ON ERRORS-IN-VARIABLES ESTIMATION WITH UNKNOWN NOISE VARIANCE RATIO

I. Markovsky [*], **A. Kukush** [**], and **S. Van Huffel** [*]

[*] *ESAT, SCD-SISTA, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[**] *Kiev National Taras Shevchenko University, Vladimirskaya st. 64, 01033, Kiev, Ukraine*

Abstract: We propose an estimation method for an errors-in-variables model with unknown input and output noise variances. The main assumption that allows identifiability of the model is clustering of the data into two clusters that are distinct in a certain specified sense. We show an application of the proposed method for system identification.

Keywords: errors-in-variables, system identification, total least squares, clustering.

1. INTRODUCTION

Corresponding to the total least squares (TLS) problem

$$\{\hat{R}_{\text{tls}}, \hat{D}_{\text{tls}}\} := \arg\min_{R,\hat{D}} \left\|D - \hat{D}\right\|_{\text{F}}^2 \quad \text{subject to} \quad RR^\top = I_p \quad \text{and} \quad R\hat{D} = 0 \tag{1}$$

is the errors-in-variables (EIV) model

$$D = \bar{D} + \tilde{D}, \quad \text{rank}(\bar{D}) = \text{row dim}(D) - p =: m. \tag{2}$$

Here $\bar{D}$ is a true value and $\tilde{D}$ is a measurement error that is modeled as a zero mean random matrix with independent and identically distributed (i.i.d.) elements. The TLS estimator (1) is maximum likelihood for the EIV model (2) if, in addition to the previous assumptions, the entries of $\tilde{D}_{ij}$ are normally distributed. The variance of $\tilde{D}_{ij}$ needs not be known, but the i.i.d. assumption is often too restrictive.

More general EIV models, where the measurement errors need not be i.i.d., have been considered in (Kukush and Van Huffel, 2004). The corresponding TLS-type problems are called *weighted* TLS problems. A key assumption in this work is that the noise covariance structure, *i.e.*, the covariance matrix of $\text{vec}(\tilde{D})$ is known up to a scaling factor. One can argue that the knowledge of the noise covariance structure up to a scalar is again restrictive in practice.

EIV models with two or more unknown noise parameters, however, are unidentifiable by second order methods, *i.e.*, there are many solutions that are not distinguishable from the second order statistics. This unidentifiability problem is well known in the context of the Frisch scheme (Frisch, 1934; De Moor, 1988). For dynamical systems a similar negative result is first proven in (Anderson, 1985).

Various additional assumptions can be imposed in order to make the EIV estimation problem with unknown noise covariance structure identifiable. An overview of methods for EIV system identification is given in (Söderström *et al.*, 2002; Söderström, 1981). In this paper we show a new assumption that allows to derive consistent parameter and noise variance estimates. The idea comes from (Wald, 1940), where a static single input single output ($m = p = 1$) EIV model is considered and the proposed estimator is the line passing through the mean values of two clusters of data points. We develop this simple idea for multi input single output static and dynamic EIV models. The key consistency assumption for the method is that the data D has as many clusters as there are unknown noise parameters. For example, the proposed method is not applicable for problems where the inputs are stationary, which is a typical assumption in much of the prior work on EIV system identification. Also in

the dynamic case, we assume that the input and output measurement noises are white and uncorrelated.

The assumption that the data can be clustered means that the true input changes character while the noise properties remain the same. This assumption can be viewed equivalently as having a set of data records from experiments with different true inputs. Such an assumption is certainly restrictive and presently we do not have specific applications in mind.

In Section 2 we describe kernel and input/output representations of static and dynamic linear models. Section 3 presents the proposed estimation method for static EIV models and states conditions for consistency. Section 4 extends the method to dynamic models and Section 5 shows simulation examples for EIV system identification. For simplicity the proposed method is applied to the special case of single output models and covariance structure known up to two unknown scalars: the input and the output noise variances. In the conclusions we discuss the extension of the method for problems involving more than two unknown parameters of the covariance matrix.

2. KERNEL AND INPUT/OUTPUT REPRESENTATIONS OF LINEAR MODELS

2.1 Static models

The data matrix $D \in \mathbb{R}^{q\times N}$ has as rows the variables of interest and as columns the observed samples of those variables. A linear static model $\mathscr{B}$ for D is a subspace of $\mathbb{R}^q$. Such a model can be represented as the kernel of a matrix $R \in \mathbb{R}^{g\times q}$, *i.e.*,

$$\mathscr{B} = \{\, d \in \mathbb{R}^q \mid Rd = 0 \,\} =: \ker(R).$$

In the numerical linear algebra literature, however, the input/output representation

$$\mathscr{B}(X) := \left\{ d = \begin{bmatrix} d_\mathrm{i} \\ d_\mathrm{o} \end{bmatrix} \begin{matrix} m \\ p \end{matrix} \;\middle|\; X^\top d_\mathrm{i} = d_\mathrm{o} \right\} \qquad (3)$$

is preferred over the kernel one because it makes explicit the input/output structure of the model. The matrix $X \in \mathbb{R}^{m\times p}$, $m+p=q$ in $\mathscr{B}(X)$ is a parameter of the model. Note that while the parameter R in a kernel representation is in general non-unique, for a fixed input/output partitioning, the parameter X is unique.

Given R, one can always find $\tilde{R}$, such that $\tilde{R}\tilde{R}^\top = I_p$ and $\ker(\tilde{R}) = \ker(R)$. Note that if R is full row rank, then the number of rows $g := \mathrm{row\,dim}(R)$ of R is equal to the number of outputs p of $\mathscr{B} = \ker(R)$. In the single output case, $RR^\top = \|R\|^2 = 1$, makes R unique.

The first m variables d_i in the input/output representation (3) are inputs, *i.e.*, they can be chosen freely. The last p variables d_o are outputs, *i.e.*, they are fixed by the input and the model. The integers m and p are invariant of the representation. Note that not every model $\mathscr{B} \subseteq \mathbb{R}^q$ admits an input/output representation with a fixed input/output partitioning. The set of models that can not be represented in the form (3), however, is non-generic in $\mathbb{R}^q$.

Corresponding to the representation (3) are the input/output partitionings of the data and measurement error matrices

$$D =: \begin{bmatrix} D_\mathrm{i} \\ D_\mathrm{o} \end{bmatrix} \begin{matrix} m \\ p \end{matrix} \qquad \text{and} \qquad \tilde{D} =: \begin{bmatrix} \tilde{D}_\mathrm{i} \\ \tilde{D}_\mathrm{o} \end{bmatrix} \begin{matrix} m \\ p \end{matrix}.$$

In Section 3 we use the input/output representation (3) and assign different noise variances to the input and output variables.

2.2 Dynamic models

In the dynamic context the given data $w_\mathrm{d} \in (\mathbb{R}^q)^T$ is a finite vector time series and the model is a subset of the data space $(\mathbb{R}^q)^T$. We consider linear time-invariant (LTI) models. The LTI model class admits a difference equation representation

$$\begin{aligned}\mathscr{B}(R) := \{\, w \in (\mathbb{R}^q)^T \mid R_0 w(t) + R_1 w(t+1) + \cdots \\ + R_l w(t+l) = 0, \quad \text{for } t = 0, \ldots, T-l \,\}. \end{aligned} \qquad (4)$$

The matrices $R_i \in \mathbb{R}^{g\times q}$ are parameters of the model. With $g = 1$ and $R_l \neq 0$, $\mathscr{B}(R)$ is a single output model.

A given LTI model $\mathscr{B}$ generically admits an input/output representation

$$\begin{aligned}\mathscr{B} = \left\{ w = \begin{bmatrix} u \\ y \end{bmatrix} \begin{matrix} m \\ 1 \end{matrix} \;\middle|\; P_0 y(t) + P_1 y(t+1) + \cdots \right. \\ + P_l y(t+l) = Q_0 u(t) + Q_1 u(t+1) + \cdots \\ \left. + Q_l u(t+l), \quad \text{for } t = 0, \ldots, T-l \right\},\end{aligned}$$

$P_i \in \mathbb{R}^{p\times p}$, $Q_i \in \mathbb{R}^{p\times m}$, and $\left(\sum_{k=0}^{l} P_k z^k\right)^{-1}\left(\sum_{k=0}^{l} Q_k z^k\right)$ is a proper rational function (transfer function of $\mathscr{B}$).

3. STATIC PROBLEMS WITH ONE OUTPUT AND TWO UNKNOWN NOISE PARAMETERS

First we consider the special case of EIV static model with one output and covariance structure known up to two scalars: input and output noise variances. This model corresponds to the classical linear system of equations $Ax \approx b$ with noises on A and b of unknown size. The model is unidentifiable without additional information on the noise covariance structure. For example, the weighted TLS method requires that the ratio of the input and output noise variances is known.

Our assumption about the noise covariance matrix is that

$$\mathbf{E}\,\tilde{D}\tilde{D}^\top = \begin{bmatrix} \mathbf{E}\,\tilde{D}_\mathrm{i}\tilde{D}_\mathrm{i}^\top & \mathbf{E}\,\tilde{D}_\mathrm{i}\tilde{D}_\mathrm{o}^\top \\ \mathbf{E}\,\tilde{D}_\mathrm{o}\tilde{D}_\mathrm{i}^\top & \mathbf{E}\,\tilde{D}_\mathrm{o}\tilde{D}_\mathrm{o}^\top \end{bmatrix} =: \begin{bmatrix} \bar{\lambda}_\mathrm{i} W_\mathrm{i} & 0 \\ 0 & \bar{\lambda}_\mathrm{o} W_\mathrm{o} \end{bmatrix}, \qquad (5)$$

where $W_\mathrm{i} \in \mathbb{R}^{m\times m}$ and $W_\mathrm{o} \in \mathbb{R}$ are *known* positive definite matrices, and $\bar{\lambda}_\mathrm{i}$ (the input variance) and $\bar{\lambda}_\mathrm{o}$ (the output variance) are *unknown* positive scalars.

Provided that the smallest eigenvalue of the sample covariance matrix $DD^\top$ has multiplicity one (the generic case), the TLS solution $\hat{R}_{\text{tls}}$ is unique and is given by the eigenvector of $DD^\top$, corresponding to the smallest eigenvalue. Alternatively, we are looking for a solution R to the nonlinear system of equations

$$R(DD^\top - \lambda I) = 0$$

corresponding to the smallest value of λ.

The TLS solution corresponds to an EIV model, in which the input/output noise variance ratio

$$\bar{\lambda}_{\text{i}}/\bar{\lambda}_{\text{o}} =: \bar{\mu}$$

is equal to one, $\{\tilde{D}_{\text{i},ij}\}$ are i.i.d., and $\{\tilde{D}_{\text{o},ij}\}$ are i.i.d. The case of a *known* noise ratio $\bar{\mu} \neq 1$, corresponds to a weighted TLS problem, which seeks for a solution to the system

$$R\left(DD^\top - \lambda_{\text{o}} \begin{bmatrix} \bar{\mu} I_m & 0 \\ 0 & 1 \end{bmatrix}\right) = 0,$$

corresponding to the smallest value of λ_{o}. The numerical solution in this case is performed via the generalized eigenvalue or singular value decomposition (instead of the ordinary eigenvalue or singular value decomposition).

Denote

$$W(\mu) := \begin{bmatrix} \mu W_{\text{i}} & 0 \\ 0 & W_{\text{o}} \end{bmatrix},$$

so that $\mathbf{E}\tilde{D}\tilde{D}^\top = \bar{\lambda}_{\text{o}} W(\bar{\mu})$. Corresponding to the EIV model and assumption (5) is the weighted TLS problem

$$\min_{\hat{R},\hat{D}} \left\| W^{-1/2}(\bar{\mu})(D - \hat{D}) \right\|_{\text{F}}^2 \quad \text{subject to} \quad R\hat{D} = 0,$$

or equivalently the nonlinear system of equations

$$R\left(DD^\top - \lambda_{\text{o}} W(\bar{\mu})\right) = 0, \tag{6}$$

where again the true input/output noise ratio $\bar{\mu}$ is assumed known. The computation can be carried out via a generalized eigenvalue or singular value decomposition. In the literature the estimator for this case is called *generalized* TLS (Van Huffel and Vandewalle, 1989).

Now consider the problem of unknown true noise ratio $\bar{\mu}$. Even if R is properly normalized, *e.g.*, $\|R\| = 1$, a solution to (6) is non-unique. This is a manifestation of a lack of identifiability of the model with unknown input and output noise variances. We propose to resolve the unidentifiability problem by considering a system of two independent estimating equations. After a permutation of the columns of D via a permutation matrix Π, define

$$D\Pi =: \begin{bmatrix} D^1 & D^2 \end{bmatrix} =: \begin{bmatrix} D_{\text{i}}^1 & D_{\text{i}}^2 \\ D_{\text{o}}^1 & D_{\text{o}}^2 \end{bmatrix} \begin{matrix} m \\ 1 \end{matrix}$$

and consider two copies of (6)

$$R\left(D^k (D^k)^\top - \lambda_{\text{o}} W(\mu)\right) = 0, \quad \text{for } k = 1, 2, \tag{7}$$

corresponding to the two parts D^1 and D^2 of the data matrix.

If the minimal singular value of $D_{\text{i}}^1 (D_{\text{i}}^1)^\top - D_{\text{i}}^2 (D_{\text{i}}^2)^\top$ is separated from zero, the problem of estimating simultaneously λ_{o}, μ, and R is identifiable (see Theorem 1 below). This condition is related to the existence of clusters in the true data $\bar{D}$. Correspondingly the clustering problem is

$$\max_{\substack{\text{permutation} \\ \text{matrix } \Pi}} \left(\min_{j=1,\ldots,m} \left| \lambda_j \left(D_{\text{i}}^1 (D_{\text{i}}^1)^\top - D_{\text{i}}^2 (D_{\text{i}}^2)^\top \right) \right| \right), \tag{8}$$

where $\lambda_1(A), \ldots, \lambda_{\dim(A)}(A)$ are the eigenvalues of A.

The estimation problem corresponding to (7) is more complicated than a generalized eigenvalue/singular value decomposition: we need to find a common generalized eigenvalue-eigenvector pair for two pairs of symmetric positive semidefinite matrices that depend on a scalar parameter. In general, (7) has no exact solution, so that an approximation is needed.

We propose the following nonlinear least squares type approximate solution:

$$\hat{\mu} = \arg\min_{\mu} \left(\left(\lambda_{\text{o}}^1 - \lambda_{\text{o}}^2\right)^2 + C \sin^2\left(\angle(R^1, R^2)\right) \right), \tag{9}$$

where $(\lambda_{\text{o}}^k, R^k)$ is the minimal generalized eigenvalue-eigenvec. pair of the pencil $\left(D^k (D^k)^\top, \text{diag}(\mu W_{\text{i}}, W_{\text{o}})\right)$, C is a regularization parameter, and $\angle(R^1, R^2)$ stands for the angle between the vectors R^1 and R^2. The first term in the cost function makes both eigenvalues close to each other, while the second term makes the corresponding eigenvectors close to each other. The regularization parameter C allows to turn the two objectives optimization problem into a one objective optimization. Our experience is that the results are rather insensitive to the value of this parameter and in the simulation examples of Section 5 we set $C = 1$.

Once the optimal estimate $\hat{\mu}$ of the noise variance ratio μ is found, the estimation problem becomes a classical generalized TLS problem. In summary, the proposed estimation procedure has three stages.

1. Cluster the data by solving (8).
2. Compute the noise variance ratio estimate $\hat{\mu}$ by solving (9) for the clusters identified on step 1.
3. Solve the standard generalized TLS problem for the estimated value of μ on step 2.

The clustering problem (8) has combinatorial complexity, so that (except for small examples) it can be solved only via heuristic methods. See (Xu and Wunsch, 2005) for a survey of clustering algorithms. In our simulation examples we use the K-means algorithm implemented in the Statistics Toolbox of MATLAB.

The optimization problem on step 2 is a rather simple one because it involves a single scalar decision variable and the search interval is lower bounded by 0. Note that each cost function evaluation involves two generalized singular value decompositions.

The procedure described above is closely related but not equivalent to the one of (Kukush *et al.*, 2005).

In (Kukush *et al.*, 2005), steps 2 and 3 of the algorithm are different.

2'. Compute the noise variance estimates $\hat{\lambda}_\mathrm{i}$ and $\hat{\lambda}_\mathrm{o}$ by solving the optimization problem

$$\min_{\lambda_\mathrm{i},\lambda_\mathrm{o}} \left(\left\| \begin{bmatrix} \mu^1 \\ \mu^2 \end{bmatrix} \right\|^2 + C \sin^2 \left(\angle(R^1, R^2) \right) \right), \quad (9')$$

where μ^k and R^k are respectively the smallest eigenvalue and corresponding eigenvector of

$$D^k (D^k)^\top - \mathrm{diag}(\lambda_\mathrm{i} W_\mathrm{i}, \lambda_\mathrm{o} W_\mathrm{o}).$$

3'. Define the estimate $\hat{R}$ as the eigenvector corresponding to the smallest eigenvalue of the matrix

$$DD^\top - \mathrm{diag}(\hat{\lambda}_\mathrm{i} W_\mathrm{i}, \hat{\lambda}_\mathrm{o} W_\mathrm{o}).$$

Problem (9) can be viewed as a modification of (9'). In general both problems are nonconvex and nonsmooth, however, (9) is univariate while (9') is bivariates and for this reason more difficult to solve. The estimator on Step 3' is called the adjusted least squares. It is equivalent to the generalized TLS estimator on step 3 with $\hat{\mu} = \hat{\lambda}_\mathrm{i} / \hat{\lambda}_\mathrm{o}$.

Statistical consistency of the estimator using steps 2' and 3' is proven in (Kukush *et al.*, 2005). Here we state the main result, specialized for the considered problem. (The result of (Kukush *et al.*, 2005) applies to multi input multi output dynamic problems as well.)

Theorem 1. Assume that:

i). there exists $\delta > 0$, such that $\mathbf{E} |\tilde{D}_{ij}|^{4+\delta}$ are uniformly bounded for all i, j

ii). $\frac{1}{N_1} \|\bar{D}_\mathrm{i}^1\|_\mathrm{F}$ and $\frac{1}{N_2} \|\bar{D}_\mathrm{i}^2\|_\mathrm{F}$ are bounded, where $N_1 := \mathrm{col\,dim}(D^1)$ and $N_2 := \mathrm{col\,dim}(D^2)$

iii). $\liminf_{N_1,N_2 \to \infty} \sigma_{\min} \left(\frac{1}{N_1} D_\mathrm{i}^1 D_\mathrm{i}^{1\top} - \frac{1}{N_2} D_\mathrm{i}^2 D_\mathrm{i}^{2\top} \right) > 0$, where $\sigma_{\min}(A)$ is the smallest singular value of A,

iv). $\liminf_{N_1 \to \infty} \frac{1}{N_1} \mathrm{trace}(X^\top W_\mathrm{i} X) > 0$
$\liminf_{N_2 \to \infty} \frac{1}{N_2} \mathrm{trace}(W_\mathrm{o}) > 0$, and

v). $\liminf_{N_k \to \infty} \frac{1}{N_k} \sigma_{\min} \left(\bar{D}_\mathrm{i}^k (\bar{D}_\mathrm{i}^k)^\top \right) > 0$, for $k = 1, 2$.

Then $\hat{\lambda}_\mathrm{i} \to \bar{\lambda}_\mathrm{i}$, $\hat{\lambda}_\mathrm{o} \to \bar{\lambda}_\mathrm{o}$, and $\hat{X} \to \bar{X}$, as $N_1, N_2 \to \infty$, almost surely.

4. APPLICATION FOR DYNAMIC MODELS

Certain system identification problems can be posed as structured TLS problems (Markovsky *et al.*, 2005*a*). The main difference with the static TLS problem (1) is that now the data matrix is block-Hankel structured. For application of the structured TLS method, however, the measurement error covariance structure should be known up to a scaling factor. In this section we address a dynamic version of the problem of Section 3.

Written in a matrix form the difference equation

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_l w(t+l) = 0,$$

becomes the structured system of equations

$$R \mathscr{H}_{l+1}(w) = 0, \quad (10)$$

where $\mathscr{H}_{l+1}(w)$ is the block-Hankel matrix with $l+1$ block rows, constructed from the time series w:

$$\mathscr{H}_{l+1}(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-l) \\ w(2) & w(3) & \cdots & w(T-l+1) \\ w(3) & w(4) & \cdots & w(T-l+2) \\ \vdots & \vdots & & \vdots \\ w(l+1) & w(l+2) & \cdots & w(T) \end{bmatrix}, \quad (11)$$

and

$$R := \begin{bmatrix} R_0 & R_1 & \cdots & R_l \end{bmatrix}.$$

With noisy data w_d and $T > q(l+1) + l$, generically (4) is not compatible, so that an approximation is needed.

The *structured* TLS problem is the dynamic equivalent of the TLS problem (1):

$$\min_{R,\hat{w}} \|w - \hat{w}\|^2 \quad \text{subject to} \quad RR^\top = 1, \quad \text{and} \quad R \mathscr{H}_{l+1}(\hat{w}) = 0. \quad (12)$$

Its solution however can not be expressed in closed form via a singular value decomposition of the data matrix as in the static case and needs nonlinear optimization methods (Markovsky *et al.*, 2004).

The structured TLS problem (12) corresponds to the EIV model $w_\mathrm{d} = \bar{w} + \tilde{w}$, where $\bar{w}$ is a trajectory of $\mathscr{B}(\bar{R})$ for some $\bar{R} \in \mathbb{R}^{1 \times q(l+1)}$, and $\tilde{w}$ is a white random sequence with covariance matrix that is equal to a multiple of the identity. If in addition the measurement errors $\tilde{w}$ are normally distributed, the structured TLS estimator is maximum likelihood.

More general *weighted structured TLS* problems allow to take into account the noise covariance matrix that is known up to a scaling factor. As in the static case, dynamic EIV problems in which the covariance structure is specified up to more than one scaling factor are unidentifiable.

Next we consider the case when the noise covariance matrix is known up to two scalars—the input and output noise variances. Accordingly,

$$\mathrm{cov}\left(\tilde{w}(t)\right) = \begin{bmatrix} \bar{\lambda}_\mathrm{i} I_m & 0 \\ 0 & \bar{\lambda}_\mathrm{o} \end{bmatrix} = \bar{\lambda}_\mathrm{o} \begin{bmatrix} \bar{\mu} I_m & 0 \\ 0 & 1 \end{bmatrix} =: \bar{\lambda}_\mathrm{o} W(\bar{\mu}). \quad (13)$$

Let

$$\mathbf{W}(\mu) := \mathrm{diag}(\underbrace{W(\mu), \ldots, W(\mu)}_{T \text{ times}}).$$

The weighted structured TLS problem corresponding to the EIV model with the noise covariance (13) is

$$\min_{R,\hat{w}} (w_\mathrm{d} - \hat{w})^\top \mathbf{W}^{-1}(\bar{\mu}) (w_\mathrm{d} - \hat{w}) \quad \text{subject to} \quad RR^\top = 1, \quad R \mathscr{H}_{l+1}(\hat{w}) = 0. \quad (14)$$

However, this problem can be solved only for a given parameter $\bar{\mu}$.

In (Markovsky *et al.*, 2005*b*; Markovsky and Van Huffel, 2005) we show that the structured TLS problem

is generically equivalent to the following nonlinear weighted least squares problem

$$\min_X \begin{bmatrix} X^\top & -1 \end{bmatrix} \mathscr{H}_{l+1}(w_{\mathrm{d}}) \Gamma^{-1}(X, \bar{\mu}) \mathscr{H}_{l+1}^\top(w_{\mathrm{d}}) \begin{bmatrix} X \\ -1 \end{bmatrix} \tag{15}$$

where $\Gamma(X, \bar{\mu})$ is a block banded and Toeplitz matrix that depends on X and $\mathbf{W}(\bar{\mu})$. In turn the optimization problem (15) can be seen equivalently as a least squares approximate solution method for the nonlinear system of equations

$$\begin{bmatrix} X^\top & -1 \end{bmatrix} \mathscr{H}_{l+1}(w_{\mathrm{d}}) \Gamma^{-1/2}(X, \bar{\mu}) = 0.$$

In order to resolve the identifiability problem, we make the assumption that the true time series $\bar{w}$ changes its behavior at time T', *i.e.*, the time series

$$\bar{w}^1 := (\bar{w}(1), \ldots, \bar{w}(T'))$$

has different mean and dispersion than the time series

$$\bar{w}^2 := (\bar{w}(T'+1), \ldots, \bar{w}(T)).$$

This assumption corresponds to the clustering assumption in the static case. Note that now the ordering of the data samples corresponds to time, so that we can not permute them as in the static case.

Under the existence of clusters, define w_{d}^1, w_{d}^2 analogously to $\bar{w}^1$, $\bar{w}^2$, and consider the system of two nonlinear equations

$$\begin{bmatrix} X_k^\top & -1 \end{bmatrix} \mathscr{H}_{l+1}(w_{\mathrm{d}}^k) \Gamma^{-1/2}(X, \mu) = 0, \quad \text{for } k = 1, 2. \tag{16}$$

For noise free data, the equations have a common solution $X^1 = X^2 = \bar{X}$ and $\mu = \bar{\mu}$. In the presence of noise, an approximate solution is needed and we propose the same criterion as in the static case

$$\hat{\mu} = \arg\min_{\mu} \left(\lambda_{\mathrm{o}}^1 - \lambda_{\mathrm{o}}^2\right)^2 + C \sin^2\left(\angle(R^1, R^2)\right), \tag{17}$$

where R^k and λ_{o}^k come from the structured TLS problems associated with (16),

$$R^k := \begin{bmatrix} (X^k)^\top & -1 \end{bmatrix},$$

and

$$\lambda_{\mathrm{o}}^k := R^k \mathscr{H}_{l+1}(w_{\mathrm{d}}^k) \Gamma^{-1}(X, \mu) \mathscr{H}_{l+1}^\top(w_{\mathrm{d}}^k) (R^k)^\top.$$

Each cost function evaluation involves solving two structured TLS problems.

In summary, the algorithm for the considered EIV identification problem is:

1. Detect a time instant T' at which w_{d} changes its behavior.
2. Solve the optimization problem (17) for the partitioning of the data found in step 1, and
3. Solve the structured TLS problem (14) for the estimated value $\hat{\mu}$ in step 2.

Note 1. The cost functions in (9), (9'), and (17) are discontinuous. Therefore, the global minimum might not exist. The minimization is performed up to certain tolerance that decreases to zero, as the number of observations increases.

5. SIMULATION EXAMPLE

Consider the EIV model (2). The covariance structure of the measurement errors $\tilde{D}$ is known up to scaling factors (the true noise variances) $\bar{\lambda}_{\mathrm{i}}$ and $\bar{\lambda}_{\mathrm{o}}$. The simulation example aims to show consistency of the estimators for the unknown parameters $\bar{\lambda}_{\mathrm{i}}$, $\bar{\lambda}_{\mathrm{o}}$, and $\bar{X}$, the parameter in an input/output representation of the true model that has generated the data $\bar{D}$.

Let $U_{N'}(l, u)$ be a matrix with N' columns, composed of independent and uniformly distributed random variables in the interval $[l, u]$. The true values $\bar{D}_{\mathrm{i}}$ and $\bar{R}$ are selected as follows:

$$\bar{D}_{\mathrm{i}} = \begin{bmatrix} U_{N'}(-0.5, 0.5) & U_{N'}(10, 20) \end{bmatrix}, \quad \bar{X}^\top = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

where N' is varied from 10 to 500. Correspondingly $\bar{D}_{\mathrm{o}} := \bar{X}^\top \bar{D}_{\mathrm{i}}$. Note that we artificially create two clusters (the first N' and the last N' columns of $\bar{D}$). In this simulation example the K-means clustering algorithm detects without errors the two clusters. The measurement errors are zero mean, independent, normally distributed with variances $\lambda_{\mathrm{i}} = 0.01$ and $\lambda_{\mathrm{o}} = 0.04$.

With this simulation setup we apply the proposed estimation method and average the results for 500 noise realizations. The average values of the noise variance estimates $\hat{\lambda}_{\mathrm{i}}$ and $\hat{\lambda}_{\mathrm{o}}$ are shown on Figure 1.

Figure 2 shows the average relative estimation error

$$e := \frac{1}{500} \sum_{k=1}^{500} \frac{\|\bar{X} - \hat{X}^{(k)}\|}{\|\bar{X}\|},$$

where $\hat{X}^{(i)}$ is the estimate of the parameter $\bar{X}$ in the ith repetition of the experiment.

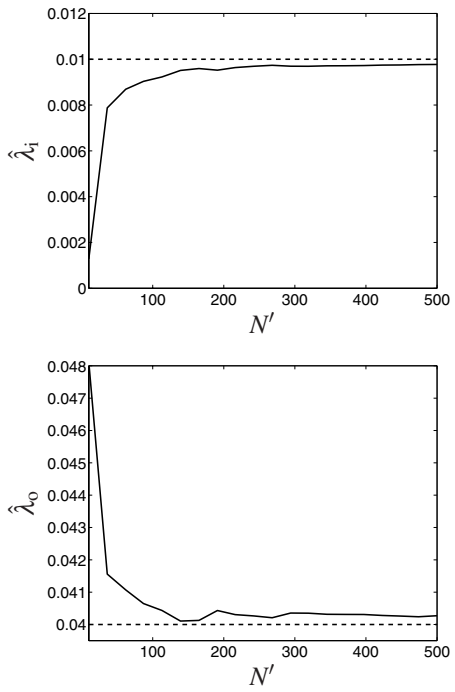

Fig. 1. Average values of the noise variance estimates $\hat{\lambda}_{\mathrm{i}}$ and $\hat{\lambda}_{\mathrm{o}}$ as a function of half the sample size N'. The dashed lines are the true values.

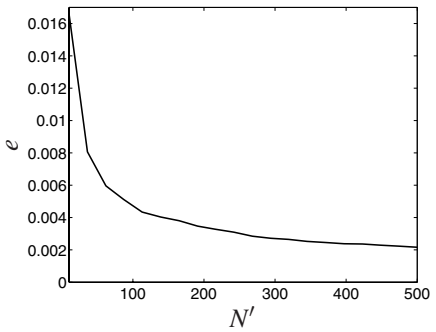

Fig. 2. Relative error of estimation e as a function of half the sample size N'.

6. CONCLUSIONS

We have considered an EIV estimation problems for single output static and dynamic systems, with measurement error covariance matrix known up to two unknown parameters. The model is identifiable under the assumption that the data has two clusters that are distinct in a specified sense. The proposed estimation method has three steps: cluster the data, solve a univariate optimization problem for the noise variance ratio, solve a standard TLS-type problem for the estimated noise variance ratio. In the static case the cost function evaluation on the second step involves solving a couple of generalized TLS problems and in the dynamic case it involves solving a couple of weighted structured TLS problems.

Identifiability of the model is recovered by constructing two estimating equations corresponding to the two clusters. The idea generalizes to problems involving more than two unknown parameters in the measurement error covariance matrix. As many clusters and corresponding estimating equations are needed as there are unknown covariance parameters. The estimation procedure in this case, however, requires solving a multidimensional optimization problem on the second step. It is an open problem what special properties (if any) this optimization problem has and how to exploit them in effective EIV estimation algorithms.

ACKNOWLEDGEMENTS

I. Markovsky is a postdoctoral researcher and S. Van Huffel is a full professor at the K.U.Leuven, Belgium. A. Kukush is a full professor at the Kiev National Taras Shevchenko Univ., Ukraine. We are grateful to Prof. H. Schneeweiss for fruitful discussions. Prof. Kukush is supported by a senior postdoctoral fellowship from the Dep. of Applied Economics of the K.U.Leuven. Our research is supported by Research Council KUL: GOA-AMBioRICS, GOA-Mefisto 666, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), G.0360.05 (EEG signal processing), research communities (ICCoS, ANMMM); IWT: PhD Grants; Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'); EU: PDT-COIL, BIOPATTERN, ETUMOUR.

REFERENCES

Anderson, B. (1985). Identification of scalar errors-in-variables models with dynamics. *Automatica* **21**, 625–755.

De Moor, B. (1988). Mathematical concepts for modeling of static and dynamic systems. PhD thesis. Dept. EE, K.U.Leuven, Belgium.

Frisch, R. (1934). Statistical confluence analysis by means of complete regression systems. Technical Report 5. Univ. of Oslo, Economics Institute.

Kukush, A. and S. Van Huffel (2004). Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$. *Metrika* **59**(1), 75–97.

Kukush, A., I. Markovsky and S. Van Huffel (2005). Estimation in a linear multivariate measurement error model with clustering in the regressor. Technical Report 05–170. Dept. EE, K.U.Leuven.

Markovsky, I. and S. Van Huffel (2005). Weighted structured total least squares. Technical Report 05–41. Dept. EE, K.U.Leuven.

Markovsky, I., J. C. Willems, S. Van Huffel, B. De Moor and R. Pintelon (2005*a*). Application of structured total least squares for system identification and model reduction. *IEEE Trans. on Aut. Control* **50**(10), 1490–1500.

Markovsky, I., S. Van Huffel and A. Kukush (2004). On the computation of the structured total least squares estimator. *Num. Lin. Alg. with Appl.* **11**, 591–608.

Markovsky, I., S. Van Huffel and R. Pintelon (2005*b*). Block-Toeplitz/Hankel structured total least squares. *SIAM J. Matrix Anal. Appl.* **26**(4), 1083–1099.

Söderström, T. (1981). Identification of stochastic linear systems in presence of input noise. *Automatica* **17**(5), 713–725.

Söderström, T., U. Soverini and K. Mahata (2002). Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing* **82**, 1139–1154.

Van Huffel, S. and J. Vandewalle (1989). Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error. *SIAM J. Matrix Anal.* **10**(3), 294–315.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.* **11**, 284–300.

Xu, R. and D. Wunsch (2005). Survey of clustering algorithms. *IEEE Trans. on Neural Networks* **13**(3), 645–678.