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 (IWR), University of Heidelberg, Germany

Introduction

Linear inverse problems in computer vision, including shape fitting and motion estimation, 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 non-zero 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 $[A+\hat{A}]x = [b+\hat{b}]$.
- complicated **correlations** arise canonically from the use of linear filters.

We, therefore, propose a maximum likelihood (ML) estimator for regression in the **more general case of arbitrary positive definite covariance**.

Errors-In-Variables (EIV) Model

Given $m, n \in \mathbb{N}$, **noisy variables**

$$A \in \mathbb{R}^{m\times n}, \qquad b \in \mathbb{R}^m \ , \tag{1}$$

and a symmetric positive definite matrix

$$\Sigma \in \mathbb{R}^{(mn+m)\times(mn+m)} \tag{2}$$

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:

1. 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

$$A = A_l + A_e \text{ and } b = b_l + b_e \ . \tag{3}$$

2. 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

$$\mathrm{vec}([A'_e, b'_e]) \sim \mathcal{N}(0, \Sigma) \ , \text{ i.e.} \tag{4}$$

$$P(\mathrm{vec}([A'_e, b'_e])) = \frac{\exp\left(-\frac{1}{2}\|\mathrm{vec}([A'_e, b'_e])\|^2_\Sigma\right)}{\sqrt{(2\pi)^{mn+m}\det\Sigma}} \ . \tag{5}$$

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

$$\forall v \in \mathbb{R}^{mn+m} : \quad \|v\|_\Sigma := \sqrt{v^T\Sigma^{-1}v} \ , \tag{6}$$

which denotes the Mahalanobis norm for given Σ.

3. The latent vector b_l linearly depends on the columns of A_l, i.e.

$$\exists x \in \mathbb{R}^n : \quad A_l x = b_l \ , \tag{7}$$

making this system of equations solvable.

Equilibrated Total Least Squares (ETLS)

From the wide range of subspace estimators, we compare directly against ETLS [7]. In ETLS, the data is equilibrated, $[A', b'] := W_L[A, b]W_R^T$, before 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. ETLS equals MLE iff $\mathrm{cov}(\mathrm{vec}(W_L[A, b]W_R^T)) = \mathbb{1}$. However,

$$\mathrm{cov}\left(\mathrm{vec}(W_L[A, b]W_R^T)\right) = (W_R \otimes W_L)\mathrm{cov}(\mathrm{vec}([A, b]))(W_R \otimes W_L)^T \ ,$$

and due to the structure of $(W_R \otimes W_L)$, covariance matrices $\mathrm{cov}(\mathrm{vec}([A, b]))$ exist such that W_L and W_R cannot be chosen to equilibrate them.

Maximum-Likelihood Estimation (MLE)

Maximum likelihood estimates $\hat{A}_e, \hat{b}_e$ of A'_e, b'_e are the most likely (w.r.t. (5)) errors satisfying (3) and (7). The solution $\hat{x}$ then follows from $(A-\hat{A}_e)\hat{x} = b-\hat{b}_e$. For an observed A,b, maximum likelihood estimation of A_l and b_l hence reduces to the **optimization problem**

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

$$\text{subject to} \quad \exists x \in \mathbb{R}^n : (A - A_e)x = b - b_e \ . \tag{9}$$

Substituting for b_e (9) in the objective function (8) yields equivalently

$$\underset{d\in\mathbb{R}^{mn+n}}{\mathrm{argmin}}\, p(d) \ , \tag{10}$$

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

$$p(\mathrm{vec}([A_e, x])) = \|\mathrm{vec}([A_e, b - (A - A_e)x])\|_\Sigma \ . \tag{11}$$

Now, 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, p may be non-convex.

Algorithm

We use two maps as a data structure to represent the multivariate polynomial p.

- For each **monomial**, all variable indices are mapped to their powers
- Monomials are mapped to their **leading coefficients**

Thereby, the first and second derivatives of p can be computed algebraically. The second partial derivatives of p are linear forms. This suggests the use of one of the following numerical optimizers.

- Newton optimization with **subspace trust region**
- **Cyclic coordinate descent**, iteratively minimizing the polynomial with respect to a single variable, globally

In principle, the minimization of p can be reduced to root-finding of univariate polynomials by the construction of a **Gröbner basis**. However, Buchberger's algorithm [2] exhibits exponential runtime in the worst case, so this approach is suitable only for very small problems with highly structured covariance.

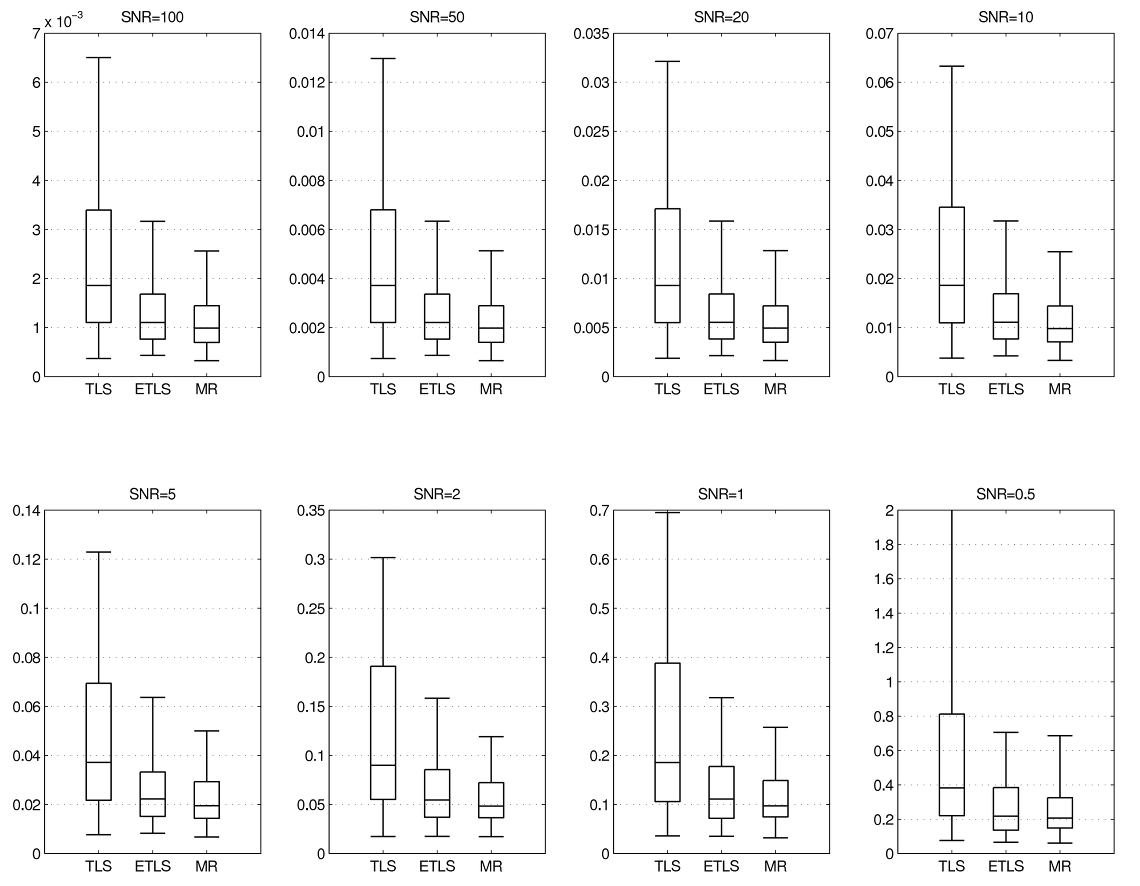
Plane Fitting

The plane fitting problem amounts to solving an over-determined linear system $Ax = b, A \in \mathbb{R}^{m\times 2}, b \in \mathbb{R}^m, x \in \mathbb{R}^2$ approximately. As **ground truth**, we draw a matrix A_l and a vector x and let $b_l = A_l x$. A_l and b_l are then perturbed,

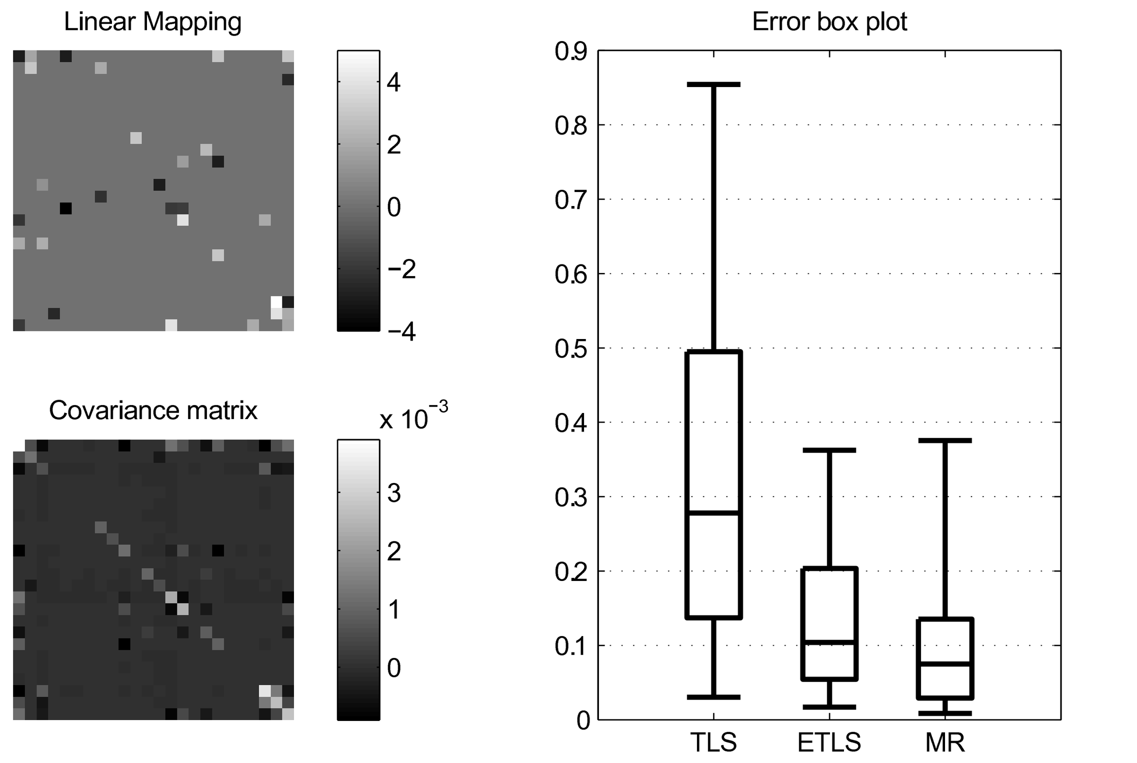
$$\mathrm{vec}([A, b]) = \mathrm{vec}([A_l, b_l]) + L\ \mathrm{vec}(D) \ , \tag{12}$$

with $\mathrm{vec}(D) \sim \mathcal{N}(0, \sigma^2)$ and $L \in \mathbb{R}^{3m\times 3m}$ such that $\Sigma = \sigma^2 LL^T$. Graphs below depict the distributions of the Euclidean distance between x and estimates $\hat{x}$, accumulated over 100 random plane fitting problems.

Random Correlation



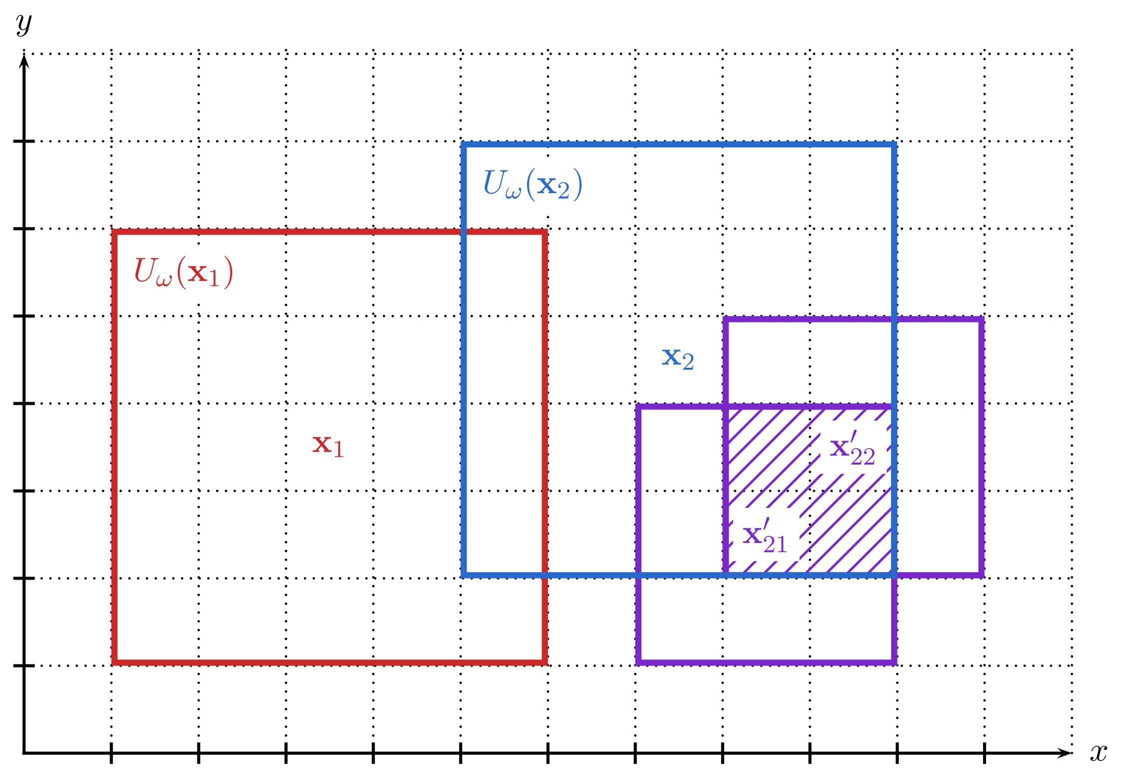
Structured Correlation



Local 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. Moreover, consider linear shift invariant filters with finite impulse response which approximate partial derivatives,

$$\partial_x g = L_x g, \quad \partial_y g = L_y g, \quad \partial_t g = L_t g \ . \tag{13}$$

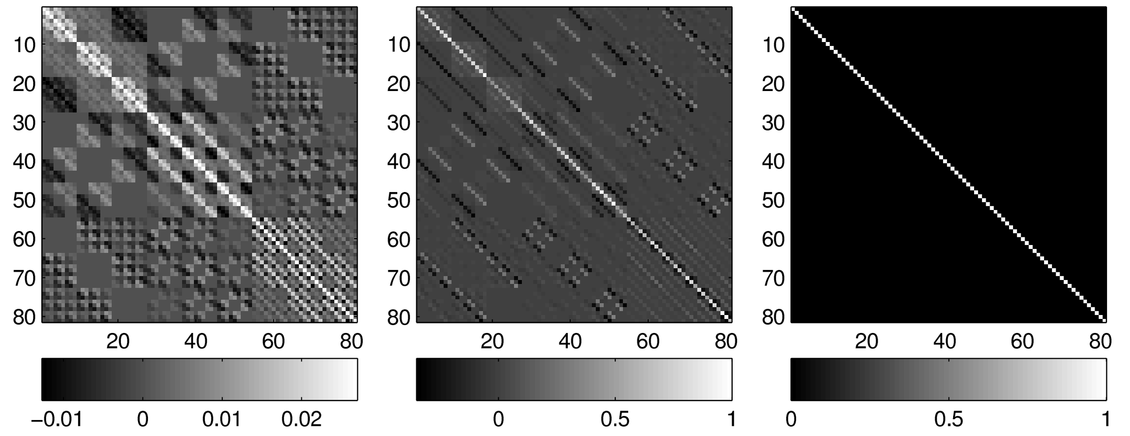


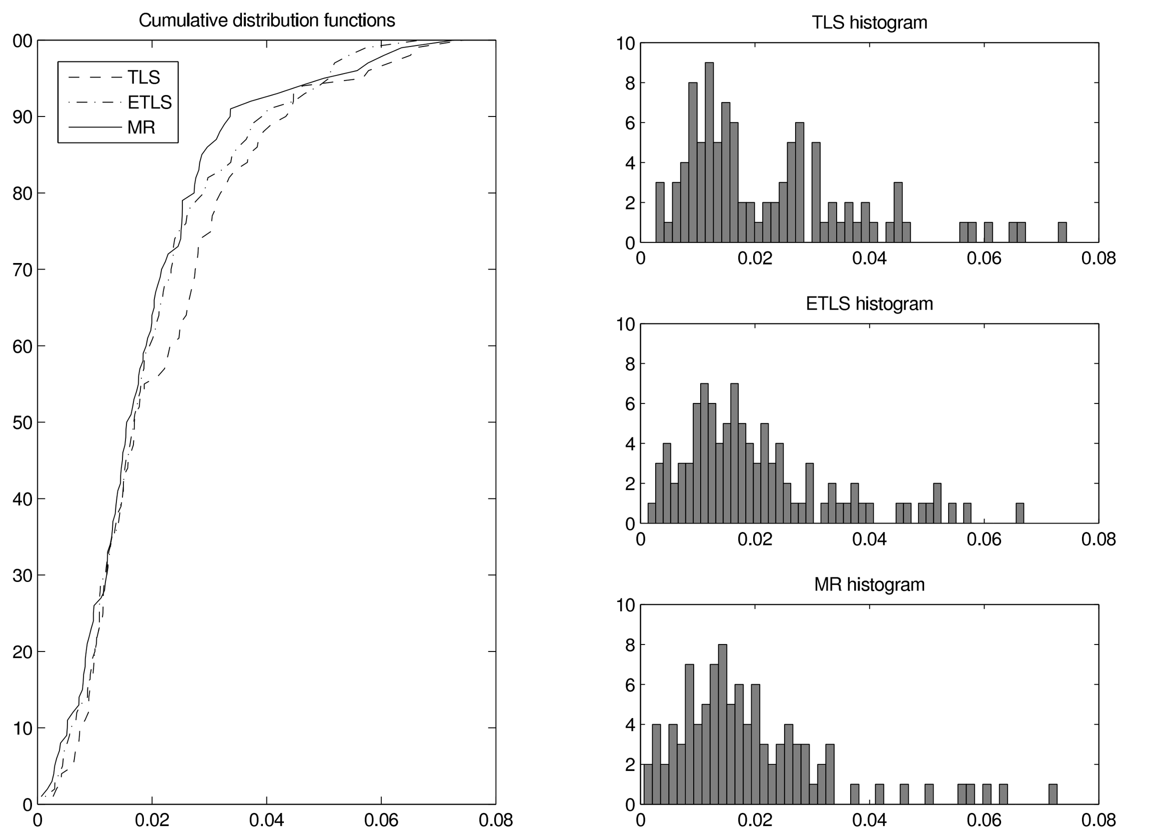
Suppose now the existence of latent gray values $g_l \in \mathbb{R}^N$ and errors $g_e \in \mathbb{R}^N$ which are realizations of random variables distributed according to $\mathcal{N}(0, \Sigma_g)$, such that $g = g_l + g_e$. Then, the brightness change constraint equation (BCCE) for locally constant displacement [4, 1] imposes the existence of optical flow $f_x, f_y \in \mathbb{R}$ which is constant with respect to the chosen patch, i.e.

$$\underbrace{\begin{pmatrix} (\partial_x g_l)_1 & (\partial_y g_l)_1 \\ \vdots & \vdots \\ (\partial_x g_l)_k & (\partial_y g_l)_k \end{pmatrix}}_{A_l} \begin{pmatrix} f_x \\ f_y \end{pmatrix} = -\underbrace{\begin{pmatrix} (\partial_t g_l)_1 \\ \vdots \\ (\partial_t g_l)_k \end{pmatrix}}_{b_l} \ . \tag{14}$$

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. (14). Statistically appropriate motion estimation now means solving the EIV problem (A, b, Σ) with

$$\Sigma = \mathrm{cov}(\mathrm{vec}([A, b])) = \begin{pmatrix} L_x\Sigma_g L_x^T & L_x\Sigma_g L_y^T & L_x\Sigma_g L_t^T \\ L_y\Sigma_g L_x^T & L_y\Sigma_g L_y^T & L_y\Sigma_g L_t^T \\ L_t\Sigma_g L_x^T & L_t\Sigma_g L_y^T & L_t\Sigma_g L_t^T \end{pmatrix} \ . \tag{15}$$





Conclusion

We propose a ML estimator for linear EIV regression under additive Gaussian noise with zero-mean and **arbitrary positive-definite covariance**. This estimator requires the minimization of a non-convex multivariate polynomial which is carried out by cyclic coordinate descent and a Newton method, respectively.

- MLE **clearly outperforms** TLS and ETLS [7] in a simulation where planes are fitted to clouds of points jittered by correlated noise.
- In optical flow estimation, ETLS is a **good approximation** to MLE (outperforming TLS).
- In very small, highly structured problems, a Gröbner basis can be constructed [2] to reduce MLE to root-finding of univariate polynomials.

References

[1] 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.
[2] 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.
[3] B. De Moor. Total least squares for affinely structured matrices and the noisy realization problem. *IEEE TSP*, 42(11):3104–3113, Nov 1994.
[4] 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.
[5] J. H. Manton, R. Mahony, and H. Yingbo. The geometry of weighted low rank approximations. *IEEE TSP*, 51(2):500–514, February 2003.
[6] 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.
[7] M. Mühlich. *Estimation in Projective Spaces and Application in Computer Vision*. PhD thesis, Johann Wolfgang Goethe Universität Frankfurt am Main, 2005.
[8] H. Scharr. Optimal filters for extended optical flow. In *Complex Motion, Lect. Notes in Comp. Sci.*, volume 3417. Springer, 2004.
[9] S. Van Huffel and P. Lemmerling, editors. *Total Least Squares and Errors-in-Variables Modeling*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
[10] 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.