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},\end{aligned} \tag{38}$$

$$\begin{aligned}E[\Delta_2\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}] &= \sum_{\alpha=1}^N E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u}\\ &= \varepsilon^2\sum_{\alpha=1}^N V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u} = \varepsilon^2\boldsymbol{N}_{\mathrm{LS}}\boldsymbol{u},\end{aligned} \tag{39}$$

where we define

$$\boldsymbol{N}_{\mathrm{LS}} \equiv \sum_{\alpha=1}^N V_0[\boldsymbol{\xi}_\alpha], \tag{40}$$

the expectation of $\Delta_2\boldsymbol{u}^\perp$ is given by

$$\begin{aligned}E[\Delta_2\boldsymbol{u}^\perp] &= \varepsilon^2\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\sum_{\alpha=1}^N(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha\\ &\quad + \varepsilon^2\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\sum_{\alpha=1}^N(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}_{\mathrm{LS}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}\\ &\quad - \varepsilon^2\bar{\boldsymbol{M}}_{\mathrm{LS}}^-\boldsymbol{N}_{\mathrm{LS}}\boldsymbol{u}.\end{aligned} \tag{41}$$

4.2 Taubin Method

The method due to Taubin (1991)[2] is to minimize, instead of (27),

$$J = \frac{\sum_{\alpha=1}^{N}(\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2}{\sum_{\alpha=1}^{N}(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = \frac{(\boldsymbol{u}, \boldsymbol{M}_{\rm LS}\boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{N}_{\rm LS}\boldsymbol{u})}. \tag{42}$$

This is a Rayleigh quotient, so it is minimized by the eigenvector of the generalized eigenvalue problem

$$\boldsymbol{M}_{\rm LS}\hat{\boldsymbol{u}} = \lambda \boldsymbol{N}_{\rm LS}\hat{\boldsymbol{u}}, \tag{43}$$

for the smallest eigenvalue. The matrix $\boldsymbol{N}_{\rm LS}$ may be singular, but we can solve (43) by reducing the number of parameters as prescribed by Chojnacki, et al. (2004, 2005) (see Appendix 4 for the procedure).

As in the case of LS, we expand (43) in the form

$$\begin{aligned}&(\bar{\boldsymbol{M}}_{\rm LS} + \Delta_1\boldsymbol{M}_{\rm LS} + \Delta_2\boldsymbol{M}_{\rm LS})(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots)\\ &\quad = (\Delta_1\lambda + \Delta_2\lambda + \cdots)\boldsymbol{N}_{\rm LS}(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots),\end{aligned} \tag{44}$$

and equate first and second order terms on both sides. We obtain

$$\bar{\boldsymbol{M}}_{\rm LS}\Delta_1\boldsymbol{u} + \Delta_1\boldsymbol{M}_{\rm LS}\boldsymbol{u} = \Delta_1\lambda\boldsymbol{N}_{\rm LS}\boldsymbol{u}, \tag{45}$$

$$\begin{aligned}&\bar{\boldsymbol{M}}_{\rm LS}\Delta_2\boldsymbol{u} + \Delta_1\boldsymbol{M}_{\rm LS}\Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{M}_{\rm LS}\boldsymbol{u}\\ &\quad = \Delta_1\lambda\boldsymbol{N}_{\rm LS}\Delta_1\boldsymbol{u} + \Delta_2\lambda\boldsymbol{N}_{\rm LS}\boldsymbol{u}.\end{aligned} \tag{46}$$

Computing the inner product with $\boldsymbol{u}$ on both sides of (45), we again find that $\Delta_1\lambda = 0$. So, the first order error $\Delta_1\boldsymbol{u}$ is again given by (34) and hence its covariance matrix $V[\Delta_1\boldsymbol{u}]$ is given by (35). In other words, *LS and the Taubin method have the same accuracy to a first approximation*.

However, the Taubin method is known to be substantially better than LS. This implies that the difference should be second-order effects. Multiplying $\bar{\boldsymbol{M}}_{\rm LS}^-$ on both sides of (46) and solving for $\Delta_2\boldsymbol{u}^\perp$ ($\equiv \bar{\boldsymbol{M}}_{\rm LS}^-\bar{\boldsymbol{M}}_{\rm LS}^-\Delta_2\boldsymbol{u}$), we obtain

$$\begin{aligned}\Delta_2\boldsymbol{u}^\perp =& -\bar{\boldsymbol{M}}_{\rm LS}^-\Delta_1\boldsymbol{M}_{\rm LS}\Delta_1\boldsymbol{u} - \bar{\boldsymbol{M}}_{\rm LS}^-\Delta_2\boldsymbol{M}_{\rm LS}\boldsymbol{u}\\ &-\Delta_2\lambda\bar{\boldsymbol{M}}_{\rm LS}\boldsymbol{N}_{\rm LS}\boldsymbol{u}\\ =& \bar{\boldsymbol{M}}_{\rm LS}^-\Delta_1\boldsymbol{M}_{\rm LS}\bar{\boldsymbol{M}}_{\rm LS}^-\Delta_1\boldsymbol{M}_{\rm LS}\boldsymbol{u} - \bar{\boldsymbol{M}}_{\rm LS}^-\Delta_2\boldsymbol{M}_{\rm LS}\boldsymbol{u}\\ &+\Delta_2\lambda\bar{\boldsymbol{M}}_{\rm LS}\boldsymbol{N}_{\rm LS}\boldsymbol{u}.\end{aligned} \tag{47}$$

Comparing this with (37), we find that an extra term $\Delta_2\lambda\bar{\boldsymbol{M}}_{\rm LS}^-\boldsymbol{N}_{\rm LS}\boldsymbol{u}$ is added. We now evaluate its expectation.

[2]Taubin (1991) studied curve fitting, which he analyzed purely from a geometric point of view without using statistical terms such as means and covariance matrices. What is shown here is a modification of his method in the present framework.

Computing the inner product with $\boldsymbol{u}$ on both sides of (46), and noting that $(\boldsymbol{u}, \boldsymbol{M}_{\rm LS}\Delta_2\boldsymbol{u})$ identically vanishes, we have

$$\Delta_2\lambda = \frac{(\boldsymbol{u}, \Delta_1\boldsymbol{M}_{\rm LS}\Delta_1\boldsymbol{u}) + (\boldsymbol{u}, \Delta_2\boldsymbol{M}_{\rm LS}\boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{N}_{\rm LS}\boldsymbol{u})}. \tag{48}$$

From (34) and the first of (31), we have

$$\begin{aligned}&E[(\boldsymbol{u}, \Delta_1\boldsymbol{M}_{\rm LS}\Delta_1\boldsymbol{u})]\\ &\quad = -E[(\boldsymbol{u}, \Delta_1\boldsymbol{M}_{\rm LS}\bar{\boldsymbol{M}}_{\rm LS}^-\Delta_1\boldsymbol{M}_{\rm LS}\boldsymbol{u})]\\ &\quad = -E[(\Delta_1\boldsymbol{M}_{\rm LS}\boldsymbol{u}, \bar{\boldsymbol{M}}_{\rm LS}^-\Delta_1\boldsymbol{M}_{\rm LS}\boldsymbol{u})]\\ &\quad = -E\Big[\Big(\sum_{\alpha=1}^{N}(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}_{\rm LS}^-\sum_{\beta=1}^{N}(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta\Big)\Big]\\ &\quad = -\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})(\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{M}}_{\rm LS}^-, \bar{\boldsymbol{\xi}}_\beta)\\ &\quad = -\varepsilon^2\sum_{\alpha=1}^{N}(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{M}}_{\rm LS}^-, \bar{\boldsymbol{\xi}}_\alpha)\\ &\quad = -\varepsilon^2\sum_{\alpha=1}^{N}{\rm tr}[(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{M}}_{\rm LS}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top]\\ &\quad = -\varepsilon^2{\rm tr}\Big[\bar{\boldsymbol{M}}_{\rm LS}^-\sum_{\alpha=1}^{N}(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big]\\ &\quad = -\varepsilon^2{\rm tr}[\bar{\boldsymbol{M}}_{\rm LS}^-\bar{\boldsymbol{M}}_{\rm LS}'] = -q\varepsilon^2,\end{aligned} \tag{49}$$

where $\bar{\boldsymbol{M}}_{\rm LS}'$ is the matrix in (36) and $q = {\rm tr}[\bar{\boldsymbol{M}}_{\rm LS}^-\bar{\boldsymbol{M}}_{\rm LS}']$. From (39), we have

$$(\boldsymbol{u}, \Delta_2\boldsymbol{M}_{\rm LS}\boldsymbol{u}) = \varepsilon^2(\boldsymbol{n}, \boldsymbol{N}_{\rm LS}\boldsymbol{u}). \tag{50}$$

Thus, the expectation of (48) is

$$E[\Delta_2\lambda] = \Big(1 - \frac{q}{(\boldsymbol{n}, \boldsymbol{N}_{\rm LS}\boldsymbol{u})}\Big)\varepsilon^2. \tag{51}$$

Hence, we obtain

$$\begin{aligned}E[\Delta_2\boldsymbol{u}^\perp] =& \varepsilon^2\bar{\boldsymbol{M}}_{\rm LS}^-\sum_{\alpha=1}^{N}(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}_{\rm LS}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha\\ &+\varepsilon^2\bar{\boldsymbol{M}}_{\rm LS}^-\sum_{\alpha=1}^{N}(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}_{\rm LS}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}\\ &-\frac{q\varepsilon^2}{(\boldsymbol{u}, \boldsymbol{N}_{\rm LS}\boldsymbol{u})}\bar{\boldsymbol{M}}_{\rm LS}^-\boldsymbol{N}_{\rm LS}\boldsymbol{u}.\end{aligned} \tag{52}$$

If the number N of data is fairly large, which is the case in many vision applications, so is $(\boldsymbol{u}, \boldsymbol{N}_{\rm LS}\boldsymbol{u}) = \sum_{\alpha=1}^{N}(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})$. Hence, the last term is much smaller than the corresponding term $\varepsilon^2\bar{\boldsymbol{M}}_{\rm LS}^-\boldsymbol{N}_{\rm LS}\boldsymbol{u}$ in (41) for LS, which explains the improved accuracy of the Taubin method as compared with LS. We now confirm this by numerical experiments.

Figure 4: 20 points on an ellipse

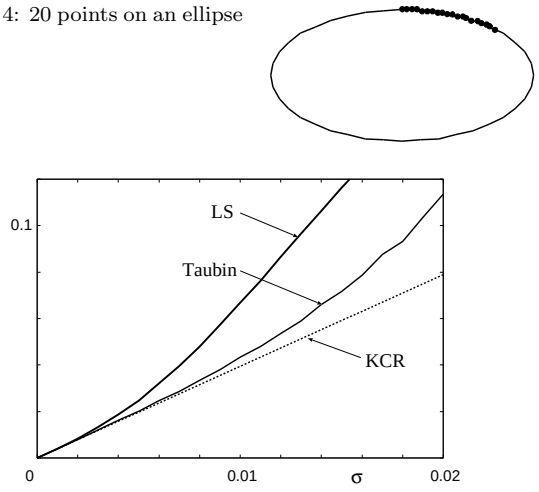


Figure 5: Noise level vs. RMS error for the ellipse data in Fig. 4: LS (*thick solid line*), Taubin (*thin solid line*), and KCR lower bound (*dotted line*)

Example 3 Figure 4 shows $N = 20$ points $(\bar{x}_\alpha, \bar{y}_\alpha)$ taken on ellipse

$$\frac{x^2}{50^2} + \frac{y^2}{100^2} = 1, \tag{53}$$

with equal intervals. From them, we generated data points (x_α, y_α) by adding Gaussian noise of mean 0 and standard deviation σ to the x and y coordinates independently. Then, we fitted an ellipse by LS and the Taubin method.

Figure 5 plots for different σ the fitting error evaluated by the following root mean square over 10,000 independent trials:

$$D = \sqrt{\frac{1}{10000}\sum_{a=1}^{10000} \|\boldsymbol{P}_{\mathbf{u}}\hat{\boldsymbol{u}}^{(a)}\|^2}. \tag{54}$$

Here, $\hat{\boldsymbol{u}}^{(a)}$ is the ath value of $\hat{\boldsymbol{u}}$. The thick and thin line are for LS and the Taubin method, respectively. The dotted line is the corresponding KCR lower bound $D_{\rm KCR} = \sqrt{{\rm tr}V_{\rm KCR}[\hat{\boldsymbol{u}}]}$ (see (17).

As we can see, the LS solution is of very low accuracy, while the Taubin solution is fairly accurate. The plots for LS and Taubin should have, at $\sigma = 0$, the same slope distinct from that of the KCR lower bound, as far as the first order error $\Delta_1\boldsymbol{u}$ is concerned. However, this effect is too weak to be visible in Fig. 5, implying that *the performance difference between LS and Taubin is mostly due to second order error* $\Delta_2\boldsymbol{u}$, in particular the last term of (41). □

4.3 Optimally Weighted Least Squares

A well known correction to LS is to appropriately weight each summand in (27) in the form

$$J = \sum_{\alpha=1}^{N} W_\alpha(\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2, \tag{55}$$

which is minimized by the unit eigenvector of

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top, \tag{56}$$

for the smallest eigenvalue. The weight W_α is determined so that the covariance matrix of the resulting estimate is as close to the KCR lower bound as possible.

Following the analysis in Sect. 4.1, we can easily see that the first order covariance matrix in (35) is now replaced by

$$V[\Delta_1\boldsymbol{u}] = \varepsilon^2 \bar{\boldsymbol{M}}^- \Big(\sum_{\alpha=1}^{N} W_\alpha(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big)\bar{\boldsymbol{M}}^-. \tag{57}$$

It is not difficult to see that this coincides with the KCR lower bound if we set

$$W_\alpha = \frac{1}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \tag{58}$$

In fact, we have

$$V[\Delta_1\boldsymbol{u}] = \varepsilon^2 \bar{\boldsymbol{M}}^- \Big(\sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big)\bar{\boldsymbol{M}}^- = \varepsilon^2 \bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^- = \varepsilon^2 \bar{\boldsymbol{M}}^-, \tag{59}$$

where we define

$$\bar{\boldsymbol{M}} \equiv \sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \tag{60}$$

Evidently, (59) equals the KCR lower bound given by (17).

However, we cannot use (58), because the true value $\boldsymbol{u}$ is unknown. So, we do iterations. Namely, we first give an appropriate initial guess of $\boldsymbol{u}$, say by LS, substitute it into (58) and compute the eigenvector of the matrix $\boldsymbol{M}$ in (56) for the smallest eigenvalue. Using the resulting solution, we update the weight W_α and iterate this process. This method is known as *optimally weighted* (*iterative*) *least squares*, or simply *reweight* (Taubin 1991). The fact that this method achieves the KCR lower bound to a first approximation was pointed out by Chernov and Lesort (2004).

We now evaluate its accuracy. After the iterations have converged, the resulting solution $\hat{\boldsymbol{u}}$ satisfies

$$\hat{\boldsymbol{M}}\hat{\boldsymbol{u}} = \lambda\hat{\boldsymbol{u}}, \tag{61}$$

where

$$\hat{\boldsymbol{M}} = \sum_{\alpha=1}^{N} \frac{\boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top}{(\hat{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})}. \tag{62}$$

Substituting $\boldsymbol{\xi}_\alpha = \bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha$, $\hat{\boldsymbol{u}} = \boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots$, and $\lambda = \Delta_1\lambda + \Delta_2\lambda + \cdots$ into (61), we have

$$\begin{aligned}&(\bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M} + \Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M})(\boldsymbol{u} + \Delta_1\boldsymbol{u} \\ &\quad + \Delta_2\boldsymbol{u} + \cdots) \\ &= (\Delta_1\lambda + \Delta_2\lambda + \cdots)(\boldsymbol{u} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots),\end{aligned} \tag{63}$$

where we put

$$\begin{aligned}\Delta_1\boldsymbol{M} &= \sum_{\alpha=1}^{N} \frac{\Delta\boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}, \\ \Delta_2\boldsymbol{M} &= \sum_{\alpha=1}^{N} \frac{\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})},\end{aligned} \tag{64}$$

$$\begin{aligned}\Delta_1^*\boldsymbol{M} &= -2\sum_{\alpha=1}^{N} \frac{(\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top, \\ \Delta_2^*\boldsymbol{M} &= -2\sum_{\alpha=1}^{N} \frac{(\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} (\Delta\boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top) \\ &\quad + \sum_{\alpha=1}^{N} \Big(-\frac{2(\Delta_2\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} + \frac{4(\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\ &\quad - \frac{(\Delta_1\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} \Big) \frac{\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}.\end{aligned} \tag{65}$$

Here, $\Delta_1^*\boldsymbol{M}$ and $\Delta_2^*\boldsymbol{M}$ are, respectively, the first and second order perturbations of $\boldsymbol{M}$ for using $\hat{\boldsymbol{u}}$ in the denominator in (62).

Equating first and second order terms on both sides of (63), we obtain

$$\bar{\boldsymbol{M}}\Delta_1\boldsymbol{u} + (\Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M})\boldsymbol{u} = \Delta_1\lambda\boldsymbol{u}, \tag{66}$$

$$\begin{aligned}&\bar{\boldsymbol{M}}\Delta_2\boldsymbol{u} + (\Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M})\Delta_1\boldsymbol{u} + (\Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M})\boldsymbol{u} \\ &\quad = \Delta_1\lambda\Delta_1\boldsymbol{u} + \Delta_2\lambda\boldsymbol{u}.\end{aligned} \tag{67}$$

Computing the inner product with $\boldsymbol{u}$ on both sides of (66) and noting that $(\boldsymbol{u}, \bar{\boldsymbol{M}}\boldsymbol{u})$, $(\boldsymbol{u}, \Delta_1\boldsymbol{M}\boldsymbol{u})$, and $(\boldsymbol{u}, \Delta_1^*\boldsymbol{M}\boldsymbol{u})$ all identically vanish, we find that $\Delta_1\lambda = 0$. Multiplying $\bar{\boldsymbol{M}}^-$ on both sides of (66) and solving for $\Delta_1\boldsymbol{u}$, we obtain as before

$$\Delta_1\boldsymbol{u} = -\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}, \tag{68}$$

whose covariance matrix $V[\Delta_1\boldsymbol{u}]$ coincides with the KCR lower bound $\varepsilon^2\bar{\boldsymbol{M}}^-$.

Multiplying $\bar{\boldsymbol{M}}^-$ on both sides of (67) and solving for $\Delta_2\boldsymbol{u}^\perp$ ($\equiv \bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\Delta_2\boldsymbol{u}$), we obtain

$$\begin{aligned}\Delta_2\boldsymbol{u}^\perp &= -\bar{\boldsymbol{M}}^-(\Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M})\Delta_1\boldsymbol{u} \\ &\quad -\bar{\boldsymbol{M}}^-(\Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M})\boldsymbol{u} \\ &= \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u} + \bar{\boldsymbol{M}}^-\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u} \\ &\quad -\bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{M}\boldsymbol{u} - \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{M}\boldsymbol{u}.\end{aligned} \tag{69}$$

Now, we compute its expectation. We first see that

$$\begin{aligned}&E[\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}] \\ &= E\Big[\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{\Delta\boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} \bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big] \\ &= \bar{\boldsymbol{M}}^- \sum_{\alpha,\beta=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta)E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})} \\ &\quad + \bar{\boldsymbol{M}}^- \sum_{\alpha,\beta=1}^{N} \frac{(\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta, E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})} \\ &= \varepsilon^2\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\ &\quad + \varepsilon^2\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}.\end{aligned} \tag{70}$$

We also see that

$$\begin{aligned}&E[\bar{\boldsymbol{M}}^-\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}] \\ &= E\Big[2\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\Delta_1\boldsymbol{M}\boldsymbol{u}, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\boldsymbol{u}\Big] \\ &= 2\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}, \bar{\boldsymbol{M}}^- E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}.\end{aligned} \tag{71}$$

The expectation $E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]$ is

$$\begin{aligned}&E[(\Delta_1\boldsymbol{M}\boldsymbol{u})(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top] \\ &= E\Big[\sum_{\alpha=1}^{N} \frac{(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} \sum_{\beta=1}^{N} \frac{(\Delta\boldsymbol{\xi}_\beta, \boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\Big] \\ &= \sum_{\alpha,\beta=1}^{N} \frac{(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})} \\ &= \varepsilon^2 \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} = \varepsilon^2 \sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} \\ &= \varepsilon^2\bar{\boldsymbol{M}}.\end{aligned} \tag{72}$$

Hence, (71) becomes

$$E[\bar{\boldsymbol{M}}^{-}\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u}] = 2\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} = 2\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u},\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \tag{73}$$

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

$$E[\bar{\boldsymbol{M}}^{-}\Delta_2\boldsymbol{M}\boldsymbol{u}] = E\Big[\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\Delta\boldsymbol{\xi}_\alpha,\boldsymbol{u})\Delta\boldsymbol{\xi}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big] = \bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u}}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = \varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = \varepsilon^2\bar{\boldsymbol{M}}^{-}\boldsymbol{N}\boldsymbol{u}, \tag{74}$$

where we define

$$\boldsymbol{N} \equiv \sum_{\alpha=1}^{N}\frac{V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \tag{75}$$

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

$$E[\bar{\boldsymbol{M}}^{-}\Delta_2^*\boldsymbol{M}\boldsymbol{u}] = E\Big[-2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\Delta_1\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\Delta\boldsymbol{\xi}_\alpha,\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}\Big] = 2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{u},E[\Delta\boldsymbol{\xi}_\alpha(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]\bar{\boldsymbol{M}}^{-}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \tag{76}$$

The expectation $E[\Delta\boldsymbol{\xi}_\alpha(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top]$ is

$$E[\Delta\boldsymbol{\xi}_\alpha(\Delta_1\boldsymbol{M}\boldsymbol{u})^\top] = E\Big[\Delta\boldsymbol{\xi}_\alpha\sum_{\beta=1}^{N}\frac{(\Delta\boldsymbol{\xi}_\beta,\boldsymbol{u})\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})}\Big] = \sum_{\beta=1}^{N}\frac{E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u}\bar{\boldsymbol{\xi}}_\beta^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})} = \frac{\varepsilon^2V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \tag{77}$$

Hence, (76) becomes

$$E[\bar{\boldsymbol{M}}^{-}\Delta_2^*\boldsymbol{M}\boldsymbol{u}] = 2\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^3} = 2\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}, \tag{78}$$

Table 2: The role of the Taubin method and renormalization

	No weight		Iterative reweight
eigenvalue problem	LS	↔	optimally weighted LS
	⇓		⇓
generalized eigenvalue problem	Taubin	↔	renormalization

which is the same as (73). Thus, the expectation of $\Delta_2\boldsymbol{u}^\perp$ in (69)

$$E[\Delta_2\boldsymbol{u}^\perp] = \varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha,V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} + \varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} - \varepsilon^2\bar{\boldsymbol{M}}^{-}\boldsymbol{N}\boldsymbol{u}. \tag{79}$$

4.4 Renormalization

We can see the similarity between (34) and (41) for (unweighted) LS and (68) and (79) for optimally weighted LS, where the (unweighted) matrix $\boldsymbol{M}_{\rm LS}$ is replaced by the weighted matrix $\boldsymbol{M}$. We have seen that the last term $-\varepsilon^2\bar{\boldsymbol{M}}_{\rm LS}^{-}\boldsymbol{N}_{\rm LS}\boldsymbol{u}$ in (41) can be reduced by using the Taubin method, replacing (29) by (43) by inserting the (unweighted) matrix $\boldsymbol{N}_{\rm LS}$.

The above comparison implies that the last term $-\varepsilon^2\bar{\boldsymbol{M}}^{-}\boldsymbol{N}\boldsymbol{u}$ in (79) may be removed if we replace the eigenvalue problem of (61) by a generalized eigenvalue problem

$$\hat{\boldsymbol{M}}\hat{\boldsymbol{u}} = \lambda\hat{\boldsymbol{N}}\hat{\boldsymbol{u}}, \tag{80}$$

by inserting the weighted matrix

$$\hat{\boldsymbol{N}} = \sum_{\alpha=1}^{N}\frac{V_0[\boldsymbol{\xi}_\alpha]}{(\hat{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})}. \tag{81}$$

Indeed, this is the idea of the *renormalization* of Kanatani (1993, 1996) (Table 2). His original idea was that the exact value $\boldsymbol{u}$ is obtained as the eigenvector of $\bar{\boldsymbol{M}}$ in (60) for eigenvalue 0. If we approximate $\bar{\boldsymbol{M}}$ by $\hat{\boldsymbol{M}}$ in (61), we have

$$\hat{\boldsymbol{M}} = \bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M} + \Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M}. \tag{82}$$

Evidently $E[\Delta_1 \boldsymbol{M}] = \boldsymbol{O}$ and $E[\Delta_1^* \boldsymbol{M}] = \boldsymbol{O}$, but we see from the second of (64) that

$$E[\Delta_2 \boldsymbol{M}] = \sum_{\alpha=1}^{N} \frac{E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = \sum_{\alpha=1}^{N} \frac{\varepsilon^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = \varepsilon^2 \boldsymbol{N}. \tag{83}$$

Hence, $\hat{\boldsymbol{M}} - \varepsilon^2 \boldsymbol{N}$ is closer to $\bar{\boldsymbol{M}}$ in expectation than $\hat{\boldsymbol{M}}$. Though we do not know ε^2 and $\boldsymbol{N}$, the latter may be approximated by $\hat{\boldsymbol{N}}$. The former is simply regarded as an unknown to be estimated. Kanatani (1993,1996) approximated it by the value c that make $\hat{\boldsymbol{M}} - c\hat{\boldsymbol{N}}$ singular, since the true value $\bar{\boldsymbol{M}}$ has eigenvalue 0. Thus, Kanatani's renormalization goes as follows:

1. Initialize $\hat{\boldsymbol{u}}$, say by LS, and let $c = 0$.
2. Solve the eigenvalue problem
$$(\hat{\boldsymbol{M}} - c\hat{\boldsymbol{N}})\boldsymbol{u} = \lambda \boldsymbol{u}, \tag{84}$$
and let $\boldsymbol{u}$ be the unit eigenvector for the eigenvalue λ closest to 0.
3. If $\lambda \approx 0$, return $\hat{\boldsymbol{u}}$ and stop. Else, let
$$c \leftarrow c + \frac{\lambda}{(\boldsymbol{u}, \hat{\boldsymbol{N}}\boldsymbol{u})}, \qquad \hat{\boldsymbol{u}} \leftarrow \boldsymbol{u}, \tag{85}$$
and go back to Step 2.

This method has been demonstrated to result in dramatic improvement over (unweighted or optimally weighted) LS in many computer vision problems including fundamental matrix computation for 3-D reconstruction and homography estimation for image mosaicing (Kanatani and Ohta 2003; Kanatani et al. 2000). We now analyze its accuracy.

After the iterations have converged, we have

$$(\hat{\boldsymbol{M}} - c\hat{\boldsymbol{N}})\hat{\boldsymbol{u}} = \boldsymbol{0}, \tag{86}$$

which is essentially (80). As before, we have the perturbation expansion

$$\Big(\bar{\boldsymbol{M}} + (\Delta_1 \boldsymbol{M} + \Delta_1^* \boldsymbol{M}) + (\Delta_2 \boldsymbol{M} + \Delta_2^* \boldsymbol{M}) + \cdots - (\Delta_1 c + \Delta_2 c + \cdots)(\boldsymbol{N} + \Delta_1^* \boldsymbol{N} + \cdots)\Big)(\boldsymbol{u} + \Delta_1 \boldsymbol{u} + \Delta_2 \boldsymbol{u} + \cdots) = \boldsymbol{0}, \tag{87}$$

where

$$\Delta_1^* \boldsymbol{N} = -2 \sum_{\alpha=1}^{N} \frac{(\Delta_1 \boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}) V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}, \tag{88}$$

which arises from the expansion of the denominator in the expression of $\hat{\boldsymbol{N}}$ (the second order perturbation $\Delta_2^* \boldsymbol{N}$ does not affect the subsequent analysis).

Equating first and second order terms on both sides of (87), we obtain

$$\bar{\boldsymbol{M}} \Delta_1 \boldsymbol{u} + (\Delta_1 \boldsymbol{M} + \Delta_1^* \boldsymbol{M} - \Delta_1 c \boldsymbol{N})\boldsymbol{u} = \boldsymbol{0}, \tag{89}$$

$$\bar{\boldsymbol{M}} \Delta_2 \boldsymbol{u} + (\Delta_1 \boldsymbol{M} + \Delta_1^* \boldsymbol{M} - \Delta_1 c \boldsymbol{N})\Delta_1 \boldsymbol{u} + (\Delta_2 \boldsymbol{M} + \Delta_2^* \boldsymbol{M} - \Delta_1 c \Delta_1^* \boldsymbol{N} - \Delta_2 c \boldsymbol{N})\boldsymbol{u} = \boldsymbol{0}. \tag{90}$$

Computing the inner product with $\hat{\boldsymbol{u}}$ on both sides of (89), we find that $\Delta_1 c = 0$ as before. Multiplying $\bar{\boldsymbol{M}}^-$ on both sides of (89) and solving for $\Delta_1 \boldsymbol{u}$, we again obtain (68). Thus, its covariance matrix $V[\Delta_1 \boldsymbol{u}]$ coincides with the KCR lower bound $\varepsilon^2 \bar{\boldsymbol{M}}^-$.

Multiplying $\bar{\boldsymbol{M}}^-$ on both sides of (90) and solving for $\Delta_2 \boldsymbol{u}^\perp$, we obtain

$$\begin{aligned}\Delta_2 \boldsymbol{u}^\perp =& -\bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M} \Delta_1 \boldsymbol{u} - \bar{\boldsymbol{M}}^- \Delta_1^* \boldsymbol{M} \Delta_1 \boldsymbol{u} - \bar{\boldsymbol{M}}^- \Delta_2 \boldsymbol{M} \boldsymbol{u} \\ & - \bar{\boldsymbol{M}}^- \Delta_2^* \boldsymbol{M} \boldsymbol{u} + \Delta_2 c \bar{\boldsymbol{M}}^- \boldsymbol{N} \boldsymbol{u} \\ =& \bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M} \bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M} \boldsymbol{u} + \bar{\boldsymbol{M}}^- \Delta_1^* \boldsymbol{M} \bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M} \boldsymbol{u} \\ & - \bar{\boldsymbol{M}}^- \Delta_2 \boldsymbol{M} \boldsymbol{u} - \bar{\boldsymbol{M}}^- \Delta_2^* \boldsymbol{M} \boldsymbol{u} + \Delta_2 c \bar{\boldsymbol{M}}^- \boldsymbol{N} \boldsymbol{u}.\end{aligned} \tag{91}$$

Comparing this with (69), we find that an extra term $\Delta_2 c \bar{\boldsymbol{M}}^- \boldsymbol{N} \boldsymbol{u}$ is added. We now evaluate its expectation.

Computing the inner product with $\boldsymbol{u}$ on both sides of (90) and noting that $(\boldsymbol{u}, \bar{\boldsymbol{M}} \Delta_2 \boldsymbol{u})$, $(\boldsymbol{u}, \Delta_1^* \boldsymbol{M} \Delta_1 \boldsymbol{u})$, and $(\boldsymbol{u}, \Delta_2^* \boldsymbol{M} \boldsymbol{u})$ all identically vanish, we have

$$\Delta_2 c = \frac{(\boldsymbol{u}, \Delta_2 \boldsymbol{M} \boldsymbol{u}) + (\boldsymbol{u}, \Delta_1 \boldsymbol{M} \Delta_1 \boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{N} \boldsymbol{u})} \tag{92}$$

We first note from the definition of $\boldsymbol{N}$ in (75) that

$$(\boldsymbol{u}, \boldsymbol{N}\boldsymbol{u}) = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = N. \tag{93}$$

The expectation of $(\boldsymbol{u}, \Delta_2 \boldsymbol{M} \boldsymbol{u})$ is

$$E[(\boldsymbol{u}, \Delta_2 \boldsymbol{M} \boldsymbol{u})] = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, \varepsilon^2 V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} = N\varepsilon^2. \tag{94}$$

The expectation of $(\boldsymbol{u}, \Delta_1 \boldsymbol{M} \Delta_1 \boldsymbol{u})$ is

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

$$
\begin{aligned}
&= -\sum_{\alpha,\beta=1}^{N} \frac{(\boldsymbol{u}, E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\beta^\top]\boldsymbol{u})(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\beta)}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\beta]\boldsymbol{u})} \\
&= -\varepsilon^2 \sum_{\alpha=1}^{N} \frac{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha)}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\
&= -\varepsilon^2 \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha)}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} \\
&= -\varepsilon^2 \sum_{\alpha=1}^{N} \frac{\mathrm{tr}[\bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top]}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} \\
&= -\varepsilon^2 \mathrm{tr}\Big[\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}\Big] \\
&= -\varepsilon^2 \mathrm{tr}[\bar{\boldsymbol{M}}^- \bar{\boldsymbol{M}}] = -\varepsilon^2 \mathrm{tr}[\boldsymbol{P}_{\mathbf{u}}] = -(p-1)\varepsilon^2, \qquad (95)
\end{aligned}
$$

Thus, from (92) we have

$$
E[\Delta_2 c] = \Big(1 - \frac{p-1}{N}\Big)\varepsilon^2, \qquad (96)
$$

and hence from (91)

$$
\begin{aligned}
E[\Delta_2 \boldsymbol{u}^\perp] = &\varepsilon^2 \bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\
&+\varepsilon^2 \bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2} \\
&-\frac{p-1}{N}\varepsilon^2 \bar{\boldsymbol{M}}^- \boldsymbol{N}\boldsymbol{u}. \qquad (97)
\end{aligned}
$$

(96) corresponds to the well known formula of unbiased estimation of the noise variance ε^2 (note that the p-dimensional unit vector $\boldsymbol{u}$ has $p-1$ degrees of freedom). As in the case of the Taubin method (see 52), the last term in (97) is much smaller than the corresponding term $\varepsilon^2 \bar{\boldsymbol{M}}^- \boldsymbol{N}\boldsymbol{u}$ in (79).

Kanatani's renormalization was at first not well understood. This was due to the generally held preconception that parameter estimation should *minimize* some cost function. People wondered what renormalization was actually minimizing. In this line of thought, Chojnacki et al. (2001) interpreted renormalization be an approximation to ML to be discussed below. We have seen, however, that optimal estimation does *not* necessarily mean minimizing a cost function and that renormalization is an effort to improve accuracy by a direct means. We will see another example in Sect. 4.6 again.

Example 4 Figure 6 is the RMS error plot corresponding to Fig. 5 using the ellipse data in Example 3. The thick solid line is for LS, the dashed line is for optimally weighted LS, and the thick solid line is for renormalization. The dotted line is for the KCR lower bound. Although the plots for optimally weighted LS and renormalization should both be tangent to that of the KCR lower bound at $\sigma = 0$, but not for LS, this is not visible from the figure, again confirming that the performance difference is mostly due to the second order error $\Delta_2\boldsymbol{u}$.

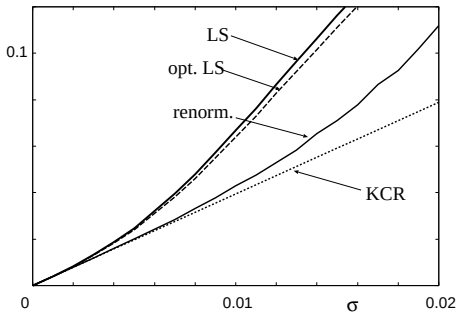


Figure 6: Noise level vs. RMS error for the ellipse data in Fig. 4: LS (*thick solid line*), optimally weighted LS (*dashed line*), renormalization (*thin solid line*), and the KCR lower bound (*dotted line*)

In fact, we can see from Fig. 6 that the accuracy gain of optimally weighted LS over the (unweighted) LS is rather small, meaning that satisfaction of the KCR lower bound in the first order is not a good indicator of high accuracy. In contrast, renormalization performs considerably better than optimally weighted LS, clearly demonstrating that the last term of (79) has a decisive influence on the accuracy. The situation is exactly parallel to the relationship between LS and the Taubin method (Fig. 5). □

4.5 Maximum Likelihood (ML)

We are assuming that we observe noisy data $\boldsymbol{\xi}_\alpha = \bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha$, $\alpha = 1, ..., N$ and that their true values $\bar{\boldsymbol{\xi}}_\alpha$ satisfy the constraint

$$
(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{u}) = 0, \qquad \alpha = 1, ..., N. \qquad (98)
$$

If the noise terms $\Delta\boldsymbol{\xi}_\alpha$ are subject to independent Gaussian of mean $\mathbf{0}$ and covariance matrix $V[\boldsymbol{\xi}_\alpha]$ $(= \varepsilon^2 V_0[\boldsymbol{\xi}_\alpha])$, maximum likelihood (ML) reduces to minimization of the Mahalanobis distance

$$
J = \sum_{\alpha=1}^{2} (\boldsymbol{\xi}_\alpha - \bar{\boldsymbol{\xi}}_\alpha, V_0[\boldsymbol{\xi}_0](\boldsymbol{\xi}_\alpha - \bar{\boldsymbol{\xi}}_\alpha)), \qquad (99)
$$

subject to (98). Introducing Lagrange multipliers, we can eliminate the constraint of (98) and rewrite (99) as

$$
J = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})}. \qquad (100)
$$

Differentiating this with respect to $\boldsymbol{u}$, we obtain

$$\nabla_{\mathbf{u}} J = \sum_{\alpha=1}^{N} \frac{2(\boldsymbol{\xi}_\alpha, \boldsymbol{u})\boldsymbol{\xi}_\alpha}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})} - \sum_{\alpha=1}^{N} \frac{2(\boldsymbol{\xi}_\alpha, \boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \quad (101)$$

Hence, the ML estimator $\hat{\boldsymbol{u}}$ is the solution of

$$\hat{\boldsymbol{M}}\hat{\boldsymbol{u}} = \hat{\boldsymbol{L}}\hat{\boldsymbol{u}}, \quad (102)$$

where $\hat{\boldsymbol{M}}$ is defined by (62) and $\hat{\boldsymbol{L}}$ is given by

$$\hat{\boldsymbol{L}} = \sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \hat{\boldsymbol{u}})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\hat{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})^2}. \quad (103)$$

Equation (102) can be solved by various numerical means. The *FNS* (*fundamental numerical scheme*) of Chojnacki et al. (2000) reduces (102) to iterative eigenvalue computing (see Appendix 5); the *HEIV* (*heteroscedastic errors-in-variable*) of Leedan and Meer (2000) reduces it to iterative generalized eigenvalue computing (see Appendix 6). We can also use the projective Gauss-Newton iterations of Kanatani and Sugaya (2006, 2007) (see Appendix 7). We now analyze the accuracy of the resulting ML estimator.

Whatever iterative scheme is used, (102) holds in the end. Its perturbation expansion is

$$(\bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M} + \Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M} + \cdots - \Delta_2\boldsymbol{L} - \Delta_2^*\boldsymbol{L})(\bar{\boldsymbol{u}} + \Delta_1\boldsymbol{u} + \Delta_2\boldsymbol{u} + \cdots) = \boldsymbol{0}, \quad (104)$$

where

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

$$\Delta_2^*\boldsymbol{L} = \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, \Delta_1\boldsymbol{u})^2 V_0[\boldsymbol{\xi}_\alpha]}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2} + 2\sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, \Delta_1\boldsymbol{u})(\Delta\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{u}})V_0[\boldsymbol{\xi}_\alpha]}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2}. \quad (105)$$

Note that (103) vanishes if $\boldsymbol{\xi}_\alpha$ and $\hat{\boldsymbol{u}}$ are replaced by $\bar{\boldsymbol{\xi}}_\alpha$ and $\boldsymbol{u}$, respectively. Hence, the 0th order term of $\boldsymbol{L}$ is $\boldsymbol{O}$. Since (103) contains the quadratic term $(\boldsymbol{\xi}_\alpha, \hat{\boldsymbol{u}})^2$, the first order perturbations $\Delta_1\boldsymbol{L}$ and $\Delta_1^*\boldsymbol{L}$ are also $\boldsymbol{O}$.

Equating first and second order terms on both sides of (105), we obtain

$$\bar{\boldsymbol{M}}\Delta_1\boldsymbol{u} + (\Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M})\bar{\boldsymbol{u}} = \boldsymbol{0}, \quad (106)$$

$$\bar{\boldsymbol{M}}\Delta_2\boldsymbol{u} + (\Delta_1\boldsymbol{M} + \Delta_1^*\boldsymbol{M})\Delta_1\boldsymbol{u} + (\Delta_2\boldsymbol{M} + \Delta_2^*\boldsymbol{M} - \Delta_2\boldsymbol{L} - \Delta_2^*\boldsymbol{L})\bar{\boldsymbol{u}} = \boldsymbol{0}. \quad (107)$$

Multiplying $\bar{\boldsymbol{M}}^-$ on both sides of (106) and solving for $\Delta_1\boldsymbol{u}$, we again obtain (68). Hence, its covariance matrix $V[\Delta_1\boldsymbol{u}]$ coincides with the KCR lower bound $\varepsilon^2\bar{\boldsymbol{M}}^-$.

Multiplying $\bar{\boldsymbol{M}}^-$ on both sides of (107) and solving for $\Delta_2\boldsymbol{u}^\perp$, we obtain

$$\begin{aligned}\Delta_2\boldsymbol{u}^\perp =& -\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\Delta_1\boldsymbol{u} - \bar{\boldsymbol{M}}^-\Delta_1^*\boldsymbol{M}\Delta_1\boldsymbol{u} - \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{M}\bar{\boldsymbol{u}} \\ & - \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{M}\bar{\boldsymbol{u}} + \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{L}\bar{\boldsymbol{u}} + \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\bar{\boldsymbol{u}} \\ =& \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{u}} + \bar{\boldsymbol{M}}^-\Delta_1^*\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{u}} \\ & - \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{M}\bar{\boldsymbol{u}} - \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{M}\bar{\boldsymbol{u}} \\ & + \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{L}\bar{\boldsymbol{u}} + \bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\bar{\boldsymbol{u}} \end{aligned} \quad (108)$$

For computing its expectation, we only need to consider the new terms $\bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{L}\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\bar{\boldsymbol{u}}$. First, we see that

$$\begin{aligned} & E[\bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{L}\bar{\boldsymbol{u}}] \\ &= \bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{u}}, E[\Delta\bar{\boldsymbol{\xi}}_\alpha\Delta\bar{\boldsymbol{\xi}}_\alpha^\top]\bar{\boldsymbol{u}})V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2} \\ &= \bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{u}}, \varepsilon^2 V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2} \\ &= \varepsilon^2\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})} = \varepsilon^2\bar{\boldsymbol{M}}^-\boldsymbol{N}\bar{\boldsymbol{u}}. \end{aligned} \quad (109)$$

For $\bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\bar{\boldsymbol{u}}$, we have

$$\begin{aligned} E[\bar{\boldsymbol{M}}^-\Delta_2^*\boldsymbol{L}\bar{\boldsymbol{u}}] =& \bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top]\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2} \\ & + 2\bar{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\bar{\boldsymbol{\xi}}_\alpha, E[\Delta_1\boldsymbol{u}\Delta\bar{\boldsymbol{\xi}}_\alpha^\top]\bar{\boldsymbol{u}})V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2}. \end{aligned} \quad (110)$$

We have already seen that the first order error $\Delta_1\boldsymbol{u}$ satisfies the KCR lower bound, so $E[\Delta_1\boldsymbol{u}\Delta_1\boldsymbol{u}^\top] = \varepsilon\bar{\boldsymbol{M}}^-$ (see (59)). On the other hand,

$$\begin{aligned} & E[\Delta_1\boldsymbol{u}\Delta\bar{\boldsymbol{\xi}}_\alpha^\top]\bar{\boldsymbol{u}} \\ &= -E[\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{u}}\Delta\bar{\boldsymbol{\xi}}_\alpha^\top]\bar{\boldsymbol{u}} \\ &= -\bar{\boldsymbol{M}}^- E\Big[\sum_{\beta=1}^{N} \frac{\Delta\boldsymbol{\xi}_\beta\bar{\boldsymbol{\xi}}_\beta^\top + \bar{\boldsymbol{\xi}}_\beta\Delta\boldsymbol{\xi}_\beta^\top}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{u}})}\bar{\boldsymbol{u}}\Delta\bar{\boldsymbol{\xi}}_\alpha^\top\Big] \\ &= -\bar{\boldsymbol{M}}^- \sum_{\beta=1}^{N} \frac{(\bar{\boldsymbol{u}}, E[\Delta\boldsymbol{\xi}_\beta\Delta\bar{\boldsymbol{\xi}}_\alpha^\top]\bar{\boldsymbol{u}})\bar{\boldsymbol{\xi}}_\beta}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{u}})} \\ &= -\varepsilon^2\bar{\boldsymbol{M}}^- \frac{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})\bar{\boldsymbol{\xi}}_\alpha}{(\bar{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})} = -\varepsilon^2\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha. \end{aligned} \quad (111)$$

Hence,

$$
\begin{aligned}
&E[\bar{\boldsymbol{M}}^{-}\Delta_2^*\boldsymbol{L}\bar{\boldsymbol{u}}]\\
&=\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2}\\
&\quad-2\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}\boldsymbol{\xi}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2}\\
&=-\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}}}{(\bar{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{u}})^2}.
\end{aligned}
\tag{112}
$$

Adding (109) and (112) to (79), we conclude that

$$
E[\Delta_2\boldsymbol{u}^{\perp}]=\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha,V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}. \tag{113}
$$

Comparing this with (79) and (97), we can see the last two terms there are removed.

Note 1. Optimally weighted LS vs. ML

There has been a widespread misunderstanding that optimally weighted LS can actually compute ML because (55) is identical to (100) if the weight W_α is chosen as in (58). However, this is not so (Chojnacki et al. 2000; Kanatani 1996).

Optimally weighted LS minimizes J in (100) for $\boldsymbol{u}$ in the numerator with $\boldsymbol{u}$ in the denominator fixed. Then, the resulting solution $\boldsymbol{u}$ is substituted into the denominator, followed by the minimization of J for $\boldsymbol{u}$ in the numerator, and this is iterated. This means that when the solution $\hat{\boldsymbol{u}}$ is obtained, it is guaranteed that

$$
\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\hat{\boldsymbol{u}}+\delta\boldsymbol{u})^2}{(\hat{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})}\geq\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\hat{\boldsymbol{u}})^2}{(\hat{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})}, \tag{114}
$$

for any infinitesimal perturbation $\delta\boldsymbol{u}$, which the convergence of optimally weighted LS means. This, however, does not guarantee that

$$
\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\hat{\boldsymbol{u}}+\delta\boldsymbol{u})^2}{(\hat{\boldsymbol{u}}+\delta\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha](\hat{\boldsymbol{u}}+\delta\boldsymbol{u}))}\geq\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\hat{\boldsymbol{u}})^2}{(\hat{\boldsymbol{u}},V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})}, \tag{115}
$$

for any infinitesimal perturbation $\delta\boldsymbol{u}$, which minimization of J really means.

Note 2. Gaussian noise assumption

There is a subtle point to be clarified. If the constraint is the "linear" form of (98), the transition from (99) to (100) is *exact*; no approximation is used. However, the crucial assumption is that the noise in $\boldsymbol{\xi}_\alpha$ is *Gaussian*, because then and only then is the likelihood proportional to $e^{-J/\text{constant}}$. If the noise is not Gaussian, minimization of J in (99) is not ML in its strict sense.

This issue arises when one linearizes the problem by changing variables in the form of $\boldsymbol{\xi}=\boldsymbol{\xi}(\boldsymbol{\xi})$ as described in Sect. 3.4. Namely, if the noise in the $\boldsymbol{x}$-space is assumed Gaussian, the corresponding noise in the transformed $\boldsymbol{\xi}$-space is no longer Gaussian. Hence, ML in the strict sense (also called "Golden Standard" (Hartley and Zisserman 2000) should minimize the Mahalanobis distance in the $\boldsymbol{x}$-space, which is different from the Mahalanobis distance in the $\boldsymbol{\xi}$-space.

If we transform the the Mahalanobis distance in the $\boldsymbol{\xi}$-space into the $\boldsymbol{\xi}$-space, extra higher-order terms are introduced in the numerator on the right-hand side of (100). The effect of perturbation in the denominator is of higher order, since the numerator is $O(\varepsilon^2)$, while the denominator is $O(1)$, justifying the computation of (18) and (19) (in fact, we experimentally confirmed that adding second order terms in the denominator have no perceptible differences in the final result). However, the small changes in the numerator exhibit perceptible differences when the noise level ε is large.

We do not go into the details of this issue by two reasons, not to mention the space limitation. First, we have confirmed by experiments that no differences are observed in the small noise range over which our numerical experiments are conducted. The second is a practical issue. If the noise is large, the Gaussian noise assumption is questionable. There may be systematic errors and outlying data; their removal is more important than optimal estimation based on the Gaussian noise assumption. Hence, it does not have much sense arguing which to adopt, ML based on Gaussian noise in the $\boldsymbol{x}$-space or ML based on Gaussian noise in the $\boldsymbol{\xi}$-space, if the Gaussian noise assumption is questionable.

The motivation of our analysis is that the result should be used in precise control such as visual sensing of robots, as mentioned in Sect. 2.5. In such a domain, useful applications are possible only when the noise is made small (e.g., to a few pixels or subpixels) by improving measurement devices and image processing operations. For such applications, the above analysis seems satisfactory.

4.6 Hyperaccuracy Fitting

(113) implies the possibility of improving the accuracy of ML further. Namely, we "subtract" (113) from the ML estimator $\hat{\boldsymbol{u}}$ (Kanatani 2006). Of course, (113) cannot be precisely computed, because it involves the true values $\bar{\boldsymbol{\xi}}_\alpha$ and $\boldsymbol{u}$. So, we plug in the data $\boldsymbol{\xi}_\alpha$ and the ML estimator $\hat{\boldsymbol{u}}$. As is well known, the unknown squared noise level ε^2 is estimated from the residual of (100) in the following form (Kanatani 1996):

$$
\hat{\varepsilon}^2=\frac{(\hat{\boldsymbol{u}},\hat{\boldsymbol{M}}\hat{\boldsymbol{u}})}{N-(p-1)}. \tag{116}
$$

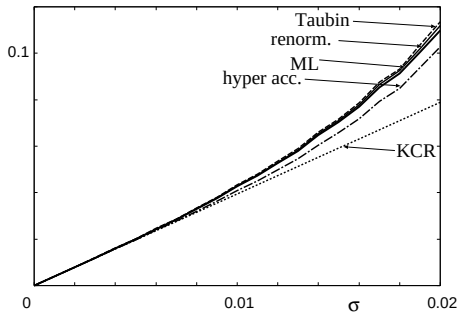
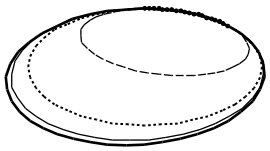


Figure 7: Noise level vs. RMS error for the ellipse data in Fig. 4: Taubin (*dashed line*), renormalization (*thin solid line*), ML (*thick solid line*), hyperaccurate correction (*chained line*), and the KCR lower bound (*dotted line*)

Figure 8: One instance of ellipse fitting: LS (*broken line*), ML (*thick solid line*), hyperaccuracy correction (*thin solid line*), true ellipse (*dotted line*)



Hence, the correction has the form

$$\tilde{\boldsymbol{u}} = N\Big[\hat{\boldsymbol{u}} - \hat{\varepsilon}^2 \hat{\boldsymbol{M}}^- \sum_{\alpha=1}^{N} \frac{(\hat{\boldsymbol{M}}^- \boldsymbol{\xi}_\alpha, V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})\boldsymbol{\xi}_\alpha}{(\hat{\boldsymbol{u}}, V_0[\boldsymbol{\xi}_\alpha]\hat{\boldsymbol{u}})^2}\Big], \qquad (117)$$

where the operation $N[\,\cdot\,]$ denotes normalization to unit norm for compensating for the parallel component $\Delta\boldsymbol{u}^{\parallel}$ (see Fig. 3).

Example 5 Figure 7 shows the RMS error plot corresponding to Figs. 5 and 6, using the ellipse data in the Example 3. The dashed line is for the Taubin method, the thin line is for renormalization, and the thick solid line is for ML; we used the FNS of Chojnacki et al. (2000) for computing ML. The dotted line indicates the KCR lower bound.

We can see that in spite of the drastic bias reduction of ML "in form" ((113)), as compared to the Taubin method ((79)) and renormalization ((97)), *ML has only comparable accuracy to the Taubin method and renormalization*. The chained line shows the result of the hyperaccurate correction of (117). We can see that the error is further reduced.

Figure 8 shows one instance of ellipse fitting ($\sigma = 0.015$). 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 the hyperaccurate correction. We can see that the fitted ellipse is closer to the true shape after the correction.

For comparing all the methods tested so far, we define the "error ratio" $D/D_{\rm KCR}$ by D in (54) divided by $D_{\rm KCR}$ and average it over the tested range of σ. Table 3 list this value for different method. □

Table 3: Average error ratio of different methods

LS	1.636
Optimally weighted LS	1.575
Taubin	1.144
Renormalization	1.133
ML	1.125
Hyperaccurate correction	1.007
KCR lower bound	1.000

5. Conclusions

We have given a rigorous accuracy analysis of various techniques for geometric fitting. We first pointed out how our problem is different from traditional statistical analysis and explained why we need a different framework. After giving general theories in our new framework, we selected typical techniques and analyzed their accuracy up to second order terms. Table 2 summarizes the first order error, its covariance matrix, and the second order bias. Conducting numerical simulations of ellipse fitting, we have observed the following:

1. LS and the Taubin method have the same error to a first approximation. However, the latter achieves much higher accuracy, because a dominant second order bias term of LS is reduced.
2. Optimally weighted LS achieves the KCR lower bound to a first approximation. However, the accuracy gain over (unweighted) LS is rather small. This is due to the existence of second order bias terms.
3. Renormalization reduces the dominant bias term of optimally weighted LS, resulting in considerable accuracy improvement.
4. ML is less biased than renormalization. Yet, the accuracy gain is rather small; ML, renormalization, and Taubin all have similar accuracy behavior.
5. By estimating and subtracting the bias term from the ML solution, we can achieve higher accuracy than ML ("hyperaccuracy").

Thus, we conclude that it is the *second order error*, not the first, that has dominant effects over the accuracy. We have also found that *not all second order terms have the same degree of influence*. The influence of individual bias terms on accuracy is very difficult to analyze in the geometric fitting framework, where all basic relationships are only implicitly described. Also, there may be a possibility that adding a special bias term may increase the accuracy, which is also very difficult to analyze. These are left for future research.

Table 4: Summary of the first order error and the second order bias

Method	First order error & covariance matrix	Second order bias
LS	$-\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}$ $\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\bar{\boldsymbol{M}}'_{\mathrm{LS}}\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}$	$\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\sum_{\alpha=1}^{N}(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha+\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\sum_{\alpha=1}^{N}(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}-\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\boldsymbol{N}_{\mathrm{LS}}\boldsymbol{u}$
Taubin	$-\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\Delta_1\boldsymbol{M}_{\mathrm{LS}}\boldsymbol{u}$ $\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\bar{\boldsymbol{M}}'_{\mathrm{LS}}\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}$	$\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\sum_{\alpha=1}^{N}(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha+\varepsilon^2\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\sum_{\alpha=1}^{N}(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}-\frac{q}{(\boldsymbol{u},\boldsymbol{N}_{\mathrm{LS}}\boldsymbol{u})}\bar{\boldsymbol{M}}^{-}_{\mathrm{LS}}\boldsymbol{N}_{\mathrm{LS}}\boldsymbol{u}$
opt. LS	$-\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u}$ $\varepsilon^2\bar{\boldsymbol{M}}^{-}$	$\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha,V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}+\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}-\varepsilon^2\bar{\boldsymbol{M}}^{-}\boldsymbol{N}\boldsymbol{u}$
renorm.	$-\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u}$ $\varepsilon^2\bar{\boldsymbol{M}}^{-}$	$\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha,V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}+\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{\xi}}_\alpha,\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u}}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}-\frac{p-1}{N}\varepsilon^2\bar{\boldsymbol{M}}^{-}\boldsymbol{N}\boldsymbol{u}$
ML	$-\bar{\boldsymbol{M}}^{-}\Delta_1\boldsymbol{M}\boldsymbol{u}$ $\varepsilon^2\bar{\boldsymbol{M}}^{-}$	$\varepsilon^2\bar{\boldsymbol{M}}^{-}\sum_{\alpha=1}^{N}\frac{(\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha,V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})\bar{\boldsymbol{\xi}}_\alpha}{(\boldsymbol{u},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{u})^2}$

Acknowledgments: The author thanks Wojciech Chojnacki of the University of Adelaide, Australia, Nikolai Chernov of the University of Alabama at Birmingham, U.S.A., Steve Maybank of the University of London, Birkbeck, U.K., and Peter Meer of the Rutgers University, U.S.A., for helpful discussions. He also thanks Yasuyuki Sugaya of Toyohashi University of Technology, Japan, and Junpei Yamada of RyuSho Industrial Co., Ltd., Japan, for helping with numerical experiments.

Appendix 1: Derivation of the KCR Lower Bound

For simplicity, we consider only the case where no intrinsic constraints exist on the data $\boldsymbol{x}_\alpha$ or the parameter $\boldsymbol{u}$, assuming that the noise is identical and isotropic Gaussian with mean 0 and variance ε^2. In other words, we assume that the probability density of each datum $\boldsymbol{x}_\alpha$ is

$$p(\boldsymbol{x}_\alpha)=\frac{1}{(\sqrt{2\pi})^n\varepsilon^n}e^{-\|\mathbf{x}_\alpha-\bar{\mathbf{x}}_\alpha\|^2/2\varepsilon^2}. \tag{118}$$

Suppose an unbiased estimator $\hat{\boldsymbol{u}}(\boldsymbol{x}_1,...,\boldsymbol{x}_N)$ is given. Its unbiasedness mean

$$E[\hat{\boldsymbol{u}}-\boldsymbol{u}]=\boldsymbol{0}, \tag{119}$$

where $E[\,\cdot\,]$ is expectation over the joint probability density $p(\boldsymbol{x}_1)\cdots p(\boldsymbol{x}_N)$. Since this density is parameterized by the true data values $\bar{\boldsymbol{x}}_\alpha$, (119) can be viewed as an equation of $\bar{\boldsymbol{x}}_\alpha$ as well as the unknown $\boldsymbol{u}$. The crucial fact is that (119) should be an *identity* in $\bar{\boldsymbol{x}}_\alpha$ and $\boldsymbol{u}$ that satisfies (1), because unbiasedness is a "property" of the estimator $\hat{\boldsymbol{u}}$ that should hold for whatever values of $\bar{\boldsymbol{x}}_\alpha$ and $\boldsymbol{u}$. Hence, (119) should be invariant to infinitesimal variation of $\bar{\boldsymbol{x}}_\alpha$ and $\boldsymbol{u}$. This means

$$\begin{aligned}
&\delta\int(\hat{\boldsymbol{u}}-\boldsymbol{u})p_1\cdots p_N d\boldsymbol{x}\\
&=-\int(\delta\boldsymbol{u})p_1\cdots p_N d\boldsymbol{x}\\
&\quad+\sum_{\alpha=1}^{N}\int(\hat{\boldsymbol{u}}-\boldsymbol{u})p_1\cdots\delta p_\alpha\cdots p_N d\boldsymbol{x}\\
&=-\delta\boldsymbol{u}+\int(\hat{\boldsymbol{u}}-\boldsymbol{u})\sum_{\alpha=1}^{N}(p_1\cdots\delta p_\alpha\cdots p_N)d\boldsymbol{x},
\end{aligned}$$

where p_α is an abbreviation of $p(\boldsymbol{x}_\alpha)$ and $\int d\boldsymbol{x}$ is a shorthand of $\int\cdots\int d\boldsymbol{x}_1\cdots\boldsymbol{x}_N$. Note that we consider variations in $\bar{\boldsymbol{x}}_\alpha$ (not $\boldsymbol{x}_\alpha$) and $\boldsymbol{u}$. Since the estimator $\hat{\boldsymbol{u}}$ is a function of the data $\boldsymbol{x}_\alpha$, it is not affected by such variations. Since the variation $\delta\boldsymbol{u}$ is independent of $\boldsymbol{x}_\alpha$, it can be moved outside the integral $\int d\boldsymbol{x}$. Also note that $\int p_1\cdots p_N d\boldsymbol{x}=1$.

The infinitesimal variation of (118) with respect to $\bar{\boldsymbol{x}}_\alpha$ is

$$\delta p_\alpha=(\boldsymbol{l}_\alpha,\delta\bar{\boldsymbol{x}}_\alpha)p_\alpha, \tag{121}$$

where we define the *score* $\boldsymbol{l}_\alpha$ by

$$\boldsymbol{l}_\alpha\equiv\nabla_{\bar{\mathbf{x}}_\alpha}\log p_\alpha=\frac{\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}_\alpha}{\varepsilon^2}. \tag{122}$$

Since (119) is an identity in $\bar{\boldsymbol{x}}_\alpha$ and $\boldsymbol{u}$ that satisfies (1), the variation (120) should vanish for arbitrary infinitesi-

mal variations $\delta\bar{\boldsymbol{x}}_\alpha$ and $\delta\boldsymbol{u}$ that are compatible with (1). If (121) is substituted into (120), its vanishing means

$$E[(\hat{\boldsymbol{u}}-\boldsymbol{u})\sum_{\alpha=1}^{N}\boldsymbol{l}_\alpha^\top\delta\bar{\boldsymbol{x}}_\alpha]=\delta\boldsymbol{u}. \tag{123}$$

The infinitesimal variation of (1) has the form

$$(\nabla_{\mathbf{x}}\bar{F}_\alpha,\delta\bar{\boldsymbol{x}}_\alpha)+(\nabla_{\mathbf{u}}\bar{F}_\alpha,\delta\boldsymbol{u})=0, \tag{124}$$

where $\nabla_{\mathbf{x}}\bar{F}_\alpha$ and $\nabla_{\mathbf{u}}\bar{F}_\alpha$ mean gradients of $F(\boldsymbol{x},\boldsymbol{u})$ with respect to $\boldsymbol{x}$ and $\boldsymbol{u}$, respectively, evaluated at $\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha$. Consider the following particular variations $\delta\bar{\boldsymbol{x}}_\alpha$:

$$\delta\bar{\boldsymbol{x}}_\alpha=-\frac{(\nabla_{\mathbf{x}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\delta\boldsymbol{u}. \tag{125}$$

Evidently, (124) is satisfied for whatever $\delta\boldsymbol{u}$. Substituting (125) into (123), we obtain

$$E\Big[(\hat{\boldsymbol{u}}-\boldsymbol{u})\sum_{\alpha=1}^{N}\boldsymbol{m}_\alpha^\top\Big]\delta\boldsymbol{u}=-\delta\boldsymbol{u}, \tag{126}$$

where we define the vectors $\boldsymbol{m}_\alpha$ by

$$\boldsymbol{m}_\alpha=\frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{x}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\boldsymbol{l}_\alpha. \tag{127}$$

Since (126) should hold for arbitrary variation $\delta\boldsymbol{u}$, we have

$$E[(\hat{\boldsymbol{u}}-\boldsymbol{u})\sum_{\alpha=1}^{N}\boldsymbol{m}_\alpha^\top]=-\boldsymbol{I}. \tag{128}$$

Hence, we have

$$E\Big[\begin{pmatrix}\hat{\boldsymbol{u}}-\boldsymbol{u}\\ \sum_{\alpha=1}^{N}\boldsymbol{m}_\alpha\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{u}}-\boldsymbol{u}\\ \sum_{\alpha=1}^{N}\boldsymbol{m}_\alpha\end{pmatrix}^\top\Big]=\begin{pmatrix}V[\hat{\boldsymbol{u}}] & -\boldsymbol{I}\\ -\boldsymbol{I} & \boldsymbol{M}\end{pmatrix}, \tag{129}$$

where we define the matrix $\boldsymbol{M}$ by

$$\begin{aligned}\boldsymbol{M}&=E\Big[\Big(\sum_{\alpha=1}^{N}\boldsymbol{m}_\alpha\Big)\Big(\sum_{\beta=1}^{N}\boldsymbol{m}_\beta\Big)^\top\Big]\\&=\sum_{\alpha,\beta=1}^{N}\frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{x}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}E[\boldsymbol{l}_\alpha\boldsymbol{l}_\beta]\frac{(\nabla_{\mathbf{x}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\\&=\frac{1}{\varepsilon^2}\frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}.\end{aligned} \tag{130}$$

In the above equation, we use the identity $E[\boldsymbol{l}_\alpha\boldsymbol{l}_\beta^\top]=\delta_{\alpha\beta}\boldsymbol{I}/\varepsilon^4$, which is a consequence of independence of the noise in each datum $\boldsymbol{x}_\alpha$.

Since the inside of the expectation $E[\,\cdot\,]$ on the left-hand side of (129) is positive semidefinite, so is the right-hand side. Hence, the following is also positive semidefinite:

$$\begin{aligned}&\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^{-1}\\ & \boldsymbol{M}^{-1}\end{pmatrix}\begin{pmatrix}V[\hat{\boldsymbol{u}}] & -\boldsymbol{I}\\ -\boldsymbol{I} & \boldsymbol{M}\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & \\ \boldsymbol{M}^{-1} & \boldsymbol{M}^{-1}\end{pmatrix}\\&\quad=\begin{pmatrix}V[\hat{\boldsymbol{u}}]-\boldsymbol{M}^{-1} & \\ & \boldsymbol{M}^{-1}\end{pmatrix}.\end{aligned} \tag{131}$$

From this, we conclude that

$$V[\hat{\boldsymbol{u}}]\succ\boldsymbol{M}^{-1}. \tag{132}$$

This result is easily generalized to the case where intrinsic constraints exist on the data $\boldsymbol{x}_\alpha$ and the parameter $\boldsymbol{u}$ and the covariance matrix $V[\boldsymbol{x}_\alpha]$ is not full rank (Kanatani 1996). In the general case, we obtain (9).

Appendix 2: Linear Approximation of ML

For simplicity, we consider only the case where no intrinsic constraints exist on the data $\boldsymbol{x}_\alpha$ or the parameter $\boldsymbol{u}$, assuming that the noise is identical and isotropic Gaussian. Substituting $\bar{\boldsymbol{x}}_\alpha=\boldsymbol{x}_\alpha-\Delta\boldsymbol{x}_\alpha$ into (12) and assuming that the noise term $\Delta\boldsymbol{x}_\alpha$ is small, we obtain the linear approximation

$$F_\alpha-(\nabla_{\mathbf{x}}F_\alpha,\Delta\boldsymbol{x}_\alpha)=0, \tag{133}$$

subject to which we want to minimize $\sum_{\alpha=1}^{N}\|\Delta\boldsymbol{x}_\alpha\|^2$. Introducing Lagrange multipliers λ_α, let

$$L=\frac{1}{2}\sum_{\alpha=1}^{N}\|\Delta\boldsymbol{x}_\alpha\|^2+\sum_{\alpha=1}^{N}\lambda_\alpha(F_\alpha-(\nabla_{\mathbf{x}}F_\alpha,\Delta\boldsymbol{x}_\alpha)). \tag{134}$$

Taking the derivative of L with respect to $\Delta\boldsymbol{x}_\alpha$ and setting it to $\boldsymbol{0}$, we have

$$\Delta\boldsymbol{x}_\alpha-\lambda_\alpha\nabla_{\mathbf{x}}F_\alpha=\boldsymbol{0}. \tag{135}$$

Hence, $\Delta\boldsymbol{x}_\alpha=\lambda_\alpha\nabla_{\mathbf{x}}F_\alpha$. Substitution of this into (133) yields

$$F_\alpha-(\nabla_{\mathbf{x}}F_\alpha,\lambda_\alpha\nabla_{\mathbf{x}}F_\alpha)=0, \tag{136}$$

from which we obtain λ_α in the form

$$\lambda_\alpha=\frac{F_\alpha}{\|\nabla_{\mathbf{x}}F_\alpha\|^2}. \tag{137}$$

Thus,

$$\begin{aligned}J&=\sum_{\alpha=1}^{N}\|\Delta\boldsymbol{x}_\alpha\|^2=\sum_{\alpha=1}^{N}\|\lambda_\alpha\nabla_{\mathbf{x}}F_\alpha\|^2\\&=\sum_{\alpha=1}^{N}\frac{F_\alpha^2}{\|\nabla_{\mathbf{x}}F_\alpha\|^4}\|\nabla_{\mathbf{x}}F_\alpha\|^2=\sum_{\alpha=1}^{N}\frac{F_\alpha^2}{\|\nabla_{\mathbf{x}}F_\alpha\|^2}.\end{aligned} \tag{138}$$

This result can easily be generalized to the case where intrinsic constraints exist on the data $\boldsymbol{x}_\alpha$ and the parameter $\boldsymbol{u}$ and the covariance matrix $V[\boldsymbol{x}_\alpha]$ is not full rank (Kanatani 1996). In the general case, we obtain (15).

Appendix 3: Covariance Matrix of ML

For simplicity, we consider only the case where no intrinsic constraints exist on the data $\boldsymbol{x}_\alpha$ or the parameter $\boldsymbol{u}$, assuming that the noise is identical and isotropic Gaussian with mean 0 and variance ε^2, so $V[\boldsymbol{x}_\alpha] = \varepsilon^2\boldsymbol{I}$. Letting $\boldsymbol{x}_\alpha = \bar{\boldsymbol{x}}_\alpha + \Delta\boldsymbol{x}_\alpha$ and replacing $\boldsymbol{u}$ by $\boldsymbol{u} + \Delta\boldsymbol{u}$ in (15), we can expand J in the form

$$J = \sum_{\alpha=1}^{N} \frac{((\nabla_{\mathbf{x}}\bar{F}_\alpha, \Delta\boldsymbol{x}_\alpha) + (\nabla_{\mathbf{u}}\bar{F}_\alpha, \Delta\boldsymbol{u}))^2}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2} + O(\varepsilon^3), \quad (139)$$

where $\nabla_{\mathbf{x}}\bar{F}_\alpha$ and $\nabla_{\mathbf{u}}\bar{F}_\alpha$ have the same meaning as in (124). Note that replacing $\nabla_{\mathbf{x}}F_\alpha$ by $\nabla_{\mathbf{x}}\bar{F}_\alpha$ by in the denominator does not affect the leading term because the numerator is $O(\varepsilon^2)$; the difference is absorbed into the remainder term $O(\varepsilon^3)$.

If we find $\Delta\boldsymbol{u}$ that minimizes (139), the ML estimator $\hat{\boldsymbol{u}}$ is given by $\boldsymbol{u} + \Delta\boldsymbol{u}$. Since the first term on the right-hand side of (139) is quadratic in $\Delta\boldsymbol{u}_\alpha$, the derivative of J with respect to $\Delta\boldsymbol{u}$ is

$$2\sum_{\alpha=1}^{N} \frac{((\nabla_{\mathbf{x}}\bar{F}_\alpha, \Delta\boldsymbol{x}_\alpha) + (\nabla_{\mathbf{u}}\bar{F}_\alpha, \Delta\boldsymbol{u}))\nabla_{\mathbf{u}}\bar{F}_\alpha}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2} + O(\varepsilon^2). \quad (140)$$

Letting this be 0, we have

$$\sum_{\alpha=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\Delta\boldsymbol{u} = -\sum_{\alpha=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{x}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\Delta\boldsymbol{x}_\alpha + O(\varepsilon^2), \quad (141)$$

from which we obtain

$$\sum_{\alpha=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\Delta\boldsymbol{u}\Delta\boldsymbol{u}^\top \sum_{\beta=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\beta)(\nabla_{\mathbf{u}}\bar{F}_\beta)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\beta\|^2} = \sum_{\alpha,\beta=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{x}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}\Delta\boldsymbol{x}_\alpha\Delta\boldsymbol{x}_\beta^\top \frac{(\nabla_{\mathbf{x}}\bar{F}_\beta)(\nabla_{\mathbf{u}}\bar{F}_\beta)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2} + O(\varepsilon^3). \quad (142)$$

Taking expectation on both sides, we obtain

$$\sum_{\alpha=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2}V[\hat{\boldsymbol{u}}]\sum_{\beta=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\beta)(\nabla_{\mathbf{u}}\bar{F}_\beta)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\beta\|^2} = \sum_{\alpha=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{x}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2} \frac{(\nabla_{\mathbf{x}}\bar{F}_\alpha)(\nabla_{\mathbf{u}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2} + O(\varepsilon^4)$$

$$= \sum_{\alpha=1}^{N} \frac{(\nabla_{\mathbf{u}}\bar{F}_\alpha)(\nabla_{\mathbf{x}}\bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}}\bar{F}_\alpha\|^2} + O(\varepsilon^4). \quad (143)$$

Note that $E[O(\varepsilon^3)] = O(\varepsilon^4)$, because odd noise terms vanish in expectation. The first term in the last expression is the KCR lower bound in this case.

This result can easily be generalized to the case where intrinsic constraints exist on the data $\boldsymbol{x}_\alpha$ and the parameter $\boldsymbol{u}$ and the covariance matrix $V[\boldsymbol{x}_\alpha]$ is not full rank (Kanatani 1996). We conclude that the covariance matrix of the ML estimator agrees with the KCR lower bound except for $O(\varepsilon^4)$.

Appendix 4: Procedure for the Taubin Method

In most vision applications, the embedded data $\boldsymbol{\xi}_\alpha$, the parameter $\boldsymbol{u}$, and the normalized covariance matrix $V_0[\boldsymbol{\xi}_\alpha]$ are decomposed in the form

$$\boldsymbol{\xi}_\alpha \begin{pmatrix} \boldsymbol{z}_\alpha \\ C \end{pmatrix}, \qquad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{v} \\ a \end{pmatrix}, \qquad V_0[\boldsymbol{\xi}_\alpha] = \begin{pmatrix} V_0[\boldsymbol{z}_\alpha] & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 \end{pmatrix}, \quad (144)$$

where C and a are constants; see (21) for ellipse fitting and (24) for fundamental matrix computation. Here, $\boldsymbol{z}_\alpha$ and $\boldsymbol{v}$ are $(p-1)$-dimensional vectors, and $V_0[\boldsymbol{z}_\alpha]$ is a $(p-1)\times(p-1)$ normalized covariance matrix of $\boldsymbol{z}_\alpha$; see (22) and (25).

For computing estimates $\hat{\boldsymbol{v}}$ and $\hat{a}$ of $\boldsymbol{v}$ and a, we define $(p-1)\times(p-1)$ matrices $\tilde{\boldsymbol{M}}_{LS}$ and $\tilde{\boldsymbol{N}}_{\mathrm{LS}}$ by

$$\tilde{\boldsymbol{M}}_{LS} = \sum_{\alpha=1}^{N} \tilde{\boldsymbol{z}}_\alpha\tilde{\boldsymbol{z}}_\alpha^\top, \qquad \tilde{\boldsymbol{N}}_{\mathrm{LS}} = \sum_{\alpha=1}^{N} V_0[\boldsymbol{z}_\alpha], \quad (145)$$

where

$$\tilde{\boldsymbol{z}}_\alpha = \boldsymbol{z}_\alpha - \bar{\boldsymbol{z}}, \qquad \bar{\boldsymbol{z}} = \frac{1}{N}\sum_{\alpha=1}^{N} \boldsymbol{z}_\alpha. \quad (146)$$

Then, (43) splits into two equations

$$\tilde{\boldsymbol{M}}_{LS}\hat{\boldsymbol{v}} = \lambda\tilde{\boldsymbol{N}}_{\mathrm{LS}}\hat{\boldsymbol{v}}, \qquad (\hat{\boldsymbol{v}}, \bar{\boldsymbol{z}}) + C\hat{a} = 0. \quad (147)$$

If we compute the $(p-1)$-dimensional unit generalized eigenvector $\hat{\boldsymbol{v}}$ of the first equation for the smallest generalized eigenvalue λ (see, e.g., Kanatani (1996) for the procedure), the second gives $\hat{a}$. Hence, $\hat{\boldsymbol{u}}$ is given by

$$\hat{\boldsymbol{u}} = N[\begin{pmatrix} \hat{\boldsymbol{v}} \\ \hat{a} \end{pmatrix}], \quad (148)$$

where $N[\,\cdot\,]$ denotes normalization to unit norm.

Appendix 5: Procedure for FNS

The FNS of Chojnacki et al. (2000) solves (102) by the following iterations:

1. Initialize $\hat{\boldsymbol{u}}$, say by LS.
2. Compute the matrix $\hat{\boldsymbol{M}}$ in (62) and the matrix $\hat{\boldsymbol{L}}$ in (103), and solve the eigenvalue problem
$$(\hat{\boldsymbol{M}} - \hat{\boldsymbol{L}})\boldsymbol{u} = \lambda \boldsymbol{u}. \tag{149}$$
Let $\boldsymbol{u}$ be the unit eigenvector for the eigenvalue λ closest to 0.
3. If $\boldsymbol{u} \approx \hat{\boldsymbol{u}}$ except for sign, stop. Else, let $\hat{\boldsymbol{u}} \leftarrow \boldsymbol{u}$ and go back to Step 2.

Later, Chojnacki et al. (2005) pointed out that convergence performance improves if we choose in Step 2 not the eigenvalue closest to 0 but the smallest one. See Kanatani and Sugaya (2006, 2007) for the comparative experiment of this effect.

Appendix 6: Procedure for HEIV

In most vision applications, the embedded data $\boldsymbol{\xi}_\alpha$, the parameter $\boldsymbol{u}$, and the normalized covariance matrix $V_0[\boldsymbol{\xi}_\alpha]$ are decomposed in the form of (144). For computing estimates $\hat{\boldsymbol{v}}$ and $\hat{a}$ of $\boldsymbol{v}$ and a, we define $(p-1)\times(p-1)$ matrices $\tilde{\boldsymbol{M}}$ and $\tilde{\boldsymbol{L}}$ by

$$\tilde{\boldsymbol{M}} = \sum_{\alpha=1}^{N} \frac{\tilde{\boldsymbol{z}}_\alpha \tilde{\boldsymbol{z}}_\alpha^\top}{(\hat{\boldsymbol{v}}, V_0[\boldsymbol{z}_\alpha]\hat{\boldsymbol{v}})},$$
$$\tilde{\boldsymbol{L}} = \sum_{\alpha=1}^{N} \frac{(\hat{\boldsymbol{v}}, \tilde{\boldsymbol{z}}_\alpha)^2 V_0[\boldsymbol{z}_\alpha]}{(\hat{\boldsymbol{v}}, V_0[\boldsymbol{z}_\alpha]\hat{\boldsymbol{v}})^2}, \tag{150}$$

where we put

$$\tilde{\boldsymbol{z}}_\alpha = \boldsymbol{z}_\alpha - \bar{\boldsymbol{z}},$$
$$\bar{\boldsymbol{z}} = \sum_{\alpha=1}^{N} \frac{\boldsymbol{z}_\alpha}{(\hat{\boldsymbol{v}}, V_0[\boldsymbol{z}_\alpha]\hat{\boldsymbol{v}})} \Bigg/ \sum_{\beta=1}^{N} \frac{1}{(\hat{\boldsymbol{v}}, V_0[\boldsymbol{z}_\beta]\hat{\boldsymbol{v}})}. \tag{151}$$

Then, (102) splits into the following two equations (Chojnacki et al. 2004, 2005):

$$\tilde{\boldsymbol{M}}\hat{\boldsymbol{v}} = \tilde{\boldsymbol{L}}\hat{\boldsymbol{v}}, \qquad (\hat{\boldsymbol{v}}, \bar{\boldsymbol{z}}) + C\hat{a} = 0. \tag{152}$$

If determine $\hat{\boldsymbol{v}}$ from the first equation, the second determines $\hat{a}$. Hence, the estimate $\hat{\boldsymbol{u}}$ is given in the form of (148). The HEIV of Leedan and Meer (2000) solves the first equation by the following iterations:

1. Initialize $\hat{\boldsymbol{v}}$, say by LS.
2. Compute the matrices $\tilde{\boldsymbol{M}}$ and $\tilde{\boldsymbol{L}}$ in (150), and solve the generalized eigenvalue problem
$$\tilde{\boldsymbol{M}}\boldsymbol{v} = \lambda \tilde{\boldsymbol{L}}\boldsymbol{v}. \tag{153}$$
Let $\boldsymbol{v}$ be the unit generalized eigenvector for the generalized eigenvalue λ closest to 1.
3. If $\boldsymbol{v} \approx \hat{\boldsymbol{v}}$ except for sign, return $\hat{\boldsymbol{v}}$ and stop. Else, let $\hat{\boldsymbol{v}} \leftarrow \boldsymbol{v}$ and go back to Step 2.

Leedan and Meer (2000) pointed out that choosing in Step 3 not the generalized eigenvalue closest to 1 but the smallest one improves the convergence performance. See Kanatani and Sugaya (2006, 2007) for the comparative experiment of this effect.

Appendix 7: Projective Gauss-Newton Iterations

Since the gradient $\nabla_{\mathbf{u}} J$ is given by (101), we can minimize the function J in (100) by Newton iterations. If we evaluate the Hessian $\nabla_{\mathbf{u}}^2 J$, the increment $\Delta\boldsymbol{u}$ in $\boldsymbol{u}$ is determined by solving

$$(\nabla_{\mathbf{u}}^2 J)\Delta\boldsymbol{u} = -\nabla_{\mathbf{u}} J. \tag{154}$$

Since $\nabla_{\mathbf{u}}^2 J$ is singular (the function J is constant in the direction of $\boldsymbol{u}$), the solution is indeterminate. However, if we use pseudoinverse and compute

$$\Delta\boldsymbol{u} = -(\nabla_{\mathbf{u}}^2 J)^- \nabla_{\mathbf{u}} J, \tag{155}$$

we obtain a solution, which is orthogonal to $\boldsymbol{u}$.

Differentiating (100) and introducing Gauss-Newton approximation (i.e., ignoring terms that contain $(\boldsymbol{u}, \boldsymbol{\xi}_\alpha)$), we see that the Hessian is nothing but the matrix $\hat{\boldsymbol{M}}$ in (62) for $\boldsymbol{u} = \hat{\boldsymbol{u}}$. In order to compute pseudoinverse, we enforce $\hat{\boldsymbol{M}}$, which is generally nonsingular, to have eigenvalue 0, using the projection matrix $\boldsymbol{P}_{\hat{\mathbf{u}}} = \boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^\top$. The iteration procedure given by Kanatani and Sugaya (2006, 2007) goes as follows:

1. Initialize $\hat{\boldsymbol{u}}$, say by LS.
2. Compute
$$\boldsymbol{u} = N[\hat{\boldsymbol{u}} - (\boldsymbol{P}_{\hat{\mathbf{u}}}\hat{\boldsymbol{M}}\boldsymbol{P}_{\hat{\mathbf{u}}})^-(\hat{\boldsymbol{M}} - \hat{\boldsymbol{L}})\hat{\boldsymbol{u}}]. \tag{156}$$
3. If $\boldsymbol{u} \approx \hat{\boldsymbol{u}}$, return $\hat{\boldsymbol{u}}$ and stop. Else, let $\hat{\boldsymbol{u}} \leftarrow \boldsymbol{u}$ and go back to Step 2.

This scheme is just as effective as FNS, HEIV, and renormalization. See Kanatani and Sugaya (2006, 2007) for the comparative experiments.

References

Akaike, H. (1977). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *16*(6), 716–723.

Amari, S., & Kawanabe, M. (1997). Information geometry of estimating functions in semiparametric statistical models. *Bernoulli*, *3*, 29–54.

Begelfor, E. & Werman, M. (2005). How to put probabilities on homographies. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(10), 1666–1670.

Bickel, P. J., Klassen, C. A. J., Ritov, Y. & Wellner, J. A. (1994). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press,

Bookstein, F. L. (1979). Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, *9*(1), 56–71.

Chernov, M. & Lesort, C. (2004). Statistical efficiency of curve fitting algorithms. *Computational Statisitics and Data Analysis*, *47*(4), 713–728.

Chojnacki, W., Brooks, M. J., & van den Hengel, A. (2001). Rationalising the renormalisation method of Kanatani. *Journal of Mathematical Imaging and Vision*, *14*(1), 21–38.

Chojnacki, W., Brooks, M. J., van den Hengel, A. & Gawley, D. (2000). On the fitting of surfaces to data with covariances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1294–1303.

Chojnacki, W., Chojnacki, W., Brooks, M. J., van den Hengel, A., & Gawley, D. (2004). From FNS to HEIV: A link between two vision parameter estimation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(2), 264–268.

Chojnacki, W., J. Brooks, M. J., van den Hengel, A. & Gawley, D. (2005). FNS, CFNS and HEIV: A unifying approach. *Journal of Mathematical Imaging and Vision*, *23*(2), 175–183.

Hartley, R., & Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press.

Kanatani, K. (1993). Renormalization for unbiased estimation. In *Proceedings of the 4th international conference on computer vision*, Berlin, Germany, May 1993 (pp. 599–606).

Kanatani, K. (1996). *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam: Elsevier. Reprinted by Dover, New York (2005).

Kanatani, K. (1998). Cramer-Rao lower bounds for curve fitting. *Graphical Models and Image Processing*, *60*(2), 93–99.

Kanatani, K. (2004). Uncertainty modeling and model selection for geometric inference. *IEEE Transactions on Pattern Analysis and Machine Intellibence.*, *26*(10), 1307–1319.

Kanatani, K. (2005). Further improving geometric fitting. In *Proceedings of the 5th international conference on 3-D digital imaging and modeling*, Ottawa, Canada, June 2005 (pp. 2–13).

Kanatani, K. (2006). Ellipse fitting with hyperaccuracy. *IEICE Tranactions on Information and Systems*, *E89-D*(10), 2653–2660.

Kanatani, K., & Ohta, N. (2003). Comparing optimal three-dimensional reconstruction for finite motion and optical flow. *Journal of Electronic Imaging*, *12*(3), 478–488.

Kanatani, K., Ohta, N., & Kanazawa, Y. (2000). Optimal homography computation with a reliability measure. *IEICE Transactions on Information and Systems*, *E83-D*(7), 1369–1374.

Kanatani, K., & Sugaya, Y. High accuracy fundamental matrix computation and its performance evaluation. *IEICE Transactions on Information and Systems*, *E90-D*(2), 579–585.

Kanatani, K., & Sugaya, Y. (2007). Performance evaluation of iterative geometric fitting algorithms. *Computational Statistics and Data Analysis*, *52*(2), 1208–1222.

Leedan, Y., & Meer, P. (2000). Heteroscedastic regression in computer vision: Problems with bilinear constraint. *Internaltional Journal of Computer Vision*, *37*(2), 127–150.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *16*(1), 1–32.

Ohta, N. (2003). Motion parameter estimation from optical flow without nuisance parameters. In *3rd international workshop on statistical and computational theory of vision*, October 2003, Nice, France: http://www.stat.ucla.edu/~sczhu/Workshops/SCTV2003.html.

Okatani, T., & Deguchi, K. (2003). Toward a statistically optimal method for estimating geometric relations from noisy data: Cases of linear relations. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, Madison, WI, USA, (Vol. 1, pp. 432–439).

Pennec, X., Fillard, P., & Ayache, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computuer Vision*, *66*(1), 41–66

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific.

Sampson, P. D. (1982). Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithms. *Computer Graphics and Image Processing*, *18*(1), 97–108.

Taubin, G. (1991). Estimation of planar curves, surfaces, and non-planar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(11), 1115–1138.