IEEE International Conference on Image Processing, Barcelona, 2003,

ERROR-IN-VARIABLES LIKELIHOOD FUNCTIONS FOR MOTION ESTIMATION

Oscar Nestares
Instituto de Óptica "Daza de Valdés" (CSIC)
Serrano 121, Madrid 28006, Spain

David J. Fleet
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304

ABSTRACT

Over-determined linear systems with noise in all measurements are common in computer vision, and particularly in motion estimation. Maximum likelihood estimators have been proposed to solve such problems, but except for simple cases, the corresponding likelihood functions are extremely complex, and accurate confidence measures do not exist. This paper derives the form of simple likelihood functions for such linear systems in the general case of heteroscedastic noise. We also derive a new algorithm for computing maximum likelihood solutions based on a modified Newton method. The new algorithm is more accurate, and exhibits more reliable convergence behavior than existing methods. We present an application to affine motion estimation, a simple heteroscedastic estimation problem.

1. INTRODUCTION

Problems with noisy linear constraints, especially with noise in all measurements, are ubiquitous in computer vision. Examples include line-fitting [1], gradient-based optical flow estimation [2, 3], linear subspace methods for the estimation of 3D translation [4], and methods for camera calibration and estimation of the fundamental matrix [5]. Such formulations are often called *error-in-variables* (EIV) problems in the statistical literature, and *total-least-squares* (TLS) in the numerical computation literature. Van Huffel and Vandewalle [6] linked both approaches, showing that the TLS solution is a Maximum Likelihood Estimator (MLE) for independent, identically distributed (IID), isotropic Gaussian noise. Recently, a MLE has been proposed for the more general case of heteroscedastic (nonidentically distributed) error-in-variable (HEIV) problems [5]. This is important since an IID assumption is often unrealistic.

One simple HEIV problem is the estimation of affine optical flow from gradient constraints of the form

$$I_x p_0 + I_x x p_1 + I_x y p_2 + I_y p_3 + I_y x p_4 + I_y y p_5 + I_t = 0, \quad (1)$$

which relate the spatiotemporal image gradient (I_x, I_y, I_t) and the spatial coordinates (x, y), with the six unknown parameters of the affine motion model, $(p_0, p_1, p_2, p_3, p_4, p_5)$. ML estimators for the affine parameters are formulated using a collection of gradient constraints from within a local image neighborhood. Measurement noise in Eq. 1 is caused by image formation, by numerical computation of the gradients, and by violations of brightness constancy. Furthermore, even if the noise affecting the gradient were isotropic and IID, the noise affecting the terms in the gradient constraint in Eq. 1 will be non-IID. In particular, the gradient measurements in Eq. 1 are scaled by the spatial position of the measurement. Accordingly, if σ^2 were the variance of each component of the gradient, the effective covariance matrix for the different terms in the linear gradient constraint in Eq. 1 would be given by

$$\mathbf{C}_{(x,y)} = \sigma^2 \begin{pmatrix} 1 & x & y & 0 & 0 & 0 & 0 \\ x & x^2 & xy & 0 & 0 & 0 & 0 \\ y & xy & y^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y & 0 \\ 0 & 0 & 0 & x & x^2 & xy & 0 \\ 0 & 0 & 0 & y & xy & y^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

Our goal in this paper is to derive likelihood functions for HEIV problems like this one. As is well-known [7], existing EIV likelihood functions are expressed as functions of both the parameters of interest (the affine parameters above), as well as the so-called *nuisance parameters*. The nuisance parameters refer to the measurement quantities in the absence of noise (the true spatial gradients in Eq. 1). As a consequence, these likelihood functions are expressed in a parameter space whose dimension increases linearly with the number of measurements. This makes it difficult to derive confidence measures, and impractical for Bayesian inference. The likelihood function we seek is a function only of the parameters of interest (e.g., the affine motion parameters), and hence it lies in a low dimensional parameter space.

In deriving the likelihood function for the general case of heteroscedastic noise we establish constraints on the nuisance parameters that must hold for one to obtain a closed-form expression for the likelihood, whose maximum is identical to that of previous MLE objective functions [5]. The resultant likelihood function can be used in several ways. First, it can lead to improved optimization procedures. Second, it can be used to obtain confidence measures based on the Cramer-Rao lower bound. Finally, as noted by Gull [8], the inclusion of suitable priors that may be different from that implicitly assumed in previous MLEs also improves the accuracy of the estimator.

2. LIKELIHOOD FUNCTION FOR LINEAR CONSTRAINTS

Consider the problem of estimating an N-dimensional variable, $\mathbf{x}$, in an over-determined linear system, $\mathbf{A}_0\mathbf{x} = 0$, where our only observations, $\mathbf{A}$, are noisy measurements of $\mathbf{A}_0$. To derive the desired likelihood function, $p(\mathbf{A}\,|\,\mathbf{x})$, we begin by formulating the likelihood function for a single linear constraint with isotropic noise. Then, we formulate the likelihood function for a single constraint with an arbitrary noise covariance. Finally, we consider the heteroscedastic case with multiple constraints.

2.1. One Linear Constraint

A single linear constraint with noisy measurements can be expressed in homogeneous coordinates with the following two equations:

$$\mathbf{a}_0{}'\mathbf{x} \;=\; 0 \tag{3}$$

$$\mathbf{a} \;=\; \mathbf{a}_0 + \mathbf{n} \tag{4}$$

where $\mathbf{a}_0$ (the nuisance parameters) are the true values for which the linear constraint holds exactly, $\mathbf{x}$ is the parameter vector we wish to estimate, $\mathbf{a}$ is the vector of noisy measurements, and $\mathbf{x}'$ denotes the transpose of $\mathbf{x}$. Our objective is to derive the likelihood function $p(\mathbf{a}\,|\,\mathbf{x})$. We do so by marginalizing the joint conditional probability of the true values and noisy measurements, $p(\mathbf{a},\mathbf{a}_0\,|\,\mathbf{x})$, over $\mathbf{a}_0$:

$$p(\mathbf{a}\,|\,\mathbf{x}) \;=\; \int_{\mathbf{a}_0} \mathrm{d}\mathbf{a}_0\, p(\mathbf{a},\mathbf{a}_0\,|\,\mathbf{x}). \tag{5}$$

Applying Bayes rule to the joint conditional probability, and noting that $\mathbf{a}$ is independent of $\mathbf{x}$ when conditioned on $\mathbf{a}_0$, we find that $p(\mathbf{a}\,|\,\mathbf{x})$ is given by

$$p(\mathbf{a}\,|\,\mathbf{x}) \;=\; \int_{\mathbf{a}_0} \mathrm{d}\mathbf{a}_0\, p(\mathbf{a}\,|\,\mathbf{a}_0)\, p(\mathbf{a}_0\,|\,\mathbf{x}). \tag{6}$$

Given Eq. 4 it follows that $p(\mathbf{a}\,|\,\mathbf{a}_0) \;=\; p_{\mathbf{n}}(\mathbf{a}-\mathbf{a}_0)$, where $p_{\mathbf{n}}(\mathbf{n})$ is the pdf of the noise. It is also clear from Eq. 6 that we need to specify a conditional prior on the nuisance parameters, i.e., $p(\mathbf{a}_0\,|\,\mathbf{x})$.

2.2. Isotropic Gaussian Noise and Prior Distributions

We first consider the case of isotropic Gaussian measurement noise, $\mathbf{n}\sim\mathcal{N}(\mathbf{0},\sigma_n^2\mathbf{I}_N)$, where $\mathbf{I}_N$ denotes the $N\times N$ identity matrix. We also consider an isotropic, zero-mean Gaussian prior for the true values, $\mathbf{a}_0\sim\mathcal{N}(\mathbf{0},\sigma_0^2\mathbf{I}_N)$. The conditional prior on $\mathbf{a}_0$, given $\mathbf{x}$, will be confined to a plane in the $\mathbf{a}_0$ space for which the constraint in Eq. 3 holds. This can be formulated as the product of the density for $\mathbf{a}_0$ and a Dirac delta function that is nonzero in the hyperplane defined by Eq. 3. Given the normal pdf for $\mathbf{a}_0$, this yields

$$p(\mathbf{a}_0\,|\,\mathbf{x}) = \frac{||\mathbf{x}||}{\sqrt{(2\pi\sigma_0^2)^{N-1}}}\exp\left(\frac{-1}{2\sigma_0^2}\mathbf{a}_0'\mathbf{a}_0\right)\delta(\mathbf{a}_0'\mathbf{x})\;, \tag{7}$$

where the scaling factor is chosen to make the pdf integrate to 1. This prior is basically equivalent to the constant power model proposed in [7], but adapted to the homogeneous parameterization we are using here. Introducing a Gaussian pdf for the noise, with this prior, one can solve the integral in Eq. 6 to obtain the following likelihood function:

$$p(\mathbf{a}\,|\,\mathbf{x}) = k\exp\!\left(\frac{-1}{2\sigma_n^2}\left[(1-\gamma)\mathbf{a}'\mathbf{a}+\gamma\frac{\mathbf{a}'\mathbf{x}\mathbf{x}'\mathbf{a}}{||\mathbf{x}||^2}\right]\right)\;, \tag{8}$$

where the factor $\gamma\equiv\frac{\sigma_0^2}{\sigma_n^2+\sigma_0^2}$ is related to the signal to noise ratio (SNR$\equiv\frac{\sigma_0^2}{\sigma_n^2}$), and $k=\gamma^{(N-1)/2}/[(2\pi)^{N/2}\sigma_0^{N-1}\sigma_n]$.

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

2.3. Arbitrary Noise and Prior Covariances

Next we consider the case of Gaussian noise with an arbitrary covariance matrix, $\mathbf{n}\sim\mathcal{N}(\mathbf{0},\sigma_n^2\mathbf{C}_n)$, with $|\mathbf{C}_n|=1$. In this case we first re-normalize the measurements, thereby mapping the problem to a new parameter space in which the noise covariance is isotropic [9]:

$$\underbrace{\mathbf{R}_n\mathbf{a}}_{\mathbf{a}_R} = \underbrace{\mathbf{R}_n\mathbf{a}_0}_{\mathbf{a}_{0_R}} + \underbrace{\mathbf{R}_n\mathbf{n}}_{\mathbf{n}_R} \tag{9}$$

where $\mathbf{R}_n$ is such that $\mathbf{C}_n^{-1}=\mathbf{R}_n'\mathbf{R}_n$.

To obtain the previous isotropic formulation one additional assumption is necessary. That is, the covariance of the re-normalized true values, $\mathbf{a}_{0_R}$, is isotropic, which will be the case only if the covariance of the true values is proportional to the covariance of the noise. This is a strong assumption that is needed mainly for mathematical convenience to keep the expression simple. For the more general case of arbitrary covariances for both the noise and the prior on the true measurements, the resulting likelihood function is not given in closed-form. Making this additional assumption we can apply the previous isotropic result to the re-normalized values, and then undo the re-normalization to obtain the following likelihood function:

$$p(\mathbf{a}\,|\,\mathbf{x}) = k\exp\!\left(\frac{-1}{2\sigma_n^2}\left[(1-\gamma)\mathbf{a}'\mathbf{C}_n^{-1}\mathbf{a}+\gamma\frac{\mathbf{a}'\mathbf{x}\mathbf{x}'\mathbf{a}}{\mathbf{x}'\mathbf{C}_n\mathbf{x}}\right]\right) \tag{10}$$

2.4. Multiple Constraints

We now generalize to the case of L independent constraints, expressed in matrix form as $\mathbf{A}_0\mathbf{x}=0$, where the rows of $\mathbf{A}_0$

are formed by the true values for each equation, $\mathbf{a}_{0_i}'$. Following the same notation, we collect the noisy measurements $\mathbf{a}_i'$ into matrix $\mathbf{A}$. Because we assume that the L constraints are independent, the complete likelihood function is the product of the individual ones:

$$p(\mathbf{A} \mid \mathbf{x}) = \prod_{i=1}^{L} p(\mathbf{a}_i \mid \mathbf{x}) \tag{11}$$

In the case of Gaussian noise with different noise covariances for each constraint (the heteroscedastic case), and assuming again prior covariances proportional to noise covariances, we obtain the following likelihood function:

$$k^L \exp\left(\sum_{i=1}^{L} \frac{-1}{2\sigma_{n_i}^2} \left[(1-\gamma_i)\mathbf{a}_i'\mathbf{C}_{n_i}^{-1}\mathbf{a}_i + \gamma_i \frac{\mathbf{a}_i'\mathbf{x}\mathbf{x}'\mathbf{a}_i}{\mathbf{x}'\mathbf{C}_{n_i}\mathbf{x}} \right] \right) \tag{12}$$

where the noise covariance at each constraint is $\sigma_{n_i}^2 \mathbf{C}_{n_i}$, and γ_i is related to the SNR for the i^{th} constraint. The negative log-likelihood function is then given by

$$\mathcal{L} = K + \sum_{i=1}^{L} \frac{\gamma_i}{2\sigma_{n_i}^2} \frac{\mathbf{x}'\mathbf{D}_i\mathbf{x}}{\mathbf{x}'\mathbf{C}_{n_i}\mathbf{x}} \tag{13}$$

where $\mathbf{D}_i = \mathbf{a}_i\mathbf{a}_i'$.

This log-likelihood function is closely related to the optimization criterion proposed in [5] for the non-intercept case. Indeed, one can show that they share the same minimum for high SNR cases (i.e., for $\gamma_i \simeq 1$ for all i). The main difference between the two formulations is the inclusion of prior information for the nuisance parameters, i.e., the variance of the true values. As a consequence this formulation should produce more accurate parameter estimates [8]. In addition, because this likelihood function is expressed in closed-form one can derive its gradient and Hessian. This facilitates the use of gradient-descent minimization procedures, and also helps one estimate confidence measures based on the Cramer-Rao lower bound.

3. MINIMIZATION OF $\mathcal{L}$

Given Eq. 13 we can derive the gradient and Hessian of $\mathcal{L}$ with respect to $\mathbf{x}$. This makes it possible to apply any of the standard methods for function minimization (gradient descent, conjugate gradient, Newton, etc.). In most cases these are guaranteed to converge to a local minima.

We developed a version of Newton's method for minimizing $\mathcal{L}$. It differs from the standard Newton minimization because of the structure of the negative log-likelihood function in the parameter space of $\mathbf{x}$. In particular, note from Eq. 13 that $\mathcal{L}$ is invariant to change in the magnitude of $\mathbf{x}$. It follows that $\mathcal{L}$ is constant in the radial direction, which of course is consistent with our use of a homogeneous coordinate system whose solution is a direction in the space of $\mathbf{x}$ (except for degenerate cases). This structure of the negative log-likelihood function causes the gradient to be orthogonal to the direction of $\mathbf{x}$, and the gradient modulus to be inversely proportional to $||\mathbf{x}||$, being 0 at the minimum direction. A gradient modulus decreasing in the radial direction means that the Hessian is non-positive definite. For this reason, we modify the standard Newton method by replacing the negative singular values of the Hessian by 0. In addition, at each iteration we re-normalize the updated value of $\mathbf{x}$ so that $||\mathbf{x}|| = 1$. This is not strictly necessary but has proven to improve numerically stability to the algorithm (so the solution does not increase or decrease without bound).

4. RESULTS

4.1. Monte-Carlo simulations

We have tested our modified Newton minimization of the negative log-likelihood function simulating estimation problems with 3 and 5 dimensional parameter vectors, and three different number of equations ($L = 10, 20, 50$) and average SNRs (0.5, 1 and 10). For each condition, we ran 400 trials.

For each trial, random values for $\mathbf{a}_0$ and $\mathbf{x}$ were generated. The measurements $\mathbf{a}$ were obtained adding Gaussian noise to $\mathbf{a}_0$, with different, random covariance for each equation. For each trial we obtained parameter estimates using our proposed HEIV Newton minimization method and using the method described by Matei and Meer [5]. For both we made comparisons with $\gamma_i = 1$ (the high SNR approximation) and with use of the true values of γ_i. The starting point for each of the iterative algorithms was the solution returned by the Total-Least-Squares (TLS) estimator.

We then measured the angular error between the estimates, $\hat{\mathbf{x}}$, and the true value $\mathbf{x}_0$, defined as $e = \arccos(\mathbf{x}_0'\hat{\mathbf{x}})$. We found that angular errors averaged over all conditions were similar for the two methods (0.54 for modified Newton and 0.55 for Matei and Meer). The slight difference in the results is due to a small fraction of trials on which the method of Matei and Meer failed to converge.

More interesting is the comparison between errors for the high SNR approximation with those obtained using the true values of γ_i. Fig. 1 compares averaged errors as a function of SNR, for estimates obtained using our modified Newton method in simulations with $L = 50$, and for 3D and 5D parameter vectors. As expected, the average error is always greater when using the high SNR approximation, and both errors tend to be equal when the approximation $\gamma_i \simeq 1$ holds (i.e., for high SNR). It is important to notice that the true values where generated using an isotropic prior, which is not in agreement with the assumption that the prior has a covariance proportional to that of the noise. Despite this fact, there is a decrease of about 5% when averaging the error for all conditions tested. This is consistent with previous findings [8].

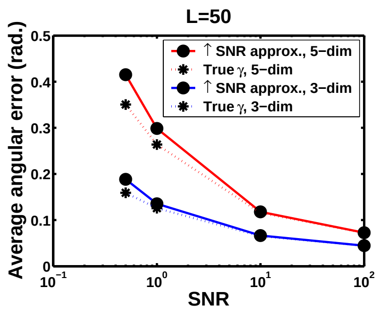

Fig. 1. Average angular errors (rad.) for $L = 50$ equations, as a function of SNR, using a high SNR approximation with $\gamma_i = 1$ (continuous line) or the true values of γ_i (dotted line), for 5- (red) and 3-dimensional (blue) parameter vector.

4.2. Affine motion estimation

We also tested the HEIV estimation method with the affine motion estimation problem described in Section 1. To illustrate the performance of the method in this particular task, we have used a noisy version of the *diverging tree* sequence [10], with a SNR close to 14 dB (see Fig. 2, left).

Gradients were computed using 5-tap derivative filters in space and time, making the noise variance similar for spatial and temporal derivatives. Taking Eq. 1 for every 5 pixels in the image (to enforce constraint independence) we estimated the parameters of the affine flow using LS and TLS estimators. Additionally, we used the covariances in Eq. 2 to estimate the parameters using the HEIV estimator. From the estimated parameters we computed the estimated optical flow, and its error with respect to the true velocity field. Fig. 2 shows the error (scaled to facilitate visualization) on the estimated flow for the TLS (center) and HEIV (right) estimators. Errors are significantly higher for the TLS estimator, which does not take into consideration the different covariance matrix of the noise at each location. The angular error of the optical flow [10] averaged over all pixels is 5.65 deg. for LS, 6.03 deg. for TLS, and 2.78 deg. for HEIV (error is higher than the one reported in [10] because of the added noise). Using the wrong noise model the TLS solution is often worse than the simple LS solution that assumes an even simpler noise model. By comparison, using the right noise model with the HEIV estimator produces significantly more accurate optical flow results.

5. CONCLUSIONS

We have derived simplified likelihood functions for HEIV problems, defined only on the parameters of interest. Under some conditions, these functions can be expressed in closed-form. This permits one to compute the gradient and Hessian of the negative log-likelihood which may play a significant role in the optimization. Here we have implemented a modified Newton method, showing more reliable convergence behavior than previous proposed algorithms in our simulations.

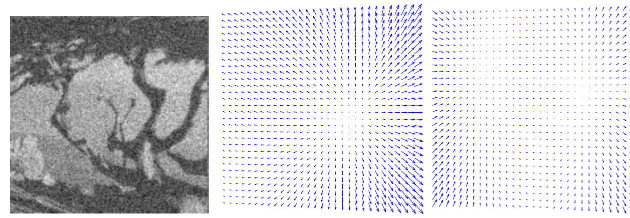

Fig. 2. Frame 20th of *diverging tree* after adding noise (left), and error of the estimated flow using TLS (center) and HEIV (right).

Deriving the likelihood function requires the introduction of a priori distribution over the true values underlying the noisy measurements. If such prior information is available then the estimates are more accurate than the corresponding MLE without prior information. This is consistent with the use of prior information reported by Gull [8].

We have presented an application to affine motion estimation where the noise affecting the measurements is non-IID, and therefore the accuracy of the estimated optical flow improves greatly when using the HEIV estimator.

We are currently working on the computation of confidence measures on the estimates based on Cramer-Rao lower bounds for the general case (for the isotropic case, see [7]), and on methods for estimating γ_i factors from the data, for situations when this information is unavailable.

6. REFERENCES

[1] Y. Kanazawa and K. Kanatani, “Optimal line fitting and reliability evaluation,” *IEICE Trans. Inf. & Syst.*, vol. E79-D, no. 9, pp. 1317–1322, 1996.

[2] B. K. Horn and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[3] R. Mester and M. Münlich, “Improving motion and orientation estimation using an equilibrated total least squares approach,” *Proc. IEEE ICIP*, Thessaloniki, 2001, pp. 929–932.

[4] A. D. Jepson and D. J. Heeger, “Linear subspace methods for recovering the translation direction,” in *Spatial Vision in Humans and Robots*, L Harris and M Jenkin, Eds., pp. 39–62. Cambridge University Press, New York, 1993.

[5] B. Matei and P. Meer, “A general method for errors-in-variables problems in computer vision,” *Proc. IEEE CVPR*, Hilton Head, 2000, pp. 18–25.

[6] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem*, SIAM, Philadelphia, 1991.

[7] O. Nestares, D. J. Fleet, and D. J. Heeger, “Likelihood functions and confidence bounds for total-least-squares problems,” *Proc. IEEE CVPR*, Hilton Head, 2000, pp. 523–530.

[8] S. F. Gull, “Bayesian data analysis: Straight-line fitting,” in *Maximum Entropy and Bayesian Methods*, J Skilling, Ed., pp. 511–518. Kluwer Academic Publishers, 1989.

[9] W. J. MacLean, “Removal of translation bias when using subspace methods,” *Proc. IEEE ICCV*, Corfu, 1999, pp. 753–758.

[10] J. Barron, D. Fleet, S. Beauchemin, and T. Burkitt, “Performance of optical flow techniques,” *Int. J. of Comp. Vis.*, vol. 12, pp. 43–77, 1994.