

# On the Fitting of Surfaces to Data with Covariances

Wojciech Chojnacki, Michael J. Brooks, *Member, IEEE*, Anton van den Hengel, and Darren Gawley

**Abstract**—We consider the problem of estimating parameters of a model described by an equation of special form. Specific models arise in the analysis of a wide class of computer vision problems, including conic fitting and estimation of the fundamental matrix. We assume that noisy data are accompanied by (known) covariance matrices characterizing the uncertainty of the measurements. A cost function is first obtained by considering a maximum-likelihood formulation and applying certain necessary approximations that render the problem tractable. A novel, Newton-like iterative scheme is then generated for determining a minimizer of the cost function. Unlike alternative approaches such as Sampson's method or the renormalization technique, the new scheme has as its theoretical limit the minimizer of the cost function. Furthermore, the scheme is simply expressed, efficient, and unsurpassed as a general technique in our testing. An important feature of the method is that it can serve as a basis for conducting theoretical comparison of various estimation approaches.

**Index Terms**—Statistical methods, surface fitting, parameter estimation, covariance matrix, maximum-likelihood, renormalization, conic fitting, fundamental matrix.

## 1 INTRODUCTION

AN important problem in computer vision is that of estimating the coefficients of a given algebraic equation that constrains a set of image feature locations. Geometrically, such an equation describes a multidimensional surface (possibly just a curve) embedded in a space of higher dimension. In some cases, candidate surfaces are subject to *ancillary constraints* not involving feature locations. The above estimation problem can thus be viewed as that of fitting a curve or surface, possibly lying within a certain manifold, through a number of points. *Conic fitting* is one particular problem of this kind [1], [2]. Two other problems are estimating coefficients of the *epipolar equation* [3] and estimating coefficients of the *differential epipolar equation* [4], [5], each involving an ancillary *cubic constraint*.

The principal equation that applies to a wide class of problems, including those specified above, takes the form

$$\theta^T u(x) = 0. \quad (1)$$

Here,  $\theta = [\theta_1, \dots, \theta_l]^T$  is a vector representing unknown parameters,  $x = [x_1, \dots, x_k]^T$  is a vector representing an element of the data (for example, the locations of a pair of corresponding points), and  $u(x) = [u_1(x), \dots, u_l(x)]^T$  is a vector with the data transformed in a problem-dependent manner such that: 1) each component of  $u(x)$  is a quadratic form in the compound vector  $[x^T, 1]^T$  and 2) one component of  $u(x)$  is equal to 1. The ancillary constraint, if it applies, may be expressed as

$$\psi(\theta) = 0 \quad (2)$$

for some scalar-valued function  $\psi$ . The estimation problem can now be stated as follows: Given a collection  $(x_1, \dots, x_n)$  of image data, determine  $\theta \neq 0$  satisfying (2) such that (1) holds with  $x$  replaced by  $x_i$  for  $1 \leq i \leq n$ . When  $n > l$  and noise is present, the corresponding system of equations is overdetermined and as such may fail to have a nonzero solution. In this situation, we are concerned with finding  $\theta$  that best fits the data in some sense. Various methods for finding such a fit have been developed. A vast class of techniques rests upon the use of *cost functions* measuring the extent to which the data and candidate estimates fail to satisfy (1). If—for simplicity—one sets aside the ancillary constraint, then, given a cost function  $J = J(\theta; x_1, \dots, x_n)$ , a corresponding estimate  $\hat{\theta}$  is defined by

$$J(\hat{\theta}) = \min_{\theta \neq 0} J(\theta; x_1, \dots, x_n).$$

A straightforward cost function is

$$J_{\text{ALS}}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \frac{\theta^T u(x_i) u(x_i)^T \theta}{\|\theta\|^2},$$

where  $\|\theta\| = (\theta_1^2 + \dots + \theta_l^2)^{1/2}$ . It leads to an *algebraic least-squares estimate*  $\hat{\theta}_{\text{ALS}}$  that can be explicitly computed. When some information about the measurement error is available, more involved cost functions can be devised. Under the assumption that the uncertainty of each data point  $x_i$  can be described by a  $k \times k$  covariance matrix  $\Lambda_{x_i}$ , a compelling case (drawing upon the principle of maximum-likelihood) may be made for adoption of the cost function

$$J_{\text{AML}}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \frac{\theta^T u(x_i) u(x_i)^T \theta}{\theta^T \partial_x u(x_i) \Lambda_{x_i} \partial_x u(x_i)^T \theta},$$

• The authors are with the Department of Computer Science, Adelaide University, SA 5005, Australia.

E-mail: {wojteck, mjb, hengel, dg}@cs.adelaide.edu.au.

Manuscript received 9 Dec. 1999; accepted 10 Aug. 2000.

Recommended for acceptance by P. Meer.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111058.

where, for any  $k$  vector  $\mathbf{y}$ ,  $\partial_x \mathbf{u}(\mathbf{y})$  denotes the  $l \times k$  matrix of the partial derivatives of the function  $\mathbf{x} \mapsto \mathbf{u}(\mathbf{x})$  evaluated at  $\mathbf{y}$ . The corresponding *approximated maximum-likelihood estimate*  $\hat{\theta}_{\text{AML}}$  cannot be expressed in closed form, but a numerical approximation to it can be calculated by employing a suitable numerical scheme. One such scheme was proposed by Sampson [6]. Kanatani [7, Chapter 12] noted that Sampson's algorithm is susceptible to statistical bias and proposed amendments in the form of the *first-order* and *second-order renormalization schemes*. While the aim of both types of renormalization was to produce better approximations to  $\hat{\theta}_{\text{AML}}$ , it emerges that these schemes do not theoretically act to compute  $\hat{\theta}_{\text{AML}}$  (see a companion paper [8]).

In this work, we evolve a fundamental Newton-like algorithm for calculating  $\hat{\theta}_{\text{AML}}$  based on the algebraic constraint to which all critical points of  $J_{\text{AML}}$  are subject. This algorithm provides a genuine means for theoretically calculating  $\hat{\theta}_{\text{AML}}$ , is elegantly expressed, and is unsurpassed in our comparative testing. As it turns out, it can also be used to establish a framework for examining theoretical relationships between various existing methods of minimizing  $J_{\text{AML}}$  [8].

## 2 SOME EXAMPLES

We begin by showing that two of the previously mentioned application problems can be written as either (1) or the system of (1) and (2).

### 2.1 Conic Fitting

Consider a plane with a coordinate system. Any point in the plane can be represented as a pair  $(m_1, m_2)$  of coordinates. An equivalent representation of  $(m_1, m_2)$  is via the vector  $\mathbf{m} = [m_1, m_2, 1]^T$ . A *conic* is any set of points  $(m_1, m_2)$  satisfying

$$am_1^2 + bm_1m_2 + cm_2^2 + dm_1 + em_2 + f = 0$$

for some real numbers  $a, b, c, d, e$  not all simultaneously zero. The same set can also be described in terms of  $\mathbf{m}$  by

$$\mathbf{m}^T \mathbf{A} \mathbf{m} = 0, \quad (3)$$

where

$$\mathbf{A} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}.$$

We can group together the coefficients  $a, b, c, d, e, f$ , obtaining the vector of *parameters*  $\boldsymbol{\theta} = [a, b, c, d, e, f]^T$ . The *variables*  $m_1, m_2$  can be represented as the vector  $\mathbf{x} = [m_1, m_2]^T$ . If, adopting the terminology of Matei and Meer [9], [10], we also introduce the vector of *basis functions* or *carriers*  $\mathbf{u}(\mathbf{x}) = [m_1^2, m_1m_2, m_2^2, m_1, m_2, 1]^T$ , then

$$\mathbf{m}^T \mathbf{A} \mathbf{m} = \boldsymbol{\theta}^T \mathbf{u}(\mathbf{x}).$$

With this identity, it is clear that (3) can be written as (1).

### 2.2 Fundamental Matrix Estimation

Consider a camera with an image plane that is equipped with a coordinate system. Any 3D element of the scene perspectively projected onto this plane gives rise to an *image*

*point* represented by coordinates  $(m_1, m_2)$  or, equivalently, by the vector  $\mathbf{m} = [m_1, m_2, 1]^T$ . A 3D point projected onto the image planes of two different cameras endowed with two separate coordinate systems gives rise to a pair of *corresponding points*. When represented by  $(\mathbf{m}, \mathbf{m}')$ , this pair satisfies the *epipolar equation*

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0, \quad (4)$$

where  $\mathbf{F} = [f_{ij}]$  is a  $3 \times 3$  *fundamental matrix* that incorporates information about the relative orientation and internal geometry of the cameras [3]. The matrix  $\mathbf{F}$  is subject to the *cubic constraint*

$$\det \mathbf{F} = 0. \quad (5)$$

Let  $\boldsymbol{\theta} = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^T$  be the vector of parameters, let  $\mathbf{x} = [m_1, m_2, m'_1, m'_2]^T$  be the vector of variables, and let

$$\mathbf{u}(\mathbf{x}) = [m_1m'_1, m_2m'_1, m'_1, m_1m'_2, m_2m'_2, m'_2, m_1, m_2, 1]^T$$

be the vector of carriers. Then,

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = \boldsymbol{\theta}^T \mathbf{u}(\mathbf{x})$$

and again, we see that (4) reduces to (1). Equation (5) can be expressed as (2), if we let

$$\psi(\boldsymbol{\theta}) = \theta_1(\theta_5\theta_9 - \theta_6\theta_8) + \theta_2(\theta_6\theta_7 - \theta_4\theta_9) + \theta_3(\theta_4\theta_8 - \theta_5\theta_7).$$

## 3 COST FUNCTIONS AND ESTIMATORS

We set out to estimate the parameter  $\boldsymbol{\theta}$  based on the available data  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ . Since (1) does not change if  $\boldsymbol{\theta}$  is multiplied by a nonzero scalar, the effective entity to be estimated is the one-dimensional ray through  $\boldsymbol{\theta}$ :

$$[\boldsymbol{\theta}] = \{t\boldsymbol{\theta} : t \text{ a nonzero scalar}\}.$$

Here, we tacitly assume that  $\boldsymbol{\theta} \neq 0$ .

Estimators of  $[\boldsymbol{\theta}]$  will be evolved from *cost functions*  $J(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)$  of special form. We shall use cost functions that are *homogeneous* in  $\boldsymbol{\theta}$ , or  $\boldsymbol{\theta}$ -homogeneous, such that

$$J(t\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = J(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)$$

for every nonzero scalar  $t$ . If—for simplicity—the ancillary constraint is set aside, then, given a  $\boldsymbol{\theta}$ -homogeneous cost function  $J$ , an *estimate*  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  based on  $J$  is defined by

$$J(\hat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta} \neq 0} J(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n).$$

This equation specifies  $\hat{\boldsymbol{\theta}}$  only to within a scalar factor. Nevertheless, the ray  $[\hat{\boldsymbol{\theta}}]$  is uniquely defined and we take this for the  $J$ -based estimate of  $[\boldsymbol{\theta}]$ . The assignment of  $\hat{\boldsymbol{\theta}}$  or  $[\hat{\boldsymbol{\theta}}]$  to  $\mathbf{x}_1, \dots, \mathbf{x}_n$  will be termed the  $J$ -based *estimator* of  $\boldsymbol{\theta}$  or  $[\boldsymbol{\theta}]$ , respectively. Henceforth, we shall deal exclusively with estimates and estimators of  $\boldsymbol{\theta}$ .

Once an estimate has been generated by minimizing a specific cost function, the ancillary constraint (if it applies) can further be accommodated via an adjustment procedure. One possibility is to use a general scheme delivering an “optimal correction” described in [7, Subsection 9.5.2]. In

what follows, we shall confine our attention to the estimation phase that precedes adjustment.

### 3.1 Algebraic Least-Squares Estimator

For each  $i = 1, \dots, n$ , let

$$A_i = \mathbf{u}(\mathbf{x}_i)\mathbf{u}(\mathbf{x}_i)^T \quad (6)$$

and let  $M = \sum_{i=1}^n A_i$ . A straightforward estimator is derived from the cost function

$$J_{\text{ALS}}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{\boldsymbol{\theta}^T M \boldsymbol{\theta}}{\|\boldsymbol{\theta}\|^2}.$$

The numerator  $\boldsymbol{\theta}^T M \boldsymbol{\theta}$  above is the sum of the squares of the algebraic distances  $|\boldsymbol{\theta}^T \mathbf{u}(\mathbf{x}_i)|$ . Accordingly, the  $J_{\text{ALS}}$ -based estimate of  $\boldsymbol{\theta}$  is termed the *algebraic least-squares (ALS) estimate* and is denoted  $\hat{\boldsymbol{\theta}}_{\text{ALS}}$ .

Since  $\hat{\boldsymbol{\theta}}_{\text{ALS}}$  is a minimizer of  $J_{\text{ALS}}$ , we have that, when  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\text{ALS}}$ , the following equation holds:

$$\partial_{\boldsymbol{\theta}} J_{\text{ALS}}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{0}^T.$$

Here,  $\partial_{\boldsymbol{\theta}} J_{\text{ALS}} = [\partial J_{\text{ALS}}/\partial \theta_1, \dots, \partial J_{\text{ALS}}/\partial \theta_l]$ . Direct computation shows that

$$[\partial_{\boldsymbol{\theta}} J_{\text{ALS}}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)]^T = 2\mathbf{X}\boldsymbol{\theta},$$

where

$$\mathbf{X}\boldsymbol{\theta} = \frac{M - J_{\text{ALS}}(\boldsymbol{\theta})\mathbf{I}_l}{\|\boldsymbol{\theta}\|^2},$$

$J_{\text{ALS}}(\boldsymbol{\theta})$  is short for  $J_{\text{ALS}}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)$ , and  $\mathbf{I}_l$  is the  $l \times l$  identity matrix. We see that  $\hat{\boldsymbol{\theta}}_{\text{ALS}}$  is an eigenvector of  $M$  with eigenvalue  $J_{\text{ALS}}(\hat{\boldsymbol{\theta}}_{\text{ALS}})$ . If  $\boldsymbol{\xi}$  is any eigenvector of  $M$  with a corresponding eigenvalue  $\lambda$ , then, necessarily,  $J_{\text{ALS}}(\boldsymbol{\xi}) = \lambda$ . Since  $J_{\text{ALS}}(\hat{\boldsymbol{\theta}}_{\text{ALS}}) \leq J_{\text{ALS}}(\boldsymbol{\xi})$ , we conclude that  $J_{\text{ALS}}(\hat{\boldsymbol{\theta}}_{\text{ALS}}) \leq \lambda$ . Thus,  $\hat{\boldsymbol{\theta}}_{\text{ALS}}$  is an eigenvector of  $M$  corresponding to the smallest eigenvalue. In practice,  $\hat{\boldsymbol{\theta}}_{\text{ALS}}$  can readily be determined by carrying out singular-value decomposition (SVD) of  $M$  (cf. [11, Section 5.7], [12, Section 2.6]).

### 3.2 Maximum-Likelihood Estimator

The ALS estimator treats all data as being equally valuable. When information about the measurement errors is available, it is desirable that it be incorporated into the estimation process. Prior knowledge of the measurement errors may indicate that some data are more reliable than others. This could be taken into account in the cost function by weighting the better data more heavily than the poorer data. Here, we present an estimator capable of informed weighting. It is based on the principle of maximum-likelihood and draws upon Kanatani's work on geometric fitting [7, chapter 7].

The measurement errors being generally unknowable, we regard the collective data  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  as a sample value taken on by an aggregate of vector-valued random variables  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ . We assume that the distribution of  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is not exactly specified, but is an element of a collection  $\{P_{\boldsymbol{\eta}} \mid \boldsymbol{\eta} \in H\}$  of candidate distributions, with  $H$  the set of all  $(n+1)$ -tuples  $\boldsymbol{\eta} = (\boldsymbol{\theta}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_n)$  such that  $\boldsymbol{\theta} \neq \mathbf{0}$  and

$$\boldsymbol{\theta}^T \mathbf{u}(\bar{\mathbf{x}}_1) = \dots = \boldsymbol{\theta}^T \mathbf{u}(\bar{\mathbf{x}}_n) = 0. \quad (7)$$

The candidate distributions are to be such that if some distribution  $P_{\boldsymbol{\eta}}$  is in effect, then each  $\mathbf{x}_i$  ( $i = 1, \dots, n$ ) is a noise-driven, fluctuating quantity around  $\bar{\mathbf{x}}_i$ .

We assume that the data come equipped with a collection  $(\Lambda_{\mathbf{x}_1}, \dots, \Lambda_{\mathbf{x}_n})$  of (known) positive definite  $k \times k$  covariance matrices. These matrices constitute repositories of prior information about the uncertainty of the data. We put the  $\Lambda_{\mathbf{x}_i}$  to use by assuming that, for each  $\boldsymbol{\eta} \in H$ ,  $P_{\boldsymbol{\eta}}$  is the unique distribution satisfying the following conditions:

- For any  $i, j = 1, \dots, n$  with  $i \neq j$ , the random vectors  $\mathbf{x}_i$  and  $\mathbf{x}_j$  (or, equivalently, the noises behind  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ) are stochastically independent.
- For each  $i = 1, \dots, n$ , the random vector  $\mathbf{x}_i$  has multivariate normal distribution with mean value vector  $\bar{\mathbf{x}}_i$  and covariance matrix  $\Lambda_{\mathbf{x}_i}$ .

Each  $P_{\boldsymbol{\eta}}$  can be described in terms of the probability density function

$$f(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n \mid \boldsymbol{\eta}) = (2\pi)^{-kn/2} \prod_{i=1}^n (\det \Lambda_{\mathbf{x}_i})^{-1/2} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\tilde{\mathbf{x}}_i - \bar{\mathbf{x}}_i)^T \Lambda_{\mathbf{x}_i}^{-1} (\tilde{\mathbf{x}}_i - \bar{\mathbf{x}}_i) \right\},$$

where  $(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n)$  is an arbitrary member of the sample space.

The collection  $\{P_{\boldsymbol{\eta}} \mid \boldsymbol{\eta} \in H\}$  is a *statistical model* of possible distributions of  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ . In this model, the random variables  $\mathbf{x}_1, \dots, \mathbf{x}_n$  are all placed on equal footing. In particular, these variables are not partitioned into a group of independent *explanatory* variables and a group of *response* variables that are functions of the members of the first group. As a result,  $\{P_{\boldsymbol{\eta}} \mid \boldsymbol{\eta} \in H\}$  falls into the category of the so-called *errors-in-variables* models [13], [14]. Furthermore, since each compound parameter  $\boldsymbol{\eta}$  is not arbitrary, but is subject to the constraint (7) that is geometric in nature, estimation based on  $\{P_{\boldsymbol{\eta}} \mid \boldsymbol{\eta} \in H\}$  is an instance of *geometric fitting* [7, chapter 7], [15].

The number  $f(\mathbf{x}_1, \dots, \mathbf{x}_n \mid \boldsymbol{\eta})$  indicates the probability (density) of obtaining  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  under the error model indexed by  $\boldsymbol{\eta}$ . As such,  $f(\mathbf{x}_1, \dots, \mathbf{x}_n \mid \boldsymbol{\eta})$  measures the match between the data and the predictions of the model. This property of the  $f(\mathbf{x}_1, \dots, \mathbf{x}_n \mid \boldsymbol{\eta})$  ( $\boldsymbol{\eta} \in H$ ) can be exploited to design an estimator of  $\boldsymbol{\eta}$  whereby the greatest confidence is given to that choice of  $\boldsymbol{\eta}$  for which the *likelihood function*  $\boldsymbol{\eta} \mapsto f(\mathbf{x}_1, \dots, \mathbf{x}_n \mid \boldsymbol{\eta})$  attains a maximum. Given the form of  $f(\mathbf{x}_1, \dots, \mathbf{x}_n \mid \boldsymbol{\eta})$ , it is immediately seen that the *maximum-likelihood estimate* thus defined is the parameter  $\hat{\boldsymbol{\eta}}_{\text{ML}}$  at which the cost function

$$J_{\text{ML}}(\boldsymbol{\eta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_i)^T \Lambda_{\mathbf{x}_i}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_i)$$

attains a minimum. Each term  $(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T \Lambda_{\mathbf{x}_i}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_i)$  in the above summation represents the squared *Mahalanobis distance* between  $\mathbf{x}_i$  and  $\bar{\mathbf{x}}_i$ . Note that the value of  $\hat{\boldsymbol{\eta}}_{\text{ML}}$  remains unchanged if the covariance matrices are multiplied by a common scalar.

The parameter  $\eta = (\theta; \bar{x}_1, \dots, \bar{x}_n)$  naturally splits into two parts:  $\pi_1(\eta) = \theta$  and  $\pi_2(\eta) = (\bar{x}_1, \dots, \bar{x}_n)$ . These parts encompass the *principal parameters* and *nuisance parameters*, respectively. We are mostly interested in the  $\pi_1$ -part of  $\hat{\eta}_{\text{ML}}$ , which we call the maximum-likelihood estimate of  $\theta$  and denote  $\hat{\theta}_{\text{ML}}$ . (As will become apparent shortly, the  $\pi_2$ -part of  $\hat{\eta}_{\text{ML}}$ , which can be viewed as the maximum-likelihood estimate of  $(x_1, \dots, x_n)$ , is a function of  $\hat{\theta}_{\text{ML}}$ .) The estimate  $\hat{\theta}_{\text{ML}}$  can be effectively determined by splitting the search for the minimizer of  $J_{\text{ML}}$  into two steps: one in which, for each principal parameter, a conditional minimizer is found by letting only the nuisance parameters vary; and one in which minimizers obtained in the first step are compared by letting the principal parameters vary. The splitting is embodied by the formula

$$\begin{aligned} \min_{\eta} J_{\text{ML}}(\eta; x_1, \dots, x_n) \\ = \min_{\theta \neq 0} \{ \min_{\eta} \{ J_{\text{ML}}(\eta; x_1, \dots, x_n) : \eta \text{ with } \pi_1(\eta) = \theta \} \}. \end{aligned}$$

For each  $\theta \neq 0$ , let  $\eta_{\theta} = (\theta; \bar{x}_1^{\theta}, \dots, \bar{x}_n^{\theta})$  be the compound parameter at which the inner minimum is attained. Then, clearly,

$$J_{\text{ML}}(\hat{\eta}_{\text{ML}}; x_1, \dots, x_n) = \min_{\theta \neq 0} J_{\text{ML}}(\eta_{\theta}; x_1, \dots, x_n).$$

If we introduce the function

$$J_{\text{ML}}^{\#}(\theta; x_1, \dots, x_n) = J_{\text{ML}}(\eta_{\theta}; x_1, \dots, x_n),$$

then  $\hat{\theta}_{\text{ML}}$  can be identified as the minimizer of  $J_{\text{ML}}^{\#}$  and we also have  $\hat{\eta}_{\text{ML}} = \eta_{\hat{\theta}_{\text{ML}}}$ , showing that determining  $\hat{\eta}_{\text{ML}}$  is tantamount to determining  $\hat{\theta}_{\text{ML}}$  and, in particular, that the maximum-likelihood estimate of  $(\bar{x}_1, \dots, \bar{x}_n)$  can be expressed in terms of  $\hat{\theta}_{\text{ML}}$  as  $(\bar{x}_1^{\hat{\theta}_{\text{ML}}}, \dots, \bar{x}_n^{\hat{\theta}_{\text{ML}}})$ .

The function  $J_{\text{ML}}^{\#}$  does not lend itself to explicit calculation, so the determination of  $\hat{\theta}_{\text{ML}}$  directly on the basis of  $J_{\text{ML}}^{\#}$  is impractical. However, as we shall see next, a tractable approximation to  $J_{\text{ML}}^{\#}$  can be derived. The new cost function will serve as a basis for determining an approximation of  $\hat{\theta}_{\text{ML}}$ .

Interestingly,  $\hat{\eta}_{\text{ML}}$  can be found by employing another method that requires no intermediary step involving  $J_{\text{ML}}^{\#}$ . As it turns out,  $\hat{\eta}_{\text{ML}}$  and a vector of certain additional unknowns (which happen to be suitable Lagrange multipliers) satisfy a system of equations that expresses, in a differential form, the fact that  $\hat{\eta}_{\text{ML}}$  is a minimizer of a constrained problem. This system can be solved numerically by employing, say, the Newton-Raphson method. Finding  $\hat{\eta}_{\text{ML}}$  (and, hence,  $\hat{\theta}_{\text{ML}}$ ) in this way constitutes the so-called *direct approach* to maximum-likelihood fitting [15], [16], [17]. In contrast, finding  $\hat{\theta}_{\text{ML}}$  (and, hence,  $\hat{\eta}_{\text{ML}}$ ) by resorting to  $J_{\text{ML}}^{\#}$  embodies the so-called *reduced approach* [17].

### 3.3 Approximated Maximum-Likelihood Estimator

To obtain an approximation of  $J_{\text{ML}}^{\#}$ , first note that the definition of  $\eta_{\theta}$  involves minimization subject to the constraints  $\theta^T u(\bar{x}_i) = 0$  ( $i = 1, \dots, n$ ). This minimization can effectively be handled with the use of Lagrange multipliers. Fix  $i = 1, \dots, n$  arbitrarily. Note that the gradient (the column vector of the partial derivatives) of  $(x_i - y)^T \Lambda_{x_i}^{-1}(x_i - y)$  with respect to  $y$  is  $-\Lambda_{x_i}^{-1}(x_i - y)$  and

the gradient of  $\theta^T u(y)$  with respect to  $y$  is  $\partial_x u(y)^T \theta$ , where  $\partial_x u(y) = [(\partial u_i / \partial x_j)(y)]_{1 \leq i \leq l, 1 \leq j \leq k}$ . Now, the constrained minimization condition implies that, evaluated at  $\bar{x}_i^{\theta}$ , these two gradients are proportional; that is, there exists a (scalar) Lagrange multiplier  $\lambda_i$  such that

$$\Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta}) = \lambda_i \partial_x u(\bar{x}_i^{\theta})^T \theta. \quad (8)$$

Multiplying both sides of this equation by the positive definite square root  $\Lambda_{x_i}^{1/2}$  of  $\Lambda_{x_i}$ , we obtain

$$\Lambda_{x_i}^{-1/2}(x_i - \bar{x}_i^{\theta}) = \lambda_i \Lambda_{x_i}^{1/2} \partial_x u(\bar{x}_i^{\theta})^T \theta,$$

where  $\Lambda_{x_i}^{-1/2}$  denotes the inverse of  $\Lambda_{x_i}^{1/2}$ . Hence,

$$\lambda_i^2 = \frac{\| \Lambda_{x_i}^{-1/2}(x_i - \bar{x}_i^{\theta}) \|^2}{\| \Lambda_{x_i}^{1/2} \partial_x u(\bar{x}_i^{\theta})^T \theta \|^2} = \frac{(x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta})}{\theta^T \partial_x u(\bar{x}_i^{\theta}) \Lambda_{x_i} \partial_x u(\bar{x}_i^{\theta})^T \theta}. \quad (9)$$

Multiplying both sides of (8) by  $(x_i - \bar{x}_i^{\theta})^T$  gives

$$(x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta}) = \lambda_i (x_i - \bar{x}_i^{\theta})^T \partial_x u(\bar{x}_i^{\theta})^T \theta. \quad (10)$$

Now, replacing the difference  $u(x_i) - u(\bar{x}_i^{\theta})$  by the first-order term  $\partial_x u(\bar{x}_i^{\theta})(x_i - \bar{x}_i^{\theta})$  in the Taylor expansion of  $y \mapsto u(y) - u(\bar{x}_i^{\theta})$  about  $\bar{x}_i^{\theta}$  and taking into account that  $\theta^T u(\bar{x}_i^{\theta}) = 0$ , we obtain

$$\theta^T u(x_i) = \theta^T \partial_x u(\bar{x}_i^{\theta})(x_i - \bar{x}_i^{\theta}). \quad (11)$$

Equation (10) can now be written as

$$(x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta}) = \lambda_i u(x_i)^T \theta.$$

This, jointly with (9), yields

$$\begin{aligned} & \left[ (x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta}) \right]^2 \\ &= \lambda_i^2 \theta^T u(x_i) u(x_i)^T \theta \\ &= \frac{(x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta}) \cdot \theta^T u(x_i) u(x_i)^T \theta}{\theta^T \partial_x u(\bar{x}_i^{\theta}) \Lambda_{x_i} \partial_x u(\bar{x}_i^{\theta})^T \theta}, \end{aligned}$$

whence, upon dividing both sides by  $(x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta})$ ,

$$(x_i - \bar{x}_i^{\theta})^T \Lambda_{x_i}^{-1}(x_i - \bar{x}_i^{\theta}) = \frac{\theta^T u(x_i) u(x_i)^T \theta}{\theta^T \partial_x u(\bar{x}_i^{\theta}) \Lambda_{x_i} \partial_x u(\bar{x}_i^{\theta})^T \theta}.$$

If we further simplify this by substituting  $\partial_x u(x_i)$  for  $\partial_x u(\bar{x}_i^{\theta})$ , we finally arrive at the following approximate expression  $J_{\text{AML}}$  for  $J_{\text{ML}}^{\#}$ :

$$J_{\text{AML}}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \frac{\theta^T u(x_i) u(x_i)^T \theta}{\theta^T \partial_x u(x_i) \Lambda_{x_i} \partial_x u(x_i)^T \theta}.$$

If, for each  $i = 1, \dots, n$ , we let

$$B_i = \partial_x u(x_i) \Lambda_{x_i} \partial_x u(x_i)^T,$$

then, recalling the definition (6),  $J_{\text{AML}}$  can be simply written as

$$J_{\text{AML}}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \frac{\theta^T A_i \theta}{\theta^T B_i \theta}. \quad (12)$$

The  $J_{\text{AML}}$ -based estimate of  $\theta$  will be called the *approximated maximum-likelihood (AML) estimate* and will be denoted  $\hat{\theta}_{\text{AML}}$ .

It should be observed that  $J_{\text{AML}}$  can be derived in an alternative manner. This can be done without recourse to the maximum-likelihood principle by, for example, using a gradient weighted approach that also incorporates covariances. Various terms may be used to describe the method that aims to minimize  $J_{\text{AML}}$ , although some of the terms may not be fully discriminating. Candidate labels include “heteroscedastic regression” [18], “weighted orthogonal regression” [13], [19], and “gradient weighted least-squares” [20].

We finally note that an alternative method of finding an approximation to  $\hat{\theta}_{\text{ML}}$  has recently been developed by Leedan and Meer [18], [21] and further generalized by Matei and Meer [9]. It is based on an application of the maximum-likelihood principle to a statistical model formed with the use of approximations to candidate distributions of  $(u(x_1), \dots, u(x_n))$ , each of which is derived from a respective candidate distribution of  $(x_1, \dots, x_n)$ .

### 3.4 Example Cost Functions

It is instructive to revisit our fitting examples from Sections 2.1 and 2.2 and present in each case the function  $J_{\text{AML}}$  in the form in which the parameters to be estimated are represented by a matrix rather than a vector. It is easily seen that, in the case of conic fitting,  $J_{\text{AML}}$  reduces to the cost function

$$J(A) = \frac{1}{2} \sum_{i=1}^n \frac{(m_i^T A m_i)^2}{m_i^T A \Lambda_{m_i} A^T m_i}$$

with the covariance matrices having the form

$$\Lambda_m = \begin{bmatrix} \Lambda_p & 0 \\ 0^T & 0 \end{bmatrix}, \quad (13)$$

where  $\Lambda_p$  is the “natural”  $2 \times 2$  covariance matrix corresponding to the vector  $p$  defined via the representation  $m = [p^T, 1]^T$  (or, equivalently, defined as the vector of coordinates  $p = [m_1, m_2]^T$ ). In the case of fundamental matrix estimation,  $J_{\text{AML}}$  reduces to Zhang’s favored *gradient weighted least-squares* cost function (see  $J_2$  in the appendix of [20]) given by

$$J(F) = \sum_{i=1}^n \frac{(m_i'^T F m_i')^2}{m_i'^T F \Lambda_{m_i'} F^T m_i' + m_i'^T F^T \Lambda_{m_i'} F m_i'},$$

where again the covariance matrices have the form given in (13).

## 4 MINIMIZING THE COST FUNCTION

Minimizing  $J_{\text{AML}}$  is a challenging problem. One strategy is to rely upon a general solver such as the Levenberg-Marquardt method [12, Section 15.5], or the downhill simplex method of Nelder and Mead [12, Section 10.4] or one of the direction set methods of Powell [12, Section 10.5]. However, such a solver will not take into account the special form of the problem and may be unduly slow. A

commonly adopted approach to minimizing functions involving fractional expressions is that ascribed to Sampson [6]. When applied to  $J_{\text{AML}}$ , Sampson’s scheme (SMP) starts the search for a minimizer by substituting an initial estimate into the denominators  $\theta^T B_i \theta$  in (12), which transforms our problem into one of minimizing a quadratic form in  $\theta$ ; the latter problem is straightforward and can be solved using SVD. This process is then repeated with the newly obtained estimate substituted into the denominators until a measure of convergence is obtained. (Note, however, that Sampson’s original scheme did not incorporate covariance matrices.)

Kanatani showed that this approach, involving a “freezing” of the denominators, is subject to systematic bias. Accordingly, he devised the technique of renormalization in which an attempt is made at each iteration to undo the bias [7, chapter 12]. We show in a companion paper [8] that the estimate  $\hat{\theta}_{\text{REN}}$  obtained via renormalization does not theoretically act to compute  $\hat{\theta}_{\text{AML}}$ . However, a case may be made that this is unimportant given that  $\hat{\theta}_{\text{REN}}$  and  $\hat{\theta}_{\text{AML}}$  are both first-order approximations to  $\hat{\theta}_{\text{ML}}$  and as such are likely to be statistically equivalent.

In this section, we derive a straightforward minimization scheme that is a genuine means of theoretically determining the minimizer of  $J_{\text{AML}}$ .

### 4.1 Variational Equation

Since  $\hat{\theta}_{\text{AML}}$  is a minimizer of  $J_{\text{AML}}$ , we have that, when  $\theta = \hat{\theta}_{\text{AML}}$ , the following equation holds:

$$\partial_{\theta} J_{\text{AML}}(\theta; x_1, \dots, x_n) = 0^T. \quad (14)$$

We term this the *variational equation*. Direct computation shows that

$$[\partial_{\theta} J_{\text{AML}}(\theta; x_1, \dots, x_n)]^T = 2X_{\theta}\theta, \quad (15)$$

where

$$X_{\theta} = \sum_{i=1}^n \frac{A_i}{\theta^T B_i \theta} - \sum_{i=1}^n \frac{\theta^T A_i \theta}{(\theta^T B_i \theta)^2} B_i. \quad (16)$$

Thus, (14) can be written as

$$X_{\theta}\theta = 0. \quad (17)$$

This is a nonlinear equation and is unlikely to admit solutions in closed form.

### 4.2 Fundamental Numerical Scheme

Closed-form solutions of the variational equation may be infeasible, so in practice  $\hat{\theta}_{\text{AML}}$  has to be found numerically. We assume that  $\hat{\theta}_{\text{AML}}$  lies close to  $\hat{\theta}_{\text{ALS}}$  so as to guarantee that, when seeded with  $\hat{\theta}_{\text{ALS}}$ , the numerical method that we are going to develop generates an estimate that coincides with  $\hat{\theta}_{\text{AML}}$ . Taking  $\hat{\theta}_{\text{ALS}}$  as an initial guess  $\theta_0$ , we construct a sequence  $\{\theta_k\}$  of successive updates. Under favorable conditions, the sequence will converge. Exploiting the assumption about the accuracy of the initial guess, we take the corresponding limit for the final estimate. In practice, the limit will be identified with the final term of  $\{\theta_k\}$  stopped after a finite number of steps. The stopping rule will be the choice of the first  $k$  such that the distance of

1. Set  $\theta_0 = \hat{\theta}_{\text{ALS}}$ .
2. Assuming  $\theta_{k-1}$  is known, compute the matrix  $X_{\theta_{k-1}}$ .
3. Compute a normalised eigenvector of  $X_{\theta_{k-1}}$  corresponding to the eigenvalue closest to zero (in absolute value) and take this eigenvector for  $\theta_k$ .
4. If  $\theta_k$  is sufficiently close to  $\theta_{k-1}$ , then terminate the procedure; otherwise increment  $k$  and return to Step 2.

Fig. 1. Fundamental numerical scheme.

some kind between  $\theta_{k-1}$  and  $\theta_k$ , say  $\|\theta_{k-1} - \theta_k\|$ , is less than a preassigned quantity.

A straightforward algorithm for numerically solving the variational equation can be derived by realizing that a vector  $\theta$  satisfies (17) if and only if it falls into the null space of the matrix  $X_\theta$ . Thus, if  $\theta_{k-1}$  is a tentative guess, then an improved guess can be obtained by picking a vector  $\theta_k$  from that eigenspace of  $X_{\theta_{k-1}}$  which most closely approximates the null space of  $X_\theta$ ; this eigenspace is, of course, the one corresponding to the eigenvalue closest to zero in absolute value. The *fundamental numerical scheme* (FNS) implementing this idea is presented in Fig. 1. The algorithm can be regarded as a variant of the Newton-Raphson method.

A commonly adopted assumption is that, when initialized sufficiently close to a solution of the variational equation, the algorithm will produce a succession of estimates converging to this solution. We shall not give a formal proof of convergence, but shall show that, when the sequence of updates converges, the limit is a solution of (17). In other words, we shall demonstrate that the method delivers a proper outcome whenever convergence holds.

Suppose that  $\{\theta_k\}$  is a sequence of normalized vectors convergent to  $\theta_\infty$  and such that

$$X_{\theta_{k-1}}\theta_k = \lambda_k\theta_k, \quad (18)$$

where  $\lambda_k$  is scalar. Let  $\{\lambda_{k_n}\}$  be any convergent subsequence of  $\{\lambda_k\}$  and let  $\lambda_\infty = \lim_{n \rightarrow \infty} \lambda_{k_n}$ . Substituting  $k_n$  for  $k$  in (18) and letting  $n \rightarrow \infty$ , we obtain

$$X_{\theta_\infty}\theta_\infty = \lambda_\infty\theta_\infty. \quad (19)$$

It immediately follows from (16) that  $\theta^T X_\theta \theta = 0$  for all  $\theta \neq 0$ . Hence,

$$0 = \theta_\infty^T X_{\theta_\infty} \theta_\infty = \lambda_\infty \|\theta_\infty\|^2,$$

showing that  $\lambda_\infty = 0$  and, in view of (19), that the limit vector  $\theta_\infty$  satisfies (17), as expected.

As a side observation, note that, by the above argument, every convergent subsequence of  $\{\lambda_k\}$  converges to zero. Since a sequence is convergent whenever each of its convergent subsequences converges to one and the same limit, we conclude that  $\{\lambda_k\}$  converges and has limit zero. Thus,  $\lim_{k \rightarrow \infty} \lambda_k = 0$  is a necessary condition for the convergence of  $\{\theta_k\}$ .

## 5 EXPERIMENTAL RESULTS

The fundamental numerical scheme and other algorithms were tested on the problems of conic fitting [1], [2], [6], [10],

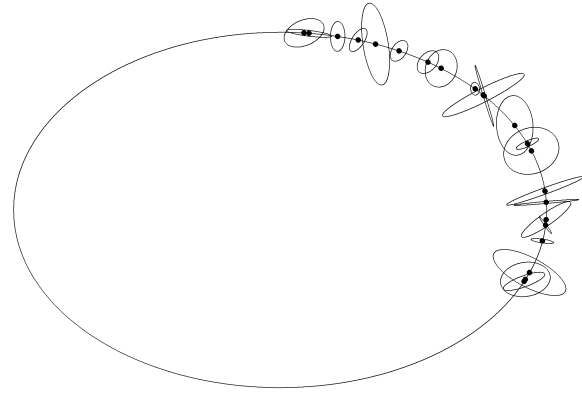


Fig. 2. True ellipse, true points, and covariance ellipses.

[22], [23], [24], [25], [26], [27], [28], [29], [30], [31] and estimating the fundamental matrix [2], [20], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43]. Experiments reported here are synthetic as they permit precise control of the conditions under which performance can be evaluated. In our tests, no special knowledge of the domain was utilized. Furthermore, no attempt was made to eliminate outliers, our experiments at this stage comparing general techniques for solving problems of the form specified in (1). In both conic fitting and fundamental matrix estimation, ancillary constraints were ignored. In particular, it was not assumed that the conic was an ellipse. In the case of fundamental matrix estimation, the cubic constraint was not enforced as a postprocess. As it happens, our (unconstrained) fundamental numerical scheme always generated fundamental matrices which, when scaled to unit Frobenius norm, turned out to have determinants smaller than  $10^{-11}$ .

### 5.1 Conic Fitting

A randomly oriented ellipse was generated such that the ratio of its major to minor axes was in the range  $[2, 3]$  and its major axis was approximately 200 pixels in length. About one third of the ellipse's boundary was chosen as the base curve and this included the point of maximum curvature of the ellipse. A set of *true points* was then randomly selected from a distribution uniform along the length of the base curve. A range of tests was then conducted. Each test was carried out with respect to an *average level of noise*,  $\sigma$ , defined as the expected value of the trace of the covariance matrices employed in the test.

For a given test, true points were perturbed randomly in accordance with their associated covariance matrices, yielding the *data points*. In general, the noise conformed to an inhomogeneous and anisotropic distribution. Fig. 2 shows a large ellipse, some selected true points, and a set of ellipses associated with these points. Each of the smaller ellipses represents a level set of the probability density function used to generate an individual datum, and as such captures graphically the nature of the uncertainty described by its covariance matrix.

The following procedure was adopted for generating covariance matrices associated with image points, prescribing anisotropic and inhomogeneous noise at a given average level  $\sigma$ . Given a point  $p$ , the scale  $\alpha$  of a prospective covariance matrix  $\Lambda_p$  was first selected from a uniform

TABLE 1  
Conic Fitting Errors (in Pixels) vs. Average Noise Level

AVERAGE NOISE LEVEL	SCHEME								
	ALS	SMP*	LM*	FNS*	TAU	SMP	REN	LM	FNS
1	2.61	2.53	2.43	2.43	2.51	1.02	1.00	1.00	1.00
2	5.66	5.48	5.03	5.03	5.25	2.17	2.05	2.05	2.05
3	8.63	8.32	7.45	7.45	7.81	3.31	3.10	3.09	3.09
4	12.09	11.73	10.07	10.07	10.55	4.50	4.15	4.12	4.12
5	15.63	15.30	12.86	12.86	13.38	5.78	5.21	5.19	5.21
6	19.98	19.20	15.51	15.51	16.38	6.89	6.07	6.05	6.05
7	23.45	22.65	17.82	17.82	18.79	8.18	7.16	7.14	7.14
8	27.08	26.41	20.12	20.12	21.21	9.61	8.23	8.12	8.22
9	30.16	29.54	22.06	22.06	23.38	10.71	9.02	8.96	8.96
10	34.74	34.10	24.60	24.60	26.03	12.21	10.13	10.10	10.10

distribution in the range  $[0, 2\sigma]$ . (Similar results were obtained using other distributions.) Next, an eccentricity parameter  $\beta$  was generated from a uniform distribution between 0 and 0.5. An intermediate covariance matrix was then formed by setting

$$\Lambda'_p = \alpha \begin{bmatrix} \beta & 0 \\ 0 & 1 - \beta \end{bmatrix}.$$

This matrix was then “rotated” by an angle  $\gamma$  selected from a uniform distribution between 0 and  $2\pi$  to generate the final covariance

$$\Lambda_p = O_\gamma \Lambda'_p O_\gamma^T$$

with

$$O_\gamma = \begin{bmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{bmatrix}.$$

Since  $\mathbf{E}[\text{Tr } \Lambda] = \mathbf{E}[\text{Tr } \Lambda'] = \mathbf{E}[\alpha] = \sigma$ , where  $\text{Tr } A$  denotes the trace of the matrix  $A$ , the noise generated by employing  $\Lambda_p$  has the average level  $\sigma$ .

Each method was then challenged to determine the coefficients of the best fitting conic. If the method was able to utilize uncertainty information, it was supplied with the appropriate covariance matrices. For each  $\sigma$ , the conic coefficients were computed 2,000 times from 60 data points generated independently in each repetition (with new true data points and covariance matrices generated at each iteration). Critically, for each set of coefficients obtained, an error measure was computed as the sum of the shortest pixel distances of each *true point* from the estimated ellipse. A composite error measure was then obtained by averaging this error over all 2,000 trials. This entire process was then repeated for different average levels of noise (with  $\sigma$  varying from 1 to 10 pixels).

Note that the error measure used in the experiment took advantage of the fact that the underlying true points were known. Were these unknown, an alternative measure might be the sum of the Mahalanobis distances from the data points to the estimated ellipses.

The methods tested were as follows:

- **ALS:** Algebraic least-squares scheme.
- **SMP\*:** Sampson’s scheme with identity matrices.
- **TAU:** Taubin-like scheme.
- **FNS\*:** FNS with identity matrices.

- **LM\*:** LM with identity matrices.
- **SMP:** Sampson’s scheme.
- **REN:** Renormalization scheme.
- **LM:** Levenberg-Marquardt scheme.
- **FNS:** Fundamental numerical scheme.

Recall that SMP is effectively Sampson’s method with covariances (see Section 4). SMP\* and FNS\* are the variants of SMP and FNS obtained by replacing true covariances by the default covariance  $\text{diag}(1, 1)$ . The ALS, SMP\*, and FNS\* methods are included so as to indicate the performance of estimators when covariance information is not available.

The REN scheme chosen was a second-order method; in tests carried out in a companion paper, it performed as well as any of the renormalization variants (see the SORIII scheme in [8]).

A Levenberg-Marquardt (LM) scheme was included so as to provide a baseline in both accuracy and timing trials; specifically, the MINPACK routine LMDER (available from <http://www.netlib.org/minpack>) was used to directly minimize  $J_{\text{AML}}$ , with the analytical derivatives of  $J_{\text{AML}}$ , as in (15), supplied so as to improve the execution time.

The TAU scheme is a one-step, noniterative method, which aims to minimize the approximation to  $J_{\text{AML}}$  given by  $(\sum_{i=1}^n \theta^T A_i \theta) / (n^{-1} \sum_{i=1}^n \theta^T B_i \theta)$ . The scheme is a variant of the method of Taubin [31] in that it incorporates covariances matrices.

The ALS method uses the LINPACK routine DSVDC (<http://www.netlib.org/linpack>) to perform SVD and the EISPACK routine RS (<http://www.netlib.org/eispack>) is used in those methods requiring computation of eigenvalues and associated eigenvectors since the matrices involved are symmetric. It should also be noted that the various iterative schemes were supplied with similar stopping conditions so as to enable fair comparison.

Table 1 shows the average error (in terms of the shortest pixel distance of a true point to the estimated ellipse) obtained for each of the methods. Fig. 3 shows the tabular data in graphical form.

We note that errors rise approximately linearly with increasing noise. As would be expected, the results show that, in the face of data contaminated with inhomogeneous and anisotropic noise, those (uninformed) methods which do not make use of covariance information perform markedly worse than those (informed) methods that do use this information. An exception to this is the informed,

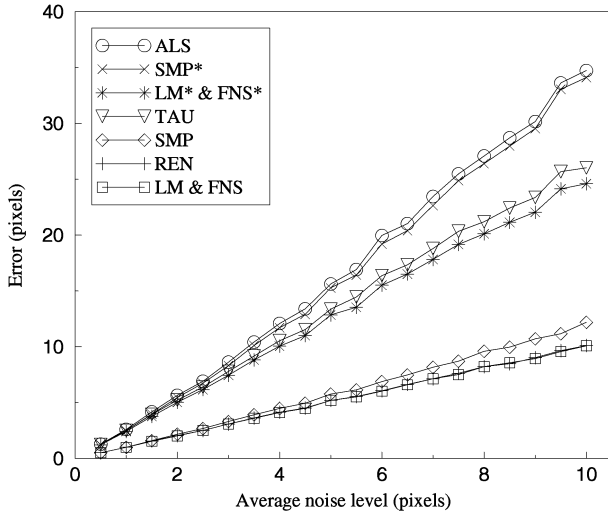


Fig. 3. Conic fitting errors (in pixels) vs. average noise level.

noniterative method of TAU, which fails even to match the uninformed FNS\*. Of the uninformed methods, LM\* and FNS\* are identical in accuracy and ahead of both ALS and SMP\*. Of the informed methods, REN, LM, FNS are clear leaders, with SMP lagging behind systematically with average errors up to 21 percent greater.

## 5.2 Fundamental Matrix Estimation

A realistic stereo camera configuration was first selected with non-co-planar optical axes and slightly differing left and right camera intrinsic parameters. Randomly chosen 3D points were then projected onto the images so as to generate many pairs of corresponding points. A range of tests was then conducted, with each test being carried out with respect to an average level of noise.

For a given test, image points were perturbed by adding inhomogeneous anisotropic noise at some average level  $\sigma$ , consistent with covariance matrices generated via the method described below. All tests involved 60 pairs of perturbed corresponding points. (Similar results were observed in trials using between 40 and 100 pairs.) Fig. 4 depicts a typical image, showing a subset of (unperturbed) image points together with associated ellipses. Each of the ellipses represents a level set of the probability density function describing the noise spread around the ellipse center and, as such, captures graphically the covariances employed.

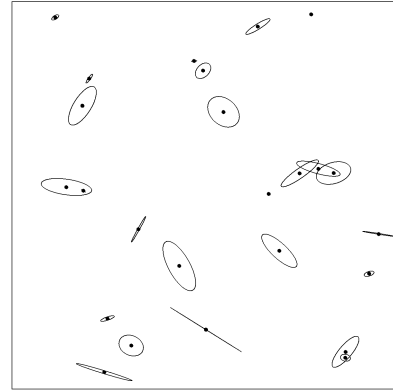


Fig. 4. True stereo image points and associated covariance ellipses.

The following procedure was adopted for generating covariance matrices associated with points of the form  $x = [p^T, p'^T]^T$ , where  $p$  and  $p'$  are corresponding points. First, covariance matrices  $\Lambda_p$  and  $\Lambda_{p'}$  were generated using the recipe given earlier. Next, covariance matrices  $\Lambda_x$  were constructed by setting

$$\Lambda_x = \begin{bmatrix} \Lambda_p & 0 \\ 0 & \Lambda_{p'} \end{bmatrix}.$$

Each method under test was then challenged to compute the fundamental matrix. If the method was able to assimilate uncertainty information, it was supplied with the appropriate covariance matrices. For each  $\sigma$ , the fundamental matrix was computed 2,000 times from a given set of 60 corresponding (true) points, but with fresh covariance matrices and perturbations generated each time. For each fundamental matrix obtained, an error measure was computed as the sum of the distances of the underlying *true points* to the epipolar lines derived from the estimated fundamental matrix, in both the left and right images. A composite error measure was then obtained by averaging this error over all 2,000 trials. This entire process was then repeated for different average levels of noise (with  $\sigma$  varying from 1 to 10 pixels in steps of 1).

As previously mentioned, the methods tested were ALS, SMP\*, LM\*, FNS\*, TAU, SMP, REN, LM, and FNS, with SMP\* and FNS\* being the SMP and FNS schemes with all covariance matrices taken to be  $\text{diag}(1, 1, 1, 1)$ . Again, the ALS, SMP\*, and FNS\* methods were included to indicate

TABLE 2  
Fundamental Matrix Estimation Errors (in Pixels) vs. Average Noise Level

AVERAGE NOISE LEVEL	SCHEME								
	ALS	SMP*	LM*	FNS*	TAU	SMP	REN	LM	FNS
1	0.290	0.286	0.279	0.279	0.284	0.214	0.210	0.210	0.210
2	0.592	0.580	0.551	0.551	0.562	0.404	0.386	0.386	0.386
3	0.946	0.934	0.865	0.865	0.878	0.676	0.634	0.634	0.634
4	1.226	1.228	1.117	1.117	1.126	0.860	0.787	0.786	0.786
5	1.629	1.625	1.485	1.485	1.501	1.126	1.020	1.019	1.019
6	2.022	2.000	1.778	1.778	1.811	1.404	1.303	1.300	1.300
7	2.401	2.376	2.022	2.022	2.053	1.701	1.482	1.481	1.481
8	2.663	2.686	2.236	2.236	2.255	1.857	1.576	1.573	1.573
9	3.136	3.118	2.523	2.523	2.565	2.116	1.781	1.779	1.779
10	3.461	3.502	2.797	2.797	2.820	2.407	1.971	1.967	1.967



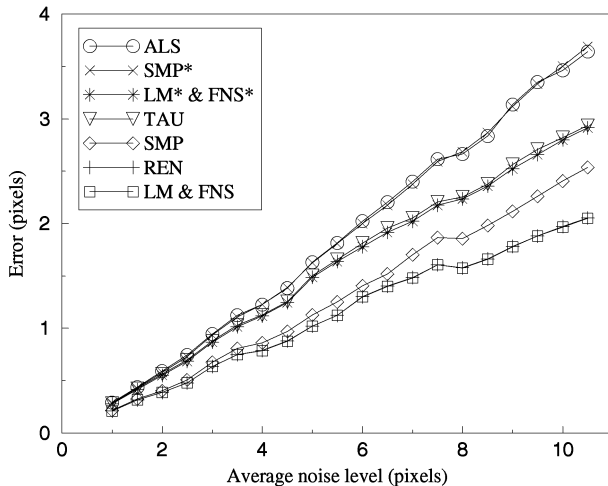


Fig. 5. Fundamental matrix estimation errors (in pixels) vs. average noise level.

the performance of the measures when covariance information was not utilized. The REN scheme chosen was the same as that used for conic fitting.

Table 2 shows the average epipolar-distance pixel errors obtained for each method. Fig. 5 depicts the tabular data in graphical form. We note again that errors mount approximately linearly with increasing noise. As would be expected, the results show that, in the face of data contaminated with inhomogeneous and anisotropic noise, those methods which do not make use of uncertainty (covariance) information perform significantly less well than those that do. Again, however, the informed, non-iterative method of TAU fails to match the uninformed FNS\*. Among the noncovariance methods, FNS\* and LM\* achieve notably better results than ALS and SMP\*. FNS, REN, and LM perform almost on par and ahead of SMP (which had errors up to 22 percent worse).

### 5.3 Timing Tests

Timing tests were conducted on the various schemes. Stopping conditions were devised so as to place similar demands upon each of the iterative methods. None of the schemes was affected either by change in noise level or the provision of default (identity) covariances in place of full covariances. Accordingly, times for LM\* and FNS\* are not given. Fig. 6 shows histogram timing data for each of the methods ALS, SMP, LM, and FNS. In each case, the left bar denotes the time taken to complete a single test, averaged over the complete suite of conic-fitting experiments; analogously, the right bar refers to the fundamental-matrix estimation experiments. REN was not included, as SORIII is a particularly slow (but accurate) form of renormalization. FNS typically converged within eight iterations in the case of conic fitting and four iterations in the case of fundamental matrix estimation. Oscillation was not found to be a practical problem. Interestingly, FNS emerges as being significantly faster than LM while generating essentially the same results. It should also be noted that LM and FNS did not speed up when their initial values were supplied by TAU rather than ALS.

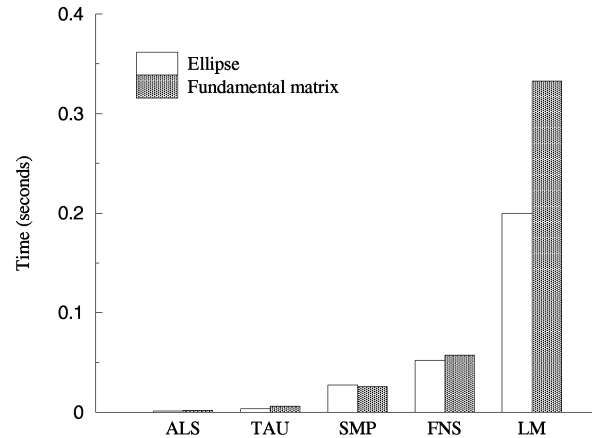


Fig. 6. Timing results for various methods.

## 6 CONCLUSION

We analyzed the problem of estimating parameters of a model described by an equation of special form. The estimation problem was handled via regression in which a surface is to be fitted to data and their associated covariances. A fundamental numerical scheme was derived that acts as a theoretical minimizer of an approximated maximum-likelihood cost function. Resting upon a novel variational equation, the scheme is simply derived and implemented, computationally efficient, and unsurpassed in our accuracy tests. The main contribution of this work can be seen as elucidatory, with a novel framework being generated for understanding the reduced approach to the maximum-likelihood estimation of parameters arising in a class of computer vision problems.

## ACKNOWLEDGMENTS

The authors are grateful for the insightful comments of Marino Ivancic, Kenichi Kanatani, Garry Newsam, Naoya Ohta, and Robyn Owens. In addition, the authors would like to thank three anonymous referees for providing suggestions that led to improvements in the presentation of the paper. This work was in part funded by the Australian Research Council and the Cooperative Research Centre for Sensor Signal and Information Processing.

## REFERENCES

- [1] F. Bookstein, "Fitting Conic Sections to Scattered Data," *Computer Graphics and Image Processing*, vol. 9, no. 1, pp. 56-71, 1979.
- [2] Z. Zhang, "Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting," *Image and Vision Computing*, vol. 15, no. 1, pp. 57-76, 1997.
- [3] O.D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, Mass.: MIT Press, 1993.
- [4] M.J. Brooks, W. Chojnacki, and L. Baumela, "Determining the Egomotion of an Uncalibrated Camera from Instantaneous Optical Flow," *J. Optical Soc. Am. A*, vol. 14, no. 10, pp. 2,670-2,677, 1997.
- [5] T. Viéville and O.D. Faugeras, "Motion Analysis with a Camera with Unknown, and Possibly Varying Intrinsic Parameters," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 750-756, June 1995.
- [6] 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.
- [7] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam: Elsevier, 1996.

- [8] W. Chojnacki, M.J. Brooks, and A. van den Hengel, "Rationalising the Renormalisation Method of Kanatani," *J. Math. Imaging and Vision*, 2001, to appear.
- [9] B. Matei and P. Meer, "A General Method for Errors-in-Variables Problems in Computer Vision," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 18-25, June 2000.
- [10] B. Matei and P. Meer, "Reduction of Bias in Maximum-Likelihood Ellipse Fitting," *Proc. 15th Int'l Conf. Pattern Recognition*, vol. 3, pp. 802-806, Sept. 2000.
- [11] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Applications*, second ed. San Diego, Calif.: Academic Press, 1985.
- [12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C*. Cambridge UK: Cambridge Univ. Press, 1995.
- [13] W.A. Fuller, *Measurement Error Models*. New York: John Wiley & Sons, 1987.
- [14] S. Van Huffel and J. Vandewalle, *Total Least Squares Problems*. Philadelphia: SIAM, 1991.
- [15] B. Triggs, "A New Approach to Geometric Fitting." Available from <http://www.inrialpes.fr/movi/people/Triggs>.
- [16] W. Förstner, "On Weighting and Choosing Constraints for Optimally Reconstructing the Geometry of Image Triplets," *Proc. Sixth European Conf. Computer Vision*, D. Vernon, ed., vol. 2, pp. 669-684, June 2000.
- [17] B. Triggs, "Optimal Estimation of Matching Constraints," *ECCV'98 Workshop 3D Structure from Multiple Images of Large-Scale Environments (SMILE)*, R. Koch and L. Van Gool, eds., June 1998. Available from <http://www.inrialpes.fr/movi/people/Triggs>.
- [18] Y. Leedan and P. Meer, "Heteroscedastic Regression in Computer Vision: Problems with Bilinear Constraint," *Int'l J. Computer Vision*, vol. 37, no. 2, pp. 127-150, 2000.
- [19] P.T. Boggs and J.E. Rodgers, "Orthogonal Distance Regression," *Contemporary Mathematics*, vol. 112, pp. 183-194, 1989.
- [20] Z. Zhang, "On the Optimization Criteria Used in Two-View Motion Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 717-729, July 1998.
- [21] Y. Leedan and P. Meer, "Estimation with Bilinear Constraints in Computer Vision," *Proc. Sixth Int'l Conf. Computer Vision*, pp. 733-738, Jan. 1998.
- [22] T. Ellis, "Ellipse Detection and Matching with Uncertainty," *Image and Vision Computing*, vol. 10, no. 2, pp. 271-276, 1992.
- [23] A. Fitzgibbon, M. Pilu, and R.B. Fisher, "Direct Least Square Fitting of Ellipses," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476-480, May 1999.
- [24] A.W. Fitzgibbon and R.B. Fisher, "A Buyer's Guide to Conic Fitting," *Proc. Sixth British Machine Vision Conf.*, D. Pycock, ed., vol. 2, pp. 513-522, Sept. 1995.
- [25] W. Gander, G.H. Golub, and R. Strebel, "Least-Square Fitting of Circles and Ellipses," *BIT*, vol. 43, pp. 558-578, 1994.
- [26] K. Kanatani, "Statistical Bias of Conic Fitting and Renormalisation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 320-326, Mar. 1994.
- [27] Y. Kanazawa and K. Kanatani, "Optimal Conic Fitting and Reliability Evaluation," *IEICE Trans. Information and Systems*, vol. E79-D, no. 9, pp. 1,323-1,328, Sept. 1996.
- [28] J. Porill, "Fitting Ellipses and Predicting Confidence Envelopes Using a Bias Corrected Kalman Filter," *Image and Vision Computing*, vol. 8, no. 1, pp. 37-41, 1990.
- [29] P.L. Rosin, "A Note on the Least Square Fitting of Ellipses," *Pattern Recognition Letters*, vol. 14, no. 10, pp. 799-808, 1993.
- [30] P.L. Rosin and G.A. West, "Nonparametric Segmentation of Curves into Various Representations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1,140-1,153, Dec. 1995.
- [31] G. Taubin, "Estimation of Planar Curves, Surfaces, and Nonplanar Space Curves Defined by Implicit Equations, with Applications to Edge and Range Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1,115-1,138, Nov. 1991.
- [32] O. Faugeras, "What Can Be Seen in Three Dimensions with an Uncalibrated Stereo Rig?" *Proc. Second European Conf. Computer Vision*, G. Sandini, ed., pp. 563-578, May 1992.
- [33] R. Hartley, "In Defense of the Eight-Point Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580-593, June 1997.
- [34] R. Hartley, R. Gupta, and T. Chang, "Stereo from Uncalibrated Cameras," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 761-764, June 1992.
- [35] K. Kanatani, "Unbiased Estimation and Statistical Analysis of 3D Rigid Motion from Two Views," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 1, pp. 37-50, Jan. 1993.
- [36] Q.-T. Luong and O.D. Faugeras, "The Fundamental Matrix: Theory, Algorithms, and Stability Analysis," *Int'l J. Computer Vision*, vol. 17, no. 1, pp. 43-75, 1996.
- [37] M. Mühlich and R. Mester, "The Role of Total Least Squares in Motion Analysis," *Proc. Fifth European Conf. Computer Vision*, H. Burkhardt and B. Neumann, eds., vol. 2, pp. 305-321, June 1998.
- [38] M. Mühlich and R. Mester, "Subspace Methods and Equilibration in Computer Vision," Technical Report XP-TR-C-21, Inst. for Applied Physics, Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany, Nov. 1999.
- [39] C.V. Stewart, "Robust Parameter Estimation in Computer Vision," *SIAM Review*, vol. 41, no. 3, pp. 513-537, 1999.
- [40] P.H.S. Torr and D.W. Murray, "The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix," *Int'l J. Computer Vision*, vol. 24, no. 3, pp. 271-300, 1997.
- [41] J. Weng, T.S. Huang, and N. Ahuja, "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 451-476, 1989.
- [42] Z. Zhang, "Determining the Epipolar Geometry and Its Uncertainty: A Review," *Int'l J. Computer Vision*, vol. 27, no. 2, pp. 161-195, 1998.
- [43] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A Robust Technique for Matching Two Uncalibrated Images through the Recovery of the Unknown Epipolar Geometry," *Artificial Intelligence*, vol. 78, pp. 87-119, 1995.



**Wojciech Chojnacki** is a professor of mathematics at the Institute of Mathematics, Cardinal Stefan Wyszyński University, Poland. Currently, he is a senior research fellow in the Department of Computer Science at Adelaide University, working on a range of problems in computer vision. His research interests include differential equations, mathematical foundations of computer vision, functional analysis, and harmonic analysis.



**Michael J. Brooks** received the PhD degree from the University of Essex in 1983. Presently, he is the Chair of Artificial Intelligence at Adelaide University. He is a program leader within the Cooperative Research Center for Sensor Signal and Information Processing and a member of the Australian Research Council's Information Technology and Electrical Engineering Large Grants Panel. He is a member of the IEEE.



**Anton van den Hengel** received the PhD degree in computer science from Adelaide University in 2000, where he is currently a lecturer in the Department of Computer Science. His research interests include structure from motion, parameter estimation theory, and commercial applications of computer vision.



**Darren Gawley** graduated with first class honors from the Department of Computer Science at Adelaide University, where he is currently working toward a PhD degree in the field of computer vision.