Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

On the Generalization of the Koopmans-Levin Estimation Method

István Vajk and Jenő Hetthéssy

***Abstract*—The paper discusses parameter estimation methods originated from the direct or indirect minimization of a performance index related to the generalized version of the Koopmans-Levin algorithm (GKL). Unlike algorithms directly minimizing an appropriate loss function, indirect estimation algorithms perform data compression into a subspace first then derive the parameter estimation from the eigenvectors spanning this subspace. The application of scaleable Hankel and Toeplitz type matrices offers a compact and uniform treatment of the various algorithms. Optimal setting of the weighting matrices applicable in the performance index to reduce the variance of the parameter estimation is also shown. A simulation study has been added to compare the performance of the presented identification algorithms.**

I. Introduction

NOWDAYS parameter estimation algorithms related to subspace methods represent an extremely popular field in identification. Just considering the 13th IFAC Symposium on System Identification (27-29 August 2003, Rotterdam, The Netherlands) more than a quarter of the papers were devoted to discuss or to refer to subspace based algorithms. The first step in classical subspace methods essentially realizes data compression. Compression of the Hankel matrices formed by the input/output samples can efficiently be performed via the numerically advanced QR and SVD algorithms. In the second step the parameter estimation is performed from the compressed matrices obtained in the first step. Several papers give excellent reviews on the subspace methods [1],[2],[3],[4]. An outstanding summary on the errors-in-variables (EIV) estimation methods can be found in [5]. ??? Further information about the PCA and subspace method for EIV model can be found in [6],[7].

In this paper data driving the identification algorithms will be assumed to be available from EIV models. The identification problem will be discussed following the concept of the GKL method [8]. In particular estimation algorithms resulting in the parameter estimation by directly minimizing an appropriate loss function and estimation algorithms following the indirect nature of the subspace methodology compressing the data set in one step and deriving the parameter estimations in another are discussed separately from each other.

The paper is organized as follows. Section II formulates the identification problem and introduces useful notations for the rest of the paper. As a direct estimation algorithm, in Section III the summary of a generalized version of the Koopmans-Levin estimation method (GKL) is given, while Section IV is devoted for subspace, i.e. indirect estimation methods aiming at estimating the eigenvectors spanning the subspace of the compressed data set. The indirect identification algorithms discussed in Section V involve weighting matrices, whose structure and settings determine the computing demand and the efficiency of the parameter estimation, respectively. Optimal selection of the weighting matrices is also discussed in Section V. As the most effective indirect parameter estimation algorithms require the knowledge of the covariance matrix of the eigenvectors, Section VI presents the related calculations to estimate the covariance of the eigenvectors. Section VII shows a simulation example to demonstrate the consequences of using various weighting matrices. Section VIII concludes the paper.

II. Problem Formulation

Consider linear, discrete-time systems with noiseless input-output observations u_k^o and y_k^o, respectively:

$$y_k^o + a_1 y_{k-1}^o + \ldots + a_m y_{k-m}^o = b_1 u_{k-1}^o + b_2 u_{k-2}^o + \ldots + b_n u_{k-m}^o \qquad (1)$$

Applying the backward shift operator

$$q^{-1} x_k = x_{k-1} \qquad (2)$$

and its polynomials

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \ldots + a_m q^{-m} \qquad (3)$$

and

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + \ldots + b_m q^{-m} \qquad (4)$$

Eq.(1) can directly be written as

$$A(q^{-1}) y_k^o - B(q^{-1}) u_k^o = 0 . \qquad (5)$$

To derive an even more compact form introduce parameter vectors by

$$\mathbf{b} = [b_m, b_{m-1}, \ldots, b_1, b_0]^T \qquad (6)$$

This work has been supported by the fund of the Hungarian Academy of Sciences for control research and the Hungarian National Research Fund (grant number: T042741)

I.Vajk is with the Department of Automation and Applied Informatics, Budapest University of Technology and Economics and HAS-BUTE Control Research Group, H 1521 Budapest, Hungary. (corresponding author to provide phone: 36-1-463-2870; fax: 36-1-463-2871; (e-mail: vajk@aut.bme.hu).

J.Hetthéssy is with the Department of Automation and Applied Informatics, Budapest University of Technology and Economics and HAS-BUTE Control Research Group, H 1521 Budapest, Hungary; (e-mail: hetthessy@aut.bme.hu)

$$\mathbf{a} = [a_m, a_{m-1}, ..., a_1, 1]^T \tag{7}$$

$$\boldsymbol{\theta} = [\mathbf{b}^T \quad -\mathbf{a}^T]^T \tag{8}$$

and an observation vector by

$$\mathbf{x}_k^o = [u_{k-m}^o, u_{k-m+1}^o, ..., u_k^o, y_{k-m}^o, y_{k-m+1}^o, ..., y_k^o]^T \tag{9}$$

Then the system equation Eq.(1) takes the following *implicit form*:

$$\boldsymbol{\theta}^T \mathbf{x}_k^o = 0 \tag{10}$$

provided that the b_0 coefficient is set to $b_0 = 0$. The extra zero element leads to a symmetrical structure in the observation vector.

Further on noisy observations will be assumed to be available for the parameter estimation:

$$u_k = u_k^o + \xi_k \qquad \text{var}(\xi_k) = \mu\sigma_u^2 \tag{11}$$

$$y_k = y_k^o + \eta_k \qquad \text{var}(\eta_k) = \mu\sigma_y^2 \tag{12}$$

where both ξ_k and η_k are independent white Gaussian noise components. The identification problem considered in this paper is to estimate the unknown a_i and b_i coefficients based on the noisy input-output records. It will be assumed that the $\text{var}(\xi_k)/\text{var}(\eta_k)$ ratio related to the noise components is *known*, however μ is *unknown*. Considering N consecutive steps the input/output samples will be arranged in several structures in the sequel. The following notations will be used:

$$\mathbf{u} = [u_1, u_2, ..., u_N]^T \tag{13}$$

$$\mathbf{y} = [y_1, y_2, ..., y_N]^T \tag{14}$$

$$\mathbf{x} = [\mathbf{u}^T, \mathbf{y}^T]^T \tag{15}$$

$$\mathbf{0}_n = [0, 0, ..., 0]^T, \qquad \dim(\mathbf{0}_n) = n \tag{16}$$

$\mathbf{0}_{n\times n}$: $n \times n$ zero matrix (17)

$\mathbf{I}_n$: $n \times n$ unity matrix (18)

$$Toeplitz(\mathbf{z}, q) = \begin{bmatrix} z_q & z_{q-1} & \cdots & z_1 \\ z_{q+1} & z_q & \cdots & z_2 \\ \cdots & \cdots & \cdots & \cdots \\ z_N & z_{N-1} & \cdots & z_{N-q+1} \end{bmatrix} \tag{19}$$

$$Hankel(\mathbf{z}, q) = \begin{bmatrix} z_1 & z_2 & \cdots & z_q \\ z_2 & z_3 & \cdots & z_{q+1} \\ \cdots & \cdots & \cdots & \cdots \\ z_{N-q+1} & z_{N-q+2} & \cdots & z_N \end{bmatrix} \tag{20}$$

where $\mathbf{z} = [z_1, z_2, ..., z_N]^T$.

III. Generalized Koopmans-Levin Method

To solve the identification problem outlined above a generalized Koopmans-Levin estimation method has been elaborated in [8]. To summarize the key steps of the identification algorithm consider the *error matrix* defined by

$$\begin{aligned} \mathbf{E}_q &= \begin{bmatrix} e_{m+1} & e_{m+2} & \cdots & e_q \\ e_{m+2} & e_{m+3} & \cdots & e_{q+1} \\ \cdots & \cdots & \cdots & \cdots \\ e_{N-q+m+1} & e_{N-q+m+2} & \cdots & e_N \end{bmatrix} \\ &= \begin{bmatrix} u_1 & u_2 & \cdots & u_q \\ u_2 & u_3 & \cdots & u_{q+1} \\ \cdots & \cdots & \cdots & \cdots \\ u_{N-q+1} & u_{N-q+2} & \cdots & u_N \end{bmatrix} \begin{bmatrix} b_m & & & \\ b_{m-1} & b_m & & \\ \cdots & b_{m-1} & \cdots & \\ b_1 & \cdots & \cdots & b_m \\ 0 & b_1 & \cdots & b_{m-1} \\ & 0 & \cdots & \cdots \\ & & \cdots & b_1 \\ & & & 0 \end{bmatrix} \\ &\quad - \begin{bmatrix} y_1 & y_2 & \cdots & y_q \\ y_2 & y_3 & \cdots & y_{q+1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{N-q+1} & y_{N-q+2} & \cdots & y_N \end{bmatrix} \begin{bmatrix} a_m & & & \\ a_{m-1} & a_m & & \\ \cdots & a_{m-1} & \cdots & \\ a_1 & \cdots & \cdots & a_m \\ 1 & a_1 & \cdots & a_{m-1} \\ & 1 & \cdots & \cdots \\ & & \cdots & a_1 \\ & & & 1 \end{bmatrix} \\ &= \mathbf{H}_q^u \mathbf{G}_q^b - \mathbf{H}_q^y \mathbf{G}_q^a \end{aligned} \tag{21}$$

Observe that all the matrices in the above expression are of Hankel and Toeplitz type, respectively. Specifically, the error matrix is an $(N-q+1)\times(q-m)$ matrix with the extreems in size by $q_{\min} = m+1$ (the error matrix reduces to an error column vector) and $q_{\max} = N$ (the error matrix reduces to an error row vector). Using the compact notations we have

$$\mathbf{E}_q = Hankel([e_{m+1}, e_{m+2}, ..., e_N], q-m) \tag{22}$$

$$\mathbf{H}_q^u = Hankel(\mathbf{u}, q) \tag{23}$$

$$\mathbf{H}_q^y = Hankel(\mathbf{y}, q) \tag{24}$$

$$\mathbf{G}_q^b = Toeplitz([\mathbf{0}_{q-m-1}, \mathbf{b}, \mathbf{0}_{q-m-1}], q-m) \tag{25}$$

$$\mathbf{G}_q^a = Toeplitz([\mathbf{0}_{q-m-1}, \mathbf{a}, \mathbf{0}_{q-m-1}], q-m) \tag{26}$$

Also, the hyper matrices

$$\mathbf{G}_q = \begin{bmatrix} \mathbf{G}_q^b \\ -\mathbf{G}_q^a \end{bmatrix} \tag{27}$$

$$\mathbf{X}_q = \begin{bmatrix} \mathbf{H}_q^u & \mathbf{H}_q^y \end{bmatrix} \tag{28}$$

allow us to rewrite Eq.(21) as

$$\mathbf{E}_q = \mathbf{X}_q \mathbf{G}_q \tag{29}$$

Using the error matrix the identification problem can be solved by minimizing the following loss function:

$$\begin{aligned} J_q &= tr\left\{(\mathbf{G}_q^T \mathbf{C}_q \mathbf{G}_q)^{-1} (\mathbf{E}_q^T \mathbf{E}_q / N_q)\right\} \\ &= tr\left\{(\mathbf{G}_q^T \mathbf{C}_q \mathbf{G}_q)^{-1} (\mathbf{G}_q^T \mathbf{D}_q \mathbf{G}_q)\right\} \end{aligned} \tag{30}$$

where

$$\mathbf{D}_q = \frac{1}{N_q} \mathbf{X}_q^T \mathbf{X}_q, \tag{31}$$

$$\mathbf{C}_q = \begin{bmatrix} \sigma_u^2 \mathbf{I}_q & \mathbf{0}_{q,q} \\ \mathbf{0}_{q,q} & \sigma_y^2 \mathbf{I}_q \end{bmatrix}. \quad (32)$$

and $N_q = N - q + 1$.

The GKL parameter estimation is obtained as a result of the following minimization:

$$\boldsymbol{\theta}_{GKL} = \arg\min_{\theta} tr\left((\mathbf{G}_q(\boldsymbol{\theta})^T \mathbf{C}_q \mathbf{G}_q(\boldsymbol{\theta}))^{-1} (\mathbf{G}_q(\boldsymbol{\theta})^T \mathbf{D}_q \mathbf{G}_q(\boldsymbol{\theta})) \right) \quad (33)$$

According to the structural scaling parameter q two special cases can be distinguished while minimizing the above loss function:

- Selecting $q = q_{\min} = m+1$ results in the classical Koopmans-Levin (KL) algorithm.
- Selecting $q = q_{\max} = N$ results in the maximum likelihood (ML) estimation.

Applying a general $m+1 \le q \le N$ value a flexible estimation strategy is obtained.

As far as the solution of the minimization procedure is concerned, for $q = m+1$ (KL case), no iteration is necessary to obtain the parameter estimation. However, for all $m+1 < q \le N$ the solution is iterative in nature. Note that increasing the value of q the estimation will be closer and closer to the ML estimation. The algorithm minimizing the loss function selecting $m+1 \le q \le N$ will be referred to as GKL (Generalized Koopmans-Levin) method.

To obtain $\boldsymbol{\theta}_{GKL}$ one possible iterative solution of the minimization is offered by the following algorithm:

$$\boldsymbol{\theta}_{k+1} = \arg\min_{\theta} \frac{tr\left(\left(\mathbf{G}_q(\boldsymbol{\theta}_k)^T \mathbf{C}_q \mathbf{G}_q(\boldsymbol{\theta}_k) \right)^{-1} (\mathbf{G}_q(\boldsymbol{\theta})^T \mathbf{D}_q \mathbf{G}_q(\boldsymbol{\theta}) \right)}{tr\left(\left(\mathbf{G}_q(\boldsymbol{\theta}_k)^T \mathbf{C}_q \mathbf{G}_q(\boldsymbol{\theta}_k) \right)^{-1} \mathbf{G}_q(\boldsymbol{\theta})^T \mathbf{C}_q \mathbf{G}_q(\boldsymbol{\theta}) \right)} \quad (34)$$

Numerically advanced solutions can be obtained utilizing the fact that the $\mathbf{D}_q$ and $\mathbf{C}_q$ matrices can be factorized according to Eq. (31) and $\mathbf{C}_q = \tilde{\mathbf{C}}_q^T \tilde{\mathbf{C}}_q$ respectively. Using the QR decompositions

$$\mathbf{X}_q / \sqrt{N_q} = \mathbf{Q}_X \mathbf{R}_X \quad (35)$$

$$\tilde{\mathbf{C}}_q \mathbf{G}_q(\boldsymbol{\theta}_k) = \mathbf{Q}_{CG}(\boldsymbol{\theta}_k) \mathbf{R}_{CG}(\boldsymbol{\theta}_k) \quad (36)$$

the minimization problem can be re-formulated by

$$\boldsymbol{\theta}_{k+1} = \arg\min_{\theta} \frac{vec(\mathbf{G}_q)^T (\mathbf{R}_{CG}^{-T}(\boldsymbol{\theta}_k) \otimes \mathbf{R}_X^T)(\mathbf{R}_{CG}^{-1}(\boldsymbol{\theta}_k) \otimes \mathbf{R}_X) vec(\mathbf{G}_q)}{vec(\mathbf{G}_q)^T (\mathbf{R}_{CG}^{-T}(\boldsymbol{\theta}_k) \otimes \tilde{\mathbf{C}}_q^T)(\mathbf{R}_{CG}^{-1}(\boldsymbol{\theta}_k) \otimes \tilde{\mathbf{C}}_q) vec(\mathbf{G}_q)} \quad (37)$$

where $\otimes$ is used to denote Kronecker product and $vec(\,)$ rearranges matrix elements to a (column) vector form. Applying the $vec(\,)$ operation to $\mathbf{G}_q$ the system parameters can simply be expressed in a linear way by

$$vec(\mathbf{G}_q) = \mathbf{T}_G \boldsymbol{\theta} \quad (38)$$

where $\mathbf{T}_G$ contains only zeros and ones. Consequently,

$$\boldsymbol{\theta}_{k+1} = \arg\min_{\theta} \frac{\boldsymbol{\theta}^T \mathbf{T}_G^T (\mathbf{R}_{CG}^{-T}(\boldsymbol{\theta}_k) \otimes \mathbf{R}_X^T)(\mathbf{R}_{CG}^{-1}(\boldsymbol{\theta}_k) \otimes \mathbf{R}_X) \mathbf{T}_G \boldsymbol{\theta}}{\boldsymbol{\theta}^T \mathbf{T}_G^T (\mathbf{R}_{CG}^{-T}(\boldsymbol{\theta}_k) \otimes \tilde{\mathbf{C}}_q^T)(\mathbf{R}_{CG}^{-1}(\boldsymbol{\theta}_k) \otimes \tilde{\mathbf{C}}_q) \mathbf{T}_G \boldsymbol{\theta}} \quad (39)$$

The above minimization can be performed via eigenvalue (EVD) or singular value (SVD) decomposition. Either way the GKL algorithm turns out to be iterative in nature.

It can be shown, however, that a slight modification of the loss function by Eq.(30) via introducing a weighting matrix results in a non-iterative version of the original GKL algorithm. The algorithm, referred to as WGKL algorithm, is derived and discussed in details in [8].

A key point in the classification of the identification methods discussed in this paper is to observe that both the GKL and WGKL algorithms represent direct methods to estimate the $\boldsymbol{\theta}$ parameter vector in the sense that a loss function is minimized directly by $\boldsymbol{\theta}$.

In the sequel another approach will be considered, namely the eigenvectors of the space spanned by the noisy input/output observations will first be estimated, then in a further step the $\boldsymbol{\theta}$ parameter vector estimation will be derived using the calculated eigenvectors ($\mathbf{V}$). The identification methods following the $\mathbf{X}_q \Rightarrow \mathbf{V} \Rightarrow \boldsymbol{\theta}$ pattern will be referred to as indirect algorithms. As the data set driving the estimation algorithm will be compressed from the space of the observations to a subspace spanned by the estimated eigenvectors, the indirect technique can also be qualified as a subspace identification method.

IV. Subspace Based Approach

The subspace parameter estimation methods, in general, follow a problem formulation using state-space models. To be in harmony with the results derived earlier in this paper, however, herewith along a transfer function approach will be followed.

The Subspace Identification (SID) algorithms minimize the loss function by Eq.(30) in two steps.

Step 1: $\mathbf{G}_{SID}$ is determined by

$$\mathbf{G}_{SID} = \arg\min_{G_q} tr\left((\mathbf{G}_q^T \mathbf{C}_q \mathbf{G}_q)^{-1} (\mathbf{G}_q^T \mathbf{D}_q \mathbf{G}_q) \right) \quad (40)$$

Observe that unlike in Eq.(33), here the $\mathbf{G}_q = \mathbf{G}_q(\boldsymbol{\theta})$ relation has not been taken into account. As far as the computing demand is concerned Eq.(40) leads to an eigenvalue-eigenvector (EVD) problem, just like Eq. (34), however, the dimension of the EVD problem to minimize Eq.(40) is larger than that of the EVD problem to minimize Eq.(33). In addition, the minimization by Eq.(33) needs an iterative procedure, while minimization by Eq. (40) is non-iterative. Considering Eq.(40) $\mathbf{G}_{SID}$ is constructed by the smallest $(q-m)$ eigenvectors belonging to the $(q-m)$ smallest eigenvalues. The minimum of the loss function by Eq.(40) turns out to be the sum of the $(q-m)$ smallest eigenvalues.

Consider a matrix $\mathbf{V}$ formed by the generalized eigenvectors belonging to the matrices $(\mathbf{D}_q, \mathbf{C}_q)$. According to the special structure of $\mathbf{C}_q$ and $\mathbf{D}_q$ numerical advantages can be achieved by SVD/GSVD decomposition. Using the singular vectors derived from the GSVD decomposition of $\mathbf{X}_q / \sqrt{N_q}$ and $\tilde{\mathbf{C}}_q$ result in the same $\mathbf{V}$ matrix. Separate the eigenvectors in $\mathbf{V}$ letting $\mathbf{V}_1$ involve $q+m$ eigenvectors belonging to the $q+m$ largest eigenvalues and $\mathbf{V}_2$ involve $q-m$ eigenvectors belonging to the $q-m$ smallest eigenvalues. Then $\mathbf{G}_{SID}$ is

$$\mathbf{G}_{SID} = \mathbf{V}_2 . \tag{41}$$

A key point is, however, that the solution of Eq. (40) is not unique in the sense that if $\mathbf{G}_{SID}$ is minimizing Eq.(40), then any

$$\mathbf{G}^*_{SID} = \mathbf{G}_{SID}\mathbf{P} \tag{42}$$

exhibits the same minimum, where $\mathbf{P}$ is an arbitrary non-singular matrix.

Step 2: solve the following overdetermined set of equations:

$$\mathbf{G}_q(\boldsymbol{\theta}) \approx \mathbf{G}_{SID}\mathbf{P} . \tag{43}$$

Observe that the eigenvectors in $\mathbf{V}_1$ and $\mathbf{V}_2$ satisfy

$$\mathbf{V}_1^T \mathbf{C}_q \mathbf{V}_2 = \mathbf{0} . \tag{44}$$

This means that in order to find the parameters Eq.(43) can be replaced by

$$\mathbf{V}_1^T \mathbf{C}_q \mathbf{G}_q(\boldsymbol{\theta}) \approx \mathbf{0} . \tag{45}$$

V. Parameter Estimation from Eigenvectors

In this section the $\mathbf{V} \Rightarrow \boldsymbol{\theta}$ mapping of the $\mathbf{X}_q \Rightarrow \mathbf{V} \Rightarrow \boldsymbol{\theta}$ pattern of the indirect identification will be presented. In particular, three parameter estimation algorithms will be discussed. It will be shown that the LS type solutions result in biased parameter estimation in a non-iterative way, however, they do not require the knowledge of the covariance matrix of the estimated eigenvectors. The other two algorithms, the non-iterative TLS and the iterative optimal solutions, both use the estimated eigenvector covariance matrix, whose calculation remains to be seen in Section VI.

A. LS Type Solution

Based on Eq.(43) one of the possible solutions for the parameter estimation is to minimize the error terms exhibited by Eq.(43) and Eq.(45) in Least Squares sense:

$$\boldsymbol{\theta}_{vec_1} = \arg\min_{\theta} \min_{P} vec(\mathbf{G}_q(\boldsymbol{\theta}) - \mathbf{V}_2\mathbf{P})^T \mathbf{W}_1 vec(\mathbf{G}_q(\boldsymbol{\theta}) - \mathbf{V}_2\mathbf{P}) \tag{46}$$

$$\boldsymbol{\theta}_{vec_2} = \arg\min_{\theta} vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1)^T \mathbf{W}_2 vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1) \tag{47}$$

Observe the weighting matrices applied in the above quadratic forms. Taking Eq.(44) into account it is seen that selecting

$$\mathbf{W}_1 = (\mathbf{I}_{q-m} \otimes \mathbf{C}_q\mathbf{V}_1)\mathbf{W}_2(\mathbf{I}_{q-m} \otimes \mathbf{V}_1^T\mathbf{C}_q) \tag{48}$$

the loss functions involved in finding $\boldsymbol{\theta}_{vec_1}$ and $\boldsymbol{\theta}_{vec_2}$ are identical. As far as the size of the weighting matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ is concerned, the $vec\{...\}$ operations in Eq.(46) and (47) require $\mathbf{W}_1$ and $\mathbf{W}_2$ to be large. The calculations required by Eq.(46) and Eq.(47) to find the parameter vector $\boldsymbol{\theta}_{vec.}$ can significantly be reduced if the weighting matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ take a special, block diagonal form:

$$\mathbf{W}_1 = \mathbf{I}_{q-m} \otimes \bar{\mathbf{W}}_1 \tag{49}$$

and

$$\mathbf{W}_2 = \mathbf{I}_{q+m} \otimes \bar{\mathbf{W}}_2 \tag{50}$$

respectively. In this case Eq.(46) and Eq.(47) can be re-formulated by

$$\boldsymbol{\theta}_{tr_1} = \arg\min_{\theta} \min_{P} tr((\mathbf{G}_q(\boldsymbol{\theta}) - \mathbf{V}_2\mathbf{P})^T \bar{\mathbf{W}}_1 (\mathbf{G}_q(\boldsymbol{\theta}) - \mathbf{V}_2\mathbf{P})) \tag{51}$$

and

$$\boldsymbol{\theta}_{tr_2} = \arg\min_{\theta} tr(\mathbf{V}_1^T\mathbf{C}_q\mathbf{G}_q(\boldsymbol{\theta})\bar{\mathbf{W}}_2\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1) \tag{52}$$

respectively.

The weighting matrices, however, can not be selected arbitrarily. Apart from some special cases the estimation obtained via minimizing the loss functions Eqs.(46), (47), (51) or (52) will all be biased. The point is that the weighting matrices leading to unbiased estimation depend on the parameters just to be identified. However, as the parameter estimation is based on compressed (in some sense filtered) set of data the bias is not significant.

B. TLS Type Solution

To derive procedures resulting in unbiased estimations the loss functions involved should be modified according to the Total Least Squares concept. Since the eigenvectors are calculated from $\mathbf{X}_q$ containing noisy observations, the calculated eigenvectors $\mathbf{V}_1$ have noisy components, as well. Denote the noise part by $\mathbf{N}_1 = \mathbf{V}_1 - E\{\mathbf{V}_1\}$ which has zero mean. Considering Eq.(47) the condition to obtain unbiased estimation is to minimize the modified loss function according to

$$\boldsymbol{\theta}_{vec} = \arg\min_{\theta} \frac{vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1)^T \mathbf{W} vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1)}{E\{vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{N}_1)^T \mathbf{W} vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{N}_1)\}} \tag{53}$$

or in another form

$$\boldsymbol{\theta}_{vec} = \arg\min_{\theta} \frac{vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1)^T \mathbf{W} vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1)}{tr\{(\mathbf{I}_{q+m} \otimes \mathbf{C}_q\mathbf{G}_q(\boldsymbol{\theta}))\mathbf{W}(\mathbf{I}_{q+m} \otimes \mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q)\,\mathrm{cov}(vec(\mathbf{N}_1))\}} \tag{54}$$

where $\mathbf{W}$ is the weighting matrix of appropriate size. Similarly, Eq.(52) needs to be modified as

$$\boldsymbol{\theta}_{tr} = \arg\min_{\theta} \frac{tr(\bar{\mathbf{W}}\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1\mathbf{V}_1^T\mathbf{C}_q\mathbf{G}_q(\boldsymbol{\theta}))}{tr(\bar{\mathbf{W}}\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q E\{\mathbf{N}_1\mathbf{N}_1^T\}\mathbf{C}_q\mathbf{G}_q(\boldsymbol{\theta}))} \quad (55)$$

The same procedure can be repeated for Eqs.(46) and (51), though the size of the weighting matrix is higher for those cases.

C. Optimal Solution

In the rest of this Section the optimal settings for the weighting matrices used in Eqs. (54) and (55) will be determined. Specifically, $\mathbf{W}_{opt}$ will be considered optimal if the matrix

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}_W) - \mathrm{cov}(\hat{\boldsymbol{\theta}}_{W_{opt}}) \quad (56)$$

remains positive semidefinite for any possible weighting matrix $\mathbf{W}$. The theoretical background can be found in [9]. An asymptotically best consistent estimation for $\hat{\boldsymbol{\theta}}$ can be obtained for (Eq. 56) by solving

$$\boldsymbol{\theta}_{vec}^* = \arg\min_{\theta} vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1)^T\mathbf{R}^{-1}vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1) \quad (57)$$

where

$$\mathbf{W}_{opt}^{-1} = \mathbf{R} = E\{vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{N}_1)vec(\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{N}_1)^T\} \quad (58)$$

Then straightforward algebraic manipulations allows Eq. (58) to be written as

$$\mathbf{R} = (\mathbf{I}_{q+m}\otimes\mathbf{G}_q^T\mathbf{C}_q)E[vec(\mathbf{N}_1)vec(\mathbf{N}_1)^T](\mathbf{I}_{q+m}\otimes\mathbf{C}_q\mathbf{G}_q)$$
$$= (\mathbf{I}_{q+m}\otimes\mathbf{G}_q^T\mathbf{C}_q)\,\mathrm{cov}(vec(\mathbf{N}_1))(\mathbf{I}_{q+m}\otimes\mathbf{C}_q\mathbf{G}_q) \quad (59)$$

In the very same way the asymptotically best consistent estimation for $\boldsymbol{\theta}$ assuming the application of a reduced size optimal weighting matrix can be determined as follows:

$$\boldsymbol{\theta}_{tr}^* = \arg\min_{\theta} tr(\mathbf{V}_1^T\mathbf{C}_q\mathbf{G}_q(\boldsymbol{\theta})\bar{\mathbf{R}}^{-1}\mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q\mathbf{V}_1) \quad (60)$$

where

$$\bar{\mathbf{R}} = \mathbf{G}_q^T(\boldsymbol{\theta})\mathbf{C}_q E\{\mathbf{N}_1\mathbf{N}_1^T\}\mathbf{C}_q\mathbf{G}_q(\boldsymbol{\theta}) \quad (61)$$

It is seen that the calculations both in Eqs.(57) and (61) need the knowledge of the covariance matrix of the eigenvectors. This is to be determined in the next Section.

As far as the selection of the $\mathbf{G}_q$ matrix as a key item in the derivation of various estimation algorithms is concerned, it should be noted, that similar results can be obtained in a straightforward manner by using appropriate Sylvester matrices.

VI. Estimation of Covariance of Eigenvectors

As the observation matrix $\mathbf{X}_q$ contains the noisy input and output samples, the covariance matrix can well be approximated using the singular value decomposition of $\mathbf{X}_q$. Consider the singular value decomposition of the matrix

$$\mathbf{X}_q\tilde{\mathbf{C}}_q^{-1}/\sqrt{N_q} \quad (62)$$

as follows:

$$\mathbf{X}_q\tilde{\mathbf{C}}_q^{-1}/\sqrt{N_q} = \mathbf{U}_s\mathbf{S}_s\mathbf{V}_s^T \quad (63)$$

Separating the dominant (largest $q+m$ singular values) and non-dominant (smallest $q-m$ singular values) components in the above decomposition by

$$\mathbf{U}_s\mathbf{S}_s\mathbf{V}_s^T = \mathbf{U}_{s1}\mathbf{S}_{s1}\mathbf{V}_{s1}^T + \mathbf{U}_{s2}\mathbf{S}_{s2}\mathbf{V}_{s2}^T \quad (64)$$

use only the dominant components to derive the following expression:

$$\mathbf{V}_1 = \tilde{\mathbf{C}}_q^{-T}\mathbf{X}_q^T\mathbf{U}_{s1}\mathbf{S}_{s1}^{-1}/\sqrt{N_q} \quad (65)$$

The covariance matrix of the eigenvectors can then be approximated using Eq.(65) as follows:

$$\mathrm{cov}(vec(\mathbf{N}_1)) \approx$$
$$\frac{1}{N_q}\{\sigma_u^2\sum_{k=1}^{N} vec(\tilde{\mathbf{C}}_q^{-T}\delta_{u_k}\mathbf{X}_q^T\mathbf{U}_{s1}\mathbf{S}_{s1}^{-1})vec^T(\tilde{\mathbf{C}}_q^{-T}\delta_{u_k}\mathbf{X}_q^T\mathbf{U}_{s1}\mathbf{S}_{s1}^{-1})$$
$$+\sigma_y^2\sum_{k=1}^{N} vec(\tilde{\mathbf{C}}_q^{-T}\delta_{y_k}\mathbf{X}_q^T\mathbf{U}_{s1}\mathbf{S}_{s1}^{-1})vec^T(\tilde{\mathbf{C}}_q^{-T}\delta_{y_k}\mathbf{X}_q^T\mathbf{U}_{s1}\mathbf{S}_{s1}^{-1})\} \quad (66)$$

To obtain Eq.(66) independence of the noise components has been taken into account and the notations

$$\delta_{u_k}\mathbf{X}_q = \frac{\partial\mathbf{X}_q}{\partial u_k} \text{ and } \delta_{y_k}\mathbf{X}_q = \frac{\partial\mathbf{X}_q}{\partial y_k} \quad (67)$$

have been introduced. Consider just one block in Eq.(66):

$$\mathrm{cov}(vec(\mathbf{N}_1))_{i,j} \approx \begin{bmatrix}\mathbf{C}_{ij}^u & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{ij}^y\end{bmatrix} \quad (68)$$

where

$$\mathbf{C}_{ij}^u \approx \frac{1}{N_q}\sum_{k=1}^{N} vec(\delta_{u_k}\mathbf{H}_q^u\mathbf{u}_i/s_i)vec^T(\delta_{u_k}\mathbf{H}_q^u\mathbf{u}_j/s_j) = \frac{1}{N_q}\mathbf{C}_{ij} \quad (69)$$

$$\mathbf{C}_{ij}^y \approx \frac{1}{N_q}\sum_{k=1}^{N} vec(\delta_{y_k}\mathbf{H}_q^y\mathbf{u}_i/s_i)vec^T(\delta_{y_k}\mathbf{H}_q^y\mathbf{u}_j/s_j) = \frac{1}{N_q}\mathbf{C}_{ij} \quad (70)$$

In (Eqs.69-70) $\mathbf{u}_i$ and s_i are the appropriate components of matrix $\mathbf{U}_{s1}$ and $\mathbf{S}_{s1}$. Introducing

$$\bar{\mathbf{u}}_i = \begin{bmatrix}\mathbf{0}_{q-1}^T & \mathbf{u}_i^T/s_i & \mathbf{0}_{q-1}^T\end{bmatrix}^T \quad (71)$$

and

$$\mathbf{T}_i = Toeplitz(\bar{\mathbf{u}}_i, q) \quad (72)$$

$\mathbf{C}_{ij}$ takes a compact, easy to code form:

$$\mathbf{C}_{ij} = \mathbf{T}_i^T\mathbf{T}_j \quad (73)$$

Finally, for the sub-blocks of the covariance matrix

$$\mathrm{cov}(vec(\mathbf{N}_1))_{i,j} \approx \frac{1}{N_q}\begin{bmatrix}\mathbf{T}_i^T\mathbf{T}_j & \mathbf{0}\\ \mathbf{0} & \mathbf{T}_i^T\mathbf{T}_j\end{bmatrix} \quad (74)$$

is obtained.

In the knowledge of the covariance matrix of the eigenvectors the $\boldsymbol{\theta}_{vec}^*$ and $\boldsymbol{\theta}_{tr}^*$ parameter estimations can be realized according to Eqs.(57) and (60). Note that both estimation algorithms are iterative, the iteration procedures – including the aspects related to the numerically robust realization - follow the concept shown for the GKL

algorithm.

VII. Simulation

The identification methods discussed in the paper have been investigated in a Monte Carlo simulation study using the following linear, continuous-time, second order process:

$$y_o = \frac{1}{0.1s^2 + 0.2s + 1} u_o \qquad (75)$$

The above process is sampled by $T_s = 0.05$ sec. Also, assume additive white noise with variance of 0.03 acting on both the input and the output. Number of samples processed by the identification algorithms is 180. As far as the excitation is concerned, the input is raised from 0 to 1 at step 20, and then set to -1 at step 100.

Using the same simulation environment several identification runs over 180 samples have been evaluated. For each identification run the q scaling factor (essentially the size of the Hankel matrices used) sweeps the range from $q=m+1=3$ to $q=10$. As an overall measure for the effectiveness of the parameter estimation the empirical standard deviation have been calculated:

$$s_k(q) = \sqrt{\frac{1}{L}\sum_{j=1}^{L}(\hat{a}_{k,q,j} - a_k)^2} \qquad (76)$$

where j denotes the index of the simulation run, $L=100$ stands for the number of simulation runs performed, while a_k $(k=1,2)$ are the discrete time system coefficients.

Results of the simulation study related to the a_1 coefficient are shown in Fig. 1. In Fig.1. the notation S-VEC and S-TR are used for the parameter estimation $\boldsymbol{\theta}^*_{vec}$ and $\boldsymbol{\theta}^*_{tr}$, respectively. Similar results have been obtained for a_2. The results are in harmony with the expectations, namely the more information is used by a parameter estimation procedure, the less variance in the estimation can be achieved. Consequently, the GKL algorithm working with the smallest amount of information delivers the most modest results. On the other hand, $\boldsymbol{\theta}^*_{vec}$ requires the highest amount of information and results in an estimation with the smallest variance.

VIII. Conclusion

Direct and indirect parameter estimation algorithms for the EIV identification problem have been discussed in the paper. The GKL algorithm representing the direct methods is a scaleable algorithm with $m+1 \le q \le N$ and only using the information capsulated in the $\mathbf{X}_q^T\mathbf{X}_q$ matrix. The size of the matrix to be inverted is $q-m$. The dimension of the eigenvalue/eigenvector or SVD problem to be solved is identical to the number of the parameters to be estimated: $2m+1$ (assuming $b_0 = 0$). The indirect algorithms require a data compression step, then the parameter estimation is based on an SVD decomposition with a dimension of $2q$. Further on, the most effective indirect algorithms need to calculate the covariance matrix of the eigenvectors/singular vectors. Unfortunately, this can not be calculated from the $\mathbf{X}_q^T\mathbf{X}_q$ matrix. The optimal weighting matrix can be obtained via matrix inversion. The computing demand is drastically increasing by q. In case of $\boldsymbol{\theta}^*_{vec}$ the size of the weighting matrix is $(q-m)*(q+m) = q^2 - m^2$. In practice, it is reasonable to avoid the inversion of matrices exceeding the size of 50×50. The calculation demand is worthwhile to reduce. Using $\boldsymbol{\theta}^*_{tr}$ the size of the weighting matrix drops significantly, namely it decreases to the value of $q-m$.

References

[1] D. Bauer,D, "Subspace algorithms", 13th IFAC Symposium on System Identification, Rotterdam, The Netherlands, 2003, pp. 1030-1041.

[2] C. Chou and M. Verheagen, "Subspace algorithms for the identification of multivariate errors-in-variables models", *Automatica,* Vol.33. No.10, 1997, pp. 1857-1869.

[3] M. Viberg, "Subspace-based methods for the identification of linear time-invariant systems", *Automatica*, Vol.31. No.12, 1995, pp. 1835-1851.

[4] P. Stoica, M. Cedervall and A. Eriksson, "Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output system", *IEEE Transactions on Signal Processing,* Vol.43, No.10. 1995, pp.2386-97.

[5] T. Söderström, U. Soverini and K. Mahata, "Perspectives on errors-in-variables estimation for dynamic systems", *Signal Processing*, 82, 2002, pp.1139-1154.

[6] W.Li and S.J.Qin, "Consistent dynamic PCA based on errors-in-variables subspace identification", J. of Process Control, Vol.11, 2001, pp. 661-678.

[7] J.Wang and S.J.Qin, "A new subspace identification approach based on principal component analysis", J. of Process Control, Vol. 12, 2002, pp.841-855.

[8] I. Vajk, "Identification methods in a unified framework", *Automatica* 2005, (to be published)

[9] T. Söderström, "System identification", Prentice Hall International (UK) Ltd, 1989.

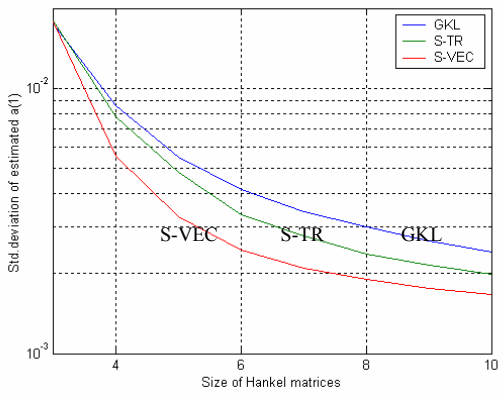

Fig. 1. Variation of the standard deviation of the a_1 parameter estimation with respect to the size of the Hankel matrices applied