# A Comparison of Ellipse Fitting Methods and Implications for Multiple-View Geometry Estimation

Zygmunt L. Szpak, Wojciech Chojnacki, Anton van den Hengel
School of Computer Science, The University of Adelaide, SA 5005, Australia
Email: {zygmunt.szpak, wojciech.chojnacki, anton.vandenhengel}@adelaide.edu.au

***Abstract*—After summarising the conceptual relationship between some old and new techniques for fitting ellipses to data, we conduct a thorough experimental comparison of the discussed methods. The newer techniques promise consistency, unbiasedness, or hyper-accuracy, but our experiments reveal that in practice the performance of these estimators is difficult to distinguish from the performance of established estimators. Since the newer techniques could potentially be extended to other estimation problems, we discuss what implications our experimental findings have for solving more general multiple-view geometry estimation problems.**

## I. Introduction

In the family of geometric parameter estimation problems, of which the most well known include fundamental matrix, homography, and trifocal tensor estimation, the task of ellipse fitting occupies a somewhat less prominent position. This is surprising, given the prevalence of ellipse-like shapes in nature and man-made environments. For example, everyday circular shapes such as road signs, tyres, and plates appear as ellipses when projected perspectively onto a plane. Any reasonable object detection system would need to be able to fit such elliptic shapes. Ellipses also appear on a microscopic scale. In biomedical image processing, certain cells and bacteria are approximated with ellipses [1]. Furthermore, there are many industrial applications where ellipse fitting is required. A particularly interesting case is the quality assurance process of steel coils. If the steel coils are too elliptic, they are classified as faulty and not shipped to a customer [2].

Aside from the numerous potential applications, there is another important reason for studying ellipse fitting, and one which provided the primary inspiration for the present work: the utility of ellipse fitting for the development of other estimation methods. This stems from the conceptual relationship between ellipse fitting and various geometric parameter estimation problems. In fact, many of the mathematical techniques that have been devised to improve ellipse fitting have subsequently been applied to fundamental matrix or homography estimation. Since an ellipse can be easily visualised, it is easier to get a comprehensive picture of how the various estimation paradigms compare with one another by concentrating on ellipse fitting. Armed with the insight gained through the analysis of this particular case, one can decide which estimation frameworks are worth extending to more complicated tasks such as trifocal tensor estimation.

At first glance, fitting an ellipse to data may seem like a simple task. The ellipse is after all a familiar shape that can be found throughout nature. However, despite its visual simplicity, ellipse fitting harbors many technical subtleties that make the problem not amenable to traditional statistical analysis. Inherent difficulties arising in the analysis of the problem have even led some researchers to ask the question "Does the best fit always exist?" [3]. There are two principal factors that complicate the ellipse fitting task: (1) ellipse fitting falls under the banner of errors-in-variables regression and (2) the estimation problem is inherently non-linear. These same factors also complicate general multiple-view geometry estimation problems that arise in computer vision.

In recent years, several new papers have addressed the ellipse fitting problem. Most papers claim various levels of optimality and the titles of some may give a reader unfamiliar with the subject the impression that all the major challenges in ellipse fitting have been solved. Despite the influx of new ellipse fitting methods, very few experimental comparisons have appeared in the literature. Those that have appeared are not comprehensive [4], [5]. The purpose of this paper is to quantify how much improvement recently proposed estimators provide and to summarise what the real challenges in ellipse fitting actually are. Another important contribution of this paper is to cast light on the conceptual similarities between various estimators, especially in the manner in which they attempt to produce unbiased estimates. We present all methods using the same notation and parametrisation, which—from a pedagogical point of view—facilitates better insight into how the estimators work.

## II. Background

A *conic* is the locus of solutions $\mathbf{x} = [m_1, m_2]^\mathsf{T}$ in the Euclidean plane $\mathbb{R}^2$ of a quadratic equation

$$am_1^2 + bm_1m_2 + cm_2^2 + dm_1 + em_2 + f = 0, \qquad (1)$$

where $a$, $b$, $c$, $d$, $e$, $f$ are real numbers such that $a^2 + b^2 + c^2 > 0$. With $\boldsymbol{\theta} = [a, b, c, d, e, f]^\mathsf{T}$ and $\mathbf{u}(\mathbf{x}) = [m_1^2, m_1m_2, m_2^2, m_1, m_2, 1]^\mathsf{T}$, equation (1) can equivalently be written as

$$\boldsymbol{\theta}^T\mathbf{u}(\mathbf{x}) = 0. \qquad (2)$$

978-1-4673-2181-5/12/$31.00 ©2012 IEEE

Any multiple of $\boldsymbol{\theta}$ by a non-zero number corresponds to one and the same conic. A conic is either an *ellipse*, or a *parabola*, or a *hyperbola* depending on whether the *discriminant* $\Delta = b^2 - 4ac$ is negative, zero, or positive. The condition $\Delta < 0$ characterising the ellipses can alternatively be written as

$$\boldsymbol{\theta}^\mathsf{T}\mathbf{F}\boldsymbol{\theta} > 0, \tag{3}$$

where

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

and $\otimes$ denotes Kronecker product. In what follows, we shall specifically be concerned with ellipses.

The task of fitting an ellipse to a set of points $\mathbf{x}_1, \ldots, \mathbf{x}_N$ can be achieved by introducing a meaningful *cost function* that characterises the extent to which any particular $\boldsymbol{\theta}$ fails to satisfy the system of copies of equation (2) associated with $\mathbf{x} = \mathbf{x}_n$, $n = 1, \ldots, N$. Once a cost function is selected, the corresponding ellipse fit results from minimising the cost function. Ideally, the minimisation should be subject to inequality (3), but often this constraint is ignored with the consequence that the corresponding fit is not an ellipse but merely a conic. In an alternate mode of operation, an estimate obtained by minimising a specific cost function can further be refined by employing an appropriate *post-hoc* correction procedure based on an independent principle like the need to obtain a statistically unbiased estimate. This post-hoc procedure need not minimise any cost function *per se*.

Below we describe a selection of estimators for ellipse fitting some of which are based on cost function minimisation and others fall into the category of post-hoc corrections. One of the simplest cost functions for ellipse fitting is the *algebraic* distance

$$J_{\mathrm{ALG}}(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^\mathsf{T}\mathbf{A}\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|^2},$$

where $\mathbf{A} = \sum_{n=1}^{N} \mathbf{u}(\mathbf{x}_n)\mathbf{u}(\mathbf{x}_n)^\mathsf{T}$. The corresponding *algebraic distance* fit $\widehat{\boldsymbol{\theta}}_{\mathrm{ALG}}$, i.e., the minimiser of $J_{\mathrm{ALG}}$, is the eigenvector of $\mathbf{A}$, unique up to a scale factor, associated with the smallest eigenvalue of $\mathbf{A}$. The algebraic distance fitting method is notorious for being numerically unstable if it is used without an appropriate pre-conditioning.

A better numerically behaved variant of $\widehat{\boldsymbol{\theta}}_{\mathrm{ALG}}$ is the *direct linear transformation* (DLT) fit $\widehat{\tilde{\boldsymbol{\theta}}}_{\mathrm{DLT}}$. It relies on a data-dependent $3 \times 3$ matrix $\mathbf{T}$ for pre-conditioning à la Hartley [6]. The transformation $\tilde{\mathbf{m}} = \mathbf{T}\mathbf{m}$, where $\mathbf{m} = [m_1, m_2, 1]^\mathsf{T}$ and $\tilde{\mathbf{m}} = [\tilde{m}_1, \tilde{m}_2, 1]^\mathsf{T}$ represent a single data point in the un-normalised and normalised coordinates, respectively, scales down the data set to a unit box with centre at the origin. In the normalised coordinates, equation (2) can be equivalently written as $\tilde{\boldsymbol{\theta}}^\mathsf{T}\mathbf{u}(\tilde{\mathbf{x}}) = 0$ with $\tilde{\mathbf{x}} = [\tilde{m}_1, \tilde{m}_2]^\mathsf{T}$ and $\tilde{\boldsymbol{\theta}}$ related to $\boldsymbol{\theta}$ via the relation

$$\tilde{\boldsymbol{\theta}} = \mathbf{E}^{-1}\mathbf{P}_{(34)}\mathbf{D}_3^{+}(\mathbf{T} \otimes \mathbf{T})^{-\mathsf{T}}\mathbf{D}_3\mathbf{P}_{(34)}\mathbf{E}\boldsymbol{\theta}.$$

Here, $\mathbf{D}_3$ is the $9 \times 6$ *duplication* matrix given by

$$\mathbf{D}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$\mathbf{P}_{(34)}$ is the *permutation* matrix for interchanging the 3rd and 4th entries of length-6 vectors, given by

$$\mathbf{P}_{(34)} = \mathrm{diag}(0,1,0) \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \mathrm{diag}(1,0,1) \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$\mathbf{E} = \mathrm{diag}(1, 2^{-1}, 1, 2^{-1}, 2^{-1}, 1)$, and, for a given matrix $\mathbf{S}$, $\mathbf{S}^{+}$ denotes the Moore–Penrose inverse of $\mathbf{S}$. Letting $\tilde{\mathbf{A}} = \sum_{n=1}^{N} \mathbf{u}(\tilde{\mathbf{x}}_n)\mathbf{u}(\tilde{\mathbf{x}}_n)^\mathsf{T}$, the minimisation of the algebraic distance cost function $J_{\mathrm{ALG}}(\tilde{\boldsymbol{\theta}}) = \|\tilde{\boldsymbol{\theta}}\|^{-2}\tilde{\boldsymbol{\theta}}^\mathsf{T}\tilde{\mathbf{A}}\tilde{\boldsymbol{\theta}}$ leads to $\widehat{\tilde{\boldsymbol{\theta}}}_{\mathrm{ALG}}$, which, of course, is the eigenvector of $\tilde{\mathbf{A}}$ corresponding to the smallest eigenvalue. The un-normalisation of $\widehat{\tilde{\boldsymbol{\theta}}}_{\mathrm{ALG}}$ finally defines $\widehat{\boldsymbol{\theta}}_{\mathrm{DLT}}$, i.e.,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{DLT}} = \mathbf{E}^{-1}\mathbf{P}_{(34)}\mathbf{D}_3^{+}(\mathbf{T} \otimes \mathbf{T})^{\mathsf{T}}\mathbf{D}_3\mathbf{P}_{(34)}\mathbf{E}\widehat{\tilde{\boldsymbol{\theta}}}_{\mathrm{ALG}}.$$

While not in an explicit fashion, Fitzgibbon *et al.* [7] effectively proposed the *direct ellipse fitting* cost function

$$J_{\mathrm{DIR}}(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^\mathsf{T}\mathbf{A}\boldsymbol{\theta}}{\boldsymbol{\theta}^\mathsf{T}\mathbf{F}\boldsymbol{\theta}}.$$

The *direct* fit $\widehat{\boldsymbol{\theta}}_{\mathrm{DIR}}$, i.e., the minimiser of $J_{\mathrm{DIR}}$, is the same as the solution of the problem

$$\min_{\boldsymbol{\theta}} \boldsymbol{\theta}^\mathsf{T}\mathbf{A}\boldsymbol{\theta} \quad \text{subject to} \quad \boldsymbol{\theta}^\mathsf{T}\mathbf{F}\boldsymbol{\theta} = 1 \tag{4}$$

and it is in this form that $\widehat{\boldsymbol{\theta}}_{\mathrm{DIR}}$ was originally introduced. The representation of $\widehat{\boldsymbol{\theta}}_{\mathrm{DIR}}$ as a solution of the problem (4) makes it clear that $\widehat{\boldsymbol{\theta}}_{\mathrm{DIR}}$ is always an ellipse. Extending the work of Fitzgibbon *et al.*, Halíř and Flusser [8] introduced a numerically stable algorithm for calculating $\widehat{\boldsymbol{\theta}}_{\mathrm{DIR}}$.

Another cost function for ellipse fitting, and more generally for conic fitting, was first proposed by Sampson [9] and next popularised, in a broader context, by Kanatani [10]. It is variously called the *Sampson*, *gradient-weighted*, and *approximated maximum likelihood* (AML) distance or cost function, and takes the form

$$J_{\mathrm{AML}}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \frac{\boldsymbol{\theta}^\mathsf{T}\mathbf{A}_n\boldsymbol{\theta}}{\boldsymbol{\theta}^\mathsf{T}\mathbf{B}_n\boldsymbol{\theta}}$$

with $\mathbf{A}_n = \mathbf{u}(\mathbf{x}_n)\mathbf{u}(\mathbf{x}_n)^\mathsf{T}$ and $\mathbf{B}_n = \partial_{\mathbf{x}}\mathbf{u}(\mathbf{x}_n)\boldsymbol{\Lambda}_{\mathbf{x}_n}\partial_{\mathbf{x}}\mathbf{u}(\mathbf{x}_n)^\mathsf{T}$ for each $n = 1, \ldots, N$. Here, for any length 2 vector $\mathbf{y}$, $\partial_{\mathbf{x}}\mathbf{u}(\mathbf{y})$ denotes the $6 \times 2$ matrix of the partial derivatives of the function $\mathbf{x} \mapsto \mathbf{u}(\mathbf{x})$ evaluated at $\mathbf{y}$, and, for each $n = 1, \ldots, N$, $\boldsymbol{\Lambda}_{\mathbf{x}_n}$ is a $2 \times 2$ symmetric *covariance* matrix describing the uncertainty of the data point $\mathbf{x}_n$ [10], [11]. The function $J_{\mathrm{AML}}$ is a first-order approximation of a genuine maximum likelihood cost function $J_{\mathrm{ML}}$ which can be evolved

$$\boldsymbol{\Delta}(\mathbf{x},\sigma) = \begin{bmatrix} 3\sigma^4 - 6\sigma^2 m_1^2 & -6\sigma^2 m_1 m_2 & \sigma^4 - \sigma^2(m_1^2 + m_2^2) & -3\sigma^2 m_1 & -\sigma^2 m_2 & -\sigma^2 \\ * & 4(\sigma^4 - \sigma^2(m_1^2 + m_2^2)) & -6\sigma^2 m_1 m_2 & -2\sigma^2 m_2 & -2\sigma^2 m_1 & 0 \\ * & * & 3\sigma^4 - 6\sigma^2 m_2^2 & -\sigma^2 m_1 & -3\sigma^2 m_2 & -\sigma^2 \\ * & * & * & -\sigma^2 & 0 & 0 \\ * & * & * & * & -\sigma^2 & 0 \\ * & * & * & * & * & 0 \end{bmatrix}$$

Fig. 1. Definition of $\boldsymbol{\Delta}(\mathbf{x},\sigma)$. The $*$'s represent the entries that can be obtained by reflection across the diagonal from the upper triangular part of the matrix.

based on the *Gaussian model of errors* in data in conjunction with the *principle of maximum likelihood*. In the case where identical independent homogeneous Gaussian noise corrupts the data points, $J_{\mathrm{ML}}(\boldsymbol{\theta})$ reduces—up to a numeric constant depending on the noise level—to the sum of orthogonal distances of the data points to the ellipse represented by $\boldsymbol{\theta}$. The *approximated maximum likelihood* fit $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$, i.e., the minimiser of $J_{\mathrm{AML}}$, satisfies the optimality condition $\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}(\boldsymbol{\theta}) = \mathbf{0}^{\mathsf{T}}$, which in an equivalent form can be written as

$$(\mathbf{M}_{\boldsymbol{\theta}} - \mathbf{N}_{\boldsymbol{\theta}})\boldsymbol{\theta} = \mathbf{0}, \tag{5}$$

where $\mathbf{M}_{\boldsymbol{\theta}} = \sum_{n=1}^{N} (\boldsymbol{\theta}^{\mathsf{T}} \mathbf{B}_n \boldsymbol{\theta})^{-1} \mathbf{A}_n$ and $\mathbf{N}_{\boldsymbol{\theta}} = \sum_{n=1}^{N} (\boldsymbol{\theta}^{\mathsf{T}} \mathbf{A}_n \boldsymbol{\theta})(\boldsymbol{\theta}^{\mathsf{T}} \mathbf{B}_n \boldsymbol{\theta})^{-2} \mathbf{B}_n$. This characterisation forms the basis for two iterative schemes for finding $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$, namely, the fundamental numerical scheme (FNS) [11] and the heteroscedastic error-in-variables (HEIV) scheme [12]. To converge, both methods require a good initial estimate and not too much noise in the data. Minimisation of $J_{\mathrm{AML}}$ can alternatively be performed by using a variant of the Levenberg–Marquardt (LM) scheme, with the advantage of avoiding possible convergence failures to which FNS and HEIV are prone.

Closely related to FNS and HEIV is Kanatani's technique of *renormalisation* [13]–[15]. While not exactly being a method for minimising $J_{\mathrm{AML}}$, it computes a close approximation of $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$. The scheme applies a specific form of statistical unbiasing in succession to produce a sequence of estimates. An update of a particular iterate is obtained—effectively—by initially solving an equation that approximates a truncated version of (5) (namely $\mathbf{M}_{\boldsymbol{\theta}}\boldsymbol{\theta} = \mathbf{0}$) and is linked with the iterate, and then converting the resulting solution, via unbiasing, into a solution of an equation approximating equation (5). The limit *renormalisation* fit $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ can be characterised as a solution of an equation that is akin to, but not identical with, equation (5)—see [16] for details.

The idea of unbiasing is at the heart of some other recent fitting techniques. All of these exploit unbiasing via a single step. The corresponding unbiased estimates are not minimisers of a cost function in any obvious way. One of these methods, due to Markovsky *et al.* [17], [18], unbiases $\widehat{\boldsymbol{\theta}}_{\mathrm{ALG}}$. The corresponding *consistent adjusted least squares* (CALS) fit is defined by $\widehat{\boldsymbol{\theta}}_{\mathrm{CALS}} = \mathbf{D}\tilde{\boldsymbol{\theta}}_{\mathrm{CALS}}$, where $\mathbf{D} = \mathrm{diag}(1,2,1,1,1,1)$ and $\tilde{\boldsymbol{\theta}}_{\mathrm{CALS}}$ is the eigenvector associated with the smallest eigenvalue of the matrix

$$\boldsymbol{\Xi}(\{\mathbf{x}_n\}_{n=1}^N, \widehat{\sigma}) = \sum_{n=1}^{N} \big(\mathbf{D}\mathbf{u}(\mathbf{x}_n)\mathbf{u}(\mathbf{x}_n)^{\mathsf{T}}\mathbf{D} + \boldsymbol{\Delta}(\mathbf{x}_n, \widehat{\sigma})\big).$$

Here, $\boldsymbol{\Delta}(\mathbf{x},\sigma)$ is a symmetric matrix defined in Fig. 1 and $\widehat{\sigma}$ is the estimate of the noise level defined by the requirement that $\boldsymbol{\Xi}(\{\mathbf{x}_n\}_{n=1}^N, \widehat{\sigma})$ should have a non-zero null space. In practice, the search for an appropriate $\widehat{\sigma}$ that satisfies the requirement can be performed using a variant of the popular bisection method for root finding [19].

Another method, namely the so-called *hyper-accurate correction* method of Kanatani, produces an unbiased variant of $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$. The corresponding *hyper-accurate* fit comes in two flavours: $\widehat{\boldsymbol{\theta}}_{\text{HYP-1}}$ [20] and $\widehat{\boldsymbol{\theta}}_{\text{HYP-2}}$ [21]. Under the assumption that $\|\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}\| = 1$, these are defined by $\widehat{\boldsymbol{\theta}}_{\text{HYP-1}} = \widehat{\boldsymbol{\theta}}_{\mathrm{AML}} - \Delta\boldsymbol{\theta}_1$ and $\widehat{\boldsymbol{\theta}}_{\text{HYP-2}} = \widehat{\boldsymbol{\theta}}_{\mathrm{AML}} - \Delta\boldsymbol{\theta}_2$ with

$$\Delta\boldsymbol{\theta}_0 = 4\widehat{\sigma}^2 \sum_{n=1}^{N} \frac{\mathbf{u}(\mathbf{x}_n)^{\mathsf{T}} \mathbf{M}^{+}_{\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}} \mathbf{B}_n \widehat{\boldsymbol{\theta}}_{\mathrm{AML}}}{(\widehat{\boldsymbol{\theta}}_{\mathrm{AML}} \mathbf{B}_n \widehat{\boldsymbol{\theta}}_{\mathrm{AML}})^2} \mathbf{M}^{+}_{\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}} \mathbf{u}(\mathbf{x}_n),$$

$$\Delta\boldsymbol{\theta}_2 = 4\widehat{\sigma}^2 \sum_{n=1}^{N} \frac{\mathbf{u}(\mathbf{x}_n)^{\mathsf{T}} \mathbf{M}^{+}_{\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}} \mathbf{u}(\mathbf{x}_n)}{(\widehat{\boldsymbol{\theta}}_{\mathrm{AML}} \mathbf{B}_n \widehat{\boldsymbol{\theta}}_{\mathrm{AML}})^2} \mathbf{M}^{+}_{\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}} \mathbf{B}_n \widehat{\boldsymbol{\theta}}_{\mathrm{AML}},$$

$$\Delta\boldsymbol{\theta}_1 = \Delta\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}_2.$$

The noise level is estimated as $\widehat{\sigma}^2 = (4(N-5))^{-1} J_{\mathrm{AML}}(\widehat{\boldsymbol{\theta}}_{\mathrm{AML}})$. Theoretically, $\widehat{\boldsymbol{\theta}}_{\text{HYP-2}}$, which is based upon a more thorough calculation of bias, should be more accurate than $\widehat{\boldsymbol{\theta}}_{\text{HYP-1}}$. The actual accuracy of both estimates is in practice indistinguishable—see the next section.

## III. Experimental Design

In line with the description from the preceding section, we choose to compare the following estimators:

- **DIR** = direct fit,
- **CALS** = consistent adjusted least squares,
- **REN** = renormalisation,
- **HYP-1** = hyper-accuracy variant 1,
- **HYP-2** = hyper-accuracy variant 2,
- **AML** = approximate maximum likelihood,
- **ML** = maximum likelihood.

To characterise the performance of an estimator we borrow concepts from machine learning literature and make a distinction between a training set and a test set. The intent is to fit an ellipse on a training set and to measure the accuracy of the fit on a corresponding test set. The reason for this distinction is

(a) 0°–180°. (b) 180°–360°.

Fig. 2. An example of two regions of interest on an ellipse generated by specifying start and end angles on a unit circle, and applying a random affine transformation.

that the quality of an ellipse fit cannot be established just by looking at the training error. For example, if the training set consists of points on a small arc of an ellipse, then an estimator may fit points on the small arc well, but the shape and size of the overall ellipse may be far from the true ellipse. To address this problem, our test set is created by sampling 1000 points at random on the true ellipse. The test error for a trial is then computed as the *root-mean-square* (RMS) *orthogonal* distance

$$\sqrt{\frac{1}{2000}\sum_{n=1}^{1000} d_n^2},$$

where $d_n$ denotes the orthogonal distance between the *estimated* ellipse and the $n$th data point lying on the *true* ellipse. This measure is nothing but the *geometric* error of the estimated ellipse with respect to the true data points. The process of computing the orthogonal distances $d_n$ is rather involved—the detailed formulae can be found in [15].

In some recent papers, instead of using the orthogonal distance on a test set to characterise the performance of an estimator, a different measure based on the *parameters* of the estimated and true ellipse was exploited [22]–[24]. Under this measure, the overall error for a thousand trials is computed as the *root-mean-square* (RMS) *parameter* distance

$$\sqrt{\frac{1}{1000}\sum_{n=1}^{1000} \|\mathbf{P}_{\underline{\boldsymbol{\theta}}_n}^{\perp}\widehat{\boldsymbol{\theta}}_n\|^2},$$

where $\mathbf{P}_{\underline{\boldsymbol{\theta}}_n}^{\perp} = \mathbf{I} - \|\underline{\boldsymbol{\theta}}_n\|^{-2}\underline{\boldsymbol{\theta}}_n\underline{\boldsymbol{\theta}}_n^{\top}$ serves to project the normalised estimate $\widehat{\boldsymbol{\theta}}_n$ onto the hyperplane tangent to the unit 5-dimensional sphere at the normalised true value $\underline{\boldsymbol{\theta}}_n$. The error for the $n$th trial is then simply the Euclidean distance between $\underline{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\theta}}_n$ measured on the hyperplane.

We caution against such an approach, because the parameter error measure is coordinate system dependent. The difference between $\underline{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ in the original coordinate system may, for example, be different from the difference between $\underline{\tilde{\boldsymbol{\theta}}}$ and $\widehat{\tilde{\boldsymbol{\theta}}}$ in a Hartley-normalised coordinate system. Furthermore, small differences between parameters do not necessarily translate into small distances when measured using orthogonal distances from points on the true ellipse. Despite these concerns, we chose to include the parameter error measure in the presentation of our results, because it has appeared in numerous other papers. We stress that, throughout the paper, when we speak of errors between parameters, we shall have a Hartley-normalised coordinate system in mind.

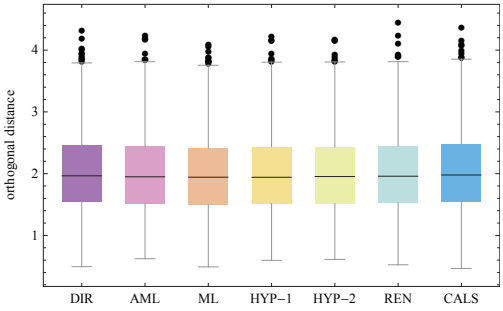
The training sets used in our experiments were all produced in the following manner. We started with a unit circle centered at the origin of our coordinate system and uniformly sampled a number of data points from a region of interest on the circle. A particular region of interest on the circle was fully specified by a start and end angle. The angles were measured counter-clockwise, with the positive $x$-axis assumed to be at $0°$. Next, a random affine transformation was applied to the circle and data points, resulting in: (1) an ellipse with random center, major/minor axes, and orientation of the axes, and (2) transformed data points that occupied a region of interest on the ellipse (see Fig. 2 for an example). Finally, independent zero-mean Gaussian noise with a specified standard deviation was added to the data points.

In addition to experiments on synthetic data, we also include experiments on real data. Specifically, we took pictures of elliptic ripples formed by water drops and manually selected a small portion of the ellipses as input data to the various estimators.

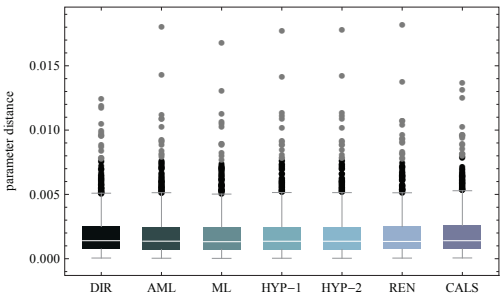
## IV. Results

In the first set of experiments, we investigated the performance of the estimators under the condition that data points be sampled from the whole ellipse (0°–360°). Even with a fairly large noise level of $\sigma = 5$ pixels, we found that there was almost no difference in the accuracy of the estimators when only a handful of data points are available (Figs. 3a and 3b). Differences only started to appear when many data points were sampled, and the results presented in Figs. 3c and 3d suggest that ML is most accurate, while DIR and REN are least accurate. Having said that, it is important to stress that the differences are not very large.

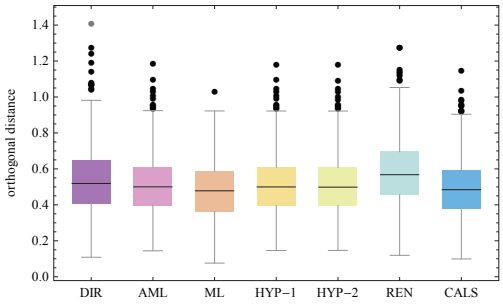
In the second set of experiments, we limited the sampling of data points to half of an ellipse (0°–180°). This simulates the case of a partially occluded ellipse. Once again, we investigated the performance of the estimators with both small and large numbers of data points. With the exception of DIR, the quality of the estimators in Fig. 4 is difficult to distinguish. A closer look at the results is given in Fig. 5 where we plot



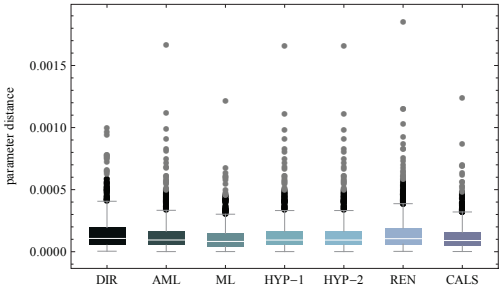

(a) Each trial sampled 15 data points with $\sigma = 5$ pixels.




(b) Each trial sampled 15 data points with $\sigma = 5$ pixels.

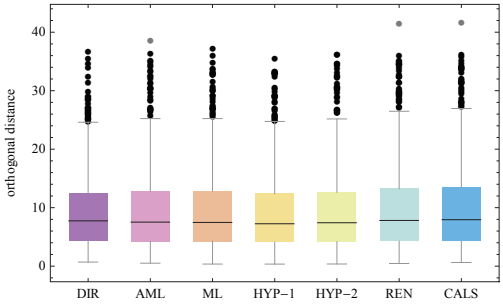



(c) Each trial sampled 250 data points with $\sigma = 5$ pixels.

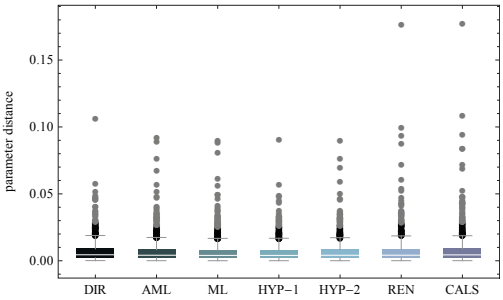



(d) Each trial sampled 250 data points with $\sigma = 5$ pixels.

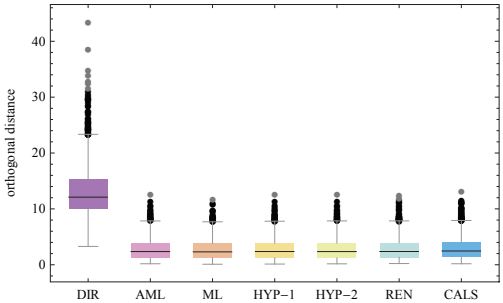
Fig. 3. Comparison of estimators based on 1000 simulations. Each trial sampled data points from the whole of the ellipse (0°–360°). Left-hand side shows orthogonal distance errors between points on a true ellipse and an estimated ellipse. Right-hand side shows errors between true ellipse parameters and estimated ellipse parameters. The box-plots depict the sample minimum, lower quartile, median (thin line), upper quartile, and sample maximum. Observations considered outliers are represented by shaded circles. A sample that lies beyond 1.5 times the inter-quartile range (the difference between the upper and lower quartiles) is shaded in black, and a sample that lies beyond 3 times that range is shaded in grey.



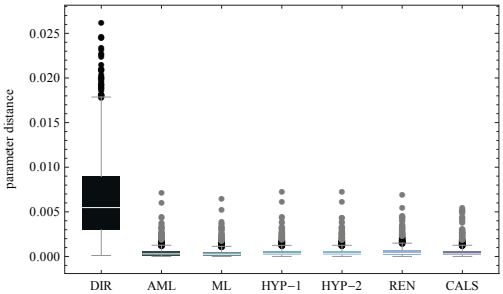

(a) Each trial sampled 15 data points with $\sigma = 5$ pixels.




(b) Each trial sampled 15 data points with $\sigma = 5$ pixels.

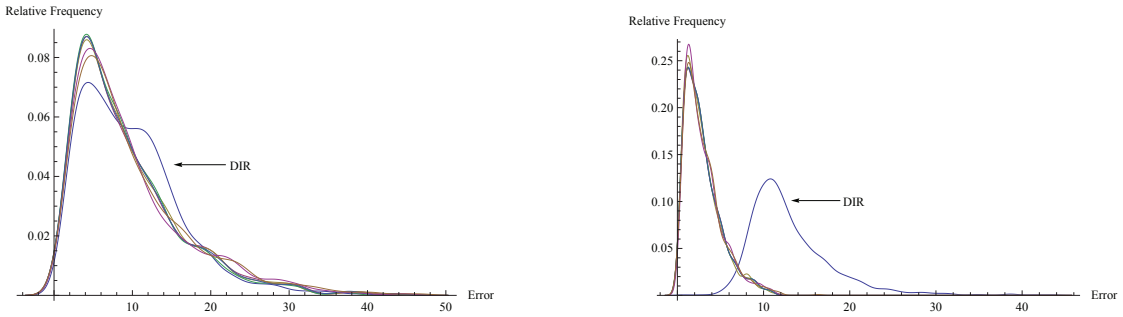



(c) Each trial sampled 250 data points with $\sigma = 5$ pixels.




(d) Each trial sampled 250 data points with $\sigma = 5$ pixels.

Fig. 4. Comparison of estimators based on 1000 simulations. Each trial sampled data points anywhere between 0°–180° on the ellipse. Left-hand side shows orthogonal distance errors between points on a true ellipse and an estimated ellipse. Right-hand side shows errors between true ellipse parameters and estimated ellipse parameters. The box-plots depict the sample minimum, lower quartile, median (thin line), upper quartile, and sample maximum. Observations considered outliers are represented by shaded circles. A sample that lies beyond 1.5 times the inter-quartile range (the difference between the upper and lower quartiles) is shaded in black, and a sample that lies beyond 3 times that range is shaded in grey.




(a) Each trial sampled 15 data points with $\sigma = 5$ pixels.

(b) Each trial sampled 250 data points with $\sigma = 5$ pixels.

Fig. 5. Comparison of estimators based on 1000 simulations. Each trial sampled data points anywhere between $0^\circ$–$180^\circ$ on the ellipse. The plot shows smoothed histograms of the root-mean square orthogonal distance errors between points on true ellipse and estimated ellipses. Apart from DIR, the performance of the estimators is almost indistinguishable. The inferior performance of DIR is clearly visible in a), and as the number of data points is increased in b) its relative performance deteriorates even more.

a smoothed histogram of the errors of each estimator. The inferior performance of DIR is clearly visible.

With HYP-1 and HYP-2, Kanatani introduced so-called 'hyper-accurate' estimators, which are capable of occasionally producing estimates that are more accurate than the gold standard ML method. In our first two sets of experiments, we could not observe any difference between AML, HYP-1, and HYP-2 in the box-whisker charts. Since the theoretical improvements in HYP-1 and HYP-2 are meant for small noise levels, our next set of experiments focused on very small noise levels. With small noise, the results reported in Figs. 6 and 7 show that the 'hyper-accurate' estimators can indeed occasionally outperform AML and ML, but it is important not to overstate the gain. The average improvement is measured in decimal points of the mean root-mean-square error, meaning that the differences between the ellipse fits would be difficult to notice visually at such small noise levels.

In our final set of synthetic experiments, we studied the behavior of the estimators when the noise level was fixed to $\sigma = 8$ pixels, while the number of data points varied. The data points were limited to half of the ellipse. The DIR estimator was once again the worst and surprisingly did not improve as the number of data points was increased. According to Fig. 8, the ML, HYP-1, HYP-2, and AML estimators all performed similarly and noticeably better than REN and CALS. However, careful study of Fig. 8 also reveals a peculiar inconsistency between results reported using orthogonal distances and parameter errors. Based on Fig. 7a, it would appear that there is no difference between REN and CALS, while on the other hand Fig. 8b shows a notable difference. This confirms our earlier assertion in Section III that small differences between parameters do not necessarily translate into small differences when measured using orthogonal distances. One cannot therefore confidently declare an estimator to be better than another based on parameter errors alone. Unfortunately, many papers are using this error measure alone to rank estimators, not just on ellipse fitting, but on homography and fundamental matrix estimation as well.

Results on real data are presented in Fig. 9. There was no visible difference between ML, HYP-1, HYP-2, REN, and AML. As usual, the DIR fit was the worst. The CALS fit was considerably better than DIR and very close to AML.

## V. Discussion

Based on our experiments, one would have to conclude that the DIR estimate is by far the worst. However, the DIR fit has one redeeming property that the other estimators do not have: it is guaranteed to always produce an ellipse fit. This is a crucial property, especially when data points are sampled from only a small portion of an ellipse. A better appreciation can be gained by looking at Fig. 10. For clarity, we show only the AML, CALS, and DIR fits, but the behavior of HYP-1, HYP-2, and REN is the same as AML. Even the gold standard ML estimate would need to employ constrained optimisation in order not to produce a fit akin to AML. In the diagram, the AML estimate produces a hyperbola instead of an ellipse. The authors of CALS recognised that their estimator was also prone to producing hyperbolas as solutions and proposed a post-hoc adjustment procedure that is meant to ensure an ellipse fit. While this post-hoc procedure works occasionally, it is also capable of producing degenerate ellipses as solutions which appear as two parallel lines in the image. It is unfortunate that the estimator which guarantees ellipse fits produces notably worse fits in general.

Another remarkable outcome of our investigations is that despite various estimators claiming to be the best, most optimal and most accurate, there is really not much difference between most of them. For example, the principal selling point of CALS is that it is the only (conditionally proven) *consistent* estimator in the family of estimators studied in this paper. Yet, from the point of view of performance, it does not seem to produce remarkably superior estimates even when the number of data points is quite large. Similarly, the hyper-accuracy of HYP-1 and HYP-2 only reveals itself at very small noise levels, and then only in decimal points. Hence, it is not clear that any of these improvements would be practically useful in the computer vision setting. In our opinion, the experiments that we have conducted hint at a more pressing

| Estimator | $\sigma = 0.2$ | $\sigma = 0.4$ | $\sigma = 0.6$ | $\sigma = 0.8$ | $\sigma = 1$ |
|---|---|---|---|---|---|
| HYP−2 | 4.19664 | 8.53906 | 12.2402 | 16.2116 | 19.1869 |
| HYP−1 | **4.18461** | **8.42429** | **11.9242** | **15.7163** | **18.2991** |
| AML | 4.20074 | 8.57776 | 12.3536 | 16.3834 | 19.4682 |
| ML | 4.20256 | 8.58908 | 12.3524 | 16.4032 | 19.4371 |

(a) Mean root-mean-square orthogonal distance errors between points on true ellipse and estimated ellipses.

| Estimator | $\sigma = 0.2$ | $\sigma = 0.4$ | $\sigma = 0.6$ | $\sigma = 0.8$ | $\sigma = 1$ |
|---|---|---|---|---|---|
| HYP−2 | 0.016603 | 0.033933 | 0.049757 | 0.062539 | 0.076324 |
| HYP−1 | **0.016548** | **0.033592** | **0.048783** | **0.061588** | **0.074284** |
| AML | 0.016623 | 0.033973 | 0.049887 | 0.06252 | 0.076585 |
| ML | 0.016653 | 0.034083 | 0.049979 | 0.06256 | 0.076617 |

(b) Root-mean-square errors between true ellipse parameters and estimated ellipse parameters.

Fig. 6. Comparison of ML, AML, HYP-1, and HYP-2 based on 1000 simulations and varying noise levels. Each trial sampled 50 data points anywhere between 0°–90° on the ellipse.

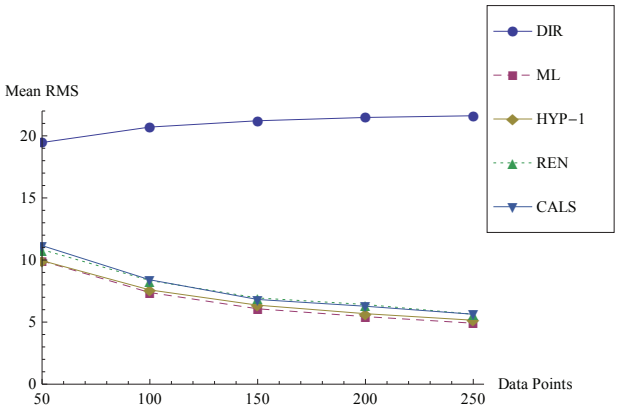
| Estimator | $\sigma = 0.2$ | $\sigma = 0.4$ | $\sigma = 0.6$ | $\sigma = 0.8$ | $\sigma = 1$ |
|---|---|---|---|---|---|
| HYP−2 | 0.249802 | 0.47224 | 0.740898 | 0.960092 | 1.17442 |
| HYP−1 | **0.249766** | 0.47233 | 0.740832 | **0.959726** | 1.17417 |
| AML | 0.249817 | **0.47219** | 0.740914 | 0.960223 | 1.17446 |
| ML | 0.249829 | 0.472219 | **0.740799** | 0.960428 | **1.17414** |

(a) Mean root-mean-square orthogonal distance errors between points on true ellipse and estimated ellipses.

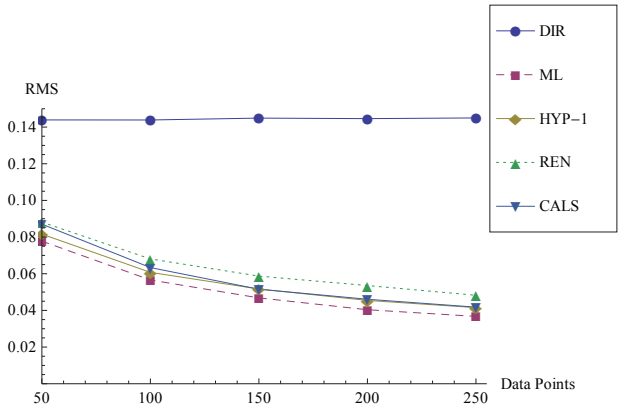
| Estimator | $\sigma = 0.2$ | $\sigma = 0.4$ | $\sigma = 0.6$ | $\sigma = 0.8$ | $\sigma = 1$ |
|---|---|---|---|---|---|
| HYP−2 | 0.001898 | 0.003622 | 0.005532 | 0.007403 | 0.009124 |
| HYP−1 | **0.001897** | 0.003623 | 0.005533 | 0.007402 | 0.009124 |
| AML | 0.001898 | 0.003622 | 0.005532 | 0.007404 | 0.009124 |
| ML | 0.001898 | **0.003622** | **0.005531** | **0.007401** | **0.00912** |

(b) Root-mean-square errors between true ellipse parameters and estimated ellipse parameters.

Fig. 7. Comparison of ML, AML, HYP-1, and HYP-2 based on 1000 simulations and varying noise levels. Each trial sampled 50 data points anywhere between 0°–180° on the ellipse.




(a) Mean root-mean-square orthogonal distance errors between points on true ellipse and estimated ellipses.




(b) Root-mean-square errors between true ellipse parameters and estimated ellipse parameters.

Fig. 8. Comparison of ML, AML, HYP-1, REN, and CALS based on 1000 trials with $\sigma = 8$ pixels. Data points were sampled anywhere between 0°-180° on the ellipse. The estimators HYP-2 and AML are omitted from the graphs because their performance was visually indistinguishable from HYP-1.

need than squeezing out more and more small accuracy gains. Namely, what is needed is a suitable replacement for the DIR estimator—one that is on par with AML but that also guarantees ellipse fits. A first attempt at such an estimator was recently proposed in [25].

## VI. Conclusion

Our investigation has revealed that most of the recently proposed estimation methods, despite being *consistent*, *unbiased*, or *hyper-accurate*, perform much the same as the well-known Sampson error (AML) estimator. Moreover, apart from the DIR estimator, they all fail to cope with ill-posed ellipse fitting problems. Unfortunately, the DIR estimator produces significantly worse fits on average. We also saw that reporting results using parameter errors alone may not reveal the full nature of the estimators. Based on these findings, we imagine that there may not be much difference between AML and the newer estimators on other tasks such as homography or fundamental matrix estimation. Experimental confirmation of this is left to future work.


## Acknowledgement

This research was partially supported by the Australian Research Council.

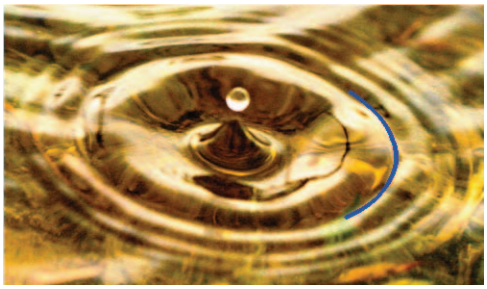

## References


[1] W. N. Gonccalves and O. M. Bruno, "Automatic system for counting cells with elliptical shape," to appear. [Online]. Available: http://arxiv.org/pdf/1201.3109.pdf

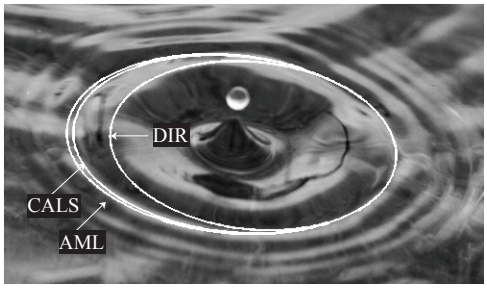
[2] D. C. H. Schleicher and B. G. Zagar, "Image processing to estimate the ellipticity of steel coils using a concentric ellipse fitting algorithm," in *Proc. Ninth Int. Conf. Signal Processing*, 2008, pp. 884–890.

[3] N. Chernov, Q. Huang, and H. Ma, "Does the best fit always exist?" to appear. [Online]. Available: http://people.cas.uab.edu/~mosya/cl/CHM.pdf

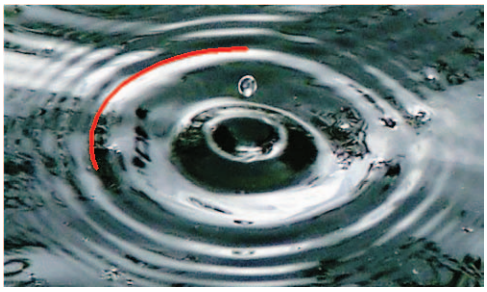
[4] P. L. Rosin, "Assessing error of fit functions for ellipses," *Graphical Models and Image Processing*, vol. 58, no. 5, pp. 494–502, 1996.



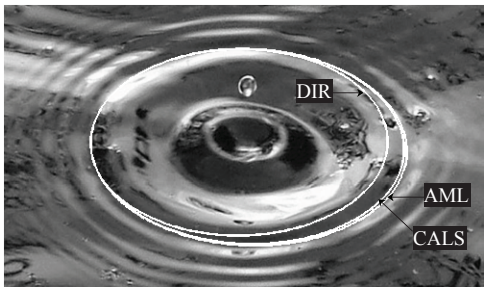
(a) Data points sampled on a small arc of a naturally formed ellipse.




(b) Estimated ellipses.



(c) Data points sampled on a small arc of a naturally formed ellipse.




(d) Estimated ellipses.

Fig. 9. Comparison of AML, CALS, and DIR estimators on real data. A small arc of the true ellipse was used as input to the estimators. The ML, HYP-1, and HYP-2 estimators are omitted from the graphs as there was no visual difference between them and the AML estimator.

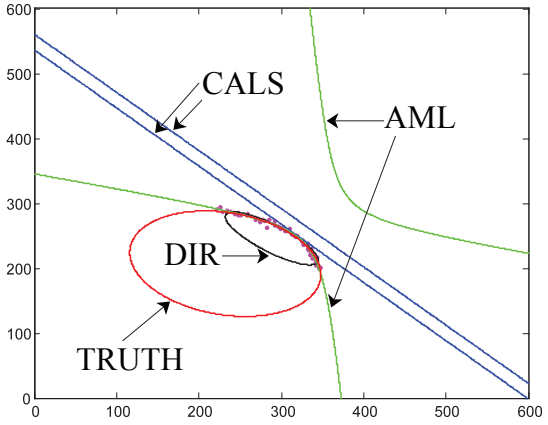



Fig. 10. Comparison of estimators for an ill-posed ellipse fitting problem. The DIR estimator always guarantees an ellipse, but the fit is very biased. The AML fit does not guarantee an ellipse and in this case produced a hyperbola. The HYP-1, HYP-2, and REN estimators are omitted from the figure because they produced similar hyperbolas. The authors of CALS detect when their fit is not an ellipse and apply a post-hoc correction procedure to ensure an ellipse fit. However, this procedure often produces a degenerate ellipse, like the one that appears in the figure as two parallel lines.


[5] ——, "Analysing error of fit functions for ellipses," *Pattern Recognition Lett.*, vol. 17, no. 14, pp. 1461–1470, 1996.

[6] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, 1997.

[7] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, 1999.

[8] R. Halíř and J. Flusser, "Numerically stable direct least squares fitting of ellipses," in *Proc. Sixth Int. Conf. in Central Europe on Computer Graphics and Visualization*, vol. 1, 1998, pp. 125–132.

[9] P. D. Sampson, "Fitting conic sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm," *Computer Graphics and Image Processing*, vol. 18, no. 1, pp. 97–108, 1982.

[10] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam: Elsevier, 1996.

[11] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, "On the fitting of surfaces to data with covariances," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1294–1303, 2000.

[12] Y. Leedan and P. Meer, "Heteroscedastic regression in computer vision: Problems with bilinear constraint," *Int. J. Computer Vision*, vol. 37, no. 2, pp. 127–150, 2000.

[13] K. Kanatani, "Renormalization for unbiased estimation," in *Proc. Fourth Int. Conf. Computer Vision*, 1993, pp. 599–606.

[14] ——, "Statistical bias of conic fitting and renormalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 3, pp. 320–326, 1994.

[15] Z. Zhang, "Parameter estimation techniques: a tutorial with application to conic fitting," *Image and Vision Computing*, vol. 15, no. 1, pp. 59–76, Jan. 1997.

[16] W. Chojnacki, M. J. Brooks, and A. van den Hengel, "Rationalising the renormalisation method of Kanatani," *J. Math. Imaging and Vision*, vol. 14, no. 1, pp. 21–38, 2001.

[17] I. Markovsky, A. Kukush, and S. V. Huffel, "Consistent least squares fitting of ellipsoids," *Numer. Math.*, vol. 98, no. 1, pp. 177–194, 2004.

[18] A. Kukush, I. Markovsky, and S. V. Huffel, "Consistent estimation in an implicit quadratic measurement error model," *Comput. Statist. Data Anal.*, vol. 47, no. 1, pp. 123–147, 2004.

[19] [Online]. Available: ftp://ftp.esat.kuleuven.ac.be/SISTA/markovsky/abstracts/02-116.html

[20] K. Kanatani, "Ellipse fitting with hyperaccuracy," in *Proc. Ninth European Conf. Computer Vision*, ser. Lecture Notes on Computer Science, vol. 1, 2006, pp. 484–495.

[21] ——, "Ellipse fitting with hyperaccuracy," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 10, pp. 2653–2660, 2006.

[22] ——, "Statistical optimization for geometric fitting: theoretical accuracy bound and high order error analysis," *Int. J. Computer Vision*, vol. 80, pp. 167–188, 2008.

[23] P. Rangarajan and K. Kanatani, "Improved algebraic methods for circle fitting," *Electron. J. Statist.*, vol. 3, pp. 1075–1082, 2009.

[24] A. Al-Sharadqah and N. Chernov, "Error analysis for circle fitting algorithms," *Electron. J. Statist.*, vol. 3, pp. 886–911, 2009.

[25] Z. Szpak, W. Chojnacki, and A. van den Hengel, "Guaranteed ellipse fitting with the Sampson distance," in *Proc. 12th European Conf. Computer Vision*, ser. Lecture Notes on Computer Science, vol. 7576, 2012, pp. 87–100.