On Errors-In-Variables Regression with Arbitrary Covariance and its Application to Optical Flow Estimation

Björn Andres, Claudia Kondermann, Daniel Kondermann,
Ullrich Köthe, Fred A. Hamprecht, Christoph S. Garbe
Interdisciplinary Center for Scientific Computing, University of Heidelberg

Abstract

Linear inverse problems in computer vision, including motion estimation, shape fitting and image reconstruction, give rise to parameter estimation problems with highly correlated errors in variables. Established total least squares methods estimate the most likely corrections $\hat{A}$ and $\hat{b}$ to a given data matrix $[A, b]$ perturbed by additive Gaussian noise, such that there exists a solution y with $[A + \hat{A}, b + \hat{b}]y = 0$. In practice, regression imposes a more restrictive constraint namely the existence of a solution x with $[A + \hat{A}]x = [b + \hat{b}]$. In addition, more complicated correlations arise canonically from the use of linear filters. We, therefore, propose a maximum likelihood estimator for regression in the general case of arbitrary positive definite covariance matrices. We show that $\hat{A}, \hat{b}$ and x can be found simultaneously by the unconstrained minimization of a multivariate polynomial which can, in principle, be carried out by means of a Gröbner basis. Results for plane fitting and optical flow computation indicate the superiority of the proposed method.

1. Introduction

Linear inverse problems in computer vision, including motion estimation, shape fitting and image reconstruction, give rise to parameter estimation problems with highly correlated errors in variables. In this section, we introduce the statistically appropriate model for this context. We discuss related work in section 2 and introduce a new estimator allowing for arbitrary correlations in the data in section 3. Comparative results for plane fitting as well as for optical flow estimation are given in section 4, and conclusions are offered in section 5.

1.1. Errors-In-Variables Model

Given noisy variables $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, m, n \in \mathbb{N}$ as well as a symmetric positive definite matrix $\Sigma \in \mathbb{R}^{(mn+m)\times(mn+m)}$ modeling the covariance of the entries of A and b, the *linear additive Gaussian errors-in-variables (EIV) model* (A, b, Σ) is specified by the following assumptions:

(a) There exist latent variables $A_l \in \mathbb{R}^{m\times n}, b_l \in \mathbb{R}^m$ and additive errors in variables $A_e \in \mathbb{R}^{m\times n}, b_e \in \mathbb{R}^m$ such that

$$A = A_l + A_e \text{ and } b = b_l + b_e \ . \tag{1}$$

(b) Let $\mathrm{vec}([A_e, b_e])$ denote the column-wise vectorization of the composite matrix $[A_e, b_e]$. Then the errors A_e, b_e are realizations of a random matrix A'_e and a random vector b'_e whose entries are normally distributed with zero mean and covariance matrix Σ

$$\mathrm{vec}([A'_e, b'_e]) \sim \mathcal{N}(0, \Sigma) \ , \text{ i.e.} \tag{2}$$

$$P(\mathrm{vec}([A_e, b_e])) = \frac{\exp\left(-\frac{1}{2}\left\|\mathrm{vec}([A_e, b_e])\right\|_\Sigma^2\right)}{\sqrt{(2\pi)^{mn+m}\det\Sigma}} \ . \tag{3}$$

with $\|\cdot\|_\Sigma : \mathbb{R}^{mn+m} \to \mathbb{R}_0^+$ such that

$$\forall v \in \mathbb{R}^{mn+m} : \quad \|v\|_\Sigma := \sqrt{v^T\Sigma^{-1}v} \ , \tag{4}$$

which denotes the Mahalanobis norm for given Σ.

(c) The latent vector b_l linearly depends on the columns of A_l, i.e.

$$\exists x \in \mathbb{R}^n : \quad A_l x = b_l \ , \tag{5}$$

making this system of equations solvable.

Under the assumption of the above EIV model, the *maximum likelihood estimates* $\hat{A}_e, \hat{b}_e$ of A'_e, b'_e are the most likely (w.r.t. eq. (3)) errors such that conditions (a) and (c) are fulfilled. The solution $\hat{x}$ to the linear system (5) then follows from $(A - \hat{A}_e)\hat{x} = b - \hat{b}_e$ which is solvable due to (a) and (c). For an observed A,b the maximum likelihood estimation (MLE) of A_l and b_l hence reduces to the optimization problem

$$\underset{A_e\in\mathbb{R}^{m\times n}, b_e\in\mathbb{R}^m}{\operatorname{argmax}} \ P(\mathrm{vec}([A_e, b_e])) \tag{6}$$

$$\text{subject to} \quad \exists x \in \mathbb{R}^n : (A - A_e)x = b - b_e \ . \tag{7}$$

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

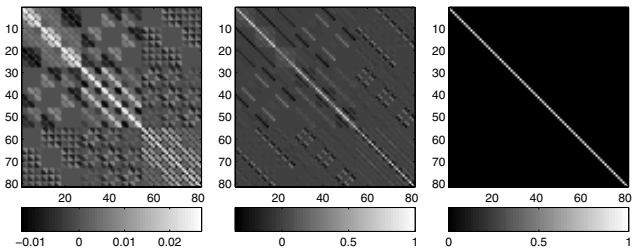

Figure 1. Locally constant optical flow estimation from $3 \times 3 \times 3$ neighborhoods is considered. Left: Covariance matrix $\Sigma \in \mathbb{R}^{81\times 81}$ of the vectors $\mathrm{vec}([\partial_x g, \partial_y g, \partial_t g])$, with structure owing to impulse response masks of linear derivative operators. Middle: Covariance matrix of ETLS equilibration $\mathrm{vec}(W_L[\partial_x g, \partial_y g, \partial_t g]W_R^T)$. Right: Unity matrix.

Due to the strict monotonicity of the exponential function in (3), the objective in (6) can be replaced by

$$\underset{A_e\in\mathbb{R}^{m\times n}, b_e\in\mathbb{R}^m}{\operatorname{argmin}} \|\mathrm{vec}([A_e, b_e])\|_\Sigma \ . \tag{8}$$

An in-depth discussion of EIV models can be found in [12, 15].

2. Related Work

This section addresses methods that solve related problems and differ from the estimator proposed in section 3 in either the objective function or the constraint. Given noisy variables $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, the *total least squares (TLS) estimator* [16] is defined as

$$\underset{A_e\in\mathbb{R}^{m\times n}, b_e\in\mathbb{R}^m}{\operatorname{argmin}} \|\mathrm{vec}([A_e, b_e])\|_{I_{mn+m}} \tag{9}$$

subject to

$$\exists y \in \mathbb{R}^{n+1} : [A - A_e, b - b_e]y = 0 \text{ and } \|y\|_2 = 1 \ , \tag{10}$$

where I_{mn+m} denotes the identity matrix of the indicated size. In fact, the Mahalanobis norm $\|\mathrm{vec}([A_e, b_e])\|_{I_{mn+m}}$ equals the Frobenius norm of the matrix $[A_e, b_e]$. Thus, TLS seeks the Frobenius minimal additive correction $[A_e, b_e]$ to $[A, b]$ such that the matrix $[A - A_e, b - b_e]$ becomes rank deficient. If $[A, b]$ results from perturbance of $[A_l, b_l]$ by independent and identically distributed Gaussian noise with zero mean, then TLS is the MLE of $[A_e, b_e]$ under the rank deficiency constraint [16]. The estimator we propose differs from TLS in two aspects: First, according to (2), we allow for arbitrary non-singular correlations in the errors. Second, we impose the constraint (5) instead of (10) stating that precisely the vector $b - b_e$ is supposed to be a linear combination of the columns of $A - A_e$. This corresponds to a restriction of the TLS formulation, where rank deficiency only requires the columns of $[A - A_e, b - b_e]$ to be linearly dependent. We thereby directly address linear regression problems, whereas TLS is in fact a subspace estimator.

Subspace estimation with arbitrary covariance is comprehensively discussed in [6, 11]. Under the restrictive assumption that the columns of the matrix $[A, b]$ are uncorrelated, subspace estimation can be reduced to TLS by weighting as detailed in [12, 17]. Our approach differs in that we consider regression rather than subspace estimation and in that we directly address polynomial minimization.

From the wide range of subspace estimators, we compare directly against *equilibrated TLS (ETLS)* which was introduced in [13] and has been applied in computer vision, e.g. in [1]. ETLS aims at linearly transforming the system $[A, b]$ such that the errors become approximately uncorrelated with equal variance. To this end, orthogonal matrices $W_L \in \mathbb{R}^{m\times m}$, $W_R \in \mathbb{R}^{(n+1)\times(n+1)}$ are employed, such that

$$[A', b'] := W_L[A, b]W_R^T \ . \tag{11}$$

Then TLS is performed on $[A', b']$ (yielding a solution $y' \in \mathbb{R}^{n+1}$), and finally $y := W_R y'$ is understood as a solution to the initial problem. Mühlich and Mester [13] derive by the perturbation theory of eigenvectors [8] conditions on W_L and W_R, from which these matrices can be computed iteratively for arbitrary covariance. The benefit of mapping such estimation problems with arbitrary covariance to TLS is that the Eckart-Young Theorem [7] for matrix approximation allows for efficient and numerically stable TLS estimation by means of the singular value decomposition (SVD) [16]. However, ETLS is only an approximation to true MLE in case of highly structured covariance matrices. In order for ETLS to be the MLE, the covariance matrix of $\mathrm{vec}(W_L[A, b]W_R^T)$ has to equal the identity matrix. However,

$$\begin{aligned} &\mathrm{cov}\left(\mathrm{vec}(W_L[A, b]W_R^T)\right) \\ = \ &(W_R \otimes W_L)\mathrm{cov}(\mathrm{vec}([A, b]))(W_R \otimes W_L)^T \end{aligned} \tag{12}$$

where $\otimes$ denotes the Kronecker product. Due to the structure of the matrix $(W_R \otimes W_L)$, covariance matrices $\mathrm{cov}(\mathrm{vec}([A, b]))$ exist such that W_L and W_R cannot be chosen to equilibrate them. An example of remaining correlation after ETLS equilibration is shown in figure 1. Nevertheless, the results reported below support ETLS as an efficient approximation to MLE.

3. Maximum Likelihood Estimator

In order to account for arbitrary covariance in the regression setting defined by the constraint (7), we propose to consider the objective function (8) and substitute for b_e according to (7) yielding

$$\underset{d\in\mathbb{R}^{mn+n}}{\operatorname{argmin}} \ p(d) \ , \tag{13}$$

with $p : \mathbb{R}^{mn+n} \rightarrow \mathbb{R}_0^+$ such that $\forall A_e \in \mathbb{R}^{m \times n} \forall x \in \mathbb{R}^n$:

$$p(\mathrm{vec}([A_e, x])) = \|\mathrm{vec}([A_e, b - (A - A_e)x])\|_\Sigma \quad . \quad (14)$$

The objective function p is a *multivariate polynomial* in the unconstrained entries of A_e and x. The largest exponent of any of the free variables is two while (mixed) terms of order one through four occur. Depending on A, b, and Σ, this polynomial may be non-convex. However, there exists a finite vector $\hat{d} \in \mathbb{R}^{mn+n}$ for which $p(\hat{d})$ equals the finite global minimum of this polynomial. This property is inherited from the Mahalanobis norm. In the following, we will use the abbreviation MR for the proposed maximum likelihood estimator, which stands for Mahalanobis regression.

3.1. Polynomial Minimization

We have reduced the problem of finding maximum likelihood estimates for the corrections $[A_e, b_e]$ to the minimization of a multivariate polynomial p. We now describe a numerical minimizer for p. Furthermore, we will show how Gröbner bases can be used to reduce the problem of multivariate polynomial minimization to univariate polynomial root finding.

3.1.1 Numerical Minimization

The multivariate polynomial to be minimized can easily be differentiated algebraically. Hence, the first two terms of the Taylor approximation of p, namely, the gradient and the Hessian, are readily available. This suggests the use of a trust region method for optimization [5]. We therefore apply the large-scale optimization routine of Matlab which employs a subspace trust region method based on the interior-reflective Newton method described in [4]. This method first estimates a two-dimensional subspace based on the gradient and the Hessian matrix. The latter is inverted by means of preconditioned conjugate gradients. Then, the resulting minimization problem is solved by a Newton step. A host of alternative minimization techniques exists. Since the second partial derivatives $\frac{\partial^2}{\partial x_j^2} p$ are linear in x_j, their roots can be evaluated efficiently. Thus, especially the cyclic coordinate descent method, which iteratively minimizes the polynomial with respect to each variable in turn, appears interesting.

3.1.2 Algebraic Minimization

In this section, we propose an algebraic approach to the minimization of the polynomial. Due to complexity, we have tested this approach only for small examples not included in this paper. We nevertheless point out the importance of this approach, which is due to the following reasons. First, the iterative numerical minimization of the multivariate polynomial is computationally expensive and prone to convergence to local minima if p is non-convex. Second, depending on the application, p may need to be minimized multiple times for equal covariance Σ but different data A,b such as is the case with optical flow estimation from patches of an image sequence. Third, Σ may arise from only a small number of coefficients induced by linear filters.

These three reasons taken together motivate the use of as much symbolic computation as possible in order to support the minimization process. Hence, in order to minimize the polynomial p over the ring of polynomials $R[x_1, ..., x_n]$ in $n \in \mathbb{N}$ indeterminates, we differentiate the polynomial algebraically with respect to all indeterminates and obtain the following system of equations

$$S = \{p_1(x_1, ..., x_n) = 0, ..., p_n(x_1, ..., x_n) = 0\} \quad . \quad (15)$$

This system can be solved by means of a Gröbner basis [3]. Let $I = \langle p_1, ..., p_n \rangle, n \in \mathbb{N}$ denote the ideal generated by the polynomials $p_1, ..., p_n$. A Gröbner basis is a special basis of I for which root-finding is simpler than for the original system of equations. This is due to the elimination property of Gröbner bases, from which follows that for $j \in \mathbb{N}_n$ a basis polynomial in only j indeterminates exists. This means that one can successively find all the solutions of the system of equations by means of root-finding of univariate polynomials.

The computation of a Gröbner basis can be carried out using Buchberger's algorithm [3]. However, since this algorithm has exponential complexity in the worst case, it may be impractical even for relatively small problems if there is little structure in the covariance matrix. However, despite this complexity one should consider the computation of Gröbner bases as a preprocessing step due to the observations mentioned above.

4. Results

In order to investigate the performance of the proposed estimator, we have applied the trust region method described above for the minimization of the multivariate polynomial (with ETLS initialization), first to the problem of fitting planes to point clouds with randomly and systematically correlated Gaussian noise (sections 4.1 4.2), and second to optical flow estimation where the covariance matrix is derived from the derivative filters applied to the image sequence (section 4.3).

4.1. Random Correlation

The plane fitting problem in 3d-space amounts to solving an over-determined linear system of equations $Ax = b, A \in \mathbb{R}^{m \times 2}, b \in \mathbb{R}^m, x \in \mathbb{R}^2$ approximately. In order to obtain test examples, we randomly choose the matrix A_l and the parameter vector x and compute the vector

$b_l = A_l x$. To generate correlated additive Gaussian noise, we first choose a matrix $L : \mathbb{R}^{3m\times 3m} \to \mathbb{R}$ with uniformly distributed entries within the range $[0, 1]$. Then, to perturb the latent data $[A_l, b_l]$, we generate uncorrelated Gaussian noise $D \in \mathbb{R}^{m\times 3}$ with given standard deviation $\sigma \in \mathbb{R}$ and compute the noisy variables by

$$\mathrm{vec}([A, b]) = \mathrm{vec}([A_l, b_l]) + L\ \mathrm{vec}(D) \qquad (16)$$

Then, the covariance matrix of the Gaussian distribution of $\mathrm{vec}([A, b])$ follows from $\Sigma = \sigma^2 LL^T$.

The error for each of these plane fitting problems is computed as the norm of the difference vector between the estimated solution $\hat{x}$ and the true solution x.

$$e := \|\hat{x} - x\|_2 \qquad (17)$$

To minimize the polynomial we employ the algorithm described in section 3.1.1. As initialization we use the result of the ETLS estimator.

To test our algorithm for the plane fitting problem we consider eight different signal to noise ratios (SNR) between 100 and 0.5. For each SNR we solve 100 randomly generated plane fitting problems (as described above) for $m = 8$ points. The error is computed according to (17) for each plane fitting problem instance. To evaluate our method we investigate the distribution of the errors for each SNR separately. In Figure 2 we compare the results of our algorithm to those obtained by TLS and ETLS. The results indicate that for every noise level the quality of the proposed MR method is far superior to the TLS solution and superior to the ETLS solution.

4.2. Systematic Correlation

We consider the same setting as in 4.1 except for the fact that we introduce stronger systematic correlation between the points by setting some entries of the mapping L manually. Figure 3 shows an example for a linear mapping, which induces systematic correlations of the noise, the corresponding covariance matrix and error box plots for the TLS, ETLS and MR method for a SNR of 10. The MR method again clearly outperforms the TLS method and is superior to the ETLS method due to its lower quantiles. The results indicate the superiority of the MR method for systematically correlated Gaussian noise.

4.3. Optical Flow Estimation

Let $N \in \mathbb{N}$ and $g \in \mathbb{R}^N$ be a vector of all gray values in a sequence of images such that any $j \in \{1, \ldots, N\}$ uniquely indicates a pixel in the sequence. Moreover, consider linear shift invariant filters with finite impulse response which approximate partial derivatives of the signal in the two spatial as well as in the temporal directions [14]. Then, the approximate derivatives $\partial_x g, \partial_y g, \partial_t g \in \mathbb{R}^m$ of a spatio-temporal

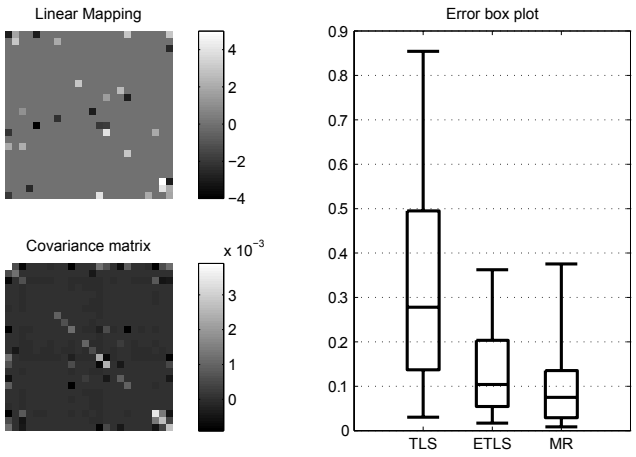

Figure 3. Results for a randomly chosen plane fitting problem in 3d-space with partially strongly correlated noise with SNR=10, left: linear mapping L for noise correlation and resulting covariance matrix C, right: error box plot for TLS, ETLS and MR

patch of size m of the sequence are obtained by matrices $L_x, L_y, L_t \in \mathbb{R}^{m\times N}$, defined by the filter coefficients, according to

$$\partial_x g = L_x g, \quad \partial_y g = L_y g, \quad \partial_t g = L_t g\ . \qquad (18)$$

Suppose now the existence of latent gray values $g_l \in \mathbb{R}^N$ and errors $g_e \in \mathbb{R}^N$ which are realizations of random variables distributed according to $\mathcal{N}(0, \Sigma_g)$, such that $g = g_l + g_e$. The covariance matrix Σ_g may, for instance, be obtained from camera calibration measurements. Then, the brightness change constraint equation (BCCE) for locally constant displacement [10, 2] imposes the existence of optical flow $f_x, f_y \in \mathbb{R}$ which is constant with respect to the chosen patch, i.e.

$$\underbrace{\begin{pmatrix} (\partial_x g_l)_1 & (\partial_y g_l)_1 \\ \vdots & \vdots \\ (\partial_x g_l)_k & (\partial_y g_l)_k \end{pmatrix}}_{A_l} \begin{pmatrix} f_x \\ f_y \end{pmatrix} = -\underbrace{\begin{pmatrix} (\partial_t g_l)_1 \\ \vdots \\ (\partial_t g_l)_k \end{pmatrix}}_{b_l}\ . \qquad (19)$$

Note that $A, A_e \in \mathbb{R}^{m\times 2}$ and $b, b_e \in \mathbb{R}^m$ can be defined by replacing g_l by g and g_e, respectively, in eq. (19). Statistically appropriate motion estimation now means solving the EIV problem (A, b, Σ) with

$$\begin{aligned} \Sigma &= \mathrm{cov}(\mathrm{vec}([A, b])) \\ &= \begin{pmatrix} L_x\Sigma_g L_x^T & L_x\Sigma_g L_y^T & L_x\Sigma_g L_t^T \\ L_y\Sigma_g L_x^T & L_y\Sigma_g L_y^T & L_y\Sigma_g L_t^T \\ L_t\Sigma_g L_x^T & L_t\Sigma_g L_y^T & L_t\Sigma_g L_t^T \end{pmatrix} . \end{aligned} \qquad (20)$$

A covariance matrix for optical flow estimation from $3 \times 3 \times 3$ neighborhoods is shown in the first graph of Figure 1.

Since the linear filters induce correlated Gaussian noise with the non-block-diagonal covariance matrix (20) in the

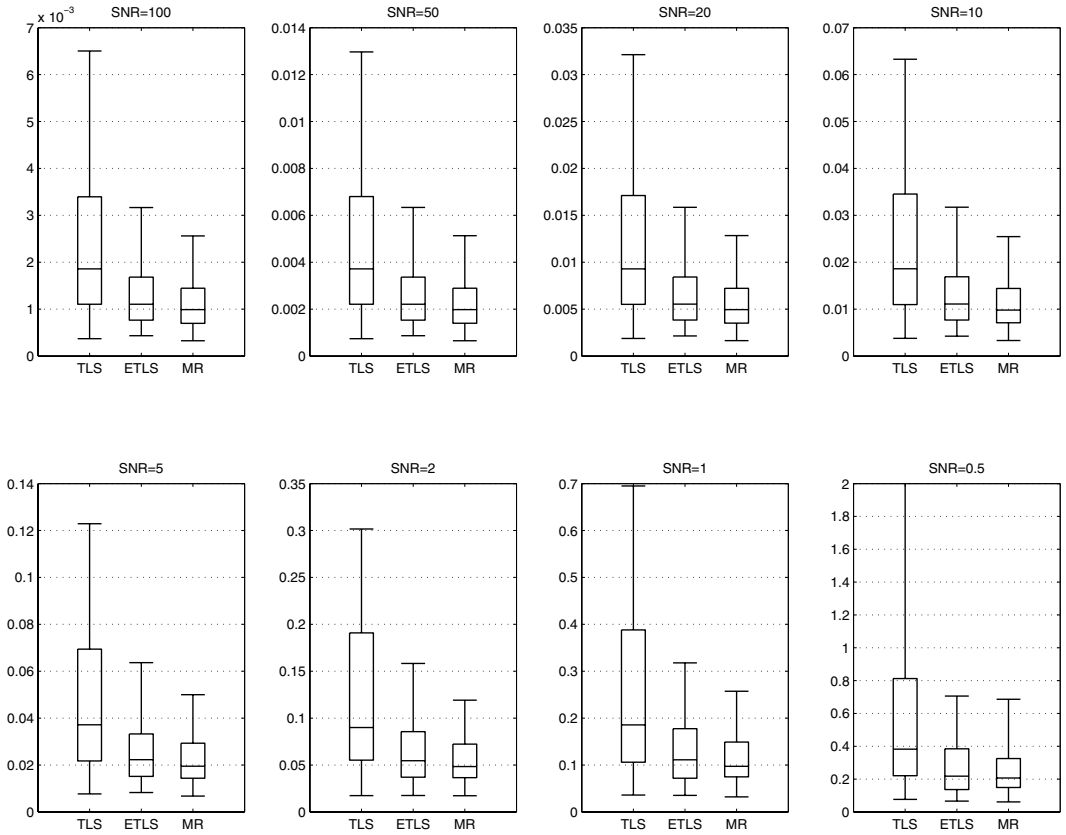

Figure 2. Error box plots of results for TLS, ETLS and MR for a randomly chosen plane fitting problem in 3d-space with randomly correlated noise for different SNRs

estimated gray value derivatives, it stands to reason to employ the MR method proposed here to obtain a maximum likelihood estimate. To examine the performance of the proposed estimator we minimize the influence from other sources of errors. In order to avoid ambiguous solutions due to the aperture problem [9], to minimize interpolation errors, and to avoid optical flow model violations, we have generated an artificial image sequence with structure on multiple scales with constant motion in the direction $[-1, -1]$ for an intensity range $[0, 1]$ as depicted in Figure 4.

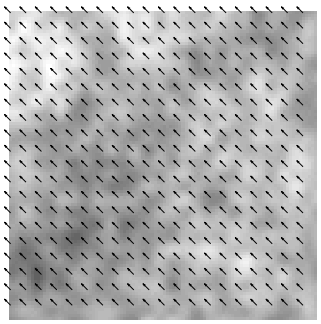

Figure 4. Randomly generated test sequence for optical flow estimation with constant motion [-1,-1]

Derivatives are computed by the isotropy optimized Scharr filters [14] of spatio-temporal size $3 \times 3 \times 3$. The size of the patches for which constant motion is assumed is set to $3 \times 3 \times 1$. The resulting covariance matrix is given by equation (20). The error of the solution is computed as the length of the difference vector between ground truth and calculated optical flow vector as indicated in (17). The sequence is perturbed by additive Gaussian noise of standard deviation 0.0005, and the optical flow is computed for 100 pixels in this sequence. Figure 5 shows the cumulative distribution functions and histograms of the errors for the TLS, the ETLS and the MR method. The results suggest that the ETLS and MR method yield lower optical flow errors than the TLS method, which is due to the violated assumption of uncorrelated noise made by TLS. Comparing the performance of the ETLS and MR method we come to the conclusion that the MR method yields slightly better results than ETLS. The reason why the results are only slightly better lies in the fact that the covariance matrix generated by the linear derivative filters is singular due to linear dependencies between filter results. Hence, we have used the Moore-Penrose pseudo inverse to compute the Mahalanobis norm.

Thus, we can conclude that for the application of plane fitting the proposed MR method yields better results than

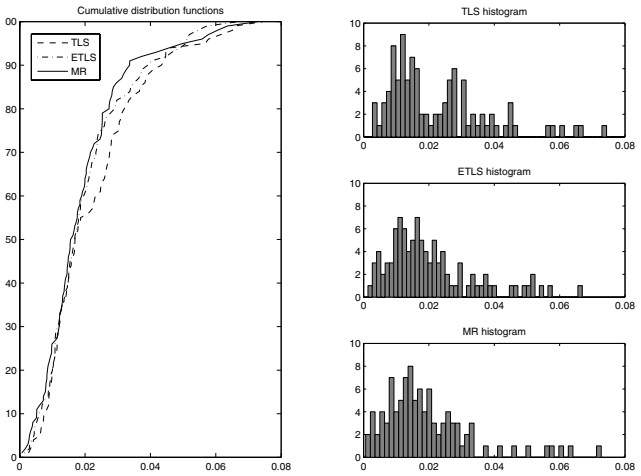

Figure 5. Cumulative distribution functions and histograms for TLS, ETLS and MR methods for optical flow error

the TLS and ETLS method. In case of optical flow estimation the results are superior to the TLS method, but only slightly superior to the ETLS method due to the singularity of the covariance matrices in optical flow estimation.

5. Conclusion and Perspectives

We have proposed a maximum likelihood estimator for linear EIV models with additive Gaussian noise of zero mean and arbitrary positive-definite covariance. Previous approaches differ in that they either consider subspace estimation or impose restrictive constraints on the covariance matrix of the multivariate normal distribution. The estimator we propose requires the minimization of a non-convex multivariate polynomial which we address by means of two iterative numerical optimizers, a Newton method and minimization by coordinate descent. We identify Gröbner bases as a tool to in principle reduce the problem algebraically to finding roots of a sequence of univariate polynomials. Towards application, we have demonstrated that the proposed estimator clearly outperforms TLS and ETLS in a regression problem, where planes are fitted to clouds of points jittered by randomly correlated noise. In addition, we have shown by the example of optical flow estimation that covariance matrices, for which the proposed estimator could prove beneficial, arise canonically in computer vision from the use of linear filters. In the case of positive semi-definite covariance matrices it is insufficient to replace Σ^{-1} by the Moore-Penrose pseudo-inverse of Σ, so a detailed analysis of this setting is subject to future research. Moreover, we expect the estimator to be particularly useful in regression residual analysis (e.g. [1]), where not only the solution x to the regression problem, but also the corrections to the data are considered.

References

[1] B. Andres, F. A. Hamprecht, and C. S. Garbe. Selection of local optical flow models by means of residual analysis. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *LNCS*, pages 72–81. Springer, 2007.

[2] M. J. Black and A. Jepson. Estimating multiple independent motions in segmented images using parametric models with local deformations. In *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 220–227, 1994.

[3] B. Buchberger. *An Algorithm for Finding the Bases Elements of the Residue Class Ring Modulo a Zero Dimensional Polynomial Ideal*. PhD thesis, University of Innsbruck, 1965.

[4] T. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6:418–445, 1996.

[5] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust-region methods*. SIAM Philadelphia, 2000.

[6] B. De Moor. Total least squares for affinely structured matrices and the noisy realization problem. *IEEE TSP*, 42(11):3104–3113, Nov 1994.

[7] C. Eckhart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrica*, 1:211–218, 1936.

[8] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore and London, 3 edition, 1996.

[9] H. Haussecker and H. Spies. Motion. In B. Jähne, H. Haussecker, and P. Geissler, editors, *Handbook of Computer Vision and Applications*, volume 2, chapter 13. Academic Press, 1999.

[10] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *DARPA Image Understanding Workshop*, pages 121–130, 1981.

[11] J. H. Manton, R. Mahony, and H. Yingbo. The geometry of weighted low rank approximations. *IEEE TSP*, 51(2):500–514, February 2003.

[12] I. Markovsky, J. C. Willems, B. De Moor, and S. Van Huffel. *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*. Number 11 in Monographs on Mathematical Modeling and Computation. SIAM, March 2006.

[13] M. Mühlich. *Estimation in Projective Spaces and Application in Computer Vision*. PhD thesis, Johann Wolfgang Goethe Universität Frankfurt am Main, 2005.

[14] H. Scharr. Optimal filters for extended optical flow. In *Complex Motion, Lect. Notes in Comp. Sci.*, volume 3417. Springer, 2004.

[15] S. Van Huffel and P. Lemmerling, editors. *Total Least Squares and Errors-in-Variables Modeling*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[16] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM, Philadelphia, 1991.

[17] P. D. Wentzell, D. T. Andrews, D. C. Hamilton, K. Faber, and B. R. Kowalski. Maximum likelihood principal component analysis. *Journal of Chemometrics*, 11:339–366, 1997.