Identifying dynamic systems with polynomial nonlinearities in the errors-in-variables context

LEVENTE HUNYADI and ISTVÁN VAJK
Budapest University of Technology and Economics
Department of Automation and Applied Informatics
1111 Budapest, Goldmann György tér 3., HUNGARY
http://www.aut.bme.hu/portal/hunyadi

Abstract: Many practical applications including speech and audio processing, signal processing, system identification, econometrics and time series analysis involve the problem of reconstructing a dynamic system model from data observed with noise in all variables. We consider an important class of dynamic single-input single-output nonlinear systems where the system model is polynomial in observations but linear in parameters, which captures a wide range of such systems. Assuming white Gaussian measurement noise that is characterized by a magnitude and a covariance structure, we propose a nonlinear extension to the generalized Koopmans–Levin method that can estimate parameters of dynamic nonlinear systems with polynomial nonlinearities given a priori knowledge on the noise covariance structure. In order to estimate noise structure, we apply a covariance matching objective function. Combining the extended Koopmans–Levin and the covariance matching approaches, an identification algorithm to estimate both model and noise parameters is proposed. The feasibility of the approach is demonstrated by Monte-Carlo simulations.

Key–Words: system identification; discrete-time dynamic systems; errors-in-variables; linearizable systems; polynomial eigenvalue problem; covariance matching

1 Introduction

The task of system identification is to build mathematical models from measured data. A discrete-time dynamic model in this context is a description of the dynamic behavior of a system or process over time, with measurements taken at equally-spaced time instants. Dynamic errors-in-variables systems where both input and output variables are observed with noise are of particular significance in applications where the quantitative description of the internal laws constituting the system or constructing a system model from observations is of interest rather than predicting future behavior. In contrast to the conventional dynamic system setup, where the error is solely attributed to the output variable and is intended to address modeling error, errors-in-variables systems have measurement error in both variables and one is not able to access their noise-free counterparts but is confined to constructing a best possible system model from noise-contaminated observations. Applications of dynamic errors-in-variables systems include speech and audio processing, signal processing, system identification, econometrics and time series analysis.

A rather general nonlinear dynamic system model takes the form

$$f\left(\theta,\ z_{0,i}\right) = 0$$

where the vector

$$z_{0,i} = \left[\begin{array}{cccccc} y_{0,i} & \dots & y_{0,i-m} & u_{0,i} & \dots & u_{0,i-m} \end{array}\right]$$

collects a range of true input $u_{0,i}$ and output $y_{0,i}$ values, $i = m+1, \dots, N$, N is the number of observations, m is the memory of the system, the vector θ encapsulates the parameters of interest, and f represents some constraint between past and present observations. This configuration is shown in Figure 1. We take a gray-box approach where the peculiarities inside the system are known up to a number of free parameters θ, i.e. the model structure determined by f and m is already available. A usual assumption that is satisfied in most applications is that the constraint f is linear in θ, i.e.

$$\theta^\top g\left(z_{0,i}\right) = 0$$

where g is a linearization of f. Given that observations $u_{0,i}$ and $y_{0,i}$ are not directly observable but contaminated with noise, the actual observations u_i and y_i satisfy $u_i = u_{0,i} + \tilde{u}_i$ and $y_i = y_{0,i} + \tilde{y}_i$, hence the

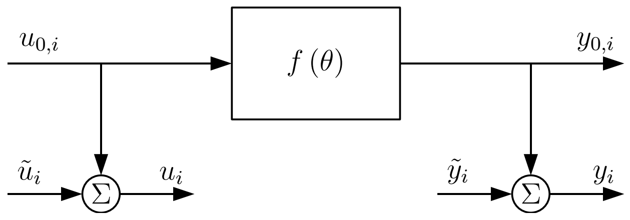

Figure 1: The basic setup for a discrete-time dynamic errors-in-variables system.

objective is to derive estimates for θ given the noisy observations u_i and y_i. The most common assumption for the sequences of $\tilde{u}_i$ and $\tilde{y}_i$ with $i = 1, \ldots, N$ is Gaussian white noise, which models measurement noise.

For the special case of identifying linear dynamic errors-in-variables systems, where the constraint f is linear not only in θ but also in observations $z_{0,i}$, a number of estimation schemes have been proposed. For single-input single-output (SISO) systems, where the system is described by the linear equation

$$\begin{aligned} y_{0,i} &+ a_1 y_{0,i-1} + a_2 y_{0,i-2} + \ldots + a_m y_{0,i-m} \\ &= b_1 u_{0,i-1} + b_2 u_{0,i-2} + \ldots + b_n u_{0,i-n} \end{aligned}$$

but neither the noise-free true input $u_{0,i}$ nor the true output $y_{0,i}$ is observable, proposed methods include bias-compensating least squares [17], the Frisch scheme [2], instrumental variable [3], higher-order statistics [12], structured total least squares [7], frequency-domain [9] and efficient maximum likelihood [15] methods, see [11] for a comprehensive survey. A recursive weighted extended least squares algorithm based on the numerically robust orthogonal Householder transformations is developed for systems identification in a noisy environment in [10], and an application of structured total least squares to dynamic GNSS (Global Navigation Satellite System) positioning problems is discussed in [16]. Other approaches to identification include neural networks [8] or genetic algorithms.

This paper deals with a nonlinear extension of the generalized Koopmans–Levin method to estimate model parameters of a dynamic system with given noise structure where the linearization g is a polynomial in terms of input and output observations, and a subsequent covariance matching objective function to estimate noise covariance structure. The types of nonlinearities in question (i.e. system structure) and system memory are assumed to be known. The Koopmans–Levin method, proposed in [6], gives a non-iterative quick estimate of the model parameters of a linear system given a priori information on the noise structure. The original method was generalized in [13] to improve estimation accuracy at the cost of increased computational complexity, incorporating as special cases the original Koopmans–Levin method and the maximum likelihood method. On the other hand, a nonlinear extension to the original Koopmans method was proposed in [14] for static systems.

The Koopmans–Levin method and its generalization are briefly described in Section 2. Section 3 combines and extends the results of [13] and [14] to nonlinear dynamic systems that comprise of polynomial nonlinearities yet are linear in model parameters. The outlined method assumes a preliminarily known noise structure, Section 4 extends the estimation method so that no such assumptions are required. In order to demonstrate the feasibility of the method, some simulation results are presented in Section 5 before the paper concludes with Section 6.

2 The generalized Koopmans–Levin method

Consider the linear SISO errors-in-variables system $G(q^{-1})$ described by the autoregressive moving average (ARMA) difference equation

$$A(q^{-1}) y_{0,i} = B(q^{-1}) u_{0,i} \tag{1}$$

where q^{-1} denotes the backward shift operator such that $q^{-1} \circ_i = \circ_{i-1}$, in which $\circ$ is a generic placeholder for a model parameter, and

$$\begin{aligned} A(q^{-1}) &= a_0 + a_1 q^{-1} + \cdots + a_m q^{-m} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + \cdots + b_m q^{-m}. \end{aligned}$$

Given the aforementioned system description, we may introduce the model parameter vector θ and the extended regressor vector z

$$\begin{aligned} \theta^\top &= \begin{bmatrix} a^\top & -b^\top \end{bmatrix} \\ a^\top &= \begin{bmatrix} a_0 & a_1 & \ldots & a_m \end{bmatrix} \\ b^\top &= \begin{bmatrix} b_0 & b_1 & \ldots & b_m \end{bmatrix} \\ z_i^\top &= \begin{bmatrix} y_i & \ldots & y_{i-m} & u_i & \ldots & u_{i-m} \end{bmatrix} \end{aligned}$$

in which the shorthand notation $z_i = z[i]$ has been used, such that $\theta^\top z_{0,i} = 0 \quad \forall i = m+1 \ldots N$ with m being the memory (or order) of the model and $\tilde{z}_i = z_i - z_{0,i}$ is the noise contribution. Notice that the system is linear in components y_i and u_i as well as in model parameters a_k and b_k with $0 \leqq k \leqq m$. Furthermore, introduce the observation sample and noise covariance matrices as

$$\begin{aligned} D &= \mathbb{E}\left(zz^\top\right) \approx \frac{1}{N-m}\sum_{i=m+1}^{N} z_i z_i^\top \\ \mu C &= \mathbb{E}\left(\tilde{z}\tilde{z}^\top\right) \approx \frac{1}{N-m}\sum_{i=m+1}^{N} \tilde{z}_i \tilde{z}_i^\top \end{aligned}$$

with μ denoting noise magnitude and C representing a normalized noise covariance matrix, or noise (covariance) structure, that expresses the relative distribution of noise between input and output.

The essence of the Koopmans–Levin method is that the (full-rank) sample covariance matrix D comprising of noisy observations can be decomposed into a (rank-deficient) noise-free component D_0 and a noise component C:

$$\theta^\top D\theta = \theta^\top D_0\theta + \theta^\top \mu C\theta = \theta^\top \mu C\theta$$

in which $\theta^\top D\theta = \theta^\top \mathbb{E}\left(zz^\top\right)\theta = \mathbb{E}\left(\theta^\top zz^\top\theta\right)$ and $\theta^\top D_0\theta = \theta^\top \mathbb{E}\left(z_0 z_0^\top\right)\theta = 0$ so that finding θ entails minimizing the objective function

$$J = \frac{1}{2}\frac{\theta^\top D\theta}{\theta^\top C\theta} \tag{2}$$

which can be effectively tackled by solving the eigenvector decomposition problem

$$\left(D - \mu C\right)\theta = 0$$

or

$$\det\left(D_0\right) = \det\left(D - \mu C\right) = 0$$

so that the model parameter vector is found by solving a generalized eigenvector problem on the matrix pair $(D,\, C)$. The problem may alternatively be formulated using matrix notation where

$$\begin{aligned} Z &= \begin{bmatrix} y_1 & \cdots & y_m & u_1 & \cdots & u_m \\ y_2 & \cdots & y_{m+1} & u_2 & \cdots & u_{m+1} \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix} \\ D &= Z^\top Z \end{aligned}$$

with Z being an $(N - m + 1) \times 2m$ matrix and $C = \mu C_\rho \otimes I_m$ denoting the noise structure such that the noise covariance matrix is known up to a multiplication by a scalar μ representing the noise magnitude, i.e.

$$\mu C_\rho = \mu \begin{bmatrix} \sin^2\rho & 0 \\ 0 & \cos^2\rho \end{bmatrix} = \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix} \tag{3}$$

in which we assume that the noise structure matrix C_ρ is preliminarily known.

One way [13] to improve the robustness of the parameter estimation approach outlined above is by instead of (2) minimizing the objective function

$$J = \frac{1}{2}\frac{1}{q-m}\operatorname{trace}\left(G_q^\top Z_q^\top \left(G_q^\top C_q G_q\right)^{-1} Z_q G_q\right)$$

where Z_q is an $(N - q + 1) \times 2q$ matrix obtained by augmenting Z with $q - m$ columns of additional past observations for both y and u; G_q is a $2q\times(q-m)$ matrix of model parameters such that $Z_{0,q}G_q = 0$; and $C_q = (\mu C_\rho) \otimes I_q$ is a diagonal covariance structure matrix of size $2q \times 2q$, and $m + 1 \leqq q \leqq N$, i.e.

$$G_\circ = \begin{bmatrix} \circ_0 & & & \\ \circ_1 & \circ_0 & & \\ \vdots & \circ_1 & & \\ \circ_m & \vdots & \ddots & \circ_0 \\ & \circ_m & & \circ_1 \\ & & \ddots & \vdots \\ & & & \circ_m \end{bmatrix}_{q,\, q-m}$$

$$G_q = \begin{bmatrix} G_y \\ -G_u \end{bmatrix}_{2q,\, q-m}$$

$$\begin{aligned} Z_{q,y} &= \begin{bmatrix} y_1 & \cdots & y_q \\ y_2 & \cdots & y_{q+1} \\ \vdots & & \vdots \\ y_{N-q+1} & \cdots & y_N \end{bmatrix} \\ Z_{q,u} &= \begin{bmatrix} u_1 & \cdots & u_q \\ u_2 & \cdots & u_{q+1} \\ \vdots & & \vdots \\ u_{N-q+1} & \cdots & u_N \end{bmatrix} \\ Z_q &= \begin{bmatrix} Z_{q,y} & Z_{q,u} \end{bmatrix} \end{aligned}$$

in which $Z_{0,q}G_q = 0$, $C_q = (\mu C_\rho) \otimes I_q$ is a diagonal covariance structure matrix of size $2q$, and $m + 1 \leqq q \leqq N$. Notice that both the model parameter vector θ and the original observation matrix Z have been extended from size m to q.

By rearranging the factors of the product within the trace operator, the above problem can be reformulated as

$$J = \frac{1}{2}\frac{1}{q-m}\operatorname{trace}\left(\left(G_q^\top C_q G_q\right)^{-1} G_q^\top D_q G_q\right) \tag{4}$$

which can be gradually approximated with the iteration scheme

$$\theta_{k+1} = \arg\min_{\theta} \frac{\operatorname{trace}\left(\gamma^{-1}(\theta_k)\delta(\theta)\right)}{\operatorname{trace}\left(\gamma^{-1}(\theta_k)\gamma(\theta)\right)}$$

with

$$\begin{aligned} \delta(\theta) &= G_q^\top(\theta) D_q G_q(\theta) \\ \gamma(\theta) &= G_q^\top(\theta) C_q G_q(\theta) \end{aligned}$$

which yields the same extreme value upon convergence. To facilitate easier computation, the above scheme is equivalent to

$$\theta_{k+1} = \arg\min_{\theta} \frac{\theta^\top T^\top \left(\gamma^{-1}(\theta_k) \otimes D_q\right) T\theta}{\theta^\top T^\top \left(\gamma^{-1}(\theta_k) \otimes C_q\right) T\theta} \quad (5)$$

where T (a sparse matrix of zeros and ones) is chosen such that $\operatorname{vec}(G_q) = T\theta$. In each iteration, minimization w.r.t. θ is attained by solving a generalized eigenvector decomposition problem on the matrix pair $(Q,\ R)$ with

$$\begin{aligned} Q &= T^\top \left(\left(G_q^\top(\theta_k) C_q G_q(\theta_k) \right)^{-1} \otimes D_q \right) T \\ R &= T^\top \left(\left(G_q^\top(\theta_k) C_q G_q(\theta_k) \right)^{-1} \otimes C_q \right) T \end{aligned}$$

where θ is the eigenvector that belongs to the smallest eigenvalue μ.

3 A nonlinear extension

In order to further generalize the Koopmans–Levin method to nonlinear systems, linear components y_i and u_i give way to the nonlinearity terms that occur in the model. Each nonlinearity term t_i (where the index denotes the time instant i) is a product of input and output variables with every possible time shift $0 \leqq k \leqq m$, each raised to a given nonnegative integer power, i.e.

$$t_i = \prod_{k=0}^{m} y_{i-k}^{p_k^y} \prod_{k=1}^{m} u_{i-k}^{p_k^u}$$

where p_k^y is the possibly zero exponent of the variable y with time shift k in the given nonlinearity term, and p_k^u is defined likewise. The zero time shift in u is not permitted to ensure a causal system.

For instance, for the nonlinearities y_i, u_i, u_i^2 and $y_i u_i$, Z_q takes the form

$$Z_q = \left[\begin{array}{cccccccccccc} y_1 & \cdots & y_q & u_1 & \cdots & u_q & u_1^2 & \cdots & u_q^2 & y_1 u_1 & \cdots & y_q u_q \\ y_2 & \cdots & y_{q+1} & u_2 & \cdots & u_{q+1} & u_2^2 & \cdots & u_{q+1}^2 & y_2 u_2 & \cdots & y_{q+1} u_{q+1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ y_{N-q+1} & \cdots & y_N & u_{N-q+1} & \cdots & u_N & & & & & & \end{array} \right]$$

where Z_q is an $(N-q+1) \times nq$ matrix, n being the number of nonlinear components (in our case, $n = 4$), and the general form for the matrix G_q becomes

$$G_q = \begin{bmatrix} G_q^1 \\ G_q^2 \\ \vdots \\ G_q^n \end{bmatrix}_{nq,\, q-m}$$

with G_q^k encapsulating the parameters for the kth nonlinearity. Notice that the matrix product $Z_{q,0} G_q$ entails that the system is still linear in parameters. However, the covariance matrix structure C_q is no longer a single diagonal matrix but is replaced by a matrix polynomial

$$C_q(\mu) = \mu C_q^{(1)} + \mu^2 C_q^{(2)} + \ldots + \mu^p C_q^{(p)}$$

in which $C_q^{(k)}$ is the kth coefficient of the matrix polynomial $C_q(\mu)$. One can use the following identities in deriving C_q:

$$\begin{aligned} \mathbb{E}\left(x_i^p\right) &= \mathbb{E}\left(x_{0,i} + n_i\right)^p \\ \mathbb{E}\left(n_i^{2p}\right) &= (2p-1)(2p-3)\ \ldots\ 1\,\sigma^{2p} \\ \mathbb{E}\left(n_i^{2p-1}\right) &= 0 \\ \mathbb{E}\left(x_i\right) &\approx \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \\ \mathbb{E}\left(x_i x_{i-\tau}\right) &= \mathbb{E}\left(x_i\right)\mathbb{E}\left(x_{i-\tau}\right) \end{aligned}$$

in which n_i is the value of a zero-mean σ^2-variance normally distributed random variable at time instant i. For example,

$$\mu \begin{bmatrix} \sigma_y^2 & 0 & 0 & 0 & \bar{y}\sigma_u^2 & \bar{y}\sigma_u^2 & \bar{u}\sigma_y^2 & 0 \\ 0 & \sigma_y^2 & 0 & 0 & \bar{y}\sigma_u^2 & \bar{y}\sigma_u^2 & 0 & \bar{u}\sigma_y^2 \\ 0 & 0 & \sigma_u^2 & 0 & 3\bar{u}\sigma_u^2 & \bar{u}\sigma_u^2 & \bar{y}\sigma_u^2 & 0 \\ 0 & 0 & 0 & \sigma_u^2 & \bar{u}\sigma_u^2 & 3\bar{u}\sigma_u^2 & 0 & \bar{y}\sigma_u^2 \\ \bar{y}\sigma_u^2 & \bar{y}\sigma_u^2 & 3\bar{u}\sigma_u^2 & \bar{u}\sigma_u^2 & 6\sigma_u^2\bar{u}^2 & 2\sigma_u^2\bar{u}^2 & 3\bar{y}\bar{u}\sigma_u^2 & \bar{y}\bar{u}\sigma_u^2 \\ \bar{y}\sigma_u^2 & \bar{y}\sigma_u^2 & \bar{u}\sigma_u^2 & 3\bar{u}\sigma_u^2 & 2\sigma_u^2\bar{u}^2 & 6\sigma_u^2\bar{u}^2 & \bar{y}\bar{u}\sigma_u^2 & 3\bar{y}\bar{u}\sigma_u^2 \\ \bar{u}\sigma_y^2 & 0 & \bar{y}\sigma_u^2 & 0 & 3\bar{y}\bar{u}\sigma_u^2 & \bar{y}\bar{u}\sigma_u^2 & \bar{y}^2\sigma_u^2 + \sigma_y^2\bar{u}^2 & 0 \\ 0 & \bar{u}\sigma_y^2 & 0 & \bar{y}\sigma_u^2 & \bar{y}\bar{u}\sigma_u^2 & 3\bar{y}\bar{u}\sigma_u^2 & 0 & \bar{y}^2\sigma_u^2 + \sigma_y^2\bar{u}^2 \end{bmatrix} - \mu^2 \begin{bmatrix} \cdots & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ & 3\sigma_u^4 & \sigma_u^4 & 0 & 0 \\ & \sigma_u^4 & 3\sigma_u^4 & 0 & 0 \\ & 0 & 0 & \sigma_y^2\sigma_u^2 & 0 \\ \cdots & 0 & 0 & 0 & \sigma_y^2\sigma_u^2 \end{bmatrix}$$

Figure 2: Example covariance matrix polynomial for the nonlinear components y_i, u_i, u_i^2 and $y_i u_i$ with $q = 2$. Entries not shown take a value of zero.

$$\begin{aligned} \mathbb{E}x^2 &= \mathbb{E}(x_0 + \tilde{x})^2 \\ &= \mathbb{E}x_0^2 + 2\mathbb{E}x_0\tilde{x} + \mathbb{E}\tilde{x}^2 = \mathbb{E}x_0^2 + \sigma^2 \\ \mathbb{E}x^4 &= \mathbb{E}(x_0 + \tilde{x})^4 \\ &= \mathbb{E}x_0^4 + 4\mathbb{E}x_0^3\tilde{x} + 6\mathbb{E}x_0^2\tilde{x}^2 + 4\mathbb{E}x_0\tilde{x}^3 + \mathbb{E}\tilde{x}^4 \\ &= \mathbb{E}x_0^4 + 6\mathbb{E}x_0^2\tilde{x}^2 + 3\sigma^4 \\ &= \mathbb{E}x_0^4 + 6\left(\mathbb{E}x^2 - \sigma^2\right)\sigma^2 + 3\sigma^4 \\ &= \mathbb{E}x_0^4 + 6\sigma^2\mathbb{E}x^2 - 3\sigma^4 \end{aligned}$$

and for terms with different time delays,

$$\begin{aligned} & \mathbb{E}\left(x_{k+1}\right)\left(x_{k-\tau}^4\right) = \\ = \; & \mathbb{E}x_{0,k+1}\left(\mathbb{E}x_{0,k-\tau}^4 + 6\sigma^2\mathbb{E}x_{0,k-\tau}^2 + 3\sigma^4\right) \\ = \; & \mathbb{E}x_{0,k+1}\left(\mathbb{E}x_{0,k-\tau}^4 + 6\sigma^2\mathbb{E}x_{k-\tau}^2 - 3\sigma^4\right) \\ = \; & \mathbb{E}x_{0,k+1}\mathbb{E}x_{0,k-\tau}^4 + \\ & \qquad + 6\sigma^2\mathbb{E}x_{k+1}\mathbb{E}x_{k-\tau}^2 - 3\sigma^4\mathbb{E}x_{k+1} \\ \approx \; & \mathbb{E}x_{0,k+1}\mathbb{E}x_{0,k-\tau}^4 + 6\sigma^2\bar{x}^3 - 3\sigma^4\bar{x}. \end{aligned}$$

A detailed description as well as a more comprehensive example on how to derive these terms is discussed in [14]. Figure 2 shows a sample covariance polynomial for $q = 2$. Observe that the matrix entries usually depend not only on noise parameters σ_u and σ_y but also on (means of) the observations themselves.

The use of covariance matrix polynomials instead of regular covariance matrices necessitates some modifications to the objective function (4) as well as the iteration scheme (5). By including the noise magnitude within the trace operator in (4) yielding

$$\operatorname{trace}\left(\left(G_q^\top \mu C_q G_q\right)^{-1} G_q^\top D_q G_q\right),$$

it is apparent that the matrix product approaches the unit matrix should the best possible model parameters and noise covariance matrix be used. In this spirit, (4) can be reformulated as

$$J = \left(\operatorname{trace}\left(\left(G_q^\top \mu C_q G_q\right)^{-1} G_q^\top D_q G_q\right) - d\right)^2 \tag{6}$$

with $d = q - m$ and m being the the order of the dynamic model to estimate where the minimum of J is attained when both model and noise magnitude estimates best match observations. Substituting the noise covariance polynomial $C_q(\mu)$ into (6), we get a parameter estimation scheme for nonlinear systems. Thus, we propose the following differentiable objective function:

$$J = \left(\operatorname{trace}\left(\gamma^{-1}\delta\right) - d\right)^2 \tag{7}$$

where

$$\begin{aligned} \delta(\theta) &= G_q^\top(\theta) D_q G_q(\theta) \\ \gamma(\theta,\, \mu) &= G_q^\top(\theta) C_q(\mu) G_q(\theta). \end{aligned}$$

As the function (7) is differentiable, a direct search utilizing the Levenberg-Marquardt method yields model parameter and noise magnitude estimates. However, the Levenberg-Marquardt method finds only local minima, making the scheme sensitive to initial values.

Iterative schemes are more robust against local minima. Modifying (5) to incorporate the covariance polynomial $C_q(\mu)$ we get

$$\bar{\theta}_{k+1} = \arg\min_{\theta,\mu} \frac{\theta^\top T^\top \left(\gamma^{-1}(\theta_k,\, \mu_k) \otimes D_q\right) T\theta}{\theta^\top T^\top \left(\gamma^{-1}(\theta_k,\, \mu_k) \otimes C_q(\mu)\right) T\theta} \tag{8}$$

where $\bar{\theta} = \left[\begin{array}{cc} \theta & \mu \end{array}\right]$ needs the solution of a polynomial eigenvalue decomposition problem

$$\Psi(\mu)\theta = \left(Q - \mu R_1 - \mu^2 R_2 - \ldots - \mu^p R_p\right)\theta = 0 \tag{9}$$

with

$$\begin{aligned} Q &= T^\top \left(\gamma^{-1}(\theta_k,\, \mu_k) \otimes D_q\right) T \\ R_i &= T^\top \left(\gamma^{-1}(\theta_k,\, \mu_k) \otimes C_q^{(i)}\right) T. \end{aligned}$$

Further reduction in the computational space is possible if a priori knowledge of the equality of certain parameters is available. Introducing the structural constraint matrix S with 0 and 1 entries such that $S\theta = \theta_R$, it is possible to augment the above equations to yield

$$\begin{aligned} Q &= S^\top T^\top \left(\gamma^{-1}(\theta_k,\ \mu_k) \otimes D_q\right) TS \\ R_i &= S^\top T^\top \left(\gamma^{-1}(\theta_k,\ \mu_k) \otimes C_q^{(i)}\right) TS \end{aligned}$$

and restrict the search for parameter estimates in θ_R with $\dim \theta_R \leqq \dim \theta$. For example, given the (non-controllable) nonlinear system

$$\begin{aligned} a x_{0,k+1} x_{0,k-\tau}^p + b x_{0,k} x_{0,k-\tau}^p + \\ + a x_{0,k+1} + b x_{0,k} + c x_{0,k-\tau} &= 0 \end{aligned}$$

with parameters a, b and c

$$S^\top = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

One way to solve a polynomial eigenvector decomposition problem is by linearization. As a result of linearization, the polynomial eigenvalue problem reduces to a generalized eigenvalue problem. In particular, (9), when subject to symmetry-preserving linearization [1] becomes for even p

$$\begin{aligned} \psi(\mu) &= \operatorname{diag}\left(\begin{bmatrix} 0 & I \\ I & R_1 \end{bmatrix}, \begin{bmatrix} 0 & I \\ I & R_3 \end{bmatrix}, \ldots,\right. \\ &\qquad \left. \ldots, \begin{bmatrix} 0 & I \\ I & R_{p-1} \end{bmatrix}\right) \\ &- \mu \operatorname{diag}\left(Q^{-1}, \begin{bmatrix} -R_2 & I \\ I & 0 \end{bmatrix}, \ldots\right. \\ &\qquad \left. \ldots, \begin{bmatrix} -R_{p-2} & I \\ I & 0 \end{bmatrix}, -R_p\right) \end{aligned}$$

and for odd p

$$\psi(\mu) = \operatorname{diag}\left(Q, \begin{bmatrix} 0 & I \\ I & R_2 \end{bmatrix}, \begin{bmatrix} 0 & I \\ I & R_4 \end{bmatrix}, \ldots\right)$$

$$-\mu \operatorname{diag}\left(\begin{bmatrix} -R_1 & I \\ I & 0 \end{bmatrix}, \begin{bmatrix} -R_3 & I \\ I & 0 \end{bmatrix}, \ldots, -R_p\right)$$

where the operator diag aligns its arguments to bring forth a block diagonal matrix. As the linearized problem has eigenvectors of dimension mp rather than m, the true polynomial eigenvector that belongs to the eigenvalue μ becomes the portion v_k of the linearized eigenvector $\psi(\mu)x = 0$ that gives the smallest normalized residual, i.e.

$$v_k = \arg\min_{v_k} \frac{\sum_k |(\Psi(\mu)v)_k|}{\sum_k |v_k|}.$$

With the iterative scheme (8) at hand, a few initial iterations can be used to seed the Levenberg-Marquardt search with appropriate initial values reducing the likelihood of (7) getting stuck in a local minimum.

4 Simultaneous model and noise parameter estimation

Contrary to the hidden assumption in the previous section, in a real-world scenario, the true noise structure C_ρ (or equivalently, a noise direction ρ that determines the ratio of input and output noise variances for a unit magnitude noise) is seldom at our disposal. As the final step of the parameter estimation method, we propose means to estimate C_ρ for white noise.

One way to parametrize noise variances, as in (3), is by writing $\sigma_u^2 = \mu \cos^2 \rho$ and $\sigma_y^2 = \mu \sin^2 \rho$ such that $C_q = C_q(\mu,\ \rho)$. Let $\hat\theta$ denote (unit-normalized) estimates obtained with a particular assumption of ρ using (8). Introduce the notations

$$\begin{aligned} \hat\delta &= G_q^\top(\hat\theta) D_q G_q(\hat\theta) \\ \hat\gamma &= G_q^\top(\hat\theta) C_q(\hat\mu,\ \rho) G_q(\hat\theta). \end{aligned}$$

Varying ρ in the range from 0 to $\frac{\pi}{2}$, one can discover the "true" value by minimizing the loss function

$$J(\rho) = \left\| \hat\delta - \hat\gamma \right\|_F \tag{10}$$

where $\|\cdot\|_F$ denotes the Frobenius norm (a technique called *covariance matching* in [15]) or the "inverted" loss function

$$J(\rho) = \operatorname{trace}\left(\hat\delta^{-1}\hat\gamma\right) \tag{11}$$

or the so-called Itakura–Saito divergence

$$J(\rho) = \operatorname{trace}(\gamma\delta^{-1}) - \log(\det(\gamma\delta^{-1})) - n \tag{12}$$

where n is the dimension of the square matrices involved. The minimum value for J in the above equations yields the optimal value for ρ.

As an alternative to the two-stage estimation procedure outlined above, a single-stage strategy might theoretically also be employed. Let

$$\begin{aligned} \delta(\theta) &= G_q^\top(\theta) D_q G_q(\theta) \\ \gamma(\theta,\ \mu,\ \rho) &= G_q^\top(\theta) C_q(\mu,\ \rho) G_q(\theta) \end{aligned}$$

and introduce the objective functions J_1, J_2 and J_3

$$\begin{aligned} J_1 &= \left(\mathrm{trace}\left(\gamma^{-1}(\theta,\, \mu,\, \rho)\delta(\theta)\right) - (q-m)\right)^2 \\ J_2 &= \left(\mathrm{trace}\left(\delta^{-1}(\theta)\gamma(\theta,\, \mu,\, \rho)\right) - (q-m)\right)^2 \\ J_3 &= \|J_d\|_F^2 = \mathrm{trace}\left(J_d^\top J_d\right) \\ J_d &= \left(\delta(\theta) - \gamma(\theta,\, \mu,\, \rho)\right). \end{aligned}$$

Notice that all are functions of θ, μ as well as ρ simultaneously. Due to their nonlinearity, they are likely to exhibit convergence to local minima if started with the wrong initial values.

5 Simulation results

In order to demonstrate the feasibility of the outlined approach, in this section we show simulation results. However, before we present the actual results, we briefly describe the simulation environment in which the data were generated.

5.1 Simulation environment

In order to produce simulation data, we used *Fræser* [4], an extensible object-oriented graphical framework for estimating errors-in-variables systems. Estimation algorithms it supports include maximum likelihood, approximated maximum likelihood, Koopmans and instrumental variable methods for static systems, as well as bias-compensating, extended instrumental variable methods, the Frisch scheme, and linear and nonlinear, regular and generalized Koopmans–Levin methods for dynamic systems. The framework is implemented in MatLab with performance-critical operations optionally performed by external C routines and rich graphical capabilities provided by Java Swing classes. It supports defining a system setup interactively with a property editor and investigating system behavior with charts and plots. System setups can be loaded and saved as M-files, while simulation results can be exported to a spreadsheet or to LaTeX. Apart from the graphical front-end, it exposes both a true object-oriented (with new-style MatLab classes) as well as a functional-style programming interface.

From a design point of view, the framework encompasses a general data flow model to generate data, feed the data to a linear or nonlinear static system or a dynamic process, contaminate the data with noise, and hand over noisy observations to an estimator to derive model and optionally noise parameters. The general data flow model is customized by plugging in the desired data generator, static or dynamic system, noise model and estimator algorithm. The pluggable components realize a common base class that provides fundamental metadata services to facilitate presentation over the user interface. The framework includes several examples on usage, including the system and estimation scheme setup that is discussed in the following section. The noise covariance matrix polynomials that are necessary for covariance-based methods are automatically generated using a built-in symbolic polynomial manipulator and written into an M-file for faster execution.

5.2 Results for a polynomial system

Let us draw our attention to an artificial yet relatively complex process described by the nonlinear relationship

$$\begin{aligned} y_{0,i} &= p_1 y_{0,i-1} + p_2 y_{0,i-2} + p_3 u_{0,i-1} + \\ &+ p_4 u_{0,i-1}^2 + p_5 y_{0,i-1} y_{0,i-2} \\ &+ p_6 u_{0,i-1} y_{0,i-1} \end{aligned}$$

comprising of both linear and polynomial terms as well as cross-correlating terms. The true parameter values are set to

$$\begin{array}{ccc} p_1 = 1.5 & p_2 = -0.7 & p_3 = 1 \\ p_4 = -0.3 & p_5 = -0.05 & p_6 = 0.1 \end{array}$$

Using the outlined nonlinear extension to the generalized Koopmans–Levin method, a simulation example of $N = 500$ samples, $q = 6$, $\sigma_u = 0.01$ and $\sigma_y = 0.01$ has been carried out to produce signals with signal-to-noise ratios of 21dB and 36dB, respectively, where

$$SNR_u[dB] = 10 \log_{10} \frac{\frac{1}{N}\sum_{i=1}^{N} u_0^2}{\frac{1}{N}\sum_{i=1}^{N}(u-u_0)^2}.$$

The parameter estimates that have been obtained with the simulation are shown in Table 1 to be rather close to their true value.

In order to illustrate the consistency of the estimation scheme, a Monte-Carlo simulation of $M = 100$ runs has been carried out. The mean values and variances of parameters thus obtained are shown in Table 2. Table 3 shows how the variance of estimates decreases as the model order q is increased. However, Tables 1, 2 and 3 assume that a noise covariance structure is preliminarily given. Figure 3 shows

term	with constraints			without constraints		
$y_{0,i}$	1.0000	±	0.0000	1.0000	±	0.0000
$y_{0,i-1}$	-1.4996	±	0.0028	-1.4998	±	0.0035
$y_{0,i-2}$	0.6996	±	0.0027	0.6999	±	0.0029
$u_{0,i}$	0.0000	±	0.0034	0.0006	±	0.0122
$u_{0,i-1}$	-1.0005	±	0.0075	-0.9990	±	0.0193
$u_{0,i-2}$	0.0000	±	0.0034	-0.0010	±	0.0150
$u_{0,i}^2$	0.0000	±	0.0034	-0.0008	±	0.0617
$u_{0,i-1}^2$	0.3019	±	0.0409	0.3109	±	0.0934
$u_{0,i-2}^2$	0.0000	±	0.0034	0.0056	±	0.0810
$y_{0,i}y_{0,i-1}$	0.0000	±	0.0034	-0.0009	±	0.0063
$y_{0,i-1}y_{0,i-2}$	0.0499	±	0.0061	0.0515	±	0.0087
$y_{0,i-2}y_{0,i-3}$	0.0000	±	0.0034	-0.0005	±	0.0041
$u_{0,i}y_{0,i}$	0.0000	±	0.0034	0.0009	±	0.0203
$u_{0,i-1}y_{0,i-1}$	-0.1008	±	0.0115	-0.1003	±	0.0317
$u_{0,i-2}y_{0,i-2}$	0.0000	±	0.0034	-0.0009	±	0.0237

Table 1: Comparison of the effectiveness of NGKL parameter estimates with a known noise structure with and without parameter equality constraints.

how to discover noise covariance structure (i.e. the noise "direction" ρ) by minimizing the distance or divergence metrics (10) and (12) over an interval to arrive at estimates for all parameters. Finally, Table 4 shows results of a comprehensive simulation in which all model and noise parameters are estimated simultaneously using the two-stage estimation scheme discussed in Section 4.

For the sake of comparison, an instrumental variable scheme, based on the bias-compensating least-squares technique for nonlinear polynomial systems (PBCLS) has been included in the table. The scheme minimizes the objective function [5]

$$J = \left\| d_{IV} - c_{IV} - (D_{IV} - C_{IV})\bar{\theta} \right\| \tag{13}$$

where

$$\bar{\theta} = (D_{IV} - C_{IV})^{\dagger}(d_{IV} - c_{IV})$$

with D_{IV} and d_{IV} being the (rectangular) covariance matrices of the regressor vector and the output vector, respectively, w.r.t. so-called instruments and $C_{IV}(\mu,\ \rho)$ and $c_{IV}(\mu,\ \rho)$ being the corresponding computed (rectangular) noise covariance matrices. Instruments include the regressor vector as well as past observations not in the system model. The notation $M^{\dagger}$ stands for the Moore-Penrose pseudoinverse $\left(M^{\top}M\right)^{-1}M^{\top}$.

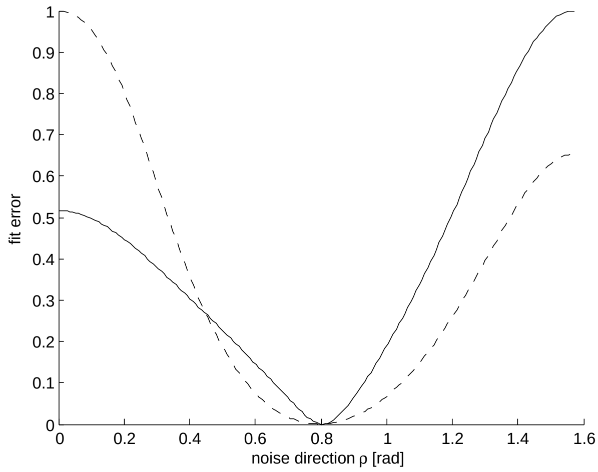

Figure 3: Discovering noise direction by successive estimation over an angle range. The Frobenius metric (10) is shown with continuous line, the Itakura–Saito metric (12) with dashed line. The vertical axis is normalized to the $[0, 1]$ range for the two metrics independently.

term	$N = 500$	$N = 1000$	$N = 2000$	$N = 4000$
$\hat{y}_{0,i-1}$	-1.49995 ± 0.00243	-1.49992 ± 0.00201	-1.49987 ± 0.00135	-1.50005 ± 0.00101
$\hat{y}_{0,i-2}$	0.69990 ± 0.00234	0.70000 ± 0.00187	0.69981 ± 0.00124	0.70012 ± 0.00088
$\hat{u}_{0,i-1}$	-1.00048 ± 0.00871	-1.00040 ± 0.00681	-1.00019 ± 0.00377	-1.00027 ± 0.00330
$\hat{u}^2_{0,i-1}$	0.30568 ± 0.04565	0.30787 ± 0.03676	0.30592 ± 0.01862	0.30946 ± 0.01703
$\hat{y}_{0,i-1}\hat{y}_{0,i-2}$	0.04983 ± 0.00676	0.04980 ± 0.00630	0.04982 ± 0.00342	0.05010 ± 0.00263
$\hat{u}_{0,i-1}\hat{y}_{0,i-1}$	-0.10175 ± 0.01249	-0.10223 ± 0.01092	-0.10115 ± 0.00677	-0.10144 ± 0.00506
$\hat{\mu}\ \left[10^{-3}\right]$	0.19777 ± 0.01267	0.19841 ± 0.00896	0.19884 ± 0.00661	0.19853 ± 0.00514

Table 2: Consistency analysis of NGKL parameter estimates with $q = 6$ and known noise structure.

nonlinearity	$q = 4$	$q = 5$	$q = 6$	$q = 7$
$\hat{y}_{0,i-1}$	-1.4995 ± 0.0021	-1.5000 ± 0.0018	-1.5001 ± 0.0019	-1.4999 ± 0.0020
$\hat{y}_{0,i-2}$	0.6995 ± 0.0020	0.7000 ± 0.0017	0.7000 ± 0.0019	0.7000 ± 0.0018
$\hat{u}_{0,i-1}$	-1.0008 ± 0.0079	-0.9994 ± 0.0065	-0.9992 ± 0.0068	-1.0001 ± 0.0063
$\hat{u}^2_{0,i-1}$	0.3053 ± 0.0422	0.3051 ± 0.0346	0.3092 ± 0.0393	0.3045 ± 0.0391
$\hat{y}_{0,i-1}\hat{y}_{0,i-2}$	0.0497 ± 0.0072	0.0505 ± 0.0058	0.0508 ± 0.0058	0.0494 ± 0.0056
$\hat{u}_{0,i-1}\hat{y}_{0,i-1}$	-0.1022 ± 0.0116	-0.1008 ± 0.0116	-0.1011 ± 0.0117	-0.1007 ± 0.0107
$\hat{\mu}\ \left[10^{-3}\right]$	0.1991 ± 0.0112	0.1970 ± 0.0095	0.1989 ± 0.0094	0.1983 ± 0.0079

Table 3: Simulation results of $N = 1000$ observations with known noise structure as the generalized Koopmans–Levin model order is increased.

nonlinearity	PBCLS (13)	Frobenius norm (10)	“inverted” (11)	Itakura-Saito (12)
$\hat{y}_{0,i-1}$	-1.5002 ± 0.0056	-1.5002 ± 0.0021	-1.4999 ± 0.0019	-1.4999 ± 0.0023
$\hat{y}_{0,i-2}$	0.7003 ± 0.0046	0.7003 ± 0.0019	0.6999 ± 0.0018	0.7000 ± 0.0021
$\hat{u}_{0,i-1}$	-0.9988 ± 0.0104	-1.0010 ± 0.0070	-1.0006 ± 0.0069	-0.9992 ± 0.0080
$\hat{u}^2_{0,i-1}$	0.2976 ± 0.3531	0.3063 ± 0.0385	0.3116 ± 0.0321	0.3100 ± 0.0412
$\hat{y}_{0,i-1}\hat{y}_{0,i-2}$	0.0503 ± 0.0059	0.0494 ± 0.0061	0.0500 ± 0.0059	0.0504 ± 0.0064
$\hat{u}_{0,i-1}\hat{y}_{0,i-1}$	-0.1008 ± 0.0341	-0.1030 ± 0.0130	-0.1022 ± 0.0115	-0.1006 ± 0.0125
$\hat{\sigma}^2_y\ \left[10^{-3}\right]$	0.1003 ± 0.0136	0.1001 ± 0.0087	0.1016 ± 0.0094	0.1000 ± 0.0091
$\hat{\sigma}^2_u\ \left[10^{-3}\right]$	0.0957 ± 0.0448	0.0969 ± 0.0121	0.0976 ± 0.0107	0.0966 ± 0.0127

Table 4: Simulation results with unknown noise structure and various methods to match covariance matrices.

6 Conclusion

After a brief description of the generalized Koopmans–Levin method for linear systems, we have introduced a nonlinear extension by augmenting the observation matrix Z_q with blocks that stand for nonlinear terms. The extension has propagated to the noise covariance matrix C_q, which would become a matrix polynomial in terms of the noise magnitude μ. A new objective function and an iteration scheme have been proposed, whose dependency on the model parameters θ inherited from the GKL method for linear systems has been extended with dependency on the noise magnitude μ. Consequently, the problem is tackled efficiently by solving a polynomial rather than a generalized eigenvector decomposition problem. Next, an optimization scheme for minimizing an error term over a bounded variable representing noise "direction" has been shown to yield a noise structure estimate. Combining these two, we have obtained a method that yields estimates for both model and noise parameters of a discrete-time dynamic errors-in-variables system that is polynomial in terms of observations. The applicability of the method has been demonstrated with simulation results.

Acknowledgments

This work has been supported by the fund of the Hungarian Academy of Sciences (HAS) for control research and the Hungarian National Research Fund (grant number T68370).

References

[1] Efstathios N. Antoniou and Stavros Vologiannidis. Linearizations of polynomial matrices with symmetries and their applications. *Electronic Journal of Linear Algebra*, 15:107–114, February 2006.

[2] Roberto Diversi, Roberto Guidorzi, and Umberto Soverini. A new criterion in EIV identification and filtering applications. In *Proc. of 13th IFAC Symposium on System Identification*, pages 1993–1998, 2003.

[3] Mats Ekman, Mei Hong, and Torsten Söderström. A separable nonlinear least-squares approach for identification of linear systems with errors in variables. In *14th IFAC Symposium on System Identification*, pages 178–183, March 2006.

[4] Levente Hunyadi. Fræser: a MatLab framework for estimating errors-in-variables systems. Available online: http:\fraeser.sourceforge.net, June 2009.

[5] Tomasz Larkowski, Jens G. Linden, Benoit Vinsonneau, and Keith J. Burnham. Novel algorithms based on conjunction of the Frisch scheme and extended compensated least squares. In *International Conference on Control*, Manchester, UK, 2008.

[6] Morris J. Levin. Estimation of a system pulse transfer function in the presence of noise. In *Proc. of Joint Automatic Control Conference*, pages 452–458, 1963.

[7] Ivan Markovsky, Alexander Kukush, and Sabine Van Huffel. On errors-in-variables with unknown noise variance ratio. In *Proc. of 14th IFAC Symposium on System Identification*, pages 172–177, March 2006.

[8] Jose De Jesus Rubio Avila, Andres Ferreyra Ramirez, and Carlos Aviles-Cruz. Nonlinear system identification with a feedforward neural network and an optimal bounded ellipsoid algorithm. *WSEAS Transactions on Computers*, 7(5):542–551, May 2008.

[9] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume. Frequency domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, 33:1073–1086, 1997.

[10] Ginalber Luiz de Oliveira Serra. Proposal of numerically robust algorithm for stochastic systems identification. *WSEAS Transactions on Electronics*, 4(5):101–108, May 2007.

[11] Torsten Söderström. Errors-in-variables methods in system identification. *Automatica*, 43:939–958, 2007.

[12] Stéphane Thil, Mei Hong, Torsten Söderström, Marion Gilson, and Hugues Garnier. Statistical analysis of a third-order cumulants based algorithm for discrete-time errors-in-variables identification. In *Proc. of 17th IFAC World Congress*, pages 420–425, Seoul, Korea, July 2008.

[13] István Vajk. Identification methods in a unified framework. *Automatica*, 41(8):1385–1393, 2005.

[14] István Vajk and Jenõ Hetthéssy. Identification of nonlinear errors-in-variables models. *Automatica*, 39:2099–2107, 2003.

[15] István Vajk and Jenõ Hetthéssy. Efficient estimation of errors-in-variables models. In *17th IFAC World Congress*, pages 1384–1389, Seoul, Korea, July 2008.

[16] He-Sheng Wang and Ann Kao. A behavioral approach to GNSS positioning and DOP determination. *WSEAS Transactions on Systems*, 7(9):854–869, September 2008.

[17] Wei Xing Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7):1142–1147, July 2002.