IEEE Conference on Computer Vision and and Pattern Recognition, Hilton Head, Vol. 1, pp. 523-530, 2000

Likelihood Functions and Confidence Bounds for Total-Least-Squares Problems

Oscar Nestares[*†] David J. Fleet[†] David J. Heeger[*]

[†]Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

[*]Department of Psychology
Stanford University
Stanford, CA 94305

Abstract

This paper addresses the derivation of likelihood functions and confidence bounds for problems involving overdetermined linear systems with noise in all measurements, often referred to as total-least-squares (TLS). It has been shown previously that TLS provides maximum likelihood estimates. But rather than being a function solely of the variables of interest, the associated likelihood functions increase in dimensionality with the number of equations. This has made it difficult to derive suitable confidence bounds, and impractical to use these probability functions with Bayesian belief propagation or Bayesian tracking. This paper derives likelihood functions that are defined only on the parameters of interest. This has two main advantages: first, the likelihood functions are much easier to use within a Bayesian framework; and second it is straightforward to obtain a reliable confidence bound on the estimates. We demonstrate the accuracy of our confidence bound in relation to others that have been proposed. Also, we use our theoretical results to obtain likelihood functions for estimating the direction of 3d camera translation.

1. Introduction

There has been growing interest in the use of Bayesian methods for computer vision research [19], with a trend away from maximum likelihood (ML) and maximum a posteriori (MAP) estimators towards the computation of full posterior probability distributions. By computing full probability distributions we can represent multimodal distributions, incorporate complex prior models, and exploit Bayesian belief propagation, both through space (with Markov random fields) and time (with Bayesian tracking). Towards this end, the computation of posterior distributions relies on the development of a generative image model (with noise), the derivation of likelihood functions, and the development (or learning) of prior distributions.

This paper concerns the derivation of probability functions (e.g., likelihood functions) when the variables of interest are partially constrained by noisy measurements. In particular, we derive likelihood functions for linear constraints, like those that occur in line-fitting problems [12], gradient-based optical flow estimation [7, 18, 21], and in linear subspace methods for the estimation of 3D translation [10].

By way of application domains, we are motivated by two problems in motion analysis, namely, the estimation of optical flow and the estimation of 3D camera translation. In gradient-based optical flow estimation, assuming brightness constancy, the velocity $\mathbf{v}(\mathbf{x},t) = (v_x(\mathbf{x},t), v_y(\mathbf{x},t))$, at position $\mathbf{x}$ and time t, is related to the spatio-temporal intensity gradient, $\vec{\nabla} I(\mathbf{x},t) = (I_x(\mathbf{x},t), I_y(\mathbf{x},t), I_t(\mathbf{x},t))'$, by the gradient constraint equation [8]:

$$\vec{\nabla} I' \mathbf{v}_h = 0 , \tag{1}$$

where $\mathbf{x}'$ denotes the transpose of $\mathbf{x}$, and $\mathbf{v}_h = (v_x, v_y, 1)'$ denotes the unknown velocity vector in homogeneous coordinates. Uncertainty is caused by noise in measuring image derivatives and by violations of brightness constancy.

Linear constraint equations also arise in subspace methods for estimating the 3D direction of camera translation [9, 10]. In this case local optical flow vectors are combined linearly to obtain a set of constraints, $\vec{\tau}$, that satisfy

$$\vec{\tau}' \mathbf{T} = 0 , \tag{2}$$

where $\mathbf{T}$ is the unknown camera translation. Because of uncertainty in optical flow, the subspace constraint measurements, $\vec{\tau}$, will be noisy.

Previous approaches to problems like these have either assumed simplified noise models, or they have avoided computing the full likelihood or posterior by formulating ML/MAP estimators instead. For example, with gradient-based optical flow estimation it is often assumed that measurements of spatial image derivatives in (1) are noiseless, while noise in temporal derivative measurements is additive and Gaussian. It is then straightforward to show that the likelihood function is Gaussian, from which one can derive

linear ML estimator. Simoncelli *et al.* [18] used this noise model, with a constant velocity model incorporating additive noise (within an image region), and a Gaussian prior that prefers slow velocities, to derive a Gaussian posterior pdf and a MAP estimator. Luettgen *et al.* [14] used this noise model, with a multiscale, Gaussian Markov random field (MRF) prior, to formulate a posterior pdf over optical flow fields.

In many cases one cannot assume such simple measurement noise models. This is true in gradient-based optical flow if one considers noise in measurements of spatial image derivatives. It is certainly true with subspace translation constraints (2), where uncertainty in optical flow affects all components of the 3D measurement vector, $\vec{\tau}$. In these cases, with linear constraints and noise in all components of the measurement vector, total-least-squares (TLS) [5] has been used as an estimator, both for optical flow [21, 1] and 3D translation direction [9, 15].

The TLS estimator is a ML estimator for independent and identically distributed (IID) additive Gaussian noise [4, 20]. However, the TLS formulation does not provide a likelihood function. Previous approaches to obtain this likelihood function, like *error-in-variables* [4], introduced the true (noiseless) values of the measurements as new parameters of the likelihood function. These parameters are often called *nuisance* parameters because we are typically not interested in estimating them. For example, the nuisance parameters in Eq. (1) are the noiseless values of the spatiotemporal derivatives and the nuisance parameters in Eq. (2) are the true values of the $\vec{\tau}$ vectors. The problem with including nuisance parameters in the likelihood function is that the dimensionality of the likelihood function then increases with the number of equations (i.e., with the number of measurements), rather than with the number of unknowns we wish to estimate. This poses a serious problem for the practical application of the likelihood function in a Bayesian calculation, and for the derivation of confidence bounds on the estimates.

In this paper, we derive a likelihood function whose dimension is equal to the number of unknowns. This simplified, low-dimensional likelihood function is easier to use in practice (e.g., propagated using belief propagation, or in Bayesian tracking). In addition, we derive the Cramer-Rao lower bound -CRLB- (a bound on the covariance of the error in the estimates). Several approximations of the CRLB have been proposed to obtain a confidence measures for ML estimators in TLS problems. We propose a different approximation of the CRLB and we show, using Monte Carlo simulations, that our approximation compares favorably with the previous approximations.

2. Likelihood Function for Linear Constraints

Consider the problem of estimating an N-dimensional variable, $\mathbf{x}$, in an over-determined linear system, $\mathbf{A}_0\mathbf{x} = \mathbf{b}_0$, where our only observations $\mathbf{A}$ and $\mathbf{b}$, are noisy measurements of $\mathbf{A}_0$ and $\mathbf{b}_0$. This problem has been studied independently from a statistical perspective, called *error-in-variables* (EIV) [4], and from a numerical analysis perspective, called *total-least-squares* (TLS) [5]. Van Huffel and Vandewalle [20] linked the results obtained in the EIV analysis with the TLS solution, showing that TLS is a maximum likelihood estimator for IID Gaussian noise.

TLS has a direct geometrical interpretation. For notational convenience, we define $\mathbf{c}_i$ as a N+1 dimensional vector formed by taking the transpose of the i-th row of the matrix $\mathbf{A}$, $\mathbf{a}_i'$, together with the i-th element of the column matrix $\mathbf{b}$, b_i, i.e., $\mathbf{c}_i \equiv (\mathbf{a}_i', b_i)'$; we also define $\mathbf{x}_h \equiv (\mathbf{x}', -1)'$. Then, TLS can be interpreted as finding the characteristic vector $\mathbf{x}_h$ of a hyperplane passing through the origin, which minimizes the sum of squared orthogonal distances between the points $\mathbf{c}_i$ and the hyperplane, given generically by $\mathbf{c}'\mathbf{x}_h = 0$. Therefore, the TLS solution $\mathbf{x}_{TLS}$ minimizes the following function:

$$\mathbf{x}_{TLS} = \arg\min_{\mathbf{x}} \frac{\mathbf{x}_h' \mathbf{C}' \mathbf{C} \mathbf{x}_h}{||\mathbf{x}_h||^2} \tag{3}$$

where $\mathbf{C} \equiv [\mathbf{A}|\mathbf{b}]$ so that the i-th row of $\mathbf{C}$ is $\mathbf{c}_i'$. Given the Singular Value Decomposition (SVD) of $\mathbf{C} = \mathbf{U}\Sigma\mathbf{V}'$, it has been shown [20] that the $\mathbf{x}_{TLS}$ which minimizes Eq. (3) is given by taking the negative of the result of normalizing the first N elements of the last column of $\mathbf{V}$ by the last element of $\mathbf{V}$.

Our goal here is to obtain a simple expression for the likelihood function on $\mathbf{x}$, given measurements $\mathbf{A}$ and $\mathbf{b}$. Our approach starts with the likelihood function formulated by the EIV analysis, which includes the nuisance parameters. Then we eliminate the nuisance parameters, resulting in a likelihood function defined only in the parameters of interest, $\mathbf{x}$, and therefore having a much lower dimensionality than the original likelihood function. One interesting result is the need of a prior probability distribution *on the true values* underlying the noisy measurements [6]. We show that certain choices of the prior and noise distributions lead to a likelihood function that is maximized by the same values of the parameters $\mathbf{x}$ that maximize the full likelihood function given by the EIV approach, which are given by the TLS solution.

2.1. One Constraint

We start by considering one equation from the above over-determined linear system, $\mathbf{a}_0'\mathbf{x} = b_0$, where $\mathbf{a}_0$ and b_0 are the true values for which the constraint holds exactly.

2.1.1. EIV Likelihood Function. Following the EIV approach [4], let the measurements be contaminated by additive noise:

$$\begin{aligned} \mathbf{a} &= \mathbf{a}_0 + \mathbf{n}_a \\ b &= b_0 + n_b \;\; = \mathbf{a}_0{}'\mathbf{x} + n_b \end{aligned} \tag{4}$$

where $(\mathbf{a}', b)'$ are the measurements, composed of signal $(\mathbf{a}_0', b_0)'$ and noise $\mathbf{n} \equiv (\mathbf{n}_a', n_b)'$. From Eq. (4) and the pdf of the noise, $p_{\mathbf{n}}(\mathbf{n})$, it is straightforward to derive the EIV likelihood function [4] as follows:

$$p(\mathbf{a}, b \,|\, \mathbf{x}, \mathbf{a}_0) \;=\; p_{\mathbf{n}}\left((\mathbf{n}_a', n_b)'\right) \tag{5}$$

where, from Eq. (4), $\mathbf{n}_a = \mathbf{a} - \mathbf{a}_0$ and $n_b = b - b_0$. (For the sake of notational simplicity, we omit the subscript on the probability functions when they are the same random variables as the arguments to the function.) In Eq. (5), $\mathbf{a}_0$ are the nuisance parameters.

The problem with the EIV formulation is that the dimension of the likelihood function grows with the number of nuisance parameters, and hence with the number of measurements. This a serious disadvantage if we are interested in estimating $\mathbf{x}$ only, in computing confidence bounds on a ML estimator for $\mathbf{x}$, or in propagating a distribution over $\mathbf{x}$ as part of a larger Bayesian calculation.

2.1.2. Integration of the Nuisance Parameters. To obtain a more compact expression for the likelihood function, our approach is to integrate over the nuisance parameters, $\mathbf{a}_0$ in Eq. (5). We formulate this using Bayes' rule as follows:

$$\begin{aligned} p(\mathbf{a}, b \,|\, \mathbf{x}) &= \int_{\mathbf{a}_0} \mathrm{d}\mathbf{a}_0 \, p(\mathbf{a}, b, \mathbf{a}_0 \,|\, \mathbf{x}) \\ &= \int_{\mathbf{a}_0} \mathrm{d}\mathbf{a}_0 \, p(\mathbf{a}, b \,|\, \mathbf{x}, \mathbf{a}_0) \, p(\mathbf{a}_0 \,|\, \mathbf{x}) \end{aligned} \tag{6}$$

where the simplified likelihood function depends on the full likelihood function in Eq. (5), and on the conditional pdf $p(\mathbf{a}_0|\mathbf{x})$, which is a conditional prior on the true values of the nuisance parameters $\mathbf{a}_0$. As with any prior probability, this conditional prior has to be specified according to the knowledge that we have about the problem. The ease with which one can solve the integral in Eq. (6) depends on both the prior and the noise distribution.

2.1.3. Prior Distributions on the Nuisance Parameters. If there is not enough *a priori* knowledge about the problem, several assumptions can be made. One idea might be to assume that $\mathbf{a}_0$ is independent of $\mathbf{x}$, and therefore the conditional prior on $\mathbf{a}_0|\mathbf{x}$ is just a prior on $\mathbf{a}_0$. However, this assumption is problematic, because it means that we do not impose any limit on the variance, or power, of $b_0|\mathbf{x}$; yet all signals of practical interest have limited power. Indeed, because of the linear relationship that relates $\mathbf{a}_0'\mathbf{x} = b_0$, it follows (assuming also that $\mathcal{E}[\mathbf{a}_0] = \mathbf{0}$, where $\mathcal{E}[\cdot]$ is the expected value) that $\mathcal{E}[b_0^2|\mathbf{x}] = \mathbf{x}'\mathcal{E}[\mathbf{a}_0\mathbf{a}_0']\mathbf{x}$. Therefore, as $||\mathbf{x}||$ increases, $\mathcal{E}[b_0^2|\mathbf{x}]$ also increases.

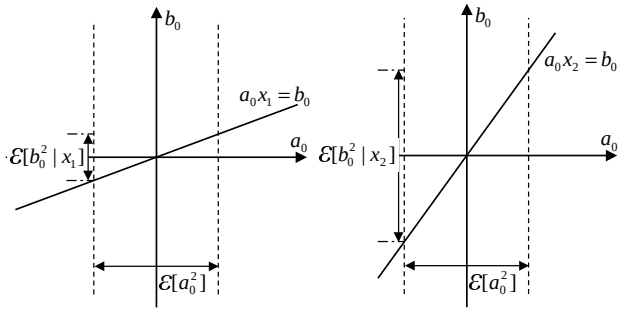

Figure 1. Illustration of the behavior of the conditional prior variances for the independent prior model.

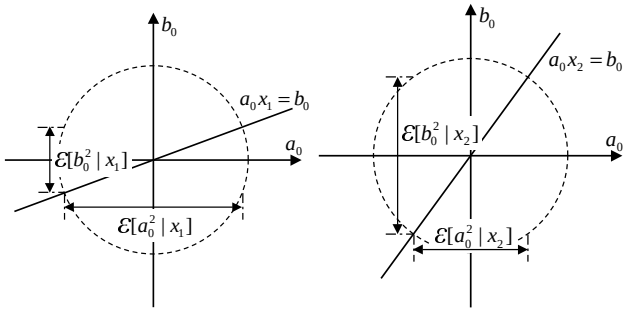

Figure 2. Illustration of the behavior of the conditional prior variances for the constant power model.

This is illustrated in Fig. 1 in 1 dimension. In these graphs, the two vertical, dashed lines denote some bounds on a potential distribution of a_0. For example, they can be the limits of a uniform distribution over a_0. Each panel corresponds to a different value of the parameter x. The variance of the conditional distribution of $a_0|x$ is independent of x, as is reflected by the same value of $\mathcal{E}[a_0^2]$ in both panels of Fig. 1. When x is given, a_0 and b_0 should lie in the line given by $a_0 x = b_0$. Therefore, as $|x|$ increases, $\mathcal{E}[b_0^2|x]$ also increases (compare the greater variance of $b_0|x_2$ in the right graph, with that of $b_0|x_1$ in the left graph).

Fig. 2 illustrates another possible model for the prior, in which the sum of the variances of a_0 and b_0, given x, remain constant. This is represented by the dashed circle, which in this case is a limit on a potential joint distribution of (a_0, b_0). As $|x|$ increases, $\mathcal{E}[b_0^2|x]$ increases (compare the greater variance for $b_0|x_2$ in the right graph, with that of $b_0|x_1$ in the left graph), but this is compensated by a decrease of $\mathcal{E}[a_0^2|x]$ (compare the smaller variance for $a_0|x_2$ in the right graph, with that of $a_0|x_1$ in the left graph), such that the sum of both variances remain constant. This last model seems more reasonable in practical applications, because the power of the total signal $(\mathbf{a}_0, b_0)$ is not allowed not grow without bound.

In what follows, we use this second (constant power) prior. Moreover, we assume a Gaussian distribution for this prior, which yields an analytical solution for the integral in Eq (6). Given $\mathbf{x}$, we know that $\mathbf{a}_0'\mathbf{x} = b_0$, and therefore the joint conditional probability $p(\mathbf{a}_0, b_0|\mathbf{x})$ is only nonzero along the hyperplane given by the constraint, $\mathbf{a}_0'\mathbf{x} - b_0$. In the example of Fig. 2, $p(\mathbf{a}_0, b_0|\mathbf{x})$ lies on the diagonal line in each panel. The joint conditional probability can thus be expressed as proportional to a multidimensional Dirac delta function, i.e. $\delta(\mathbf{a}_0'\mathbf{x} - b_0)$. To make this pdf integrable, we multiply the delta function by an isotropic Gaussian in $(\mathbf{a}_0, b_0)$, such that the final pdf is also isotropic. Intuitively, this is a soft version of the circles in Fig. 2. Thus, the power of the total signal $(\mathbf{a}_0, b_0)$ remains constant independent of $\mathbf{x}$. Letting $\mathbf{c}_0 \equiv (\mathbf{a}_0', b_0)'$ and $\mathbf{x}_h \equiv (\mathbf{x}', -1)'$ for notational convenience, the joint conditional prior is then given by:

$$p(\mathbf{a}_0, b_0|\mathbf{x}) \;=\; \frac{k_1}{(2\pi\sigma_0^2)^{N/2}}\delta(\mathbf{c}_0'\mathbf{x}_h)\;\exp\left(\frac{-\mathbf{c}_0'\mathbf{c}_0}{2\sigma_0^2}\right) \quad (7)$$

where N is the dimension of the parameter vector $\mathbf{x}$. It turns out that $k_1 = ||\mathbf{x}_h||$ (where $||\cdot||$ denotes the modulus (2-norm) of a vector) to make the integral of the pdf equal to 1 (see Appendix I).

Finally, to obtain the desired prior $p(\mathbf{a}_0|\mathbf{x})$, we integrate the joint conditional prior in Eq. (7) over b_0, resulting in:

$$p(\mathbf{a}_0|\mathbf{x}) \;=\; \frac{||\mathbf{x}_h||}{(2\pi\sigma_0^2)^{N/2}}\exp\left(\frac{-\mathbf{a}_0'\Sigma_{\mathbf{x}}^{-1}\mathbf{a}_0}{2\sigma_0^2}\right) \quad (8)$$

where $\Sigma_{\mathbf{x}} = (\mathbf{I}_N + \mathbf{x}\mathbf{x}')^{-1} = \mathbf{I}_N - \mathbf{x}\mathbf{x}'/||\mathbf{x}_h||^2$, and where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Eq. (8) provides the conditional prior on the nuisance parameters needed in Eq (6).

2.1.4. Likelihood function for Gaussian Noise. With the Gaussian conditional prior in Eq. (8), and the full likelihood function in Eq. (5), we can now return to Eq. (6) to obtain the desired likelihood function. For the special case of isotropic Gaussian noise, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I}_{N+1})$, with $\mathbf{c} \equiv (\mathbf{a}', b)'$ for notational convenience, one can derive the following analytical solution for the likelihood:

$$p(\mathbf{a}, b|\mathbf{x}) = k_2\,\exp\left(-\frac{1}{2\sigma_n^2}\left[(1-\gamma)\mathbf{c}'\mathbf{c} + \gamma\frac{\mathbf{c}'\mathbf{x}_h\mathbf{x}_h'\mathbf{c}}{||\mathbf{x}_h||^2}\right]\right) \quad (9)$$

where $k_2 = \gamma^{N/2}/[(2\pi)^{(N+1)/2}\sigma_0^N\sigma_n]$. The factor $\gamma \equiv \frac{\sigma_0^2}{\sigma_n^2+\sigma_0^2}$ is related with the signal to noise ratio (SNR). If the SNR is high ($\sigma_0^2 \gg \sigma_n^2$), then $\gamma \simeq 1$.

For non-Gaussian priors or non-Gaussian noise it may be necessary to use numerical integration to approximate the likelihood function.

2.2. Multiple Constraints

We now generalize to the case of L constraints, expressed in matrix form as $\mathbf{A}_0\mathbf{x} = \mathbf{b}_0$. The rows of $\mathbf{A}_0$ are formed by the true values for each equation, $\mathbf{a}_{0_i}'$, and $\mathbf{b}_0$ is a column vector containing the true values b_{0_i}. Following a similar notation, we collect the noisy measurements $\mathbf{a}_i'$ into matrix $\mathbf{A}$, and b_i into column vector $\mathbf{b}$.

If the L constraints are independent, we can express the complete likelihood as the product of the individual likelihoods:

$$p(\mathbf{A}, \mathbf{b}\,|\,\mathbf{x}) = \prod_{i=1}^{L} p(\mathbf{a}_i, b_i\,|\,\mathbf{x}) \quad (10)$$

In the case of IID Gaussian noise, and identical Gaussian conditional priors for every constraint, the pdf in (10) simplifies to

$$k_2^L\,\exp\left(-\frac{1}{2\sigma_n^2}\left[(1-\gamma)\mathrm{tr}(\mathbf{C}\mathbf{C}') + \gamma\frac{\mathbf{x}_h'\mathbf{C}'\mathbf{C}\mathbf{x}_h}{||\mathbf{x}_h||^2}\right]\right) \quad (11)$$

where $\mathbf{C} \equiv [\mathbf{A}|\mathbf{b}]$, and $\mathrm{tr}(\cdot)$ is the trace of a matrix. One important characteristic of this likelihood function is that, although it has been derived considering a particular form for the conditional prior distribution of the nuisance parameters, its maximum occurs at the estimate of $\mathbf{x}$ given by TLS [20], independent of γ. This becomes clear by comparing the error function that minimizes the TLS solution in Eq. (3) with the likelihood function in Eq. (11). In addition, this estimate also maximizes the original likelihood function in Eq. (5) of the EIV approach for IID Gaussian noise.

The particular choice of the conditional joint prior in Eq. (7) has led to a likelihood function whose maximum is the same as the TLS estimator. However, different choices for this prior can lead to likelihood functions whose maxima are not the same as the TLS solution, and which, in principle, produce more accurate estimates than the TLS estimator, because of the introduction of prior knowledge.

3. Confidence Bounds

The likelihood function derived in the previous Section can be used in Bayesian computations. However, there are situations in which it is enough to give a ML estimate. In those cases, it is important to provide confidence bounds on the accuracy of the estimates. A standard procedure is to give the Cramer-Rao lower bound (CRLB) on the variance of the estimates. The CRLB is given by the inverse of the expected value of the Hessian of the negative log-likelihood function, evaluated at the true values of the parameters (for unbiased estimators). This lower bound is achieved by the Maximum Likelihood estimator.

One important consequence of having derived the likelihood in Eq. (11) is that we can derive the CRLB for the TLS solution easily. The first result towards this end is that the Hessian of the negative log-likelihood function in Eq. (11)

can be shown to be

$$\mathbf{H} = \frac{\gamma}{\sigma_n^2 ||\mathbf{x}_h||^2} \left(\mathbf{M} - \frac{1}{||\mathbf{x}_h||^2} (\mathbf{x}_h' \mathbf{D} \mathbf{x}_h) \mathbf{I}_N + \frac{4}{||\mathbf{x}_h||^4} \left[(\mathbf{x}_h' \mathbf{D} \mathbf{x}_h) \mathbf{x} - ||\mathbf{x}_h||^2 (\mathbf{M}\mathbf{x} - \mathbf{A}'\mathbf{b}) \right] \mathbf{x}' \right) \quad (12)$$

where $\mathbf{M} \equiv \mathbf{A}'\mathbf{A}$, and $\mathbf{D} \equiv \mathbf{C}'\mathbf{C}$. This Hessian is relevant because it will be used to approximate the CRLB, providing an accurate confidence bound (see below).

The CRLB is obtained by taking the expected value of $\mathbf{H}$ with respect to the noisy measurements, $\mathbf{A}$ and $\mathbf{b}$, inverting the result, and evaluating it at the true value of the parameters $\mathbf{x}_0$. It can be shown that this results in:

$$\mathcal{E}[\mathbf{H}]^{-1} = \frac{1}{\gamma} \sigma_n^2 ||\mathbf{x}_{0_h}||^2 \mathbf{M}_0^{-1} \quad (13)$$

where $\mathbf{M}_0 \equiv \mathbf{A}_0' \mathbf{A}_0$. Although this CRLB corresponds to a likelihood function obtained using a specific prior for the true values, it is equivalent to the asymptotic (i.e., for a large number of equations) covariance matrix of the estimates found by Gallo [3, 20]. In addition, if the SNR is high ($\gamma \simeq 1$), the CRLB in Eq. (13) reduces to the approximations found by Koopmans [13], and later by Kanatani [11], that were derived without knowing the likelihood function.

3.1. Approximations of the CRLB

The CRLB depends on the true values, $\mathbf{A}_0$ and $\mathbf{x}_0$, on the variance of the noise σ_n^2, and on γ. However, in practical estimation problems we do not know these values, and therefore it is necessary to approximate the CRLB. Usually, the true value of $\mathbf{x}$ is approximated by its estimate, $\tilde{\mathbf{x}}$. The power of the noise is estimated as the square of the smallest singular value of the augmented data matrix, $[\mathbf{A}|\mathbf{b}]$, normalized by the number of equations [20] (we call this estimate $\tilde{\sigma}_n^2$). Because we have obtained the Hessian of the likelihood function, Eq. (12), we can use it directly to approximate the CRLB:

1. We propose here the direct use of the inverse of the Hessian in Eq. (12), $\mathbf{C}_1 = \mathbf{H}^{-1}$, that depends only on the measurements. In doing so, we are effectively approximating the expected value in Eq. (13) by a single realization.

Interestingly, several different approximations of the CRLB in Eq. (13) have been proposed. We define here two of the approximations of the CRLB, which we compare with the proposed $\mathbf{C}_1$:

2. Ohta [17] proposes the direct use of $\mathbf{M}$, composed by the noisy measurements, as an approximation of $\mathbf{M}_0$ in Eq. (13), and assuming high SNR ($\gamma \simeq 1$); this results in the approximation $\mathbf{C}_2 = \tilde{\sigma}_n^2 ||\tilde{\mathbf{x}}_h||^2 \mathbf{M}^{-1}$.
3. Van Huffel and Vandewalle [20] propose to correct $\mathbf{M}$ to account for the power of the noise, and with $\gamma \simeq 1$; this produces the approximation $\mathbf{C}_3 = \tilde{\sigma}_n^2 ||\tilde{\mathbf{x}}_h||^2 (\mathbf{M} - L\tilde{\sigma}_n^2 \mathbf{I}_N)^{-1}$.

$\mathbf{C}_2$ and $\mathbf{C}_3$ are high SNR approximations to the CRLB, taken as $\gamma \to 1$. Because the proposed approximation $\mathbf{C}_1$ accounts for the factor γ, we also considered generalizations of the last two approximations, including the factor γ as in Eq. (13):

4. Generalized version of $\mathbf{C}_2$, that is, $\mathbf{C}_4 = (1/\gamma)\mathbf{C}_2$.
5. Generalized version of $\mathbf{C}_3$, that is, $\mathbf{C}_5 = (1/\gamma)\mathbf{C}_3$.

3.2. Monte Carlo Evaluation

To determine which of these 5 approximations to the error bound is more accurate, we ran an extensive series of Monte Carlo simulations. We found that $\mathbf{C}_1$ is the most accurate approximation to the CRLB.

We simulated a 2D estimation problem (i.e., N=2). For each trial of the simulation, we drew a random set of L independent vectors $\{\mathbf{a}_{0_i}\}, 1 \leq i \leq L$, from the same Gaussian conditional prior compatible with the true value of the parameter vector $\mathbf{x}_0$, given by Eq. 8. The corresponding true values of b_{0_i} are given by the linear constraint, $b_{i_0} = \mathbf{a}_{0_i}' \mathbf{x}_0$. The power of the signal was in all cases $\sigma_0^2 = 1$. Then, we added IID Gaussian noise of variance σ_n^2 to the true values, to obtain the noisy observations.

For each trial we estimated $\mathbf{x}$ using the TLS estimator $\tilde{\mathbf{x}}$. We then obtained an error $\mathbf{e} = \tilde{\mathbf{x}} - \mathbf{x}_0$, and the five confidence bounds ($\mathbf{C}_j, 1 \leq j \leq 5$) defined above [1]. To obtain a measure of the accuracy of each confidence bound, we used $\mathbf{C}_j$ to whiten the error $\mathbf{e}$, i.e., we obtain a normalized error $\mathbf{e}_{w_j} = \mathbf{Q}_j^{-1}\mathbf{e}$, such that $\mathbf{Q}_j$ is a matrix that satisfies $\mathbf{C}_j = \mathbf{Q}_j \mathbf{Q}_j'$ (e.g., the Cholesky decomposition, or the eigenvector matrix multiplied by the square root of the eigenvalue matrix). Therefore, if the confidence bound is equal the covariance of the error ($\mathbf{C}_j = \mathcal{E}[\mathbf{e}\mathbf{e}']$), then the covariance of the whitened error has to be the identity matrix ($\mathcal{E}[\mathbf{e}_{w_j} \mathbf{e}_{w_j}'] = \mathbf{I}_N$).

We measured the accuracy of the different confidence bounds by running a large number of simulations, and computing the sample covariance of the whitened errors. Better bounds should yield sample covariances that are close to the identity matrix. To summarize the deviation of the sample covariances from the identity matrix, we have computed the norm of the singular values of the difference between the sample covariance and the identity matrix. This norm (E) should be close to 0 when the sample covariance is close to the identity.

[1] in $\mathbf{C}_1$, $\mathbf{C}_4$ and $\mathbf{C}_5$ we have used the true power of the signal, σ_0^2, and the estimated power of the noise, $\tilde{\sigma}_n^2$, to compute γ.

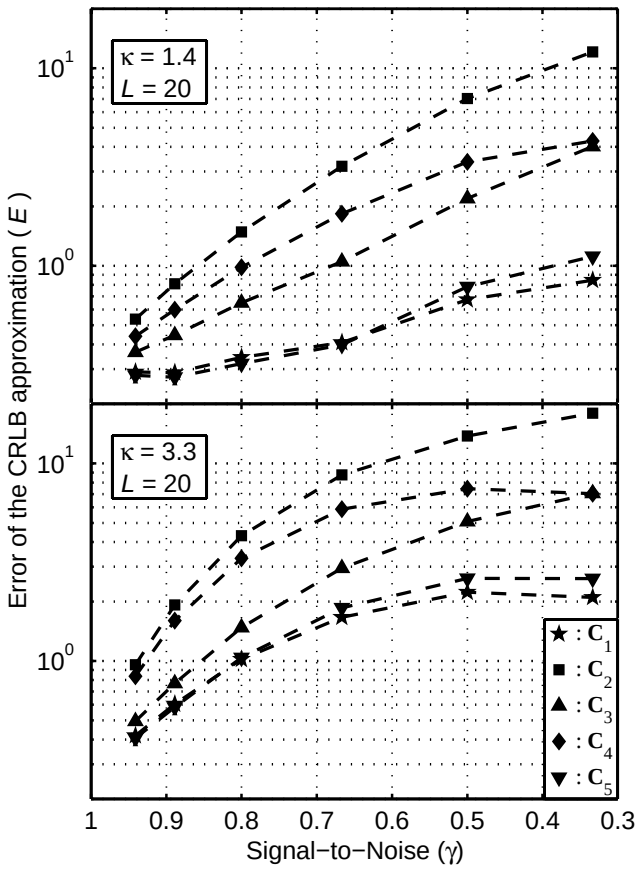

Figure 3. Comparison of the accuracy of the five different CRLB approximations. Top panel, well-conditioned system. Bottom panel, poorly conditioned system. Both panels plot the error (E) of the CRLB approximation as a function of the signal-to-noise parameter (γ).

We ran 10,000 simulations, for two different data set sample sizes (small, $L = 20$, and large, $L = 500$) and two different condition numbers (κ) of $\mathbf{A}_0$ (a well-conditioned system of equations, $\kappa = 1.4$, and a poorly-conditioned system, $\kappa = 3.3$). The poorly-conditioned system was obtained by modifying the covariance matrix of the conditional prior on $\mathbf{a}_0|\mathbf{x}$ (i.e., in this case we were sampling from a different prior than that used to compute the likelihood), such that the condition number of the resulting systems was on average the desired one, and at the same time forcing the total signal power to be unchanged ($\sigma_0^2 = 1$). In all simulations the true value of the parameter was $\mathbf{x}_0 = (0.75, 0.75)'$. The power of the additive noise (σ_n^2) was varied between 1/16 and 2 (i.e., γ between 0.94 and 0.33).

The graphs in Figure 3 plot the error of the CRLB approximations (E) as a function of the signal-to-noise parameter (γ), for the two cases having a small sample size ($L = 20$). As expected, the high SNR approximations ($\mathbf{C}_2$ and $\mathbf{C}_3$), suffer a large degradation as γ diminishes, but even for γ close to 1, results are worse than the corrected versions ($\mathbf{C}_4$ and $\mathbf{C}_5$). Also as expected, $\mathbf{C}_5$ is generally better than $\mathbf{C}_4$ because $\mathbf{M}$ is corrected by the power of the noise (likewise, $\mathbf{C}_3$ is better than $\mathbf{C}_2$). One problem with the corrected version of $\mathbf{M}$ (used to compute both $\mathbf{C}_3$ and $\mathbf{C}_5$), however, is that as the SNR decreases, we found that the inverse of $(\mathbf{M} - L\tilde{\sigma}_n^2\mathbf{I}_N)$ was not always positive definite, which is inconsistent with the definition of a covariance matrix (we have omitted those cases in the computation of the sample covariance matrices of the normalized errors).

The accuracy obtained using $\mathbf{C}_1$ (the inverse of the Hessian) is similar or better than that obtained using any of the other confidence bounds. If γ is not known, one can still use a version of $\mathbf{C}_1$ with $\gamma = 1$; we found that this approximation was still better than $\mathbf{C}_3$. In addition, the inverse of the Hessian is, by definition, positive definite, which is a great advantage with respect to bounds ($\mathbf{C}_3$ and $\mathbf{C}_5$) that use the corrected version of $\mathbf{M}$.

Finally, for the large sample size ($L = 500$, graphs not shown here), the accuracy of all 5 CRLB approximations improved dramatically, especially for $\mathbf{C}_1$ and $\mathbf{C}_5$, which in this case gave $E \approx 0$ for all the conditions tested.

4. Application to Estimation of the 3D Direction of Translation

The approach that we have described can be applied to a number of problems in computer vision. In particular, it can be applied to optical flow estimation. Indeed, TLS has already been used in optical flow estimation [21, 1], but without providing confidence bounds on the accuracy of the estimates, nor in a Bayesian framework.

Here we apply this method to derive likelihood functions for the direction of camera translation, from probability distributions of optical flow. We have extended the linear subspace method for recovering the direction of translation proposed by Jepson and Heeger [9, 10], to accept, as input, probability distributions of optical flow.

The linear subspace algorithm linearly combines optical flow vectors from different locations to obtain a set of three dimensional vectors $\{\vec{\tau}_i\}$, that are orthogonal to the direction of camera translation:

$$\vec{\tau}_i'\mathbf{T} = 0, \quad i = 1, \ldots, N \tag{14}$$

where $\mathbf{T}$ is the direction of translation of the camera, and N the number of equations [2]. Because the $\vec{\tau}_i$ vectors are a linear combination of a large number of optical flow vectors, it can be assumed that the noise affecting $\vec{\tau}_i$ is Gaussian. Thus, we can use a generalized version of Eq. (9) for an arbitrary noise covariance, to formulate the likelihood function for each individual constraint. However, we can not

[2] Typically, if the number of locations where the optical flow is available is M, it is possible to obtain $N = M - 6$ equations [10].

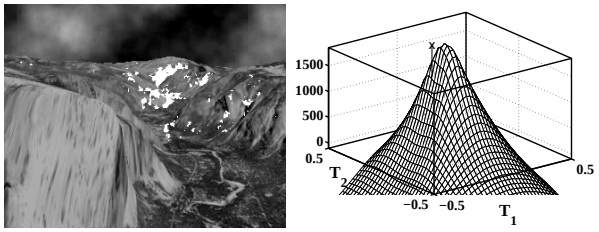

Figure 4. Left: Central frame of the *Yosemite* sequence, with the selected pixels displayed in white. Right: Likelihood (in arbitrary units) as a function of the two first components of the direction of translation (X is the true direction of translation).

use Eq. (11) because the covariance of the noise is different for each $\vec{\tau}_i$. Here we evaluate numerically the likelihood for each constraint, and then, assuming independence between constraints, we apply Eq. (10) to obtain the total likelihood.

4.1. Results

We tested this probabilistic formulation with the synthetic *Yosemite* fly-through image sequence, where the true translation velocity is known. For this purpose we have used the probability distributions of optical flow generated by an optical flow estimation method described in [16], for the central frame of the sequence. This method gives an estimate of the optical flow and of its associated covariance matrix at every pixel.

We have selected a reduced number of pixels (about 1300 with the highest confidence, shown in white in the left image of Fig. 4) to avoid an excessive computational cost. From this set of selected pixels, we generated all the possible $\vec{\tau}_i$ vectors (the number of pixels minus 6).

The resulting likelihood distribution is depicted in the right graph of Fig. 4, as a function of the two first components of $\mathbf{T}$. The maximum of this distribution is located at $\mathbf{T}_{max} = (0.07, 0.15, 0.98)$. The angular error with respect to the true translation velocity $\mathbf{T}_0 = (0, 0.17, 0.98)$ (marked with an X) is 4.5^o. We also used the original algorithm [10] to obtain a least squares estimate of the direction of translation from the same set of pixels, resulting in a $\mathbf{T}_{est} = (0.16, 0.13, 0.98)$, whose angular error ($9.7^o$) is twice the error obtained with the probabilistic method. In addition, we have obtained a full likelihood function, from which confidence measures can be extracted.

5. Conclusion

We have presented a method for computing low dimensional likelihood functions in cases were the variables of interest are linearly constrained by noisy measurements. Such likelihood functions are highly valuable both as part of a Bayesian calculation, and as a means for establishing confidence bounds on ML estimators.

The simplified likelihood functions require the introduction of prior probabilities *on the true values* underlying the noisy measurements. Different prior models can lead to substantially different likelihood functions, and therefore, to different estimators that should be more accurate than the TLS solution, because of the introduction of some prior information.

In this paper we have used a Gaussian, constant power prior, which results in the simplified likelihood functions in Eq. (9), in the case of a single constraint, and in Eq. (11), for multiple constraints. This last function has two interesting properties. First, its maximum is the same as the maximum found by TLS. Second, the Cramer-Rao lower bound for this simplified likelihood function is equivalent to previous approximations that were derived without knowing the likelihood function.

In addition, the simplified likelihood functions have two advantages with respect to previous results. (1) We can approximate the CRLB using the inverse of the Hessian of the negative log-likelihood, given in Eq. (12). We have shown that this approximation is both more accurate and stable (it is always positive definite) than previous approximations. (2) The simplified likelihood function can be used as part of a larger Bayesian calculation very efficiently (without the need of including the nuisance parameters).

We have also applied the formulation to compute probability distributions for the direction of camera translation. Starting with the linear subspace method, we introduced full probability distributions for the estimated optical flow and we end up with a full pdf for the direction of translation. The estimate we obtained using this method is more accurate than the estimate given by the original method [10], and the new method computes as additional information the full distribution of the likelihood function.

Acknowledgments. ON acknowledges financial support by a Spanish Ministry of Education (MEC)/Fulbright Fellowship. We thank Horst Haussecker and Allan Jepson for providing useful comments on this research.

References

[1] A. Bab-Hadiashar and D. Suter. Robust optica flow computation. *International Journal of Computer Vision*, 29(1):59–77, 1998.

[2] R. Bracewell. *Two–dimensional imaging*. Prentice–Hall, Inc., Englewoods Cliffs, NJ, 1995.

[3] P. P. Gallo. *Properties of estimators in error-in-variables models*. PhD thesis, University of North Carolina, 1982.

[4] L. J. Gleser. Estimation in a multivariate "error in variables" regression model: Large sample results. *The Annals of Statistics*, 9(1):24–44, 1981.

[5] G. H. Golub and C. F. V. Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17(6):883–893, 1980.
[6] S. F. Gull. Bayesian data analysis: Straight-line fitting. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, pages 511–518. Kluwer Academic Publishers, 1989.
[7] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
[8] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.
[9] A. D. Jepson and D. J. Heeger. A fast subspace algorithm for recovering rigid motion. In *Proceedings IEEE Workshop on Visual Motion*, pages 124–131, Princenton, 1991.
[10] A. D. Jepson and D. J. Heeger. Linear subspace methods for recovering the translation direction. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots*, pages 39–62. Cambridge University Press, New York, 1993.
[11] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, Amsterdam, 1996.
[12] Y. Kanazawa and K. Kanatani. Optimal line fitting and reliability evaluation. *IEICE Trans. Inf. & Syst.*, E79-D(9):1317–1322, 1996.
[13] T. C. Koopmans. *Linear Regression Analisys of Economic Time Series*. DeErven F. Bohn, N.V. Haarlem, The Netherlands, 1937.
[14] M. R. Luettgen, C. Karl, and A. S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Trans. on Image Processing*, 3(1):41–64, 1994.
[15] W. J. MacLean. Removal of translation bias when using subspace methods. In *Proc. of the IEEE International Conference on Computer Vision*, pages 753–758, Corfu, Greece, Sept 1999.
[16] O. Nestares and R. Navarro. Probabilistic multichannel optical flow analysis based on a multipurpose visual representation of image sequences. In *Proceedings of the SPIE, vol. 3644*, pages 429–440, San Jose, CA, Jan 1999.
[17] N. Ohta. Optical flow detection using a general noise model. *IEICE Trans. Inf. & Syst.*, E79-D(7):951–957, 1996.
[18] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 310–315, Mauii, Hawaii, Jun 1991.
[19] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Publishers, Boston, 1989.
[20] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1991.
[21] J. Weber and J. Malik. Robust computation of optical flow in a multi–scale differential framework. *International Journal of Computer Vision*, 14:67–81, 1995.

Appendix I

Here we derive the value of k_1 in Eq. (7) that makes the pdf integrate to 1. First, we note the following property to convert integrals with multidimensional Dirac delta functions into surface integrals [2]:

$$\int_{\mathbf{c}_0} \mathrm{d}\mathbf{c}_0 \; \delta(g(\mathbf{c}_0, \mathbf{x}))\, p(\mathbf{c}_0) \; = \int_{S \equiv \{\mathbf{c}_0 \,|\, g(\mathbf{c}_0, \mathbf{x}) = 0\}} \mathrm{d}\mathbf{s} \; \frac{p(\mathbf{c}_0)}{||\vec{\nabla}_{\mathbf{c}_0} g(\mathbf{c}_0, \mathbf{x})||} \tag{15}$$

where, $\vec{\nabla}_{\mathbf{c}_0}$ is the gradient operator with respect to $\mathbf{c}_0$, and $\mathrm{d}\mathbf{s}$ is a surface differential. Then, it follows that:

$$k \; = \; \frac{1}{\int_{\mathbf{c}_0} \mathrm{d}\mathbf{c}_0 \; \frac{\delta(\mathbf{c}_0' \mathbf{x}_h)}{(2\pi\sigma_0^2)^{N/2}} \; \exp\left(\frac{-\mathbf{c}_0' \mathbf{c}_0}{2\sigma_0^2}\right)} = ||\mathbf{x}_h|| \tag{16}$$

where $||\cdot||$ denotes the modulus (2-norm) of a vector.