
Available from: http://dx.doi.org/10.3182/20120711-3-be-2027.00391

Accessed from: http://hdl.handle.net/1959.13/1326326
An identification method for
Errors-in-Variables systems using incomplete data

Rodrigo Carvajal, Ramón Delgado, Juan C. Agüero, Graham C. Goodwin.

School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW, 2308, Australia.
Email addresses: {Rodrigo.Carvajal,Ramon,Delgado}@uon.edu.au, {Juan.Agüero,Graham.Goodwin}@newcastle.edu.au

Abstract: In this paper we develop a novel identification algorithm for Errors-in-Variables systems (represented in transfer function form) using incomplete data. We propose a Maximum Likelihood formulation in the frequency domain that considers a restricted frequency range from the available measurements. We compare the proposed technique with the traditional frequency domain system identification technique applied to Errors-in-Variables systems.

Keywords: System Identification, Errors-in-Variables systems, Maximum Likelihood, Frequency Domain Identification, Incomplete Data.

1. INTRODUCTION

Errors in Variables (EIV) systems, also known as “measurement error models”, are systems where both the input and output of the system are measured with error (Söderström [2007], Fuller [1978], Cheng and Van Ness [1999]). This kind of system has been studied in a wide range of applications, such as chemistry, chemical engineering, earth sciences, econometrics, management science, mechanical engineering, finance, ecology, time series analysis, etc (Söderström [2007]).

A general EIV system can be represented as in Fig. 1, where $G(z)$ is a linear dynamic system, $u(t)$ and $y(t)$ are the measured input and output, respectively, and $e_u(t)$ and $e_y(t)$ are the input and output noise. In some cases, prior information regarding the system may be available, which may help to identify the system.

In general, some difficulties arise in EIV system identification. Perhaps the main one is that traditional methods in system identification might not be suitable for EIV systems, since they consider output noise only. Hence, traditional methods typically produce non-consistent and/or biased estimates (see e.g. Adcock [1877, 1878], Cheng and Van Ness [1999], Fuller [1978]). In terms of identifiability, EIV systems are generally not uniquely identifiable from second order properties (Anderson and Deistler [1984], Agüero and Goodwin [2008]) i.e. there is more than one model that has the same input-output spectra. In order to overcome this identifiability difficulty, it is common to make some extra assumptions such as known noise variances (or known noise variance ratio), known order of the polynomials defining the model, and some poles or zeros that are not shared by the different transfer functions that define the system (see e.g. Agüero [2005], Söderström [2007], Agüero and Goodwin [2008], Hjalmarsson, Mårtensson, Rojas and Söderström [2011], Bottegal, Picci, and Pinzoni [2011]).

It is well known that static EIV systems are not identifiable, unless some prior knowledge is available. This leads to algorithms such as Total Least Squares (TLS) (see e.g. Golub and Van Loan [1980], Van Huffel and Vandewalle [1991]) and Orthogonal Regression (see e.g. Cheng and Van Ness [1999]). However, in Solo [1986], it has been pointed out that dynamic EIV systems have different characteristics (when compared with static EIV systems) that modify their identifiability properties. For example, when $L_0(z^{-1}) = L_1(z^{-1}) = L_2(z^{-1}) = 1$, and $G(z^{-1})$ has at least one pole, then, the system is identifiable (see Theorem 14(2) in Agüero and Goodwin [2008]).

On the other hand, the system order is usually not well defined, especially for sampled data models. This extra constraint also limits the applicability of most of the algorithms found in the System Identification literature. Moreover, the needed assumptions (in order to overcome the identifiability difficulty) generate estimates that are not valid in all of the frequency range. This has motivated the development of algorithms that are valid in a restricted frequency range, in which the assumptions (or conditions) are satisfied (see e.g. Goodwin et al [2008]). This idea of using a specific frequency range (for estimation) has previously been used in Econometrics and Statistics (see e.g. Hannan and Robinson [1973], Engle [1974, 1986]) and in Engineering (see e.g. Pintelon and Schoukens [2001], Yuz and Goodwin [2008], Pintelon and Schoukens [2007], Agüero, Yuz, and Goodwin [2008], Agüero, Yuz, and Delgado [2010]). Moreover, the frequency range selection generates the use of a portion of the available data (in the frequency domain), and this leads to an incomplete data problem.

In this paper, we will explore frequency domain system identification applied to EIV systems, focusing on the classical approach (see e.g. Pintelon and Schoukens [2007]), and the corresponding approach following Agüero,

![Fig. 1. Errors in variable system.](image-url)

---

\[ e(t) \quad L_0(z^{-1}) \quad u_0(t) \quad G(z^{-1}) \quad p_0(t) \]
\[ e_u(t) \quad L_1(z^{-1}) \quad u(t) \quad e_y(t) \quad L_2(z^{-1}) \quad y(t) \]
Tang, Yuz, Delgado and Goodwin [2012]. We show that our approach generates a different parametrization of the problem, generating a different likelihood function. This is illustrated by an example.

The remainder of the paper is structured as follows: In Section 2 we present a general framework for EIV systems. In Section 3, we present a general Maximum Likelihood (ML) estimation framework for identification with incomplete data in the frequency domain. In Section 4 we illustrate the benefits of our approach via an example. Finally, in Section 5, we present conclusions.

2. SYSTEM DESCRIPTION

Consider the EIV system shown in Fig. 1, which can be described as (see e.g. Agüero and Goodwin [2008], Söderström [2007]):

\[
u(t) = u_0(t) + L_1(z^{-1})e(u)(t) \tag{1}
\]

\[
y(t) = G(z^{-1})y_0(t) + L_2(z^{-1})e(y)(t) \tag{2}
\]

where, \(G(z^{-1}), L_2(z^{-1}), L_1(z^{-1}),\) and \(L_2(z^{-1})\) are (causal) transfer functions, \(u_0(t) = L_0(z^{-1})e(t)\) is the noise-free input signal, \(e(t), e(u)(t),\) and \(e(y)(t)\) are zero mean Gaussian white noise sequences with covariances \(\sigma_n^2, \sigma_u^2,\) and \(\sigma_y^2\), respectively. Note that, in this paper, \(u(t)\) and \(y(t)\) are taken to be scalar signals. The plant transfer function is assumed to be (see e.g. Söderström [2007]):

\[
G(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} \tag{3}
\]

where

\[
A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_n z^{-n},
\]

\[
B(z^{-1}) = b_1 + \cdots + b_n z^{-n+1}.
\]

The system filters \(L_0(z^{-1}), L_1(z^{-1}), L_2(z^{-1})\) are given by:

\[
L_0(z^{-1}) = C(z^{-1}) D(z^{-1})^{-1},
\]

where

\[
C(z^{-1}) = 1 + c_1 z^{-1} + \cdots + c_n z^{-n},
\]

\[
D(z^{-1}) = 1 + d_1 z^{-1} + \cdots + d_n z^{-n}.
\]

\[
L_1(z^{-1}) = \frac{K(z^{-1})}{M(z^{-1})},
\]

where

\[
K(z^{-1}) = 1 + k_1 z^{-1} + \cdots + k_n z^{-n},
\]

\[
M(z^{-1}) = 1 + m_1 z^{-1} + \cdots + m_n z^{-n},
\]

\[
L_2(z^{-1}) = \frac{F(z^{-1})}{H(z^{-1})},
\]

where

\[
F(z^{-1}) = 1 + f_1 z^{-1} + \cdots + f_n z^{-n},
\]

\[
H(z^{-1}) = 1 + h_1 z^{-1} + \cdots + h_n z^{-n}.
\]

For the remainder of this paper, we consider the parameter \(\theta\) as the vector that contains the coefficients that define the filters \(G, L_0, L_1,\) and \(L_2\).

The noise-free input, \(u_0(t),\) is assumed to be persistently exciting, having spectra different than zero for a sufficient number of frequencies. Moreover, from Fig. 1 and (9)–(14), the noise sequences \(u(t)\) and \(y(t),\) as well as \(u_0(t),\) are autoregressive-moving-average (ARMA) processes, and mutually independent.

---

1 Here, \(z\) should be understood as the forward shift operator.

The EIV system (see Fig. 1) can be represented as a system with a two-dimensional output vector and three mutually independent white noise signals as:

\[
\begin{bmatrix}
\zeta(t)
\end{bmatrix} = \begin{bmatrix}
(y(t))
\end{bmatrix} = \begin{bmatrix}
L_0(z^{-1})G(z^{-1})L_2(z^{-1})
0
0
\end{bmatrix} \begin{bmatrix}
e(t)
e(y)(t)
e(u)(t)
\end{bmatrix},
\]

with

\[
\begin{bmatrix}
e(t)
e(y)(t)
e(u)(t)
\end{bmatrix} \sim \mathcal{N}(0, \Sigma_c), \quad \Sigma_c = \text{diag}(\sigma_n^2, \sigma_u^2, \sigma_y^2)
\]

3. MAXIMUM LIKELIHOOD SYSTEM IDENTIFICATION IN THE FREQUENCY DOMAIN

Frequency domain identification has been widely utilized in the literature (see e.g. Hannan and Robinson [1973], Engle [1974], Pintelon and Schoukens [2001, 2007], Agüero, Yuz, and Goodwin [2007], Yuz and Goodwin [2008]). On the other hand, an important issue in ML frequency-domain identification is that, if the formulations are inconsistent, then time-domain ML does not yield the same estimates as frequency-domain ML. However, time-domain and frequency-domain ML are equivalent when consistent assumptions are applied regarding the initial and final conditions (i.e., transient response) (Agüero, Yuz, Goodwin, and Delgado [2010]). Given this equivalence, a dual time-frequency estimation is proposed in Agüero, Tang, Yuz, Delgado and Goodwin [2012]. In this paper, we extend the approach presented in Agüero, Tang, Yuz, Delgado and Goodwin [2012] to EIV systems written in transfer functions form (see Fig. 1). Indeed, frequency domain identification of EIV systems has been proposed previously in the literature (see e.g. Pintelon and Schoukens [2007]). The key difference between the earlier method and the algorithm presented in this paper corresponds to the way the transient signal is treated, as shown in the sequel.

In order to transform the available data from time domain to frequency domain, we apply the discrete Fourier transform (DFT), given by

\[
Y_k = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} y(t) e^{-j\omega_k t}; \quad k = 0, \ldots, N - 1,
\]

where \(\omega_k = \frac{2\pi k}{N}\). We use capital letters, e.g. \(Y_k\), to represent the Fourier transform of the signal \(y(t)\). In the same way, we define \(U_k\) to represent the Fourier transform of the signal \(u(t)\).

For the development of a ML estimator in the frequency domain, it is convenient to treat the measurements (input and output signals) as one signal, \(\tilde{\xi}\). This is known as a ‘symmetric’ approach (see e.g. Söderström [2007]), where

\[
\tilde{\xi} = [\zeta(0)^T, \ldots, \zeta(N-1)^T]^T.
\]

For future reference, \(\tilde{\xi}\) is taken to be distributed Normal, with mean \(\tilde{\mu}\) and covariance matrix \(\Sigma\).

Given (17), the DFT of \(\tilde{\xi}\) is given by:

\[
\begin{bmatrix}
\tilde{Z}_k
\end{bmatrix} = \begin{bmatrix}
Y_k
\end{bmatrix}, \quad k = 0, 1, \ldots, N - 1.
\]

From (16), it is possible to understand the DFT as a linear transformation of the time-domain data, from \(\mathbb{R}^N\) to \(\mathbb{C}^N\), of the form \(\tilde{Y} = F\tilde{y}\), where \(F\) is the DFT matrix. The elements of \(F\) are given by:

\[
F_{jk} = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \tilde{\xi}_i e^{-j\omega_{kj}},
\]

where \(\omega_{kj} = \frac{2\pi k j}{N}\) for \(j, k = 0, 1, \ldots, N - 1\).
where the covariance matrix of $\varphi$ variances and the transient terms. The corresponding log-
\[ (20) \]

Given (20), it is possible to define $\tilde{Z}$ as a linear transformation of the measurement vector $\tilde{z}$, as:
\[ \tilde{Z} = (F \otimes I_2) \tilde{z}, \quad (21) \]
where $I_2$ is the $(2 \times 2)$ identity matrix, and $A \otimes B$ represents the Kronecker product between matrices $A$ and $B$.

We next review the application of the classical frequency domain ML identification method (see e.g. McKeel and Ljung [1997], Pintelon and Schoukens [2001]) to EIV systems (Pintelon and Schoukens [2007]).

3.1 Classical Approach

In the classical (or traditional) approach for frequency domain ML system identification, the initial condition effects and the leakage effects are taken into account by modelling them as deterministic signals. This approach considers the fact that there is only a finite number of measurements (see e.g., Schoukens, Rolain, and Pintelon [2010]). Specifically, in the frequency domain, after the application of the DFT, the inclusion of the plant ($T_G$), signal ($T_u$, $T_v$), and noise ($T_u$, $T_v$) transient terms, yield the following model (see e.g. Pintelon and Schoukens [2007]):
\[ \tilde{Z}_k = \begin{bmatrix} L_0(e^{j\omega_k}) G(e^{j\omega_k}) & L_2(e^{j\omega_k}) & 0 \\ L_0(e^{j\omega_k}) & 0 & L_1(e^{j\omega_k}) \end{bmatrix} \tilde{E}_k + \tilde{Z}_k, \quad (22) \]
with
\[ \tilde{Z}_k = \begin{bmatrix} T_G(e^{j\omega_k}) & T_0(e^{j\omega_k}) + G(e^{j\omega_k}) T_L(e^{j\omega_k}) \\ T_L(e^{j\omega_k}) & T_0(e^{j\omega_k}) + G(e^{j\omega_k}) T_L(e^{j\omega_k}) \end{bmatrix}, \]
\[ \tilde{E}_k = \begin{bmatrix} E_k \\ E_k^{(n)} \end{bmatrix}, \sim \mathcal{N}(0, \text{diag}(\sigma_0^2, \sigma_1^2, \sigma_2^2)), \]
\[ \tilde{Z}_k \]
Thus, the transient terms are included as parameters in the identification of the system.

Remark 1. The effect of the transient terms is modelled as a deterministic signal. Hence, its value corresponds to the mean value of $\tilde{Z}_k$ (see e.g. Pintelon and Schoukens [2007]).

\[ \nabla \]

Given the frequency domain data representation (22), the ML estimate is given by (see e.g. Pintelon and Schoukens [2001])
\[ \hat{\theta}_{\text{Classic}} = \arg \max_{\theta} \log p(\tilde{Z}|\theta), \quad (23) \]
where $\theta$ is the parameter vector that includes $\theta$, the noise variances and the transient terms. The corresponding log-likelihood function $\log p(\tilde{Z}|\theta)$ is given by:
\[ \log p(\tilde{Z}|\theta) = \sum_{k \in K} \log \det C_{\tilde{Z}_k} + \sum_{k \in K} (\tilde{Z}_k - \tilde{Z}_k) H_k C_{\tilde{Z}_k}^{-1} (\tilde{Z}_k - \tilde{Z}_k), \quad (24) \]
and $K \subseteq \{0, 1, ..., N/2\}$. The matrix $C_{\tilde{Z}_k}$ corresponds to the covariance matrix of $\tilde{Z}_k$, given by
\[ C_{\tilde{Z}_k} = E[(\tilde{Z}_k - \tilde{Z}_k)(\tilde{Z}_k - \tilde{Z}_k)^H] = \begin{bmatrix} S_k^{(YY)} & S_k^{(UY)} \\ S_k^{(UY)} & S_k^{(UU)} \end{bmatrix}, \quad (25) \]
where
\[ S_k^{(YY)} = |G(e^{j\omega_k})|^2 |L_0(e^{j\omega_k})|^2 \sigma_0^2 + |L_2(e^{j\omega_k})|^2 \sigma_2^2, \]
\[ S_k^{(UY)} = |L_0(e^{j\omega_k})|^2 \sigma_0^2 + |L_1(e^{j\omega_k})|^2 \sigma_1^2, \]
\[ S_k^{(UU)} = G(e^{j\omega_k}) |L_0(e^{j\omega_k})|^2 \sigma_0^2, \]
and $S_k^{(UU)}$ is the complex conjugate of $S_k^{(UY)}$ (for more details see e.g. Pintelon and Schoukens [2007]).

We next explore frequency domain ML EIV system identification following the ideas presented in Agüiero, Yuz, Goodwin, and Delgado [2010].

3.2 Alternative approach

In our alternative approach, the effect of the transients is parametrized by the initial conditions of the system. In this way, we can re-define the system as:
\[ \hat{\zeta}(t) = \begin{bmatrix} L_0(z^{-1}) G(z^{-1}) & L_2(z^{-1}) & 0 \\ L_0(z^{-1}) & 0 & L_1(z^{-1}) \end{bmatrix} \tilde{e}(t), \quad (26) \]
\[ \hat{\zeta}(t) \sim \mathcal{N}(\mu_{\beta}, \Gamma_{\beta}), \quad (27) \]
where $\hat{\zeta}(t) = [e(t), c_y(t), c_u(t)]^T$ and $\bar{x}_0$ is the system initial condition. In the sequel, we give details of how to obtain the ML estimates given the alternative configuration.

3.2.1 Obtaining frequency domain data from time domain data

It is well known that the DFT of a real signal provides redundant information (see e.g. Oppenheim, Schafer, and Buck [1999]). On the other hand, the frequency domain data can be expressed in terms of its real and imaginary parts, by using the following real transformation (see e.g. Agüiero, Yuz, Goodwin, and Delgado [2010] for details):
\[ \tilde{Z}^R = \begin{bmatrix} \Re{\tilde{Z}_0}, \Im{\tilde{Z}_0}, \Re{\tilde{Z}_1}, \Im{\tilde{Z}_1}, \ldots, \Re{\tilde{Z}_{N/2}}, \Im{\tilde{Z}_{N/2}} \end{bmatrix}^T, \quad (28) \]
if $N$ is even.
\[ \tilde{Z}^R = \begin{bmatrix} \Re{\tilde{Z}_0}, \Re{\tilde{Z}_1}, \Im{\tilde{Z}_0}, \Im{\tilde{Z}_1}, \ldots, \Re{\tilde{Z}_{N/2}}, \Im{\tilde{Z}_{N/2}} \end{bmatrix}^T, \quad (29) \]
if $N$ is odd.

and $\Re{\{}$ and $\Im{\{}$ represents “the real and imaginary part of”, respectively. Note that $M_R$ in (28) is a square, full-rank, real, unitary matrix, and $K = \{N/2\}$ is the largest integer less than or equal to $N/2$.

For the frequency domain data, $\tilde{Z}^R$, the ML estimate in the frequency domain is given by (see e.g. Agüiero, Tang, Yuz, Delgado and Goodwin [2012])
\[ \hat{\beta}_{\text{ML-Freq}} = \arg \min_{\beta} \log p(\tilde{Z}^R|\beta). \quad (30) \]
Note that $\tilde{Z}^R$ contains all the information in the data since there exists a bijective linear transformation from $\tilde{z}$ to $\tilde{Z}^R$.

3.2.2 Frequency range selection

An embellishment of frequency domain ML estimation is that it is possible to choose a specific range of frequencies, in which the particular model best explains the data. This is useful when a low complexity model is used, or there exist small errors between the system and the hypothesised model (see Goodwin et al [2008]). Frequency range selection implies that not all the available data will be used. This leads to a problem with incomplete data.

A selection of data in the frequency domain can be understood as a linear transformation, denoted by the matrix $S_F$. This selection corresponds to the non-invertible mapping
\[ \tilde{y} = S_F \tilde{Z}^R. \quad (31) \]
This selection in the frequency domain can also be understood as a linear transformation of the available time-domain data, \( \tilde{z} \), as \( \tilde{g} = M\tilde{z} \). Then, we can define the (modified) ML estimate as follows.

**Lemma 1.** Given that \( \tilde{z} \) is distributed Normal, with mean \( \tilde{\mu}(\beta) \) and covariance matrix \( \Sigma(\beta) \), we have that:

\[
\tilde{g}(\beta) \sim N(M\tilde{\mu}(\beta), M\Sigma(\beta)M^T),
\]

and the ML estimator in the frequency domain is given by

\[
\hat{\beta}_{\text{ML-Freq}} = \arg \max_{\beta} p(\tilde{g}|\beta).
\]

**Proof:** First, we notice that, given \( \tilde{g} = M\tilde{z} \), then \( E[\tilde{g}] = ME[\tilde{z}] = M\tilde{\mu}(\beta) \). Next, in the same way, we calculate

\[
E[(\tilde{g} - M\tilde{\mu}(\beta))(\tilde{g} - M\tilde{\mu}(\beta))^T] = ME[(\tilde{z} - \tilde{\mu}(\beta))(\tilde{z} - \tilde{\mu}(\beta))^T]M^T = M\Sigma(\beta)M^T.
\]

Finally, since a linear transformation of a Gaussian distributed random vector yields another Gaussian distributed random vector, we obtain (32), and the ML estimator in the frequency domain, is obtained by maximizing (32) as in (33).

In order to obtain the likelihood function \( p(\tilde{g}|\beta) \), it is necessary to determine the covariance matrix \( \Sigma(\beta) \). Notice that if the initial condition of the system is modelled as a random vector, \( \Sigma(\beta) \) depends on \( x_0 \) (see e.g. Mittnik [1990] and Reinsel [1997]). Moreover, the initial condition is reflected in the mean (vector) \( \tilde{\mu}(\beta) \), since it propagates according to the system characteristics.

### 3.2.3 Computation of the covariance matrix \( \Gamma(\beta) \)

The system model in (26) can be expressed as a general vector ARMA model, since (26) is expressed in terms of transfer functions. Hence, we have (see e.g. Reinsel [1997]):

\[
\zeta(t) = \sum_{n=0}^{p-1} \Phi_n \zeta(t-n) - \sum_{q=1}^{1} \Theta_q \varepsilon(t-n).
\]

Given \( \tilde{z} = [\zeta(0)^T, \ldots, \zeta(N-1)^T]^T \), and defining \( \varepsilon = [\varepsilon(0)^T, \ldots, \varepsilon(N-1)^T]^T \), the model in (34) can be represented in terms of the initial condition as (see e.g. Reinsel [1997]):

\[
\Phi(\theta) \tilde{z} = \Theta(\theta) \varepsilon + F(\theta) x_0,
\]

where \( \Phi(\theta) = (I_N \otimes I_2) - \sum_{n=1}^{p} \Phi_n (I_n \otimes \Phi_{n-1}) \), \( \Theta(\theta) = (I_N \otimes I_2) - \sum_{q=1}^{1} \Theta_q (I_q \otimes \Theta_{q-1}) \), and \( L \) denotes the \( N \times N \) lag matrix that has ones on the (sub)diagonal directly below the main diagonal and zeros elsewhere. The matrix \( F(\theta) \) includes the dynamics of the system and represents the way the initial condition propagates throughout the system in the time domain. For the remainder of the paper, we consider that the initial condition is a deterministic signal. Note that an equivalent description is a stochastic signal with covariance matrix equal to the zero matrix.

If the covariance matrix of the driving noise signals is given by \( \Gamma_x = E[\varepsilon \varepsilon^T] = I_N \otimes \Sigma_e \), then the covariance matrix of \( \tilde{z} \) is given by (see e.g. Reinsel [1997]):

\[
\Sigma(\beta) = \text{Cov}(\tilde{z}) = \Phi^{-1}(\theta)(I_N \otimes \Sigma_e) \Theta^T.
\]

**Remark 2.** The covariance matrix \( \Sigma(\beta) \) can be determined by obtaining the individual covariance matrices \( \Gamma_x(t) \) and \( \zeta(t) \), which depend on the system characteristics. If the initial condition of the system is modelled as a random vector, \( \Gamma_x(t) \) also depends on \( x_0 \) (see e.g. Mittnik [1990] and Reinsel [1997]).

It is possible to see from (23)–(25), that the computation of the (log)likelihood function is simpler in the Classical approach when compared to the alternative method proposed here. However, this does not mean that the Classical approach leads to a function that is easier to optimize. This will be illustrated by an example in the following section.

**Remark 3.** From (35), it can be seen that, if the initial conditions are deterministic (i.e. are not stochastic), they have no effect on the determination of the covariance matrix \( \Sigma(\beta) \).

**Remark 4.** The determination of the initial condition should be constrained by the system dynamics. One way to obtain an initial condition is based on the signal \( \tilde{\zeta} \). However, this choice, in general, results in redundant initial conditions and, hence, the estimation of more parameters than needed. This situation will be illustrated by an example in the following section.

### 4. NUMERICAL EXAMPLE

In this section we present several numerical examples that illustrate the characteristics of the proposed approach. First, we present the differences between the proposed approach and the classical approach. Next, the performance of our proposed approach is shown via an example. Finally, we illustrate the effects of undermodelling on both approaches.

Consider the EIV system as in (26) with

\[
G(z^{-1}) = \frac{1}{1 - az^{-1}}, \quad a = 0.7,
\]

and the noise covariances

\[
\sigma_0^2 = 1, \quad \sigma_1^2 = 0.1, \quad \sigma_2^2 = 0.1.
\]

For the remainder of the paper, the system to be analysed is (partially) determined by (37) and (38).

#### 4.1 Determination of the initial conditions

For simplicity, assume that \( L_0(z^{-1}) = L_1(z^{-1}) = L_2(z^{-1}) = 1 \). From (15) and (37), the output equation of the EIV system is given by

\[
y(t) = \frac{1}{1 - az^{-1}} e(t) + e^{(y)}(t).
\]

It is possible to define an initial condition \( \tilde{x}_0 \) by considering two values: an initial condition for \( y \), and an initial condition for \( e^{(y)} \). However, this definition of \( \tilde{x}_0 \) has redundant information and leads to the challenge of estimating more parameters than actually needed. Indeed, it is possible to consider the signal \( x(t) = \frac{1}{1 - az^{-1}} e(t) \). This choice implies that the current sample of \( e \) is determined by the current and past samples of \( x \). In the same way, the current sample of \( y \) is determined by the current sample of \( e \) and the current and past samples of \( x \). Finally, for the initial condition, \( \tilde{x}_0 \), is only needed for one past sample, defined by \( x(t) \), instead of two. Hence, the initial condition is a scalar, denoted by \( x_0 \).

**Remark 5.** A similar analysis could be performed for a more general system. This is straightforward, although special attention must be paid to the case when there are multiple poles at \( z = 0 \).

**Remark 6.** In the Classical approach, it is necessary to evaluate the logarithm of the determinant of the covariance

\(^2 In a more computationally efficient way.
matrix (23). In this example, the determinant of the covariance matrix is given by:

$$\det(C_Z) = |G(e^{j\omega_k})|^2\sigma_\alpha^2\sigma_\beta^2 + |G(e^{j\omega_k})|^2\sigma_\beta^2, \quad \sigma_\alpha^2 = 1 - |\alpha|^2,$$

where

$$|G(e^{j\omega_k})|^2 = \frac{1}{1 - a e^{j\omega_k}}.$$ 

From the last expression, the log-likelihood function in the Classical approach exhibits a discontinuity at $a = 1$. This poses a problem, since most of the properties of the ML estimator are built upon assumptions (amongst others) the likelihood function is a twice-differentiable function of the parameters throughout its range (see e.g. Stuart, Ord, and Arnold [1999]). Moreover, this discontinuity has the potential to generate numerical difficulties in the optimization of the likelihood function.

\[\nabla\]

4.2 Performance comparison

In this section, we present several examples in order to show the benefits of our approach. For simplicity, we focus on the impact of the estimation of the initial condition in our approach. Regarding the Classical approach, the transient terms $T_u$, $T_y$, and $T_L$ are equal to zero. The transient term $T_G$ is given by $T_G = N^{-1/2}e^{j\omega_k}\alpha$, where $\alpha$ is the difference between the initial and the final condition, $\alpha = x_0 - x_N$ (see e.g. Cauberghe, Guillaume, Pintelon, and Verboven [2006], Agüero, Yuz, Goodwin, and Delgado [2010]). Thus, for the Classical approach, we focus on the estimation of $\alpha$.

Figure 2 shows a plot of the negative log-likelihood function versus parameters, for both, the proposed and the classical approach, for the system defined by (37) and (38). We see that both cost functions have different sensitivity with respect to the transient term. From Fig. 2, it can be seen that not estimating $\alpha$ has a greater impact on the estimation than not estimating $x_0$.

In order to carry out a performance comparison, we consider $N = 100$ measurement points and $N_{sim} = 100$ Monte-Carlo simulations. For this comparison, we set $x_0 = 0$. In the estimation procedure, we assume also $x_0 = 0$, and estimate $\alpha$. The empirical mean and variance of the estimates over the Monte-Carlo simulations is shown in Table 1. From Table 1, it is clearly seen that the two approaches provide very similar estimates, although the two estimation procedures are different.

Next, we consider the case where undermodelling is present. In particular, we assume that the samples are generated with a system that is more complex than the nominal model. In this simulation identification procedure, the undermodelling consists of modifying some of the transfer functions by adding extra dynamics to the system, given by $H(z^{-1}) = 0.5/(1 - 0.5z^{-1})$, which corresponds to one of the transfer functions of the EIV system (15). Thus, we consider the following scenarios:

I The nominal model has the same order as the true system.

\begin{table}[h]
\centering
\caption{Empirical mean and empirical standard deviation of the parameters from $N_{sim} = 100$ Monte-Carlo simulations.}
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
\textbf{True values} & \textbf{pole} & \textbf{gain} & $\sigma_\alpha^2$ & $\sigma_\beta^2$ & $\alpha$ \\
\hline
\hline
\textbf{Proposed approach} & Mean & 0.7 & 1 & 1 & 0.1 & 0.1 & - \\
& Variance & 0.0686 & 0.0003 & 0.1069 & 0.0154 & 0.0079 & - \\
\hline
\textbf{Classical approach} & Mean & 0.6972 & 1.0030 & 1.0154 & 0.0991 & 0.0979 & -0.0023 \\
& Variance & 0.0387 & 0.0714 & 0.1711 & 0.0512 & 0.0518 & 0.1124 \\
\hline
\end{tabular}
\end{table}

II The nominal model has undermodeling in $G(z^{-1})$, i.e.
\[y_0(t) = G(z^{-1})H(z^{-1})u_0(t)\]

III The nominal model has undermodeling in $L_0(z^{-1})$, i.e.
\[u_0(t) = L_0(z^{-1})H(z^{-1})e(t)\]
Table 2. Estimated values from $N_{\text{sim}} = 100$ Monte-Carlo simulations.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Proposed</th>
<th>Classical</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$\hat{a}$</td>
<td>$\tilde{a}$</td>
</tr>
<tr>
<td></td>
<td>0.7007</td>
<td>0.0024</td>
</tr>
<tr>
<td>II</td>
<td>0.8307</td>
<td>0.0137</td>
</tr>
<tr>
<td>III</td>
<td>0.5890</td>
<td>0.1074</td>
</tr>
<tr>
<td>IV</td>
<td>0.0691</td>
<td>-0.0142</td>
</tr>
<tr>
<td>V</td>
<td>0.7025</td>
<td>0.0296</td>
</tr>
<tr>
<td>VI</td>
<td>0.7168</td>
<td>-0.1025</td>
</tr>
<tr>
<td>VII</td>
<td>0.6796</td>
<td>-0.0287</td>
</tr>
</tbody>
</table>

Standard deviation

<table>
<thead>
<tr>
<th>Scenario</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.0242</td>
</tr>
<tr>
<td>II</td>
<td>0.0429</td>
</tr>
<tr>
<td>III</td>
<td>0.0498</td>
</tr>
<tr>
<td>IV</td>
<td>0.0254</td>
</tr>
<tr>
<td>V</td>
<td>0.0225</td>
</tr>
<tr>
<td>VI</td>
<td>0.0426</td>
</tr>
<tr>
<td>VII</td>
<td>0.0364</td>
</tr>
</tbody>
</table>

IV The nominal model has undermodeling in $L_1(z^{-1})$, i.e. $\tilde{u}(t) = L_1(z^{-1})H(z^{-1})e^{(\omega)}(t)$. 
V The nominal model has undermodeling in $L_2(z^{-1})$, i.e. $\tilde{y}(t) = L_2(z^{-1})H(z^{-1})e^{(\omega)}(t)$.
VI The nominal model has undermodeling in $G(z^{-1})$, and the estimation only considers the range of frequencies $[0, \pi/8]$. 
VII The nominal model has undermodeling in $L_0(z^{-1})$, and the estimation only considers the range of frequencies $[0, \pi/8]$.

All the scenarios are generated with the same driving noise realizations, and with zero initial conditions.

Table 2 shows the minimum values of the negative log-likelihood function for both approaches.

Scenarios I, IV and V exhibit the same estimates. This suggest that, for both approaches, the undermodelling in $L_1$ and $L_2$ has little impact on the identification of $G$. On the other hand, scenarios II and III exhibit estimates that are far from to the true values. This suggest that the undermodelling in $G$ or $L_0$ affects the estimation of $G$. However, scenarios VI and VII exhibit estimates that are closer to the true values than the estimates obtained in scenarios II and III, respectively. This suggest that using a reduced range of frequencies in the estimation may help to obtain better estimates, when undermodelling is present.

5. CONCLUSION

In this paper, we have presented a novel approach for identification of errors-in-variable systems, based on frequency-domain identification. We have derived an ML-based procedure, where the cost function can be obtained in terms of the individual covariance matrices of the measurements, for different lags. Numerical examples have been presented in order to establish a comparison between our proposal and the classical method. The examples show that our proposal exhibits good performance from data in a reduced frequency range.

REFERENCES

J. C. Agüero, System identification methodologies incorporating constraints, PhD. Thesis. The University of Newcastle, Australia, 2005.