

IDENTIFICATION OF NONLINEAR ERRORS-IN-VARIABLES MODELS

István Vajk, Jenő Hetthéssy

*Department of Automation and Applied Informatics
Budapest University of Technology and Economics
H-1521 Budapest, Hungary
fax : (361) 463-2871 and e-mail : vajk@aut.bme.hu*

Abstract: The paper is about a generalization of a classical eigenvalue-decomposition method originally developed for errors-in-variables linear system identification to handle an important class of nonlinear problems. A number of examples are presented to call the attention to the most critical part of the procedure turning the identification problem to a generalized eigenvalue-eigenvector calculation problem with symmetrical matrices. The elaborated method generates consistent parameter estimation. Simulation results demonstrate the effectiveness of the proposed algorithm.

Copyright © 2002 IFAC

Keywords: errors-in-variables model, nonlinear system identification, eigenvalue-eigenvector decomposition.

1. INTRODUCTION

The problem of identifying a model from noisy input and output measurements is known as the errors-in-variables (EIV) problem. Considering linear systems a number of methods were elaborated in the last few years to handle the EIV problem. E.g. Castaldi et al. (1996), Chou et al. (1997) and Heij et al. (1999) presented results for the EIV model based identification. The nonlinear problem has been exposed in Amemiya et al. (1988) and Vandersteen (1997).

The central problem discussed in this paper is an extension of one of the EIV methods available for linear models to handle nonlinear problems. The Koopmans method (Koopmans, 1937) is a well-known classical method, which has been serving as a firm theoretical basis for a number of identification algorithms for decades. Using the Koopmans algorithm Levin (Levin, 1964, Fernando et al., 1985) proposed methods for identification of EIV dynamic linear system. This algorithm can be considered the ancient version of the nowadays-popular subspace

algorithms. As a major difference Levin's algorithm is based on eigenvalue-eigenvector calculations, while the subspace algorithms are using singular value decomposition resulting in improved numerical robustness.

The goal of this paper is to extend the capability of the eigenvalue-eigenvector calculation based algorithms to handle nonlinear system identification problems. The presented algorithm exhibits procedural similarities with the approach followed in Vandersteen (1997), however, the a priori conditions and – consequently – the elaborated algorithm are completely different.

The paper is organized as follows. First an overview of the Koopmans - Levin method will be given for linear systems. Then a simple example will be used as a vehicle to show the key idea proposed to handle the nonlinear case. The steps used in the introductory example will be generalized and further examples will demonstrate the proper treatment of the recorded data in the critical phase of the proposed algorithm. Simulation results will conclude the paper.

2. EIGENVALUE-EIGENVECTOR METHOD – THE LINEAR CASE

Considering discrete-time linear systems assume that a set of noiseless observations $z_{i,t}^o$ results in a model output by the following linear combination of

$$y_t^o = a_1 z_{1,t}^o + a_2 z_{2,t}^o + \dots + a_m z_{m,t}^o, \quad (1)$$

where t denotes the discrete time instants ($t=0,1,2,\dots$) and the a_i parameters are to be estimated. Collecting the unknown parameters and the noiseless observations in parameter and observation vectors, respectively,

$$\theta^T = [a_1 \ a_2 \ \dots \ a_m] \quad (2)$$

and

$$\mathbf{z}_t^{oT} = [z_{1,t}^o \ z_{2,t}^o \ \dots \ z_{m,t}^o] \quad (3)$$

form the model equation explicitly by

$$y_t^o = \theta^T \mathbf{z}_t^o \quad (4)$$

The model equation can be reformulated in an *implicit form* of $\mathcal{G}^T \mathbf{x}_t^o = 0$, where

$$\mathcal{G}^T = [-1 \ \theta^T] = [-1 \ a_1 \ a_2 \ \dots \ a_m] \quad (5)$$

and

$$\mathbf{x}_t^{oT} = [y_t^o \ \mathbf{z}_t^{oT}] = [y_t^o \ z_{1,t}^o \ z_{2,t}^o \ \dots \ z_{m,t}^o]. \quad (6)$$

are extended parameter and noise-free observation vectors, respectively. The main consequence of using an implicit model is that each variable plays the same role, no distinction is made among input and output signals.

The application of the well known Least Squares (LS) estimation technique leads to unbiased estimation for the a_i parameters if the set of the x_t measured records available for the estimation is such that

$$\mathbf{x}_t^T = [y_t^o + \varepsilon_t \ \mathbf{z}_t^{oT}] = [y_t^o \ z_{1,t}^o \ z_{2,t}^o \ \dots \ z_{m,t}^o] \quad (7)$$

where ε_t is white noise. Koopmans (Koopmans, 1937) generalized the LS estimation scheme by developing a maximum likelihood (ML) estimation assuming extended observation vectors. In his model *all* components of the observation vector may be corrupted by additive noise according to $\mathbf{x}_t = \mathbf{x}_t^o + \mathbf{n}_t$, where \mathbf{n}_t is Gaussian white noise with covariance matrix $\mu\mathbf{C}$.

Assuming N samples find an estimation of the parameter vector \mathcal{G} . With no a priori information available there is no better estimation than the maximum likelihood (ML) estimation. To find this estimation we need to set up the likelihood function of the observations:

$$\begin{aligned} & \text{prob}(\mathbf{x}_1, \dots, \mathbf{x}_N | \mathcal{G}, \mathcal{G}^T \mathbf{x}_1^o = 0, \dots, \mathcal{G}^T \mathbf{x}_N^o = 0) = \\ & \text{const} \exp\left(-\frac{1}{2} \sum_{t=1}^N (\mathbf{x}_t - \mathbf{x}_t^o)^T (\mu\mathbf{C})^{-1} (\mathbf{x}_t - \mathbf{x}_t^o)\right) \end{aligned} \quad (8)$$

where

$$\mathcal{G}^T \mathbf{x}_t^o = 0 \quad t = 1..N. \quad (9)$$

Now the ML estimation can be obtained by maximizing the likelihood function in terms of \mathcal{G} and \mathbf{x}_t^o .

With the constraint of $\mathcal{G}^T \mathbf{x}_t^o = 0$ the above maximization is equivalent to minimize the loss function

$$J = \frac{1}{2} \frac{\mathcal{G}^T \left(\sum_{t=1}^N \mathbf{x}_t \mathbf{x}_t^T \right) \mathcal{G}}{\mathcal{G}^T \mathbf{C} \mathcal{G}} \quad (10)$$

by \mathcal{G} with no constraints, which in turn needs to solve

$$\mathbf{W} \mathcal{G} - \mathbf{C} \mathcal{G} \frac{\mathcal{G}^T \mathbf{W} \mathcal{G}}{\mathcal{G}^T \mathbf{C} \mathcal{G}} = \mathbf{0}, \quad (11)$$

where

$$\mathbf{W} = \sum_{t=1}^N \mathbf{x}_t \mathbf{x}_t^T. \quad (12)$$

According to the above equation the optimal value of \mathcal{G} can be determined as a solution of the generalized eigenvalue calculation problem defined by

$$(\mathbf{W} - \mu\mathbf{C}) \mathcal{G} = \mathbf{0}. \quad (13)$$

More exactly, the optimal value of the parameter estimation will be equal to the eigenvector related to the smallest eigenvalue found, namely

$$\min_{\mathcal{G}} J = \mu_1 \quad (14)$$

holds.

The original Koopmans method relates to static linear models. Levin, however, applied the above algorithm for linear dynamic EIV models. In this case the observation vector is composed from the delayed recorded input and output samples of the process. Otherwise the parameter vector is still calculated via solving a generalized eigenvalue-eigenvector calculation problem.

The advantage of the Koopmans-Levin method is clearly seen from the above equations, namely the consistent estimation is based on the relatively easy calculation of eigenvectors and eigenvalues of *symmetrical* matrices. Namely for symmetrical matrices of $\mathbf{W} = \mathbf{\Phi}^T \mathbf{\Phi}$ and $\mathbf{C} = \mathbf{\Gamma}^T \mathbf{\Gamma}$ the generalized eigenvalue calculation for the (\mathbf{W}, \mathbf{C}) pair can be traced back to a generalized singular value decomposition for $(\mathbf{\Phi}, \mathbf{\Gamma})$ representing a far more robust numerical algorithm. The presented methods can be qualified as direct strategies in the sense that no iteration is required to obtain the parameter estimation.

3. EIGENVALUE-EIGENVECTOR METHOD – THE NONLINEAR CASE

In this section the above-presented algorithm will be generalized for nonlinear EIV identification. First a simple example will be shown to demonstrate the way the nonlinear extension is performed. Then a nonlinear eigenvalue-eigenvector algorithm will be

presented. Finally some technical hints will be given for the calculation.

3.1 An introductory example

Just to get started let us consider how to fit a second order parabolic function to a set of u_1, \dots, u_N and y_1, \dots, y_N input/output observation records. Working with the standard assumptions the input/output records are assumed to be corrupted by independent, normal distribution white noise with zero mean and variance of $\mu\sigma_u^2$ and $\mu\sigma_y^2$, respectively. As far as the noiseless observations u_1^o, \dots, u_N^o and y_1^o, \dots, y_N^o are concerned the constraint by

$$y_t^o = \theta u_t^{o^2} \quad t=1,2,\dots \quad (15)$$

holds expressing the second order parabolic relation, though the θ parameter is not known. The maximum likelihood estimation for the unknown θ parameter can be obtained by minimizing the loss function

$$J = \frac{1}{2} \sum_{t=1}^N \begin{bmatrix} y_t - y_t^o & u_t - u_t^o \end{bmatrix} \begin{bmatrix} 1/\sigma_y^2 & 0 \\ 0 & 1/\sigma_u^2 \end{bmatrix} \begin{bmatrix} y_t - y_t^o \\ u_t - u_t^o \end{bmatrix} + \sum_{t=1}^N \lambda_t (y_t^o - \theta u_t^{o^2}) \quad (16)$$

The straightforward minimization would lead to a set of $3N+1$ equations allowing to find the $3N+1$ unknown variables (i.e. $u_1^o, u_2^o, \dots, u_N^o, y_1^o, y_2^o, \dots, y_N^o, \lambda_1^o, \lambda_2^o, \dots, \lambda_N^o, \theta$), including the θ parameter itself. It is obvious, however, that with the increasing number of observations this simple approach will not work in practice. Consequently, to derive a feasible solution a different approach should be applied.

The approach to be followed hereafter aims to turn the parameter estimation problem to a generalized eigenvalue-eigenvector calculation problem. It is noted, however, that the elaborated method will not give a maximum likelihood estimation. All we can say is that the estimation will be consistent and the underlying method will be adequate for practical applications.

Coming back to the problem to fit a second order parabolic function to the input/output records consider the following matrix constructed by the noisy (available) observations

$$\mathbf{D} = \frac{1}{N} \sum_{t=1}^N \begin{bmatrix} y_t \\ u_t^2 \end{bmatrix} \begin{bmatrix} y_t & u_t^2 \end{bmatrix}, \quad (17)$$

and find the expected value of \mathbf{D} . The components of $E\{\mathbf{D}\}$ can be calculated as follows:

$$E\left\{\frac{1}{N} \sum_{t=1}^N y_t^2\right\} = \frac{1}{N} \sum_{t=1}^N y_t^{o^2} + \mu\sigma_y^2 \quad (18)$$

$$E\left\{\frac{1}{N} \sum_{t=1}^N y_t u_t^2\right\} = \frac{1}{N} \sum_{t=1}^N y_t^o u_t^{o^2} + \frac{1}{N} \sum_{t=1}^N y_t^o \mu\sigma_u^2 \quad (19)$$

$$E\left\{\frac{1}{N} \sum_{t=1}^N u_t^4\right\} = \frac{1}{N} \sum_{t=1}^N u_t^{o^4} + \frac{6}{N} \sum_{t=1}^N u_t^{o^2} \mu\sigma_u^2 + 3\mu^2 \sigma_u^4 \quad (20)$$

All the right hand side expressions above contain sums made out of the noiseless signals *not* available for calculations. To isolate the terms with unknown variables we are going to use the following evaluations valid for the expected values:

$$E\left\{\frac{1}{N} \sum_{t=1}^N y_t\right\} = \frac{1}{N} \sum_{t=1}^N y_t^o \quad (21)$$

$$E\left\{\frac{1}{N} \sum_{t=1}^N u_t^2\right\} = \frac{1}{N} \sum_{t=1}^N u_t^{o^2} + \mu\sigma_u^2 \quad (22)$$

Substituting $\frac{1}{N} \sum_{t=1}^N y_t^o$ and $\frac{1}{N} \sum_{t=1}^N u_t^{o^2}$ back to the expressions of the $E\{\mathbf{D}\}$ components results in

$$\frac{1}{N} \sum_{t=1}^N y_t^2 = E\left\{\frac{1}{N} \sum_{t=1}^N y_t^2\right\} - \mu\sigma_y^2 \quad (23)$$

$$\frac{1}{N} \sum_{t=1}^N y_t^o u_t^{o^2} = E\left\{\frac{1}{N} \sum_{t=1}^N y_t u_t^2\right\} - \mu\sigma_u^2 E\left\{\frac{1}{N} \sum_{t=1}^N y_t\right\} \quad (24)$$

$$\begin{aligned} \frac{1}{N} \sum_{t=1}^N u_t^4 &= \\ E\left\{\frac{1}{N} \sum_{t=1}^N u_t^4\right\} &- 6\mu\sigma_u^2 E\left\{\frac{1}{N} \sum_{t=1}^N u_t^2\right\} + 3\mu^2 \sigma_u^4 \end{aligned} \quad (25)$$

Arranging the above components into a vector-matrix form:

$$\begin{aligned} \frac{1}{N} \sum_{t=1}^N \begin{bmatrix} y_t^o \\ u_t^{o^2} \end{bmatrix} \begin{bmatrix} y_t^o & u_t^{o^2} \end{bmatrix} &= E\left\{\frac{1}{N} \sum_{t=1}^N \begin{bmatrix} y_t \\ u_t^2 \end{bmatrix} \begin{bmatrix} y_t & u_t^2 \end{bmatrix}\right\} \\ &- E\left\{\begin{bmatrix} \mu\sigma_y^2 & \mu\sigma_u^2 \frac{1}{N} \sum_{t=1}^N y_t \\ \mu\sigma_u^2 \frac{1}{N} \sum_{t=1}^N y_t & 6\mu\sigma_u^2 \frac{1}{N} \sum_{t=1}^N u_t^2 - 3\mu^2 \sigma_u^4 \end{bmatrix}\right\} \end{aligned} \quad (26)$$

Now introducing

$$\mathbf{D}_o = \frac{1}{N} \sum_{t=1}^N \begin{bmatrix} y_t^o \\ u_t^{o^2} \end{bmatrix} \begin{bmatrix} y_t^o & u_t^{o^2} \end{bmatrix} \quad (27)$$

and

$$\mathbf{H}(\mu) = \begin{bmatrix} \mu\sigma_y^2 & \mu\sigma_u^2 \frac{1}{N} \sum_{t=1}^N y_t \\ \mu\sigma_u^2 \frac{1}{N} \sum_{t=1}^N y_t & 6\mu\sigma_u^2 \frac{1}{N} \sum_{t=1}^N u_t^2 - 3\mu^2 \sigma_u^4 \end{bmatrix} \quad (28)$$

we have

$$\mathbf{D}_o = E\{\mathbf{D} - \mathbf{H}(\mu)\}. \quad (29)$$

Observe that \mathbf{D}_o is singular as a result of the linear in parameters dependency introduced by $y_t^o = \theta u_t^{o^2}$:

$$\mathbf{D}_o = \begin{pmatrix} \frac{1}{N} \sum_{t=1}^N u_t^{o^4} & \theta \\ \theta & 1 \end{pmatrix} \quad (30)$$

Multiplication by

$$\mathcal{G} = [-1 \quad \theta^T]^T \quad (31)$$

leads to

$$\mathbf{D}_o \mathcal{G} = E\{\mathbf{D} - \mathbf{H}(\mu)\} \mathcal{G} = \mathbf{0} . \quad (32)$$

The operation instructed by the expected value relates to calculated mean values, thus the approximation by

$$E\left\{\frac{1}{N} \sum (...)\right\} \approx \frac{1}{N} \sum (...)\quad (33)$$

allows the parameter estimation reformulated as

$$(\mathbf{D} - \mathbf{H}(\mu)) \mathcal{G} = \mathbf{0} . \quad (34)$$

In other words the parameter estimation problem is equivalent to a generalized eigenvalue-eigenvector problem in the sense that the smallest generalized eigenvalue μ is looked for satisfying

$$\det(\mathbf{D} - \mathbf{H}(\mu)) = 0 , \quad (35)$$

while the estimation $\hat{\mathcal{G}}$ will be obtained as the eigenvector belonging to this eigenvalue.

In the sequel, the introductory example just discussed will be directly generalized for more complex linear in parameters nonlinear models.

3.2 Generalization

In contrast to the linear equation given by $\mathcal{G}^T \mathbf{x}_t^o = 0$ discussed in Section 2 nonlinear process models will now be considered. Denoting the set of the noiseless observations by \mathbf{x}_t^o and the noisy observations by \mathbf{x}_t , and further on assuming white noise conditions with zero mean and $\mu \mathbf{C}$ covariance we have

$$\mathbf{x}_t \sim N(\mathbf{x}_t^o, \mu \mathbf{C}) . \quad (36)$$

For the noiseless observations a nonlinear constraint is assumed to exist, which relation is assumed to be linear in parameters:

$$\mathcal{G}^T \mathbf{g}(\mathbf{x}_t^o) = 0 \quad (37)$$

Further on we assume that the elements of $\mathbf{g}(\mathbf{x}_t^o)$ are polynomials of the observations. Based on the noisy observations \mathbf{x}_t for $t=1,2,\dots,N$ the goal is to find an estimation for \mathcal{G} .

Calculating $\mathbf{g}(\mathbf{x}_t)$ from the available observations consider the following matrix

$$\mathbf{D} = \frac{1}{N} \sum_{t=1}^N \mathbf{g}(\mathbf{x}_t) \mathbf{g}(\mathbf{x}_t)^T \quad (38)$$

Now find $E\{\mathbf{D}\}$

$$E\{\mathbf{D}\} = E\left\{\frac{1}{N} \sum_{t=1}^N \mathbf{g}(\mathbf{x}_t) \mathbf{g}(\mathbf{x}_t)^T\right\} \quad (39)$$

by repeating the very same steps applied in the introductory example. It can be shown that

$$E\{\mathbf{D}\} = \mathbf{D}_o - E\{\mathbf{H}(\mu)\} \quad (40)$$

where

$$\mathbf{D}_o = \frac{1}{N} \sum_{t=1}^N \mathbf{g}(\mathbf{x}_t^o) \mathbf{g}(\mathbf{x}_t^o)^T . \quad (41)$$

However, it is difficult to give a general analytical form for $\mathbf{H}(\mu)$. Anyway, we have

$$\mathbf{D}_o = E\{\mathbf{D} - \mathbf{H}(\mu)\} . \quad (42)$$

Due to the constraint by $\mathcal{G}^T \mathbf{g}(\mathbf{x}_t^o) = 0$ matrix \mathbf{D}_o is singular, meaning that

$$\mathbf{D}_o \mathcal{G} = E\{\mathbf{D} - \mathbf{H}(\mu)\} \mathcal{G} = \mathbf{0} \quad (43)$$

or

$$\det(E\{\mathbf{D} - \mathbf{H}(\mu)\}) = 0 . \quad (44)$$

As it has already been shown, approximating the expected value of a calculated mean by the calculated mean itself allows to drop the expected value operation and results in

$$\mathbf{D}_o(\mathbf{x}_1^o, \dots, \mathbf{x}_N^o) \mathcal{G} = (\mathbf{D}(\mathbf{x}_1, \dots, \mathbf{x}_N) - \mathbf{H}(\mu \mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N)) \mathcal{G} = \mathbf{0} \quad (45)$$

Again, the parameter estimation problem has been turned to a generalized eigenvalue-eigenvector problem.

Due to the polynomial construction of $\mathbf{g}(\cdot)$ the matrix \mathbf{H} is a polynomial like function of μ :

$$\mathbf{H} = \mu \mathbf{H}_1 + \mu^2 \mathbf{H}_2 + \dots + \mu^m \mathbf{H}_m \quad (46)$$

where m is the maximum order in the elements of $\mathbf{g}(\cdot)$, and the matrices \mathbf{H}_i are

$$\mathbf{H}_i = \mathbf{H}_i(\mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N) . \quad (47)$$

Note that for the original Koopmans method

$$\mathbf{H}(\mu \mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N) = \mu \mathbf{C} \quad (48)$$

holds.

3.3 Practical hints

As it has been pointed out earlier it is quite difficult to derive a general analytical form for $\mathbf{H}(\mu \mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N)$. However, to demonstrate the technique how to find $\mathbf{H}(\mu \mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N)$ consider a few examples.

In the examples we apply the following properties valid for normal distribution:

$$\begin{aligned} \mu_{2k} &= 1 * 3 * \dots * (2k-1) \sigma^{2k} , \\ \mu_{2k-1} &= 0 \quad (k=1,2,\dots) \end{aligned} \quad (49)$$

where

$$\mu_r \equiv E\{(\xi - E\{\xi\})^r\} \quad (50)$$

is defined as the central moment.

Example 1

Find the following expected value:

$$E\left\{\frac{1}{N} \sum_{t=1}^N y_t^6\right\} = ? \quad (51)$$

Straightforward calculations give

$$\begin{aligned} E\left\{\frac{1}{N} \sum_{t=1}^N y_t^6\right\} &= E\left\{\frac{1}{N} \sum_{t=1}^N (y_t^o + \eta)^6\right\} = \\ &= \binom{6}{6} 15 \mu^3 \sigma^6 + \binom{6}{4} 3 \mu^2 \sigma^4 \frac{1}{N} \sum_{t=1}^N y_t^{02} \\ &+ \binom{6}{2} \mu \sigma^2 \frac{1}{N} \sum_{t=1}^N y_t^{04} + \binom{6}{0} \frac{1}{N} \sum_{t=1}^N y_t^{06} \end{aligned} \quad (52)$$

where the summations can further be expressed by μ and σ resulting in the following matrix form:

$$\begin{bmatrix} 1 \\ E\left\{\frac{1}{N}\sum_{t=1}^N y_t^2\right\} \\ E\left\{\frac{1}{N}\sum_{t=1}^N y_t^4\right\} \\ E\left\{\frac{1}{N}\sum_{t=1}^N y_t^6\right\} \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \mu \sigma^2 & \begin{pmatrix} 2 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 4 \\ 4 \end{pmatrix} 3\mu^2 \sigma^4 & \begin{pmatrix} 4 \\ 2 \end{pmatrix} \mu \sigma^2 & \begin{pmatrix} 4 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 6 \\ 6 \end{pmatrix} 15\mu^2 \sigma^4 & \begin{pmatrix} 6 \\ 4 \end{pmatrix} 3\mu^2 \sigma^4 & \begin{pmatrix} 6 \\ 2 \end{pmatrix} \mu \sigma^2 & \begin{pmatrix} 6 \\ 0 \end{pmatrix} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1}{N}\sum_{t=1}^N y_t^{o2} \\ \frac{1}{N}\sum_{t=1}^N y_t^{o4} \\ \frac{1}{N}\sum_{t=1}^N y_t^{o6} \end{bmatrix} \quad (53)$$

According to the special triangular structure of the coefficient matrix the above equation can easily be solved and we obtain:

$$\begin{aligned} \frac{1}{N}\sum_{t=1}^N y_t^{o6} = & E\left\{\frac{1}{N}\sum_{t=1}^N y_t^6\right\} - 15\mu^2 \sigma^4 E\left\{\frac{1}{N}\sum_{t=1}^N y_t^4\right\} \\ & + 45\mu^2 \sigma^4 E\left\{\frac{1}{N}\sum_{t=1}^N y_t^2\right\} + 75\mu^3 \sigma^6 \end{aligned} \quad (54)$$

This means that for a matrix \mathbf{D} with entries like

$$D_{i,j} = \frac{1}{N}\sum_{t=1}^N y_t^6 \quad (55)$$

the appropriate entry of matrix \mathbf{H} becomes:

$$H_{i,j} = 15\mu^2 \sigma^4 \frac{1}{N}\sum_{t=1}^N y_t^4 - 45\mu^2 \sigma^4 \frac{1}{N}\sum_{t=1}^N y_t^2 - 75\mu^3 \sigma^6 \quad (56)$$

Example 2

Consider a matrix \mathbf{D} given by

$$D_{i,j} = \frac{1}{N}\sum_{t=1}^N u_t^2 y_t^2 \quad (57)$$

then to derive \mathbf{H} the following set of equations should be solved:

$$\begin{bmatrix} 1 \\ E\left\{\frac{1}{N}\sum_{t=1}^N u_t^2\right\} \\ E\left\{\frac{1}{N}\sum_{t=1}^N u_t y_t\right\} \\ E\left\{\frac{1}{N}\sum_{t=1}^N y_t^2\right\} \\ E\left\{\frac{1}{N}\sum_{t=1}^N u_t^2 y_t^2\right\} \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ \mu \sigma_u^2 & 1 & & & \\ \mu c_{uy} & & 1 & & \\ \mu \sigma_y^2 & & & 1 & \\ \mu^2 \sigma_u^2 \sigma_y^2 + 2\mu^2 c_{uy}^2 & \mu \sigma_y^2 & \mu c_{uy} & \mu \sigma_u^2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1}{N}\sum_{t=1}^N u_t^{o2} \\ \frac{1}{N}\sum_{t=1}^N u_t^o y_t^o \\ \frac{1}{N}\sum_{t=1}^N y_t^{o2} \\ \frac{1}{N}\sum_{t=1}^N u_t^{o2} y_t^{o2} \end{bmatrix} \quad (58)$$

where c_{uy} stands for the cross-correlation between u_t and y_t .

Example 3

Assume we have random variables \mathbf{x}_t with normal distribution. Considering transcendent functions $f(\mathbf{x}_t)$ of \mathbf{x}_t it is noted that $E\{f(\mathbf{x}_t)\}$ may or may not exist. As an example we state that $E\left\{\frac{1}{N}\sum_{t=1}^N \ln(y_t)\right\}$ does not exist. The proof of this

statement is straightforward by expanding $\ln(y_t)$ into Taylor series and substituting the central moments.

Summary of the proposed algorithm

Based on the discussions so far and evaluating the examples presented in details earlier a procedural algorithm can be

given to generate the matrix \mathbf{H} provided a \mathbf{D} matrix of

$$\mathbf{D} = \frac{1}{N}\sum_{t=1}^N \mathbf{g}(\mathbf{x}_t) \mathbf{g}(\mathbf{x}_t)^T \quad (59)$$

i.e. a matrix with entries like

$$D_{i,j} = \frac{1}{N}\sum_{t=1}^N g_i(\mathbf{x}_t) g_j(\mathbf{x}_t) \quad (60)$$

and further on assuming noise conditions to be normal:

Step 1: Find

$$E\left\{\frac{1}{N}\sum_{t=1}^N g_i(\mathbf{x}_t) g_j(\mathbf{x}_t)\right\} \quad (61)$$

expressed by the noiseless observations and the noise parameters (mean and variance).

Step 2: When finding the expected value for the noiseless observations use the calculated mean of the noisy observations as approximations (see the examples shown previously). Set up a set of equations based on the relation between the calculated means of the noisy and noiseless observations.

Step 3: Solve the equations set up in Step 2 for

$$\frac{1}{N}\sum_{t=1}^N g_i(\mathbf{x}_t^o) g_j(\mathbf{x}_t^o) \quad (62)$$

Step 4: The $H_{i,j}$ entries can be calculated by approximating the expected values of sums involved with the sums themselves.

Step 5: The following equation should be solved numerically for μ_{\min} :

$$\det(\mathbf{D}(\mathbf{x}_1, \dots, \mathbf{x}_N) - \mathbf{H}(\mu \mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N)) = 0 \quad (63)$$

Step 6: Finally, using the fundamental relation of the algorithm

$$(\mathbf{D}(\mathbf{x}_1, \dots, \mathbf{x}_N) - \mathbf{H}(\mu_{\min} \mathbf{C}, \mathbf{x}_1, \dots, \mathbf{x}_N))\hat{\mathbf{g}} = \mathbf{0} \quad (64)$$

$\hat{\mathbf{g}}$ can be determined. The presented algorithm results in a consistent, but not ML estimation. The effectiveness of the algorithm can be further improved by an appropriate weighting sequence applied in the composition of matrix \mathbf{D} .

4. SIMULATION RESULTS

In the sequel, two numerical examples will be shown to demonstrate the effectiveness of the elaborated algorithm.

4.1 Identification of a parabolic curve

Consider $N=100$ samples of the following system:

$$y_i = a_2 u_i^2 + a_1 u_i + a_0 \quad (65)$$

with measurements corrupted by normal noise conditions of $\sigma_u = \sigma_y = 0.2$. In Fig. 1 the noisy observations, the LS estimation, the true system, as well as the results obtained by applying the proposed algorithm are given. It is seen that the true system and the nonlinear EIV model based identification derived in this paper are almost cover each other.

4.2 Identification of an elliptic curve

Consider $N=1000$ samples of the following system:

$$a_0 y_i^2 + a_1 y_i u_i + a_2 u_i^2 + a_3 y_i + a_4 u_i + a_5 = 0 \quad (66)$$

with measurements corrupted by normal noise conditions of $\sigma_u = \sigma_y = 1$. In Fig. 2 it is seen that now we have a really high noise/signal ratio, this is why 1000 samples were selected for the identification. In spite of the noisy environment the result obtained by the new algorithm is capable to almost completely reconstruct the contour determined by the noiseless observations.

5. CONCLUSION

The goal of this paper is to extend the capability of the eigenvalue-eigenvector calculation based algorithms to handle nonlinear system identification problems. A new identification algorithm for nonlinear linear in parameter errors-in-variables models has been presented in the paper. The algorithm is based on generalized eigenvalue-eigenvector calculations in the form of

$$(\mathbf{D} - \mathbf{H}(\mu))\mathbf{g} = \mathbf{0} \quad (67)$$

The vector composed by the unknown system parameters then is associated with the smallest eigenvalue found. Assuming sufficient input excitation the parameter estimation is consistent and it is direct in the sense that the estimation is obtained in one step rather than a result of iterations.

Acknowledgements: This work is supported by the fund of the Hungarian Academy of Sciences for control research and the Hungarian National Research Fund (grant number: T29815)

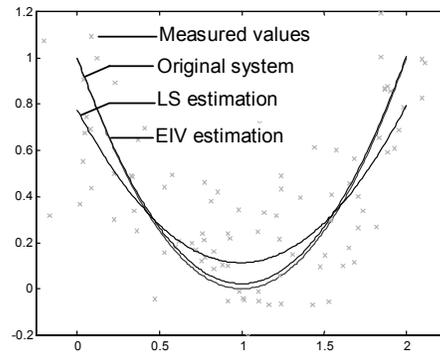


Fig. 1. Identification of a parabolic curve

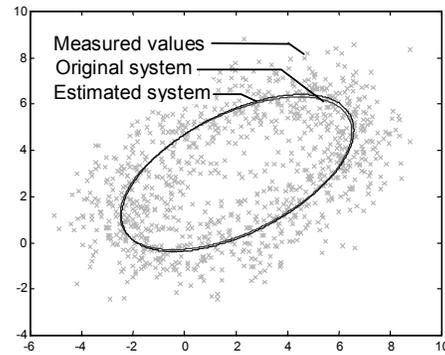


Fig. 2. Identification of an elliptic curve

REFERENCES

- Amemiya, Y., W.A. Fuller (1988) Estimation for the nonlinear functional relationship. *The Annals of Statistics*, Vol.16, No.1, 147-160.
- Castaldi P., Soverini U. (1996) Identification of dynamic errors-in-variables models. *Automatica*, Vol.32, No.4, 631-636.
- Chou C.T., Verhaegen M. (1997). Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, Vol.33, No.10. 1857-1869.
- Fernando K.V., Nicholson H. (1985) Identification of linear systems with input and output noise: the Koopmans-Levin method. *IEE Proceedings*, Vol.132.Pt.D, No.1, 30-36.
- Heij C., Scherrer W. (1999). Consistency of system identification by global total least squares. *Automatica* Vol.35, 993-1008.
- Koopmans, T. (1937) Linear regression analysis of economic time series. DeErven F.Bohn, N.V. Haarlem, The Netherlands.
- Levin M.J. (1964) Estimation of a system pulse transfer function in the presence of noise. *IEEE Trans. on Automatic Control*, 229-235
- Vandersteen, G. (1997) Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework. PhD thesis, Vrije Universiteit, Brussel.