Approximated Maximum Likelihood Estimation and the Fundamental Numerical Scheme

A. van den Hengel[†], M. Brooks, W. Chojnacki, D. Gawley

Department of Computer Science
University of Adelaide
Adelaide, SA 5005, Australia
{mjb | wojtek | hengel | dg}@cs.adelaide.edu.au

† also of ESAT-PSI/VISICS
Katholieke Universiteit Leuven

Abstract. A number of important problems within computer vision can be couched in terms of the need to estimate the value of a parameter vector on the basis of a set of image-based measurements. There have been recent advances in the techniques available to solve such problems, particularly in those based on a statistical approach. Unfortunately, these new techniques have not been as widely used as is desirable. We present in this paper an introduction to one such technique with the goal of making it more accessible to a general audience. The particular method is that of approximated maximum likelihood estimation using the fundamental numerical scheme developed by the authors. Experimental results are presented in the case of fundamental matrix estimation and the performance of the method is compared against that of a number of other estimators.

1 Introduction

Many problems in computer vision may be expressed in terms of the process of estimating a parameter vector on the basis of a number of image-based measurements. Accordingly, much effort has gone into the development of sophisticated techniques for generating estimates of parameters, some using covariance information characterising uncertainty in the data [1–10]. Maximum likelihood estimation is at the heart of a number of these methods. Unfortunately the perceived complexity of a number of maximum likelihood methods has meant that they have not been widely utilised. In this paper we aim to provide a clear introduction to approximated maximum likelihood estimation as implemented by the fundamental numerical scheme [11]. We also provide a simple guide to implementing the scheme and present results which show that the improvement in the estimates recovered is significant enough to justify the extra effort. The particular version of the maximum likelihood estimation process presented is similar to the method proposed by Kanatani [1] and arose out of an investigation of his work. It must be noted that this is not the only form of maximum likelihood estimation within the domain, but it is none the less an important form of the method.

In order to illustrate the operation of the approximated maximum likelihood estimation method we consider its application to the problem of estimating the fundamental matrix (see also [8, 12–18]).

2 The Problem Form

Problems of the form that we are interested in can be explained in terms of estimating the coefficients of a given algebraic equation that constrains a set of image feature locations. *Conic fitting* is one particular problem of this kind [19, 20] as is *3D to 2D camera projection* estimation [21] which forms part of the camera calibration process. Two others are estimating coefficients of the *epipolar equation* [22], and estimating coefficients of the *differential epipolar equation* [23, 24], each involving an ancillary constraint.

The principal equation that applies to this class of problems, including those specified above, may be represented as follows

$$\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}) = 0. \tag{1}$$

Here $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_l]^T$ is a vector representing unknown parameters; $\boldsymbol{x} = [x_1, \ldots, x_k]^T$ is a vector representing an element of the data; and $\boldsymbol{u}(\boldsymbol{x}) = [u_1(\boldsymbol{x}), \ldots, u_l(\boldsymbol{x})]^T$ is a vector with the data transformed in a problem-dependent manner such that: (i) each component of $\boldsymbol{u}(\boldsymbol{x})$ is a quadratic form in the compound vector $[\boldsymbol{x}^T, 1]^T$, (ii) one component of $\boldsymbol{u}(\boldsymbol{x})$ is equal to 1. It is important to note that in problems of the required form the parameter vector is defined only up to a scale factor and therefore that equation (1) is zero-homogeneous in $\boldsymbol{\theta}$, so multiplying $\boldsymbol{\theta}$ by a scalar has no effect on its value.

In problems of this form there is often an ancillary constraint. If such a constraint applies it may be expressed as

$$\psi(\boldsymbol{\theta}) = 0 \tag{2}$$

for some scalar-valued function ψ.

2.1 Estimating the Fundamental Matrix

A 3D point in a scene undergoing perspective projection onto the image plane of a camera gives rise to an image point represented by a pair of coordinates (m_1, m_2), or equivalently, by the vector $\boldsymbol{m} = [m_1, m_2, 1]^T$. 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 $(\boldsymbol{m}, \boldsymbol{m}')$, this pair satisfies the *epipolar equation* [22]

$$\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m} = 0, \tag{3}$$

where $\boldsymbol{F} = [f_{ij}]$ is a 3×3 *fundamental matrix* that incorporates information about the relative orientation and internal geometry of the cameras [22]. The matrix $\boldsymbol{F}$ is subject to the *rank-2* constraint $\det \boldsymbol{F} = 0$.

The term $\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m}$ above expands to

$$\begin{aligned}\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m} = & m_1 f_{1,1} m'_1 + m_2 f_{1,2} m'_1 + m'_1 f_{1,3} + m_1 f_{2,1} m'_2 + m_2 f_{2,2} m'_2 \\ & + m'_2 f_{2,3} + m_1 f_{3,1} + m_2 f_{3,2} + f_{3,3}.\end{aligned}$$

The measurement vector is thus

$$\boldsymbol{x} = [m_1, m_2, m'_1, m'_2]^T \tag{4}$$

and the function $\boldsymbol{u}(\boldsymbol{x})$ is

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

Equation (3) can thus be rewritten in the form of equation (1). The rank-2 constraint can be also be seen as a function of $\boldsymbol{\theta}$ and thus may be expressed in the form of equation (2).

3 Maximum Likelihood Estimation

Maximum likelihood estimation, like any estimation method, is concerned with selecting the model which best fits a given set of data. A classical example of a model fitting problem is that of fitting a line through a set of scattered points. The model parameters in this case might be the gradient and x-intercept of the line and the fitting process requires determining the values of these parameters such that the line best fits the points. The model thus describes a particular form the data may take, but, in doing so, may also represent some information about the process by which the data was generated. The set of all possible models from which the selection is made, is usually constrained to some class of functions, or distributions. In the line fitting example, this is the set of all lines, in the case of estimating the fundamental matrix it is the set of all possible fundamental matrices.

3.1 A general problem formulation

Estimation problems of the form we are interested in generally have a set of measurements on the basis of which the estimate is calculated, such as the set of correspondences for the estimation of the fundamental matrix. We label this set of measured data points $\mathcal{S} = \{\boldsymbol{x}_i | \, i = 1 \dots n\}$. If we parameterise the space of all possible models by the vector $\boldsymbol{\theta}$, then maximum likelihood estimation becomes the problem of selecting the model $\boldsymbol{\theta}$ under which the observed data set $\mathcal{S}$ is most likely to occur. The probability that our observed data $\mathcal{S}$ will occur given a particular model $\boldsymbol{\theta}$ is represented by the conditional probability $p(\mathcal{S}|\boldsymbol{\theta})$. We thus seek

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\mathcal{S}|\boldsymbol{\theta}).$$

By analysing the probabilities we see that this is equivalent to the following

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left((\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i)^T \boldsymbol{\Lambda}^{-} (\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i) \right)^2 .$$

This maximum likelihood estimate is statistically optimal when the variance of each $\boldsymbol{x}_i$ is the same, and the errors in the $\boldsymbol{x}_i$ are uncorrelated. The term

$$(\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i)^T \boldsymbol{\Lambda}^{-} (\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i) \tag{5}$$

represents a measure of the distance between the observed data point $\boldsymbol{x}_i$ and its expected value $\boldsymbol{x}(\boldsymbol{\theta})_i$, and $\boldsymbol{\Lambda}$ is a covariance matrix capturing the uncertainty of the measurement $\boldsymbol{x}$. Equation (5) therefore represents the Mahanalobis distance between the observed data point and its expected value. The Mahanalobis distance is a distance measure which takes into account the variance of each element of the vector $(\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i)$. However if the variance of each of the elements of the vectors is expected to be the same then $\boldsymbol{\Lambda}$ can be taken to be the identity matrix and can thus be ignored. In this case the maximum likelihood estimate corresponds to the model which minimises the sum of the squares of the Euclidean distances between the data elements and their expected values. A cost function representing the quality of such an estimate would thus be

$$\begin{aligned} J_{\mathrm{MLE}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) &= \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i)^T \boldsymbol{I}^{-} (\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i) \\ &= \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{x}(\boldsymbol{\theta})_i\|^2. \end{aligned} \tag{6}$$

Note that this cost function can only be associated with the maximum likelihood estimate when $\boldsymbol{\Lambda} = \boldsymbol{I}$. Typically the assumption is made that this is the case when no information about the individual distributions of the errors is available, but there is no reason to expect them to differ significantly. It is this case that we consider in this paper. For a more detailed introduction to probability and covariances as they apply to maximum likelihood estimation see Refs. [25–28].

Let us now define the manifold $\mathcal{M}_{\boldsymbol{\theta}}$ of all possible data which satisfy equation (1) for a particular $\boldsymbol{\theta}$. If

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}) \tag{7}$$

then the manifold $\mathcal{M}_{\boldsymbol{\theta}}$ can be defined as the set of all $\boldsymbol{x}$ consistent with $\boldsymbol{\theta}$. That is, the manifold describes the set of all vectors $\boldsymbol{x}$ for which $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = 0$ so

$$\mathcal{M}_{\boldsymbol{\theta}} = \{\boldsymbol{x} : f_{\boldsymbol{\theta}}(\boldsymbol{x}) = 0\}. \tag{8}$$

The estimation problem can now be seen as that of fitting a manifold to a set of data points.

3.2 The Fundamental Matrix Example

In terms of our example estimation problem the fundamental matrix $\boldsymbol{F}$ represents the model $\boldsymbol{\theta}$. We wish to maximise $p(\mathcal{S}|\boldsymbol{F})$, the probability that the measured correspondences $\mathcal{S}$ would occur given a pair of cameras with key parameters constrained by $\boldsymbol{F}$. We therefore minimise the sum of the squares of the distances between the elements of $\mathcal{S}$ and their expected values given a particular $\boldsymbol{F}$. If

$$f_{\boldsymbol{F}}(\{\boldsymbol{m}, \boldsymbol{m}'\}) = \boldsymbol{m}'^{T} \boldsymbol{F} \boldsymbol{m} \tag{9}$$

then the manifold containing these expected values $\mathcal{M}_{\boldsymbol{F}}$ can be defined as the set of all correspondences $\boldsymbol{x} = \{\boldsymbol{m}, \boldsymbol{m}'\}$ consistent with $\boldsymbol{F}$.

3.3 The Distance to the Manifold

We have seen that the maximum likelihood estimate is that which minimises the sum of squares of distances between each datum and its expected value. This is the sum represented in equation (6). It is useful to assume that there is some underlying set of 'true' data which the vectors $\boldsymbol{x}_i$ are measurements of. In the case of estimating the Fundamental matrix this would be the set of 'true' correspondences. If we represent an element of this set as $\bar{\boldsymbol{x}}$ then the corresponding element of the set of observed data $\boldsymbol{x}$ can be written as $\boldsymbol{x} = \bar{\boldsymbol{x}} + \Delta\bar{\boldsymbol{x}}$ where $\Delta\bar{\boldsymbol{x}}$ is the error in the estimate. Letting $\bar{\boldsymbol{\theta}}$ represent the true parameter vector, we see that $f_{\bar{\boldsymbol{\theta}}}(\bar{\boldsymbol{x}}) = 0$, but in general that $f_{\bar{\boldsymbol{\theta}}}(\boldsymbol{x}) \neq 0$. This is due to the fact that the observed measurement vector $\boldsymbol{x}$, having been contaminated by noise (represented by $\Delta\bar{\boldsymbol{x}}$), does not, in general, lie on the manifold $\mathcal{M}_{\bar{\boldsymbol{\theta}}}$.

Unfortunately it is not possible to recover the true data vector $\bar{\boldsymbol{x}}$ from its measurement $\boldsymbol{x}$ because $\Delta\bar{\boldsymbol{x}}$ is unknown. The true parameter vector is similarly not recoverable. It is possible, however, to recover an estimate of this vector from a large enough set $\mathcal{S}$ of measurement vectors $\mathcal{S} = \{\boldsymbol{x}_i : i = 1 \ldots n\}$. The number of measurements necessary depends on the nature of the estimation problem.

The distance in equation (6) is that between a data vector and its expected value for a particular parameter vector $\boldsymbol{\theta}$. The manifold $\mathcal{M}_{\boldsymbol{\theta}}$ represents the set of all data vectors corresponding to that value of the parameter vector. The expected value of the data vector, which we label $\tilde{\boldsymbol{x}}$, is therefore a member of the set described by this manifold. It is sensible to assume that the expected value corresponding to a particular measurement vector to be the point on the manifold closest to that vector. On this formulation, the distance referred to in equation (6) is the shortest distance between the data vector and the manifold. The maximum likelihood estimate of $\boldsymbol{\theta}$ is the value which corresponds to the manifold $\mathcal{M}_{\boldsymbol{\theta}}$ for which the sum of the squares of these distances is minimal (see Figure 1). For many problems, including those mentioned in the introduction, the manifold $\mathcal{M}_{\boldsymbol{\theta}}$ takes the form of a generalised conic section. The estimation process can thus be seen as a conic fitting problem.

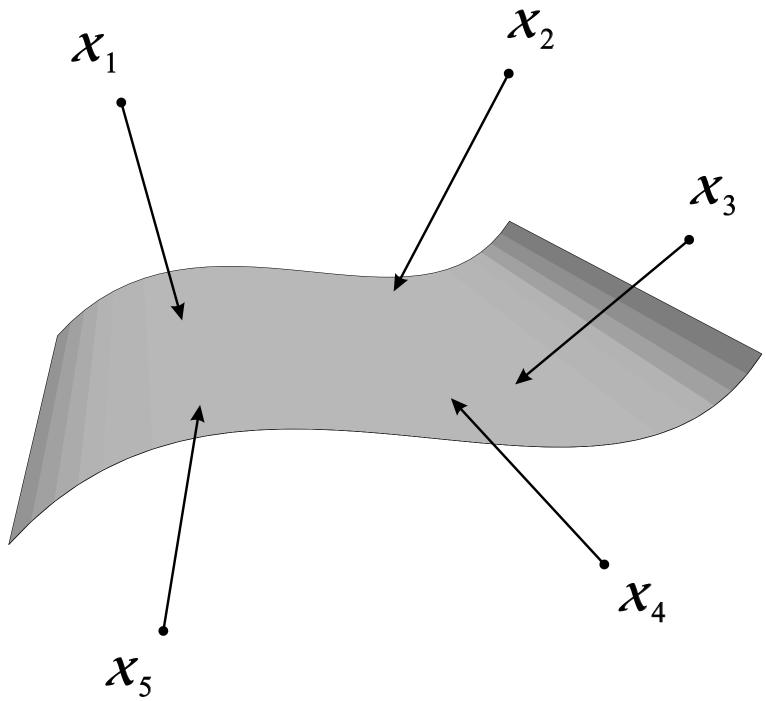

Fig. 1. The manifold $\mathcal{M}_{\boldsymbol{\theta}}$

3.4 The Advantages of Image-Based Measurements

There are many expressions which could be taken as a measure of the distance between a measurement and its expected value. We could take the sum of the squares of any one of these as the basis for a cost function. We could, for example, take $\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x})$, and this is in fact the basis of the popular algebraic least squares method described below in Section 6.2. If this distance measure is used, however, the variances of the elements of the resulting vector will depend on the value of the measurement $\boldsymbol{u}(\boldsymbol{x})$ because the transformation from $\boldsymbol{x}$ to $\boldsymbol{u}(\boldsymbol{x})$ is not Euclidean. So even if the variances of the image-based quantities $\boldsymbol{x}_1$ are the same the variances of the distance measures $\boldsymbol{u}(\boldsymbol{x}_i)$ will not be equal. The least squares solution is optimal, however, only when the variances of the individual distance measurements are the same. There are methods for compensating for this problem. See, for example, Ref. [29].

An advantage of taking an image-based quantity as a measure of the distance between a measurement and its expected value is that it this is the frame in which the noise is added. The fact that the noise is an artifact of the measurement process means that it is an image-based quantity and thus is best represented as such. The primary advantage of taking an image-based quantity, however, is that this is the means by which geometrically meaningful quantities may be measured. In the case of the problem of estimating the Fundamental Matrix, for instance, it allows the distances to epipolar lines or the reprojection error to be measured.

4 The Direct Method

One approach to finding the manifold which best fits a set of measured points is to parameterise the problem by the location of the expected values of the data $\tilde{\boldsymbol{x}}_i$ and the vector $\boldsymbol{\theta}$. Numerical minimisation can then be used to find the $\tilde{\boldsymbol{x}}_i$ and $\boldsymbol{\theta}$ corresponding to its lowest value of $J_{\mathrm{MLE}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$.

This is the approach taken by the Gold Standard method of estimating the Fundamental Matrix as described in Ref. [21]. The parameter vector $\boldsymbol{\theta}$ in this case is taken to be the elements of the projection matrix of the right camera $\boldsymbol{P_2}$, and the projection matrix of the left is taken to be $\boldsymbol{P_1} = [I, \mathbf{0}]$. This parameterisation allows the manifold of expected values of the data to be more easily expressed. It is equivalent to the Fundamental Matrix parameterisation in that any value of one parameter may easily be converted into a value of the other although there is not a one-to-one correspondence. The expected values $\tilde{\boldsymbol{x}}_i$ are parameterised by the 3D location of the corresponding scene point. Calculating the expected values themselves then requires only the projection of the 3D points onto the image planes according to $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. The parameter space over which numerical minimiser takes place is of dimension $3n + 12$ where n is the number of correspondences. This minimisation can thus be a slow process.

5 The Approximated Maximum Likelihood Method

The alternative to direct minimisation is to try to eliminate the variables $\tilde{\boldsymbol{x}}_i$ from the minimisation process. In this hope we seek a measure of the distance from the point $\boldsymbol{x}$ to the point $\tilde{\boldsymbol{x}}$ which is the closest point on the manifold $\mathcal{M}_{\boldsymbol{\theta}}$. The aim, however, is to find this distance without explicitly determining the location of the point $\tilde{\boldsymbol{x}}$. If this is possible we can remove the corresponding parameters from our minimisation process. This constitutes a significant reduction in the parameter space.

The Taylor series expansion of $\boldsymbol{f_\theta}(\boldsymbol{x})$ about the point $\tilde{\boldsymbol{x}}$ can be written as

$$\boldsymbol{f_\theta}(\boldsymbol{x}) = \boldsymbol{f_\theta}(\tilde{\boldsymbol{x}}) + \boldsymbol{f'_\theta}(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}) + O((\boldsymbol{x} - \tilde{\boldsymbol{x}})^2). \tag{10}$$

We have defined $\tilde{\boldsymbol{x}}$ to lie on the manifold $\mathcal{M}_{\boldsymbol{\theta}}$ so we know that $\boldsymbol{f_\theta}(\tilde{\boldsymbol{x}}) = 0$, thus, if the observed flow is sufficiently close to the correct flow we can assume that

$$\boldsymbol{f_\theta}(\boldsymbol{x}) = \boldsymbol{f'_\theta}(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}). \tag{11}$$

The vector that marks the shortest Euclidean distance between a point and a surface always strikes that surface at right angles, so the vector from $\boldsymbol{x}$ to $\tilde{\boldsymbol{x}}$ is perpendicular to the surface of $\mathcal{M}_{\boldsymbol{\theta}}$ and parallel to its gradient $\boldsymbol{f'_\theta}(\boldsymbol{x})$, therefore

$$|\boldsymbol{f'_\theta}(\boldsymbol{x})(\boldsymbol{x} - \tilde{\boldsymbol{x}})| = ||\boldsymbol{f'_\theta}(\boldsymbol{x})|| \; ||(\boldsymbol{x} - \tilde{\boldsymbol{x}})||. \tag{12}$$

On combining 11 and 12 we see that

$$||\boldsymbol{x} - \tilde{\boldsymbol{x}}|| = \frac{|f_{\boldsymbol{\theta}}(\boldsymbol{x})|}{||f'_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}})||}$$

which is the required Euclidean distance. Unfortunately we do not a priori know $||f'_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}})||$ but we can calculate $||f'_{\boldsymbol{\theta}}(\boldsymbol{x})||$. If, as we have assumed, $\boldsymbol{x}$ is sufficiently close to $\tilde{\boldsymbol{x}}$ then we may reasonably assume that

$$||f'_{\boldsymbol{\theta}}(\boldsymbol{x})|| \approx ||f'_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}})||\,, \tag{13}$$

so we can approximate the distance by

$$\frac{|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)|}{||f'_{\boldsymbol{\theta}}(\boldsymbol{x}_i)||}. \tag{14}$$

The cost function indicating the quality of a particular estimate is thus given by

$$J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_i^n \frac{|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)|^2}{||f'_{\boldsymbol{\theta}}(\boldsymbol{x}_i)||^2}. \tag{15}$$

We label the parameter vector $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ which minimises the cost function the Approximate Maximum Likelihood Estimate. By linearising $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ we have now arrived at an approximation to the distance between the measured data vector and the manifold with no reference to the point $\tilde{\boldsymbol{x}}$. In minimising J_{AML} we thus no longer need include parameters corresponding to the set of expected values which reduces the dimension of the search dramatically. In the case of the Fundamental Matrix this is a reduction from $3n + 12$ parameters (where n represents the number of elements in the data set) to 8, or 7 if the auxiliary constraint is used. It is important to note here that J_{AML} as given above is applicable only to the case in which the covariances of the elements of the measurement vectors may be taken to be the same. This is the same as the condition that $\boldsymbol{\Lambda} = \boldsymbol{I}$ which allows J_{MLE} to be expressed in the form of equation (6).

5.1 Geometric interpretation

Figure 2 gives a geometric interpretation of this approximated distance measure. Recall that we seek the distance between our data point $\boldsymbol{x}$ and the manifold $\mathcal{M}_{\boldsymbol{\theta}}$. The observed data element $\boldsymbol{x}$ has an associated algebraic residual (from equation (1)) of $f_{\boldsymbol{\theta}}(\boldsymbol{x})$. We now define a manifold $\mathcal{N}_{k,\boldsymbol{\theta}}$ of all points with the same residual k:

$$\mathcal{N}_{k,\boldsymbol{\theta}} = \{\boldsymbol{z} : f_{\boldsymbol{\theta}}(\boldsymbol{z}) = k\}.$$

Note that $\mathcal{M}_{\boldsymbol{\theta}} = \mathcal{N}_{0,\boldsymbol{\theta}}$

In (10) we constructed a linearisation of $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ about $\boldsymbol{x}$. This corresponds to finding the equation of the plane tangent to the manifold $\mathcal{N}_{f(\boldsymbol{x}),\boldsymbol{\theta}}$ at the same point. If we label the linearised $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ as $g_{\boldsymbol{\theta}}(\boldsymbol{x})$ then the value calculated by

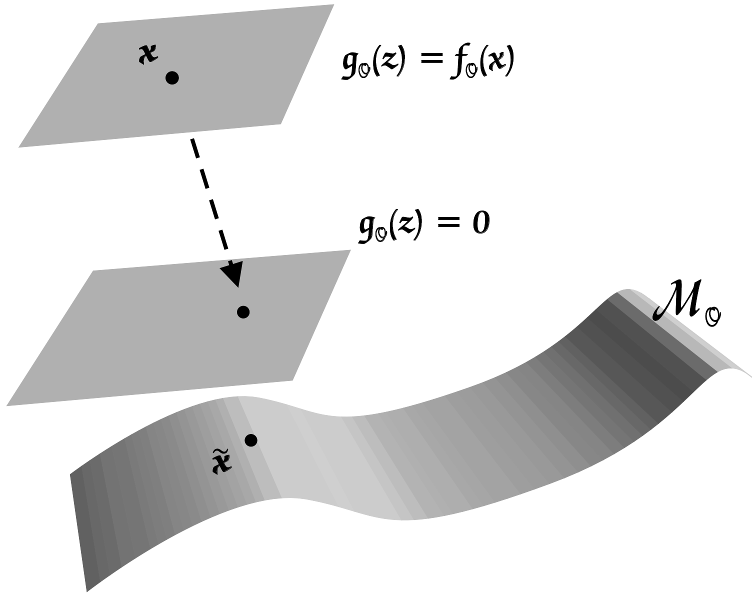

Fig. 2. The first order approximation to the distance to the manifold

equation (13) represents the distance of the point $\boldsymbol{x}$ from the plane $g_{\boldsymbol{\theta}}(\boldsymbol{z}) = 0$. The approximation that takes place in (13) is equivalent to assuming that the plane $g_{\boldsymbol{\theta}}(\boldsymbol{z}) = 0$ is tangential to the manifold $\mathcal{M}_{\boldsymbol{\theta}}$ at $\tilde{\boldsymbol{x}}$. The expression in equation (14) is thus a first order approximation to the distance between the point and the manifold.

5.2 Finding the Approximate Maximum Likelihood Estimate

In order to find the parameter vector which minimises J_{AML}, and in light of equation (7) we rewrite equation (15) thus

$$J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}_i) \boldsymbol{u}(\boldsymbol{x}_i)^T \boldsymbol{\theta}}{\boldsymbol{\theta}^T \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i) \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i)^T \boldsymbol{\theta}}$$

where

$$\partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{y}) = \begin{bmatrix} \frac{\partial u_1}{\partial x_1}(\boldsymbol{y}) & \cdots & \frac{\partial u_1}{\partial x_k}(\boldsymbol{y}) \\ \cdots & \cdots & \cdots \\ \frac{\partial u_l}{\partial x_1}(\boldsymbol{y}) & \cdots & \frac{\partial u_l}{\partial x_k}(\boldsymbol{y}) \end{bmatrix}.$$

For more information on this process, and the case in which the variances of the measurements are not taken to be equal, see Ref. [29, 30].

If we let

$$\boldsymbol{A}_i = \boldsymbol{u}(\boldsymbol{x}_i) \boldsymbol{u}(\boldsymbol{x}_i)^T \text{ and } \boldsymbol{B} = \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i)\, \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i)^T, \tag{16}$$

then J_{AML} can be simply written as

$$J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}.$$

As a minimiser of J_{AML}, $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ satisfies

$$\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \mathbf{0}^T, \tag{17}$$

where $\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}$ denotes the row vector of the partial derivatives of J_{AML} with respect to $\boldsymbol{\theta}$. We term this the *variational equation*. Direct computation shows that

$$[\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)]^T = 2\boldsymbol{X}_{\boldsymbol{\theta}} \boldsymbol{\theta},$$

where $\boldsymbol{X}_{\boldsymbol{\theta}}$ is the symmetric matrix

$$\boldsymbol{X}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}_i}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}} - \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}}{(\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta})^2} \boldsymbol{B}_i. \tag{18}$$

Thus (17) can be written as

$$\boldsymbol{X}_{\boldsymbol{\theta}} \boldsymbol{\theta} = \mathbf{0}. \tag{19}$$

This is a non-linear equation and is unlikely to admit solutions in closed form.

6 Numerical schemes

The nature of the variational equation means that in practice $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ has to be found numerically. One possibility involves the use of a generic numerical minimisation technique such as Levenberg-Marquardt or the Newton-Raphson method. This process is complicated by the fact that the cost function is zero-homogeneous in the parameter vector. This means that we are actually trying to estimate a projective quantity, and that gradient-based techniques will need to take into account the fact that the Hessian of the cost function becomes singular at the solution. The fundamental numerical scheme can be regarded as a variant of the Newton-Raphson method designed specifically for this situation.

6.1 The Fundamental Numerical Scheme

A vector $\boldsymbol{\theta}$ satisfies (19) if and only if it falls into the null space of the matrix $\boldsymbol{X}_{\boldsymbol{\theta}}$. Thus, if $\boldsymbol{\theta}_{k-1}$ is a tentative guess, then an improved guess can be obtained by picking a vector $\boldsymbol{\theta}_k$ from that eigenspace of $\boldsymbol{X}_{\boldsymbol{\theta}_{k-1}}$ which most closely approximates the null space of $\boldsymbol{X}_{\boldsymbol{\theta}}$; this eigenspace is, of course, the one corresponding to the smallest eigenvalue. The fundamental numerical scheme implementing this idea is presented in Figure 3. This method requires initialisation with a suitable $\boldsymbol{\theta}_0$. Such an initial estimate may come from any source, but generally is taken to be the standard Algebraic Least Squares Estimate.

1. Set $\boldsymbol{\theta}_0$ to a good initial estimate.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ is known, compute the matrix $\boldsymbol{X}_{\boldsymbol{\theta}_{k-1}}$.
3. Compute a normalised eigenvector of $\boldsymbol{X}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the eigenvalue closest to zero and take this eigenvector for $\boldsymbol{\theta}_k$.
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Fig. 3. Fundamental numerical scheme.

6.2 The Algebraic Least Squares Estimator

A straightforward cost function may be derived from equation (1)

$$J_{\mathrm{ALS}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \left(\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}_i) \right)^2 = \sum_{i=1}^{n} \boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}$$

Here each summand $\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}$ is the square of the *algebraic distance* $|\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}_i)|$. The $\boldsymbol{\theta}$ for which J_{ALS} is minimal is termed the *algebraic least squares* (ALS) *estimate* and is denoted $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$. It is uniquely determined, up to a scalar factor, by an eigenvector of $\sum_{i=1}^{n} \boldsymbol{A}_i$ associated with the smallest eigenvalue. For reasons of numerical stability this is found in practice by performing singular-value decomposition (SVD) of a matrix with rows set to $\boldsymbol{u}(\boldsymbol{x}_i)$ for each measurement vector in $\mathcal{S}$.

7 Implementing the Fundamental Numerical Scheme

We now present a step by step approach to implementing the fundamental numerical scheme for a particular problem. We use the problem of estimating the Fundamental Matrix on the basis of a set of corresponding points as an example.

The first step is to form the matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$. The matrix $\boldsymbol{A}_i$ is constructed from the expansion of the particular version of equation (1) applicable to the problem at hand. We have shown the expansion of the epipolar equation and the form of $\boldsymbol{u}(x)$ for this problem in section 2.1. Constructing the $\boldsymbol{A}_i$ thus requires only the formation of the matrices $\boldsymbol{u}(\boldsymbol{x}_i)\boldsymbol{u}(\boldsymbol{x}_i)^T$ for each data vector.

The matrix $\boldsymbol{B}_i$ is defined such that $\boldsymbol{B}_i = \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i) \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i)^T$. In the case of the Fundamental Matrix

$$\begin{aligned} \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x})^T &= [(\partial u_j / \partial x_i)(\boldsymbol{x})]_{1 \le i \le 4,\, 1 \le j \le 9} \\ &= \begin{bmatrix} m_1' & 0 & 0 & m_2' & 0 & 0 & 1 & 0 & 0 \\ 0 & m_1' & 0 & 0 & m_2' & 0 & 0 & 1 & 0 \\ m_1 & m_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_1 & m_2 & 1 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The algebraic least square process is used to calculate an initial estimate $\boldsymbol{\theta}_0$. The matrix $\boldsymbol{X}_{\boldsymbol{\theta}}$ is constructed according to equation (18) from the matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ and the vector $\boldsymbol{\theta}$. The eigenvectors of this matrix are then determined, and the vector corresponding to the smallest eigenvalue is selected as our new estimate. This process is repeated until the estimates converge.

8 Experiments

We now compare the performance of the fundamental numerical scheme (FNS) to that of the algebraic least squares estimator (ALS) and the Gold Standard method (GS). FNS was implemented using the EISPACK routine RS for computation of eigenvalues and associated eigenvectors of symmetric matrices. The ALS method uses the LINPACK routine DSVDC to perform SVD. For the GS scheme, the MINPACK routine LMDER was used. The selection of an appropriate stopping condition is essential when comparing iterative estimation methods. The stopping conditions chosen for FNS and GS were that the change in the value of the residual fell below 10^{-4}. This decision was made on the basis of the fact that J_{AML} is an approximation of J_{MLE}. The GS method was initialised with the FNS solution because in all tests this method converged to the same point as when GS was initialised with the ALS solution. The rank 2 constraint was enforced by performing an singular value decomposition, setting the smallest singular value to 0, and multiplying back out. The initial values for parameters in the GS method were found by triangulating image points using a polynomial root finding method as described in [31].

Our experiments proceeded as follows. A realistic stereo camera configuration was first selected with non co-planar optical axes, and slightly differing left and right camera intrinsic parameters. 3D points were then projected onto the images so as to generate many pairs of corresponding points. A range of tests was then conducted, each carried out with respect to an *average level of noise*, σ. For a given test, each image point was perturbed by adding zero-mean Gaussian noise of standard deviation σ independently to each of the two coordinates. Each method was then supplied with these noisy matching points. For each σ, the estimation process was repeated 100 times from a specific set of 50 corresponding points, with new perturbations being generated each time.

The estimates of the three schemes, ALS, FNS & GS, were compared using two error measures. The first is the average distance between the given data points and the closest points satisfying the epipolar equation. This is the same measure that the GS method directly minimises. Unsurprisingly, the results presented in Figure 4, show that the GS has the lowest errors, and FNS having lower values than ALS.

The second error metric used is the distance of the true data points from the estimated epipolar lines. This measure may be expressed as

$$J_{\mathrm{EL}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{1}{n} \sum_{i}^{n} d(\bar{\boldsymbol{m}}'_i, \boldsymbol{F}\bar{\boldsymbol{m}}_i)^2 + d(\bar{\boldsymbol{m}}_i, \boldsymbol{F}^T \bar{\boldsymbol{m}}'_i)^2.$$

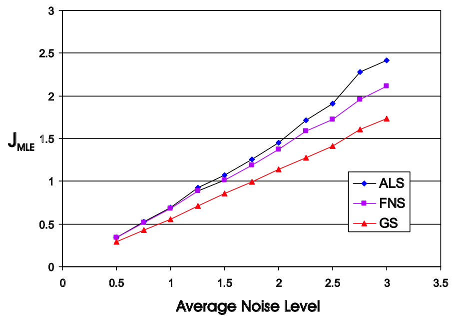

Fig. 4. J_{MLE} vs. average noise level

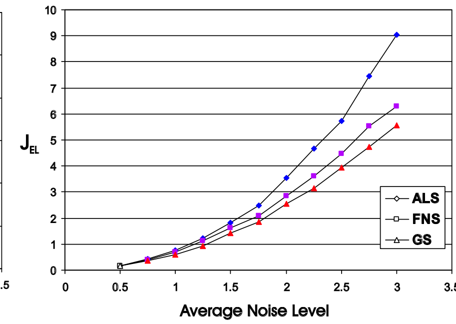

Fig. 5. J_{EL} vs. average noise level

Given that ultimately we desire an estimator which produces estimates as close as possible to the truth we should use the truth as the basis for comparison when it is available. The advantage of using the distance to epipolar lines is that it is a geometrically significant metric which is not directly minimised by any on the methods. The results for this measure are shown in Figure 5.

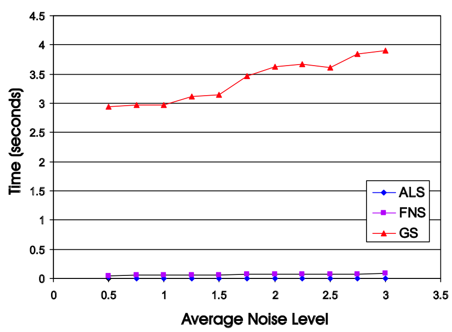

Fig. 6. Time vs. average noise level

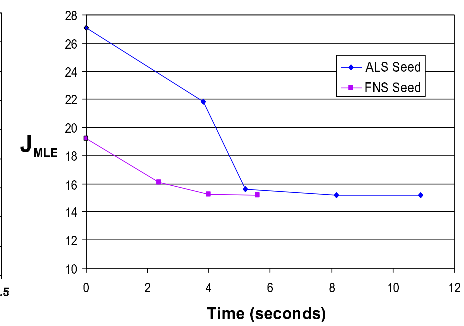

Fig. 7. J_{MLE} vs. time

The timing tests conducted on the various schemes show a dramatic difference in the processing required. These results show that the FNS method is capable of producing good estimates in very small execution times. We have shown above that the additional implementation effort necessary to calculate perform FNS over ALS is minimal. The fundamental numerical scheme thus provides a significant benefit in terms of improvement in the quality of estimates at little additional computational cost.

Figure 7 compares the progress of GS when initialised with FNS and ALS. The value of J_{MLE} (which is the entity being minimised), is shown for each iteration of the Levenberg–Marquardt minimisation, plotted against the elapsed time of computation. Although this graph shows the results for only one set of data, we found that these results were indicative of the general performance of GS. These results show that the GS estimates are identical for both seeds, however convergence is significantly faster when FNS is used as the seed. In the trials we have run the execution time of the FNS seeded method has been approximately half of that of the ALS seeded method and often significantly less. The fundamental numerical scheme thus serves as a useful estimator in its own right, but also as a valuable method of reducing the execution time of the gold standard method.

References

1. Chojnacki, W., Brooks, M.J., van den Hengel, A.: Fitting surfaces to data with covariance information: Fundamental methods applicable to computer vision. Technical Report TR 99-03, Department of Computer Science, University of Adelaide, South Australia (1999) Available from `http://www.cs.adelaide.edu.au/~mjb`.
2. Kanatani, K.: Statistical Optimization for Geometric Computation: Theory and Practice. Elsevier, Amsterdam (1996)
3. Leedan, Y., Meer, P.: Estimation with bilinear constraints in computer vision. In: Proceedings of the Sixth International Conference on Computer Vision, Bombay India, January 4–7, 1998, New Delhi, IEEE Computer Society, Narosa Publishing House (1998) 733–738
4. Matei, B., Meer, P.: Reduction of bias in maximum likelihood ellipse fitting. In: Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, September 3–8, 2000, IEEE Computer Society Press, Los Alamitos, CA (2000)
5. Matei, B., Meer, P., Tyler, D.: Performance assessment by resampling: rigid motion estimators. In Bowyer, K.W., Phillips, P.J., eds.: Empirical Evaluation Techniques in Computer Vision. IEEE Computer Society Press, Los Alamitos, CA (1998) 72–95
6. Matei, B., Meer, P.: Bootstrapping a heteroscedastic regression model with application to 3D rigid motion evaluation. In Triggs, B., Zisserman, A., Szeliski, R., eds.: Vision Algorithms: Theory and Practice. Volume 1883 of Lecture Notes in Computer Science., International Workshop on Vision Algorithms, Corfu, Greece, September, 1999, Springer-Verlag (2000) 236–252
7. Mühlich, M., Mester, R.: Subspace methods and equilibration in computer vision. Technical Report XP-TR-C-21, Institute for Applied Physics, Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany (1999)
8. Mühlich, M., Mester, R.: The role of total least squares in motion analysis. In Burkhardt, H., Neumann, B., eds.: Computer Vision—ECCV'98, Fifth European Conference on Computer Vision, Freiburg, Germany. (1998) 305–321
9. Triggs, B.: A new approach to geometric fitting. Available from http://www.inrialpes.fr/movi/people/Triggs (1998)
10. Triggs, B.: Optimal estimation of matching constraints. In Koch, R., Gool, L.V., eds.: 3D Structure from Multiple Images of Large-Scale Environments, European Workshop, SMILE'98, Freiburg, Germany, June 6–7, 1998, Proceedings. Volume 1506 of Lecture Notes in Computer Science., Berlin, Springer-Verlag (1998) 30–32

11. Chojnacki, W., Brooks, M.J., van den Hengel, A.: On the fitting of surfaces to data with covariances. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000)
12. Hartley, R.: In defense of the eight-point algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 580–593
13. Leedan, Y., Meer, P.: Heteroscedastic regression in computer vision: problems with bilinear constraint. International Journal of Computer Vision **37** (2000) 127–150
14. Luong, Q.T., Faugeras, O.D.: The fundamental matrix: theory, algorithms, and stability analysis. International Journal of Computer Vision **17** (1996) 43–75
15. Kanatani, K.: Unbiased estimation and statistical analysis of 3-D rigid motion from two views. IEEE Transactions on Pattern Analysis and Machine Intelligence **15** (1993) 37–50
16. Torr, P.H.S., Murray, D.W.: The development and comparison of robust methods for estimating the fundamental matrix. International Journal of Computer Vision **24** (1997) 271–300
17. Zhang, Z.: Determining the epipolar geometry and its uncertainty: a review. International Journal of Computer Vision **27** (1998) 161–195
18. Zhang, Z.: On the optimization criteria used in two-view motion analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998) 717–729
19. Bookstein, F.: Fitting conic sections to scattered data. Computer Graphics and Image Processing **9** (1979) 56–71
20. Zhang, Z.: Parameter estimation techniques: a tutorial with application to conic fitting. Image and Vision Computing **15** (1997) 57–76
21. Hartley, R., A.Zisserman: Multiple view geometry in computer vision. Cambridge University Press (2000)
22. Faugeras, O.D.: Three-Dimensional Computer Vision: A Geometric Viewpoint. The MIT Press, Cambridge, Mass. (1993)
23. Brooks, M.J., Chojnacki, W., Baumela, L.: Determining the egomotion of an uncalibrated camera from instantaneous optical flow. Journal of the Optical Society of America A **14** (1997) 2670–2677
24. Viéville, T., Faugeras, O.D.: Motion analysis with a camera with unknown, and possibly varying intrinsic parameters. In: Proceedings of the Fifth International Conference on Computer Vision, Cambridge, MA, June 20–23, 1995, IEEE Computer Society Press, Los Alamitos, CA (1995) 750–756
25. Stuart, A., Ord, J.K.: 8. In: Kendall's Advanced Theory of Statistics. 5th edn. Volume 1. Charles Griffin and Co, London (1987)
26. Gelb, A., Kasper, J.F., Nash, R.A., Price, C.F., Sutherland, A.A.: Applied Optimal Estimation. MIT Press, Cambridge, Massachusetts (1974)
27. Bar-Shalom, Y., Li, X.: Estimation and Tracking: Principles, Techniques and Software. Artech House, Norwood, MA (1993)
28. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. In: Numerical Recipes in C. Second edn. Cambridge University Press (1992)
29. Torr, P.H.S., Murray, D.W.: The development and comparison of robust methods for estimating the fundamental matrix. International Journal of Computer Vision **24** (1997) 271–300
30. van den Hengel, A.: Robust estimation of structure from motion in the uncalibrated case. PhD thesis, Adelaide University (2000)
31. Hartley, R., Sturm, P.: Triangulation. Computer Vision and Image Understanding **68** (1997) 146–157