Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Efficient Estimation of Errors-in-Variables Models

István Vajk, Jenő Hetthéssy

Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Budapest, Hungary, (e-mail: {vajk,hetthessy}@aut.bme.hu)

Abstract: The paper addresses the discrete-time linear process identification problem assuming noisy input and output records available for the parameter estimation. The efficient algorithms are derived for the simultaneous estimation of the process and noise parameters. Implementation techniques based on matrix and polynomial decompositions are given in details resulting in estimation algorithms with reduced computation demand. The presented algorithms deliver the parameter estimation in an iterative way as a result of a generalized eigenvalue-eigenvector or a generalized singular value decomposition.

1. INTRODUCTION

Various approaches of system identification with noisy input-output records have been with us for decades. It is enough to refer to the works of Koopmans, Frisch, Levin or Aoki. This classical field has been discussed with new techniques by a number of authors recently. Excellent reviews can be found in a survey paper (Söderström, 2007) or in regular papers (Diversi et al., 2007), as well. In addition, the structural total least squares technique (Markovsky et al., 2005), the PCA approach (Wang and Qin, 2002), as well as several topics elaborated on the 14th IFAC Symposium on System Identification (Markovsky, et al. 2006), (Ekman, et al. 2006) should be mentioned as recent results. The particular aim of this paper is to present an efficient simultaneous estimation procedure for the process and noise parameters in EIV environment. The new algorithm results in the estimation via iterative generalized eigenvalue-eigenvector (EVD) or generalized singular value decomposition (SVD). Unlike the widely used gradient techniques, the EVD/SVD techniques deliver several local minima while minimizing the loss function associated with the performance of the parameter estimation.

The paper is organized as follows. Section 2 formulates the identification problem in EIV environment. In Section 3 the maximum likelihood estimation is discussed as a minimization problem of an appropriate loss function. Section 4 is devoted to discuss efficient realizations for the parameter estimation. Three options are presented: the method using matrix decompositions reduces the computation demand without reducing the performance of the parameter estimation, while the method using polynomial decompositions introduces some approximations via the autoregressive filtering. The third method is using an approximation related to the gradient of the loss function. Section 5 presents an important extension of the EIV model identification, namely a method is derived to identify not only the process parameters, but the noise coefficients, too. In Section 6 a simulation example is given to illustrate the behavior of the extended algorithm.

2. PROBLEM FORMULATION

In the paper single-input single-output (SISO) linear time-invariant discrete time systems will be studied. Describe the noise-free process by

$$A(q^{-1})y_k^o = B(q^{-1})u_k^o \ , \tag{1}$$

where

$$q^{-1}x_k = x_{k-1} \tag{2}$$

is the backward shift operator, further on

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

and

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

Eq. (1) can be transformed to an implicit form of

$$\boldsymbol{\theta}^T \mathbf{x}_k^o = 0 \ , \tag{5}$$

where

$$\mathbf{x}_k^o = [y_{k-m}^o, y_{k-m+1}^o, ..., y_k^o, u_{k-m}^o, u_{k-m+1}^o, ..., u_k^o]^T \tag{6}$$

is the complete observation vector and

$$\boldsymbol{\theta} = [\mathbf{a}^T \quad -\mathbf{b}^T]^T \tag{7}$$

is the parameter vector with

$$\mathbf{a} = [a_m, a_{m-1}, ..., a_1, 1]^T \quad \text{and} \quad \mathbf{b} = [b_m, b_{m-1}, ..., b_1, b_0]^T , \tag{8}$$

provided that the b_0 coefficient is set to $b_0 = 0$. Note that b_0 has just been introduced to ensure a symmetrical structure in the observation vector. Later on **a** and **b** with reversed sequencing as

$$\mathbf{a}^* = [1, a_1, ..., a_{m-1}, a_m]^T \quad \text{and} \quad \mathbf{b}^* = [b_0, b_1, ..., b_{m-1}, b_m]^T \tag{9}$$

will also be applied.

To complete the EIV model, according to Fig. 1, additive noise components will be taken into account and N pair of noisy observations will be assumed to be available to estimate the unknown a_i and b_i coefficients:

$$u_k = u_k^o + \tilde{u}_k \qquad \text{var}(\tilde{u}_k) = \sigma_u^2 = \mu \cos^2 \varphi \tag{10}$$

$$y_k = y_k^o + \tilde{y}_k \qquad \text{var}(\tilde{y}_k) = \sigma_y^2 = \mu \sin^2 \varphi \tag{11}$$

Regarding the identification problem just outlined, the final goal is to derive a parameter estimation algorithm assuming that both the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ ratio and the noise intensity parameter μ are *unknown*. In Eqs.(10) and (11) observe the parameterization of the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio by φ, indicating the noise 'direction'. As far as the noise covariance matrix is concerned the notation by

$$\mathbf{C}_{yu} = \mu\bar{\mathbf{C}}_{\varphi}^{T}\bar{\mathbf{C}}_{\varphi} = \mu\mathbf{C}_{\varphi} \tag{12}$$

will be used in the sequel, where

$$\bar{\mathbf{C}}_{\varphi} = \begin{bmatrix} \sin\varphi & 0 \\ 0 & \cos\varphi \end{bmatrix}. \tag{13}$$

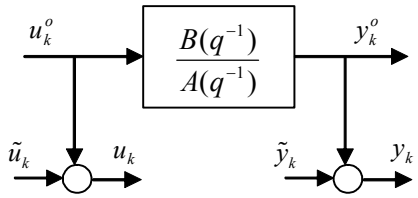

Fig. 1. Errors-in-variables model

To prepare the mathematical treatment of the identification problem a few notations need to be introduced at this point:

$$\mathbf{0}_n = [0,0,...,0]^T, \qquad \dim(\mathbf{0}_n)=n, \tag{14}$$

$$\mathbf{0}_{n,n}: \quad n\times n \text{ zero matrix}, \tag{15}$$

$$\mathbf{I}_n: \quad n\times n \text{ unity matrix}, \tag{16}$$

$$Toeplitz(\mathbf{v},q) = \begin{bmatrix} v_q & v_{q-1} & \cdots & v_1 \\ v_{q+1} & v_q & \cdots & v_2 \\ \cdots & \cdots & \ddots & \cdots \\ v_N & v_{N-1} & \cdots & v_{N-q+1} \end{bmatrix}, \tag{17}$$

$$Hankel(\mathbf{v},q) = \begin{bmatrix} v_1 & v_2 & \cdots & v_q \\ v_2 & v_3 & \cdots & v_{q+1} \\ \cdots & \cdots & \ddots & \cdots \\ v_{N-q+1} & v_{N-q+2} & \cdots & v_N \end{bmatrix} \tag{18}$$

where $\mathbf{v} = [v_1, v_2, ..., v_N]^T$.

Though only SISO systems are considered in the paper, all the results shown in the following sections can be generalized for multiple-input multiple-output (MIMO) systems.

3. MAXIMUM LIKELIHOOD ESTIMATION

The N pair of the input/output samples available for the parameter estimation will be arranged in several structures in the sequel. Starting the discussion with the *maximum likelihood (ML) parameter estimation* assume that the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio is known and in order to derive a compact form let us put all the noisy and noise-free observations into long observation vectors, respectively:

$$\mathbf{u} = [u_1, u_2, ..., u_N]^T, \quad \mathbf{u}_o = [u_1^o, u_2^o, ..., u_N^o]^T \tag{19}$$

$$\mathbf{y} = [y_1, y_2, ..., y_N]^T, \quad \mathbf{y}_o = [y_1^o, y_2^o, ..., y_N^o]^T, \tag{20}$$

$$\mathbf{x} = [\mathbf{y}^T, \mathbf{u}^T]^T, \qquad \mathbf{x}^o = [\mathbf{y}_o^T, \mathbf{u}_o^T]^T, \tag{21}$$

then introduce an appropriate

$$\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{G}_a & -\mathbf{G}_b \end{bmatrix}^T \tag{22}$$

matrix containing the model parameters such that vector equation

$$\mathbf{G}^T\mathbf{x}^o = \mathbf{0} \tag{23}$$

involves the scalar equations

$$\boldsymbol{\theta}^T\mathbf{x}_k^o = 0 \text{ for } m+1 \le k \le N. \tag{24}$$

Both $\mathbf{G}_a$ and $\mathbf{G}_b$ turn out to be matrices of Toeplitz type:

$$\mathbf{G}_a = Toeplitz([\mathbf{0}_{N-m-1}; \mathbf{a}^*; \mathbf{0}_{N-m-1}], N-m) \tag{25}$$

and

$$\mathbf{G}_b = Toeplitz([\mathbf{0}_{N-m-1}; \mathbf{b}^*; \mathbf{0}_{N-m-1}], N-m), \tag{26}$$

respectively. As far as the likelihood function is concerned, if the noise components are of Gaussian distribution then the conditional distribution of the measurements is

$$prob\left(\mathbf{x}\middle|\boldsymbol{\theta}\right) = const.\exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^o)^T(\mu\mathbf{C})^{-1}(\mathbf{x}-\mathbf{x}^o)\right) \tag{27}$$

where

$$\mathbf{C} = \mathbf{C}_{\varphi} \otimes \mathbf{I}_N. \tag{28}$$

Here $\otimes$ denotes Kronecker product. Taking the constraint by $\mathbf{G}^T\mathbf{x}^o = \mathbf{0}$ into account finding the maximum of the likelihood function is equivalent to minimize the following loss function by $\boldsymbol{\theta}$:

$$J(\boldsymbol{\theta}) = \frac{1}{2(N-m)}\mathbf{x}^T\mathbf{G}(\boldsymbol{\theta})\left(\mathbf{G}(\boldsymbol{\theta})^T\mathbf{C}\mathbf{G}(\boldsymbol{\theta})\right)^{-1}\mathbf{G}(\boldsymbol{\theta})^T\mathbf{x}. \tag{29}$$

Note that the above expression of the loss function can be re-arranged as

$$J(\boldsymbol{\theta}) = \frac{1}{2(N-m)}tr\left\{\left(\mathbf{G}(\boldsymbol{\theta})^T\mathbf{C}\mathbf{G}(\boldsymbol{\theta})\right)^{-1}\left(\mathbf{G}(\boldsymbol{\theta})^T\mathbf{D}\mathbf{G}(\boldsymbol{\theta})\right)\right\}, \tag{30}$$

where the notation of

$$\mathbf{D} = \mathbf{x}\mathbf{x}^T \tag{31}$$

has been used for the data matrix.

To find the minimum of the loss function by Eq. (29) consider its derivative with respect to the *i*-th entry of the $\boldsymbol{\theta}$ vector (θ_i)

$$\begin{aligned}(N-m)\frac{\partial J}{\partial\theta_i} &= \mathbf{x}^T\mathbf{G}[\mathbf{G}^T\mathbf{C}\mathbf{G}]^{-1}\delta_i\mathbf{G}^T\mathbf{x} - \\ &\mathbf{x}^T\mathbf{G}[\mathbf{G}^T\mathbf{C}\mathbf{G}]^{-1}\delta_i\mathbf{G}^T\mathbf{C}\mathbf{G}[\mathbf{G}^T\mathbf{C}\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{x}\end{aligned} \tag{32}$$

where $\delta_i\mathbf{G}$ is a shorthand notation for

$$\delta_i\mathbf{G} \triangleq \partial\mathbf{G}/\partial\theta_i. \tag{33}$$

To proceed the necessary condition

$$\partial J/\partial\theta_i = 0 \tag{34}$$

will be investigated. Having as many equations as parameters to be determined, in principle there might be a way to solve the above set of equations.

Once the process parameters have been estimated the noise gain μ can be calculated as follows:

$$\hat{\mu} = \mathbf{x}^T \mathbf{G}(\hat{\boldsymbol{\theta}}) \left(\mathbf{G}(\hat{\boldsymbol{\theta}})^T \mathbf{C} \mathbf{G}(\hat{\boldsymbol{\theta}}) \right)^{-1} \mathbf{G}(\hat{\boldsymbol{\theta}})^T \mathbf{x} / (N - m - n_p) . \quad (35)$$

Note that the above relation delivers a consistent unbiased estimation for $\hat{\mu}$, where n_p denotes the number of the estimated process parameters.

However, only iterative solutions can be looked for to solve Eq.(34), because $\mathbf{G}$ depends on $\boldsymbol{\theta}$. MATLAB, for example, offers tools to solve a set of nonlinear equations, however, in practice these tools work effectively only if favorable initial conditions are set. In this paper EVD or SVD algorithms will be considered rather than classical gradient algorithms. Namely unlike traditional local minimization algorithms (e.g. the Levenberg-Marquardt method), the EVDS/SVD algorithms can find several local minima.

Elaborating on Eq. (34) leads to

$$(\mathbf{D}^* - \mathbf{C}^*)\boldsymbol{\theta} = \mathbf{0} , \quad (36)$$

where the matrices $\mathbf{D}^*$ and $\mathbf{C}^*$ have the following entries:

$$D^*_{i,j} = \mathbf{x}^T \delta_j \mathbf{G} [\mathbf{G}^T \mathbf{C} \mathbf{G}]^{-1} \delta_i \mathbf{G}^T \mathbf{x} \quad (37)$$

$$C^*_{i,j} = \mathbf{x}^T \mathbf{G} [\mathbf{G}^T \mathbf{C} \mathbf{G}]^{-1} \delta_i \mathbf{G}^T \mathbf{C} \delta_j \mathbf{G} [\mathbf{G}^T \mathbf{C} \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{x} . \quad (38)$$

While calculating the $C^*_{i,j}$ elements, considerable reduction in the related computation demand can be achieved if $C^*_{i,j}$ is substituted with its expected value (this can be a rather good approximation for large *N)*:

$$E\{C^*_{i,j}\} = \mu \, tr\{[\mathbf{G}^T \mathbf{C} \mathbf{G}]^{-1} \delta_i \mathbf{G}^T \mathbf{C} \delta_j \mathbf{G}\} = \mu C^o_{i,j} , \quad (39)$$

where $C^o_{i,j}$ is the *i,j* element of $\mathbf{C}^o$. Accepting the proposed approximation Eq. (36) takes the following form:

$$(\mathbf{D}^* - \mu \mathbf{C}^o)\boldsymbol{\theta} = \mathbf{0} . \quad (40)$$

Consequently, Eq.(40) can be solved as an iterative EVD problem:

$$(\mathbf{D}^*(\boldsymbol{\theta}_\ell) - \mu_{\ell+1} \mathbf{C}^o(\boldsymbol{\theta}_\ell))\boldsymbol{\theta}_{\ell+1} = \mathbf{0} . \quad (41)$$

According to Eq.(41) in step ℓ $\boldsymbol{\theta}_\ell$ is given, $\mathbf{D}^*(\boldsymbol{\theta}_\ell)$ and $\mathbf{C}^o(\boldsymbol{\theta}_\ell)$ can be calculated, then the solution of the EVD problem results in $\boldsymbol{\theta}_{\ell+1}$ and $\mu_{\ell+1}$.

Remark: if $\mathbf{D}^*$, $\mathbf{C}^*$ and $\mathbf{C}^o$ are available in a decomposed form by

$$\mathbf{D}^* = \bar{\mathbf{D}}^{*T} \bar{\mathbf{D}}^* \quad \mathbf{C}^* = \bar{\mathbf{C}}^{*T} \bar{\mathbf{C}}^* \quad \mathbf{C}^o = \bar{\mathbf{C}}^{oT} \bar{\mathbf{C}}^{*o} , \quad (42)$$

then the generalized EVD decomposition of $(\mathbf{D}^*, \mathbf{C}^*)$ and $(\mathbf{D}^*, \mathbf{C}^o)$ can be calculated in a numerically advanced way performing the generalized SVD decomposition of $(\bar{\mathbf{D}}^*, \bar{\mathbf{C}}^*)$ and $(\bar{\mathbf{D}}^*, \bar{\mathbf{C}}^o)$.

Analyzing the size of the matrices involved in the solution discussed above, one can conclude that the above proposed solution can hardly be applied for practical identification problems feeding the algorithms with a number of records. The next Section is devoted to derive parameter estimation algorithms avoiding the direct inversion of the matrix $\mathbf{G}^T \mathbf{C} \mathbf{G}$.

4. EFFICIENT REALIZATIONS

Summing up the main result derived in the previous Section a maximum likelihood estimation algorithm has been elaborated for process parameter identification in an EIV environment. At the same time Eqs. (36-38) show that the implementation of the algorithm may have high computation demand. Specifically, the straightforward calculation of the gradient by Eq.(32) requires $O(N^3)$ flops (floating point operations) in each iteration step, where N characterizes the size of the matrices involved in the calculations related to the estimation algorithm. Consequently, the complexity of the identification algorithm can be reduced by reducing the base in N^3 or by reducing the exponent in N^3. The first option is known as the generalization of the Koopmans-Levin method and it reduces the size of the matrices involved in the calculations (Vajk, 2005). The second option exhibits an efficient calculation of the gradient $\partial J / \partial \theta_i$. This alternative will be discussed in details in the rest of this Section.

One way to reduce the calculation effort is to utilize the special banded structure for the $\mathbf{G}^T \mathbf{C} \mathbf{G}$ matrix. In the light of this concept three alternative solutions will be shown below. All of three algorithms significantly reduce the required calculations, namely the evaluation of $\partial J / \partial \theta_i$ will need only *O(N)* flops.

4.1. Reduction of the calculation effort via matrix decomposition

The key idea behind the reduction is the utilization of the special banded structure of the $\mathbf{G}^T \mathbf{C} \mathbf{G}$ matrix:

$$\boldsymbol{\Gamma} = \mathbf{G}^T \mathbf{C} \mathbf{G} = Toplitz(\mathbf{0}_{N-2m-1}, [\gamma_m \ldots \gamma_1 \; \gamma_0 \; \gamma \ldots \gamma_m], \mathbf{0}_{N-2m-1}) \quad (43)$$

The above $(N-m)\times(N-m)$ $\boldsymbol{\Gamma}$ matrix can be decomposed via Cholesky transformation to the product of an $\mathbf{L}$ lower and a $\mathbf{U} = \mathbf{L}^T$ upper band-limited triangular matrix: $\mathbf{G}^T \mathbf{C} \mathbf{G} = \mathbf{L}\mathbf{U}$. Alternatively, avoiding the square root operation the Bierman decomposition can be applied to have $\mathbf{G}^T \mathbf{C} \mathbf{G} = \bar{\mathbf{L}} \bar{\mathbf{D}} \bar{\mathbf{U}}$, where $\bar{\mathbf{L}}$ is a lower and $\bar{\mathbf{U}}$ is an upper triangular matrix, while $\bar{\mathbf{D}}$ is a diagonal matrix. Either way, the separated matrices can be stored as matrices in a compact storage form with size of $(N-m)\times(m+1)$. Moreover, the number of the operations required by the matrix decompositions is proportional to N. The elements of the $\boldsymbol{\Gamma}$ matrix can be determined by the following polynomial relations:

$$\gamma(q) = \sin^2 \varphi \; A(q)A(q^{-1}) + \cos^2 \varphi \; B(q)B(q^{-1}) . \quad (44)$$

Observe the symmetrical structure in $\gamma(q)$:

$$\gamma(q) = \gamma_m q^{-m} + ... + \gamma_1 q^{-1} + \gamma_o + \gamma_1 q + ... + \gamma_m q^m . \quad (45)$$

In the sequel the case of the Cholesky decomposition will be considered and the goal of the procedures is to derive the matrices $\bar{\mathbf{D}}^*$ and $\bar{\mathbf{C}}^*$ to activate the SVD decomposition. This matrix decomposition algorithm will be presented in details. As the fundamental point of the algorithm, first the steps of the $\mathbf{\Gamma} = \mathbf{G}^T\mathbf{C}\mathbf{G} = \mathbf{L}\mathbf{U}$ decomposition should be performed:

1. Find $\gamma_i \;\; i = 0,1,...,m$ based on the knowledge available for **a**, **b** and the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio using Eq.(44).

2. Find $\mathbf{U}$ from $\mathbf{\Gamma} = \mathbf{G}^T\mathbf{C}\mathbf{G} = \mathbf{L}\mathbf{U}$ using the compact storage form. Note that applying the Cholesky decomposition $\mathbf{L}$ can automatically be generated once $\mathbf{U}$ has been determined.

Having the $\mathbf{\Gamma} = \mathbf{G}^T\mathbf{C}\mathbf{G} = \mathbf{L}\mathbf{U}$ decomposition performed the next phase is to find the $\bar{\mathbf{D}}^*$ matrix in the following steps:

1. Construct $\mathbf{Z}_1 = \left[... \;\; \delta_i \mathbf{G}^T\mathbf{x} \;\; ... \right]$ from the block Hankel matrices of the input-output records:

$$\mathbf{Z}_1 = \left[Hankel(\mathbf{y}, m+1) \;\; Hankel(\mathbf{u}, m+1) \right] \quad (46)$$

2. Calculate $\mathbf{Z}_2 = \mathbf{L}^{-1}\mathbf{Z}_1$. For $\mathbf{L}$ is a lower band limited triangular matrix, the number of the operations required to calculate $\mathbf{Z}_2$ is proportional to N.

3. To achieve a reduction in the computing complexity consider the orthogonal-triangular decomposition by $\mathbf{Z}_2 = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q}$ is a unitary matrix ($\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$), while $\mathbf{R}$ is an upper triangular matrix. Consequently, we have

$$\mathbf{D}^* = \mathbf{Z}_2^T\mathbf{Z}_2 = \mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R} = \mathbf{R}^T\mathbf{R} . \quad (47)$$

Derive now $\tilde{\mathbf{R}}$ from $\mathbf{R}$ by discarding all the rows in $\mathbf{R}$ constructed by 0 elements only. Then $\mathbf{D}^* = \bar{\mathbf{D}}^{*T}\bar{\mathbf{D}}^* = \tilde{\mathbf{R}}^T\tilde{\mathbf{R}}$ results in a way to find $\bar{\mathbf{D}}^*$ via reduced computing complexity.

Summing up $\bar{\mathbf{D}}^* = Q\_R(\mathbf{Z}_2)$ has been derived, where $Q\_R(...)$ denotes the operation performing the procedure discussed above. Note that assuming sufficient excitation the $\tilde{\mathbf{R}}$ matrix is a quadratic matrix.

As the next phase, to find $\bar{\mathbf{C}}^*$ proceed with the algorithm as follows:

1. Determine the error vector:

$$\mathbf{z}_1 = \mathbf{G}^T\mathbf{x} . \quad (48)$$

2. Using the compact form of the Cholesky decomposition calculate $\mathbf{z}_3 = [\mathbf{G}^T\mathbf{C}\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{x}$ in two phases:

$$\mathbf{z}_2 = \mathbf{L}^{-1}\mathbf{z}_1 \quad \text{and} \quad \mathbf{z}_3 = \mathbf{U}^{-1}\mathbf{z}_2 . \quad (49)$$

3. Using $\mathbf{z}_3$ construct $\mathbf{Z}_4 = [... \, \delta_i \mathbf{G}[\mathbf{G}^T\mathbf{C}\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{x} \, ...]$ $i = 1..m+1$ as a Toeplitz matrix:

$$\mathbf{Z}_4 = Toeplitz([\mathbf{0}_m, \mathbf{z}_3, \mathbf{0}_m], m+1) . \quad (50)$$

4. Again, to reduce the size of the $\mathbf{Z}_4$ matrix the $Q\_R(...)$ operation discussed earlier is applied:

$$\mathbf{Z}_5 = Q\_R(\mathbf{Z}_4) .$$

5. Finally derive $\bar{\mathbf{C}}^*$ as

$$\bar{\mathbf{C}}^* = \bar{\mathbf{C}}_\varphi \otimes \mathbf{Z}_5 . \quad (51)$$

All the steps of the algorithm shown above need operations in a number whose maximum is proportional to *N*. Consequently, each step of the minimization procedure based on the above algorithm needs *O(N)* flops.

4.2. Reduction of calculation effort via polynomial decomposition

The number of the operations within an iteration step can further be decreased applying polynomial decomposition instead of the matrix decomposition. To do so, instead of performing an LU decomposition consider the following decomposition of the $\mathbf{\Gamma}$ matrix:

$$\mathbf{\Gamma} = \mathbf{G}^T\mathbf{C}\mathbf{G} = \sin^2\varphi \, \mathbf{G}_a^T\mathbf{G}_a + \cos^2\varphi \, \mathbf{G}_b^T\mathbf{G}_b = \mathbf{G}_c^T\mathbf{G}_c , \quad (52)$$

where the structure of $\mathbf{G}_c$ is similar to that of $\mathbf{G}_a$ or $\mathbf{G}_b$. Equivalently, the underlying polynomial decomposition can be written as

$$\sin^2\varphi \, A(q)A(q^{-1}) + \cos^2\varphi \, B(q)B(q^{-1}) = C(q)C(q^{-1}) . \quad (53)$$

Then the application of a whitening filter by $1/C(q^{-1})$ allows us to avoid the matrix inversion. As the fundamental point of the algorithm the detailed sequence of the polynomial decomposition algorithm will be presented below:

1. There is no change in the initial step. Find $\gamma_i \;\; i = 0,1,...,m$ based on the knowledge available for **a**, **b** and the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio.

2. Consider z-transform by $\gamma(z) = \gamma(q)\big|_{q=z}$ and factorize $\gamma(z)$ as

$$\gamma(z) = C(z)C(z^{-1}) , \quad (54)$$

where the polynomial $C(z)$ contains all the roots of $\gamma(z)$ within the unit circle, while $C(z^{-1})$ contains all the roots of $\gamma(z)$ outside the unit circle. Then use the coefficients of $C(z)$ and $C(z^{-1})$ in polynomials of the backward shift operator $C(q) = C(z)\big|_{z=q}$ and $C(q^{-1}) = C(z^{-1})\big|_{z=q}$, respectively.

Having the $\gamma(z) = C(z)C(z^{-1})$ decomposition performed the next phase is to find the $\bar{\mathbf{D}}^*$ matrix:

1. Apply the $1/C(q^{-1})$ filter for the input-output records:

$$u_k^F = \frac{1}{C(q^{-1})}u_k \quad \text{and} \quad y_k^F = \frac{1}{C(q^{-1})}y_k \,. \tag{55}$$

2. Using the filtered input-output records construct the following block Hankel matrices:

$$\mathbf{Z}_F = \left[Hankel(\mathbf{y}_F, m+1) \;\; Hankel(\mathbf{u}_F, m+1) \right] \tag{56}$$

where $\mathbf{y}_F$ and $\mathbf{u}_F$ are built up by filtered records according to Eq.(55).

3. Finally calculate $\bar{\mathbf{D}}^*$ according to

$$\bar{\mathbf{D}}^* = Q\_R(\mathbf{Z}_F) \,. \tag{57}$$

As far as the calculation of $\bar{\mathbf{C}}^*$ is concerned, steps from 1 to 5 shown for the matrix decomposition algorithm can be used, except Step 2:

Instead of using the compact form of the Cholesky decomposition calculate $\mathbf{z}_3 = [\mathbf{G}^T\mathbf{C}\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{x}$ applying the following sequence of filters:

$$z_{2k} = \frac{1}{C(q^{-1})}z_{1k} \quad \text{and} \quad z_{3k} = \frac{1}{C(q)}z_{2k} \,, \tag{58}$$

where the first filter is a simple autoregressive filter, which is followed by a reverse filtering. Each of the above filters requires operations whose number is proportional to N.

4.3. Reduction of calculation effort using expected value

The calculation effort can further be reduced by applying Eq.(39) concerning the expected value of $\mathbf{C}^*$ and find $\bar{\mathbf{C}}^o$, where $\mathbf{C}^o = \bar{\mathbf{C}}^{To}\bar{\mathbf{C}}^o$ assuming $N \to \infty$, i.e.

$$\mathbf{C}^o = \frac{1}{\mu}E\{(\frac{1}{C(q^{-1})}\mathbf{n}_k)(\frac{1}{C(q^{-1})}\mathbf{n}_k)^T\} \tag{59}$$

is looked for where $\mathbf{n}_k$ represents the additive noise component connected to $\mathbf{x}_k^o$. The steps of the calculation are as follows:

1. Applying a polynomial decomposition and a series expansion construct

$$\mathbf{v} = [v_m \;\; \ldots \;\; v_1 \;\; v_0 \;\; v_1 \;\; \ldots \;\; v_m] \,, \tag{60}$$

where

$$\frac{1}{C(q^{-1})}\frac{1}{C(q)} = \ldots v_m q^{-m} + \ldots + v_1 q^{-1} + v_0 + v_1 q + \ldots + v_m q^m + \ldots \tag{61}$$

2. Construct

$$\mathbf{Q} = Toeplitz(\mathbf{v}, m+1) \,. \tag{62}$$

3. Calculate the Cholesky decomposition of $\mathbf{Q}$:

$$\mathbf{Q} = \bar{\mathbf{Q}}^T\bar{\mathbf{Q}} \,. \tag{63}$$

4. Finally calculate $\bar{\mathbf{C}}^o$ according to

$$\bar{\mathbf{C}}^o = \bar{\mathbf{C}}_\varphi \otimes \bar{\mathbf{Q}} \,. \tag{64}$$

The above equations show that the number of the operations required by the calculation of $\bar{\mathbf{C}}^o$ is independent of the number of the samples driving the parameter estimation algorithm.

Using an iterative EVD/SVD algorithm based on the matrices derived along the above steps both $\boldsymbol{\theta}$ and μ can efficiently be estimated. In practice a few (say 3-10) iteration steps lead to rather accurate results. Note that the above calculations can be used not only to support an EVD-SVD type estimation, but they can contribute to efficiently run gradient type search or Levenberg-Marquardt algorithms, as well.

Up to this point it has been assumed that the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio is available for the parameter estimation algorithm. With this assumption the iterative ML algorithms discussed so far result in unbiased and efficient estimation both for $\boldsymbol{\theta}$ and μ. In the next Section the scope of the study will be extended to estimate $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio, as well.

5. ESTIMATION OF THE NOISE RATIO

Unlike in the previous sections, herewith below it will be assumed that the $\mathrm{var}(\tilde{u}_k)/\mathrm{var}(\tilde{y}_k)$ noise ratio is *not* available for the parameter estimation algorithm. In fact, a number of methods have already been presented on the simultaneous estimation of the process and noise parameters. In this Section a procedure will be derived to estimate the variances of the noise components. Specifically, this procedure is based on matching of the covariance matrices and it can be considered as a generalization of the algorithms presented in (Diversi et al., 2003).

To start the derivation find $E\{\mathbf{G}_s(\boldsymbol{\theta})^T\mathbf{D}_s\mathbf{G}_s(\boldsymbol{\theta})\}$, where the matrix $\mathbf{D}_s$ is constructed from the noisy input-output records

$$\mathbf{D}_s = \mathbf{X}_s^T\mathbf{X}_s/(N-s+1) \,. \tag{65}$$

Here

$$\mathbf{X}_s = \left[Hankel(\mathbf{y}, s) \;\; Hankel(\mathbf{u}, s) \right] \tag{66}$$

and the matrix $\mathbf{G}_s(\boldsymbol{\theta})$ is still constructed similar to Eqs.(22,24-25) using the real process parameters:

$$\mathbf{G}_s = \left[\mathbf{G}_s^a \;\; -\mathbf{G}_s^b \right]^T, \tag{67}$$

where

$$\mathbf{G}_s^a = Toeplitz([\mathbf{0}_{s-m-1}; \mathbf{a}^*; \mathbf{0}_{s-m-1}], s-m) \tag{68}$$
$$\mathbf{G}_s^b = Toeplitz([\mathbf{0}_{s-m-1}; \mathbf{b}^*; \mathbf{0}_{s-m-1}], s-m) \;\; .$$

Using the covariance matrix of the actual noise components the expected value can be expressed as

$$E\{\mathbf{G}_s(\boldsymbol{\theta})^T\mathbf{D}_s\mathbf{G}_s(\boldsymbol{\theta})\} = \mu\mathbf{G}_s(\boldsymbol{\theta})^T\mathbf{C}_s\mathbf{G}_s(\boldsymbol{\theta}) \,, \tag{69}$$

where the structure of matrix $\mathbf{C}_s$ is given by

$$\mathbf{C}_s = \mathbf{C}_\varphi \otimes \mathbf{I}_s \,. \tag{70}$$

In the above notation the index s relates to the size of the Hankel matrices applied. Recalling the law of large numbers the expected value in Eq.(69) can be approximated by the algebraic average:

$$\mathbf{G}_s(\boldsymbol{\theta})^T \mathbf{D}_s \mathbf{G}_s(\boldsymbol{\theta}) \approx \mu \mathbf{G}_s(\boldsymbol{\theta})^T \mathbf{C}_s \mathbf{G}_s(\boldsymbol{\theta}) \,. \tag{71}$$

While an estimation procedure for the noise coefficients will be derived in the sequel, the left side (calculated covariance) and the right side (covariance reflecting the assumption made on μ and φ) of Eq.(71) will be forced to be as close to each other as possible. This condition will be called *covariance matching*. Moreover, both sides will be calculated using the latest available estimations. To characterize the modeling error define the error matrix by

$$\mathbf{P}(\varphi) = \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi))^T \mathbf{D}_s \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi)) - \hat{\mu}(\varphi) \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi))^T \mathbf{C}_s(\varphi) \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi)) \tag{72}$$

and achieve covariance matching via minimizing some norm of $\mathbf{P}$. One option is to minimize $tr(\mathbf{W}_1 \mathbf{P}^T \mathbf{W}_2 \mathbf{P})$ with appropriate $\mathbf{W}_1$ és $\mathbf{W}_2$ weighting matrices. A special choice can be the Frobenius norm of $\mathbf{P}$. This is the norm which will be used in the rest of the derivation.

Once selecting a particular φ value and having an estimation for $\hat{\boldsymbol{\theta}}(\varphi)$ and $\hat{\mu}(\varphi)$ the noise variances can be estimated by minimizing $\|\mathbf{P}\|_F$, where the notation $\|...\|_F$ has been used for the Frobenius norm. Specifically, the following minimization results in a direct estimation for φ:

$$\hat{\varphi} = \arg\min_{\varphi} \Big\| \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi))^T \mathbf{D}_s \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi)) - \hat{\mu}(\varphi) \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi))^T \mathbf{C}_s(\varphi) \mathbf{G}_s(\hat{\boldsymbol{\theta}}(\varphi)) \Big\|_F \tag{74}$$

Finally, the estimated φ value leads to the following solution in terms of the noise variances:

$$\hat{\sigma}_u^2 = \hat{\mu}(\hat{\varphi}) \cos^2(\hat{\varphi}) \quad \text{and} \quad \hat{\sigma}_y^2 = \hat{\mu}(\hat{\varphi}) \sin^2(\hat{\varphi}) \,. \tag{75}$$

6. SIMULATION EXAMPLE

In this Section a simulation example will demonstrate the properties of the identification methods discussed in the previous sections. Consider a noise-free second order process given by

$$y_k^o - 1.5 y_{k-1}^o + 0.7 y_{k-2}^o = 2 u_{k-1}^o + u_{k-2}^o \,. \tag{76}$$

Let the noise-free process input be an ARMA(1,1) process:

$$u_k^o - 0.5 u_{k-1}^o = e_k + 0.7 e_{k-1} \,, \tag{77}$$

where $E\{e_k e_l\} = \delta_{k-l}$. The noise components used in the simulation study have been generated setting φ=45°. The length of the input-output records used for the identification has been $N = 10000$. The aim of the simulation is to analyze the effect of various noise conditions while identifying the system parameters and the $\text{var}(\tilde{u}_k) / \text{var}(\tilde{y}_k)$ noise ratio.

The Frobenius norm by Eq. (72) is shown in Fig. 2 as a function of the a priori assumed φ values. The minimum in this case gives estimation fairly close to 45°. Further simulation examples verified the effectiveness of the proposed algorithms. In all cases the estimations for the noise coefficients were found acceptable.

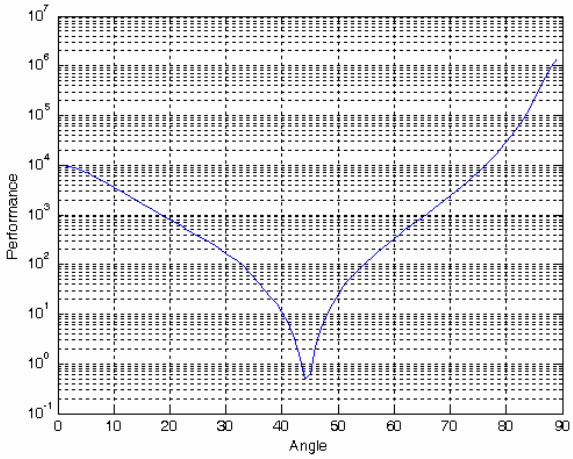

Fig.2. The covariance matching performance with respect to the assumed angle for φ=45° and s=15

Acknowledgements

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: T68370).

REFERENCES

Diversi,R., R.Guidorzi, U.Soverini (2003). A new criterion in EIV identification and filtering applications. *13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands. 1993-8.

Diversi,R., R.Guidorzi, U.Soverini (2007). Maximum likelihood identification of noisy input-output models. *Automatica*, **43**, 464-472.

Ekman,M., M.Hong, T.Söderström (2006). A separable nonlinear least-squares approach for identification of linear systems with errors in variables. *14th IFAC Symp. on System Identification*, Newcastle, Australia. 178-183.

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

Markovsky,I., A.Kukush, S.Van Huffel (2006). On errors-in-variables estimation with unknown noise variance ratio. *14th IFAC Symp. on System Identification*, Newcastle, Australia. 172-177.

Söderström,T. (2007). Errors-in-variables methods in system identification. *Automatica*, **43**,939-958.

Vajk,I. (2005). Identification methods in a unified framework. *Automatica*, **41**,1385-1393.

Wang,J, S.J.Qin (2002). A new subspace identification approach based on principal component analysis. *Journal of Process Control*, **12**, 841-855.