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.Aguero,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

Fig. 1. Errors in variable system.

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], Robinson [1974, 1986]) and in Engineering (see e.g. Pintelon and Schoukens [2001], Yuz and Goodwin [2008], Pintelon and Schoukens [2007], Agüero, Yuz, and Goodwin [2007], Agüero and Goodwin [2008], Agüero, Yuz, Goodwin, 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,
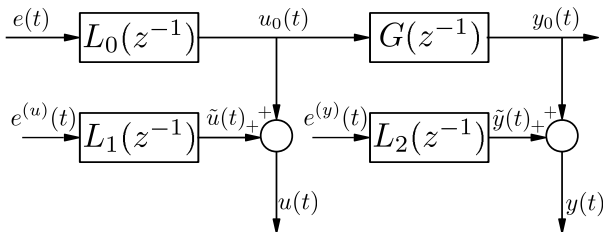
Tang, Yuz, Delgado and Goodwin [2012]. We show that our approach generates a different parametrization of the problem, generating a different likelihood function. This is illustrated by an example.

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]):

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

$$y(t) = G(z^{-1})u_0(t) + L_2(z^{-1})e^{(y)}(t) \tag{2}$$

where, $G(z^{-1})$, $L_0(z^{-1})$, $L_1(z^{-1})$, and $L_2(z^{-1})$ are (causal) transfer functions [1], $u_0(t) = L_0(z^{-1})e(t)$ is the noise-free input signal, $e(t)$, $e^{(u)}(t)$, and $e^{(y)}(t)$ are zero mean Gaussian white noise sequences with covariances σ_0^2, σ_1^2, and σ_2^2, 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

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}, \tag{4}$$

$$B(z^{-1}) = b_1 + \cdots + b_{n_b} z^{-n_b+1}. \tag{5}$$

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

$$L_0(z^{-1}) = \frac{C(z^{-1})}{D(z^{-1})}, \tag{6}$$

where

$$C(z^{-1}) = 1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}, \tag{7}$$

$$D(z^{-1}) = 1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d}, \tag{8}$$

$$L_1(z^{-1}) = \frac{K(z^{-1})}{M(z^{-1})}, \tag{9}$$

where

$$K(z^{-1}) = 1 + k_1 z^{-1} + \cdots + k_{n_k} z^{-n_k}, \tag{10}$$

$$M(z^{-1}) = 1 + m_1 z^{-1} + \cdots + m_{n_m} z^{-n_m}, \tag{11}$$

and

$$L_2(z^{-1}) = \frac{F(z^{-1})}{H(z^{-1})}, \tag{12}$$

where

$$F(z^{-1}) = 1 + f_1 z^{-1} + \cdots + f_{n_f} z^{-n_f}, \tag{13}$$

$$H(z^{-1}) = 1 + h_1 z^{-1} + \cdots + h_{n_h} z^{-n_h}. \tag{14}$$

For the remainder of this paper, we consider the parameter θ 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 $\tilde{u}(t)$ and $\tilde{y}(t)$, 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:

$$\vec{\zeta}(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} L_0(z^{-1})G(z^{-1}) & L_2(z^{-1}) & 0 \\ L_0(z^{-1}) & 0 & L_1(z^{-1}) \end{bmatrix} \begin{bmatrix} e(t) \\ e^{(y)}(t) \\ e^{(u)}(t) \end{bmatrix}, \tag{15}$$

with

$$\begin{bmatrix} e(t) \\ e^{(y)}(t) \\ e^{(u)}(t) \end{bmatrix} \sim \mathcal{N}(\vec{0}, \Sigma_e), \ \Sigma_e = \mathrm{diag}(\sigma_0^2, \sigma_2^2, \sigma_1^2)$$

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

$$Y_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y(t) e^{-j\omega_k t} \quad ; k = 0, \ldots, N-1, \tag{16}$$

where $\omega_k = \frac{2\pi}{N} k$. 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, $\vec{\zeta}$. This is known as a 'symmetric' approach (see e.g. Söderström [2007]), where

$$\vec{\zeta} = [\vec{\zeta}(0)^T, ..., \vec{\zeta}(N-1)^T]^T. \tag{17}$$

For future reference, $\vec{\zeta}$ is taken to be distributed Normal, with mean $\vec{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Given (17), the DFT of $\vec{\zeta}$ is given by:

$$\boldsymbol{\vec{Z}} = [\vec{Z}_0^T, ..., \vec{Z}_{N-1}^T]^T, \tag{18}$$

where

$$\vec{Z}_k = \begin{bmatrix} Y_k \\ U_k \end{bmatrix}, \quad k = 0, 1, ..., N-1. \tag{19}$$

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 $\vec{Y} = \mathcal{F}\vec{y}$, where $\mathcal{F}$ is the DFT matrix. The elements of $\mathcal{F}$ are given by:

$$\mathcal{F}^{(k,l)} = \frac{1}{\sqrt{N}} e^{-j2\pi\frac{(k-1)(l-1)}{N}}, \qquad k,l = 1, 2, ..., N. \quad (20)$$

Given (20), it is possible to define $\vec{\boldsymbol{Z}}$ as a linear transformation of the measurement vector $\vec{\mathfrak{z}}$, as:

$$\vec{\boldsymbol{Z}} = (\mathcal{F} \otimes \mathbf{I}_2)\,\vec{\mathfrak{z}}, \quad (21)$$

where $\mathbf{I}_2$ is the $(2{\times}2)$ identity matrix, and $\boldsymbol{A}{\otimes}\boldsymbol{B}$ represents the Kronecker product between matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

We next review the application of the *classical* frequency domain ML identification method (see e.g. McKelvey 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 ($\mathcal{T}_G$), signal ($\mathcal{T}_L$), and noise ($\mathcal{T}_u$, $\mathcal{T}_y$) transient terms, yield the following model (see e.g. Pintelon and Schoukens [2007]):

$$\vec{Z}_k = \begin{bmatrix} L_0(\mathrm{e}^{j\omega_k})G(\mathrm{e}^{j\omega_k}) & L_2(\mathrm{e}^{j\omega_k}) & 0 \\ L_0(\mathrm{e}^{j\omega_k}) & 0 & L_1(\mathrm{e}^{j\omega_k}) \end{bmatrix} \vec{\mathrm{E}}_k + \vec{\mathcal{Z}}_k, \quad (22)$$

with

$$\vec{\mathcal{Z}}_k = \begin{bmatrix} \mathcal{T}_G(\mathrm{e}^{j\omega_k}) + \mathcal{T}_y(\mathrm{e}^{j\omega_k}) + G(\mathrm{e}^{j\omega_k})\mathcal{T}_L(\mathrm{e}^{j\omega_k}) \\ \mathcal{T}_L(\mathrm{e}^{j\omega_k}) + \mathcal{T}_u(\mathrm{e}^{j\omega_k}) \end{bmatrix},$$

$$\vec{\mathrm{E}}_k = \begin{bmatrix} E_k \\ E_k^{(y)} \\ E_k^{(u)} \end{bmatrix} \sim \mathcal{N}(\vec{0}, \mathrm{diag}(\sigma_0^2, \sigma_2^2, \sigma_1^2)),$$

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 $\vec{Z}_k$ (see e.g. Pintelon and Schoukens [2007]). $\triangledown$

Given the frequency domain data representation (22), the ML estimate is given by (see e.g. Pintelon and Schoukens [2001])

$$\hat{\vartheta}_{\text{Classic}} = \arg\max_\vartheta \log p(\vec{\boldsymbol{Z}}|\vartheta), \quad (23)$$

where ϑ is the parameter vector that includes θ, the noise variances and the transient terms. The corresponding log-likelihood function $\log p(\vec{\boldsymbol{Z}}|\vartheta)$ is given by;

$$\log p(\vec{\boldsymbol{Z}}|\vartheta) = \sum_{k\in\boldsymbol{K}} \log\det C_{\vec{Z}_k} + \sum_{k\in\boldsymbol{K}} (\vec{Z}_k - \vec{\mathcal{Z}}_k)^H C_{\vec{Z}_k}^{-1}(\vec{Z}_k - \vec{\mathcal{Z}}_k),$$
$$(24)$$

and $\boldsymbol{K} \subseteq \{0, 1, ..., N/2\}$. The matrix $C_{\vec{Z}_k}$ corresponds to the covariance matrix of $\vec{Z}_k$, given by

$$C_{\vec{Z}_k} = \boldsymbol{E}[(\vec{Z}_k - \vec{\mathcal{Z}}_k)(\vec{Z}_k - \vec{\mathcal{Z}}_k)^H] = \begin{bmatrix} S_k^{(YY)} & S_k^{(YU)} \\ S_k^{(UY)} & S_k^{(UU)} \end{bmatrix}, \quad (25)$$

where

$$S_k^{(YY)} = |G(\mathrm{e}^{j\omega_k})|^2 |L_0(\mathrm{e}^{j\omega_k})|^2 \sigma_0^2 + |L_2(\mathrm{e}^{j\omega_k})|^2 \sigma_2^2,$$
$$S_k^{(UU)} = |L_0(\mathrm{e}^{j\omega_k})|^2 \sigma_0^2 + |L_1(\mathrm{e}^{j\omega_k})|^2 \sigma_1^2,$$
$$S_k^{(YU)} = G(\mathrm{e}^{j\omega_k})|L_0(\mathrm{e}^{j\omega_k})|^2 \sigma_0^2,$$

and $S_k^{(UY)}$ is the complex conjugate of $S_k^{(YU)}$ (for more details see e.g. Pintelon and Schoukens [2007]).

We next explore frequency domain ML EIV system identification following the ideas presented in Agüero, 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:

$$\vec{\zeta}(t) = \begin{bmatrix} L_0(z^{-1})G(z^{-1}) & L_2(z^{-1}) & 0 \\ L_0(z^{-1}) & 0 & L_1(z^{-1}) \end{bmatrix} \vec{e}(t), \quad (26)$$

$$\vec{\zeta}(t) \sim \mathcal{N}(\vec{\mu}_\beta, \Gamma_\beta),$$
$$\beta = [\theta^T, \sigma_0^2, \sigma_1^2, \sigma_2^2, \vec{x}_0^T]^T, \quad (27)$$

where $\vec{e}(t) = [e(t), e_y(t), e_u(t)]^T$ and $\vec{x}_0$ 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. Agüero, Yuz, Goodwin, and Delgado [2010] for details):

$$\vec{\boldsymbol{Z}}^R = M_R \vec{\mathfrak{z}}, \quad (28)$$

where

$$\vec{\boldsymbol{Z}}^R = \begin{cases} [\vec{Z}_0^T, \sqrt{2}\Re\{\vec{Z}_1^T\}, \sqrt{2}\Im\{\vec{Z}_1^T\}, ..., \vec{Z}_K^T]^T, \\ \text{if } N \text{ is even.} \\ [\vec{Z}_0^T, \sqrt{2}\Re\{\vec{Z}_1^T\}, \sqrt{2}\Im\{\vec{Z}_1^T\}, ..., \sqrt{2}\Re\{\vec{Z}_K^T\}, \sqrt{2}\Im\{\vec{Z}_K^T\}]^T, \\ \text{if } N \text{ is odd.} \end{cases}$$
$$(29)$$

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

For the frequency domain data, $\vec{\boldsymbol{Z}}^R$, the ML estimate in the frequency domain is given by (see e.g. Agüero, Tang, Yuz, Delgado and Goodwin [2012])

$$\hat{\beta}_{\text{ML-Freq}} = \arg\min_\beta p(\vec{\boldsymbol{Z}}^R|\beta). \quad (30)$$

Note that $\vec{\boldsymbol{Z}}^R$ contains all the information in the data since there exists a bijective linear transformation from $\vec{\mathfrak{z}}$ to $\vec{\boldsymbol{Z}}^R$.

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 $\mathcal{S}_F$. This selection corresponds to the non-invertible mapping

$$\vec{\mathfrak{g}} = \mathcal{S}_\mathcal{F}\vec{\boldsymbol{Z}}^R. \quad (31)$$

This selection in the frequency domain can also be understood as a linear transformation of the available time-domain data, $\vec{\mathfrak{z}}$, as $\vec{g} = \mathcal{M}\vec{\mathfrak{z}}$. Then, we can define the (modified) ML estimate as follows.

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

$$\vec{g}|\beta \sim \mathcal{N}(\mathcal{M}\vec{\boldsymbol{\mu}}(\beta), \mathcal{M}\boldsymbol{\Sigma}(\beta)\mathcal{M}^T), \qquad (32)$$

and the ML estimator in the frequency domain is given by

$$\hat{\beta}_{\text{ML-Freq}} = \arg\max_\beta p(\vec{g}|\beta), \qquad (33)$$

Proof: First, we notice that, given $\vec{g} = \mathcal{M}\vec{\mathfrak{z}}$, then $\boldsymbol{E}[\vec{g}] = \mathcal{M}\boldsymbol{E}[\vec{\mathfrak{z}}] = \mathcal{M}\vec{\boldsymbol{\mu}}(\beta)$. Next, in the same way, we calculate $\boldsymbol{E}[(\vec{g} - \mathcal{M}\vec{\boldsymbol{\mu}}(\beta))(\vec{g} - \mathcal{M}\vec{\boldsymbol{\mu}}(\beta))^T] = \mathcal{M}\boldsymbol{E}[(\mathfrak{z} - \vec{\boldsymbol{\mu}}(\beta))(\mathfrak{z} - \vec{\boldsymbol{\mu}}(\beta))^T]\mathcal{M}^T = \mathcal{M}\boldsymbol{\Sigma}(\beta)\mathcal{M}^T$. 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 $p(\vec{g}|\beta)$, it is necessary to determine the covariance matrix $\boldsymbol{\Sigma}(\beta)$. Notice that if the initial condition of the system is modelled as a random vector, $\boldsymbol{\Sigma}(\beta)$ depends on $\vec{x}_0$ (see e.g. Mittnik [1990] and Reinsel [1997]). Moreover, the initial condition is reflected in the mean (vector) $\vec{\boldsymbol{\mu}}(\beta)$, 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]):

$$\vec{\zeta}(t) - \sum_{n=0}^{p-1} \Phi_n \vec{\zeta}(t-n) = \vec{e}(t) - \sum_{n=0}^{q-1} \Theta_n \vec{e}(t-n). \qquad (34)$$

Given $\vec{\mathfrak{z}} = [\vec{\zeta}(0)^T, ..., \vec{\zeta}(N-1)^T]^T$, and defining

$$\vec{\mathrm{E}} = [\vec{e}(0)^T, ..., \vec{e}(N-1)^T]^T,$$

the model in (34) can be represented in terms of the initial condition as (see e.g. Reinsel [1997]):

$$\Phi(\theta)\vec{\mathfrak{z}} = \Theta(\theta)\vec{\mathrm{E}} + F(\theta)\vec{x}_0, \qquad (35)$$

where $\Phi(\theta) = (I_N \otimes I_2) - \sum_{n=1}^{p}(\mathrm{L}^n \otimes \Phi_{n-1})$, $\Theta(\theta) = (I_N \otimes I_2) - \sum_{n=1}^{p}(\mathrm{L}^n \otimes \Theta_{n-1})$, 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 $\Gamma_{\mathrm{E}} = \boldsymbol{E}[\vec{\mathrm{E}}\vec{\mathrm{E}}^T] = I_N \otimes \Sigma_e$, then the covariance matrix of $\vec{\mathfrak{z}}$ is given by (see e.g. Reinsel [1997]):

$$\boldsymbol{\Sigma}(\beta) = \text{Cov}(\vec{\mathfrak{z}}\vec{\mathfrak{z}}) = \Phi^{-1}[\Theta(I_N \otimes \Sigma_e)\Theta^T]. \qquad (36)$$

Remark 2. The covariance matrix $\boldsymbol{\Sigma}(\beta)$ can be determined [2] by obtaining the *individual* covariance matrices $\Gamma_\beta(l) = \text{Cov}(\vec{\zeta}(t), \vec{\zeta}(t+l))$, which depend on the system characteristics. If the initial condition of the system is modelled as a random vector, $\Gamma_\beta(l)$ also depends on $\vec{x}_0$ (see e.g. Mittnik [1990] and Reinsel [1997]). ▽

---

[2] In a more computationally efficient way.

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 $\boldsymbol{\Sigma}(\beta)$. ▽

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 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 $\vec{\zeta}$. 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 algorithm is shown via an example. Finally, we illustrate the effects of undermodelling on both approaches.

Consider the EIV system as in (26) with

$$G(z^{-1}) = \frac{1}{1 - az^{-1}}, \, a = 0.7, \qquad (37)$$

and the noise covariances

$$\sigma_0^2 = 1, \, \sigma_1^2 = 0.1, \, \sigma_2^2 = 0.1. \qquad (38)$$

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

$$y(t) = \frac{1}{1 - az^{-1}}e(t) + e^{(y)}(t). \qquad (39)$$

It is possible to define an initial condition $\vec{x}_0$ by considering two values: an initial condition for y, and an initial condition for $e^{(y)}$. However, this definition of $\vec{x}_0$ 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, $\vec{x}_0$, 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

Table 1. Empirical mean and empirical standard deviation of the parameters from $N_{sim} = 100$ Monte-Carlo simulations..

		pole	gain	σ_0^2	σ_1^2	σ_2^2	α
	True values	0.7	1	1	0.1	0.1	-
Proposed approach	Mean	0.6991	1.0025	0.9972	0.0988	0.0977	-
	Variance	0.0360	0.0608	0.1609	0.0454	0.0979	-
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

matrix (23). In this example, the determinant of the covariance matrix is given by:

$$\det(C_{\bar{Z}_k}) = |G(\mathrm{e}^{j\omega_k})|^2 \sigma_0^2 (\sigma_0^2 + \sigma_1^2) + \sigma_2^2 (\sigma_0^2 + \sigma_1^2) - |G(\mathrm{e}^{j\omega_k})|^2 \sigma_0^4,$$

where

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

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. ▽

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 $\mathcal{T}_u$, $\mathcal{T}_y$, and $\mathcal{T}_L$, are equal to zero. The transient term $\mathcal{T}_G$ is given by $\mathcal{T}_G = \frac{N^{-1/2}\mathrm{e}^{j\omega_k}}{(\mathrm{e}^{j\omega_k} - a)}\alpha$, where α 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 α.

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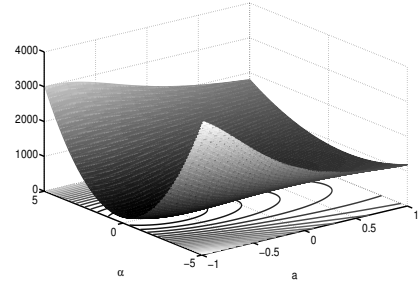
I The nominal model has the same order as the true system.
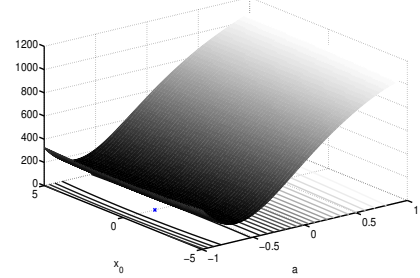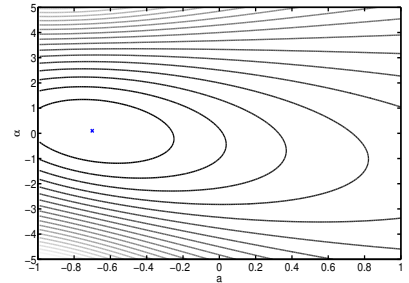

(a) Proposed approach
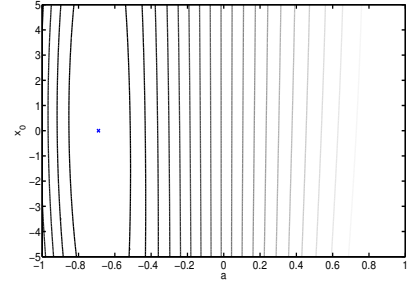
(b) Proposed approach

(c) *Classical* frequency domain approach

(d) *Classical* frequency domain approach

Fig. 2. Negative log-likelihood function for both approaches: dual time-frequency and classical frequency domain.

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 $N_{sim} = 100$
Monte-Carlo simulations.

Mean Values				
	Proposed		Classical	
Scenario	$\hat{a}$	$\hat{x}_0$	$\hat{a}$	$\hat{\alpha}$
I	0.7007	0.0324	0.7018	-0.0001
II	0.8307	0.0137	0.8485	0.0001
III	0.5890	0.1074	0.5789	0.0044
IV	0.6961	-0.0042	0.6958	-0.0119
V	0.7025	0.0296	0.7034	0.0075
VI	0.7168	-0.1025	0.7265	-0.0155
VII	0.6796	-0.0287	0.6754	0.0023
Standard deviation				
I	0.0242	0.5081	0.0247	0.1159
II	0.0429	0.5942	0.0437	0.1058
III	0.0498	1.2679	0.0510	0.1077
IV	0.0254	0.4331	0.0256	0.0999
V	0.0225	0.3980	0.0233	0.0946
VI	0.0426	1.3262	0.0398	0.1249
VII	0.0364	1.6091	0.0358	0.1202

IV The nominal model has undermodeling in $L_1(z^{-1})$, i.e.
$\tilde{u}(t) = L_1(z^{-1})H(z^{-1})e^{(u)}(t)$.

V The nominal model has undermodeling in $L_2(z^{-1})$, i.e.
$\tilde{y}(t) = L_2(z^{-1})H(z^{-1})e^{(y)}(t)$.

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 iden-
tification of errors-in-variable systems, based on frequency-
domain identification. We have derived an ML-based pro-
cedure, 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

R. J. Adcock, Note on the methods of least squares, *The Analyst*,
4(6):183–184, 1877.

R. J. Adcock, A problem in least squares, *The Analyst*, 5(2):53–54,
1878.

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

J. C. Agüero, J. I. Yuz, and G. C. Goodwin, Frequency domain iden-
tification of MIMO state space models using the EM algorithm,
in *Proc. of European Control Conference,* 2007.

J. C. Agüero and G. C. Goodwin, Identifiability of errors in variables
dynamic systems, *Automatica,* 44(2):371–382, 2008.

J. C. Agüero, J.I. Yuz, G. C. Goodwin, and R. A. Delgado, On the
equivalence of time and frequency domain maximum likelihood
estimation *Automatica,* 46(2):260–270, 2010.

J. C. Agüero, W. Tang, J. I. Yuz, R. A., and G. C. Goodwin,
Dual time-frequency domain system identification provisionally
accepted for publication, *Automatica,* 2012.

B. D. O. Anderson and M. Diestler, Identifiability in dynamic errors-
in-variables model, *Journal of Time Series Analysis,* 5(1):1–13,
1984.

G. Bottegal, G. Picci, and S. Pinzoni, On the identifiability of errors-
in-variables models with white measurement errors, *Automatica*,
47(3):545–551, 2011.

B. Cauberghe, P. Guillaume, R. Pintelon, and P. Verboven,
Frequency-domain subspace identification using FRF data from
arbitrary signals, *Journal of Sound and Vibration*, 290(2006):
555–571, 2006.

C. L. Cheng and J. W. Van Ness, *Kendall's library of statistics:
Statistical regression with measurement error*, vol. 6, Edward
Arnold, 1999.

R. F. Engle, Band Spectrum Regression, *International Economic
Review*, 15(1):1–11, 1974.

W. A. Fuller, *Measurement error models*, Wiley, 1987.

G. H. Golub and C. F. Van Loan, An analysis of the total least
squares problem, *SIAM Journal of Numerical Analysis*, 17(6):
883–893, 1980

G. C. Goodwin, J. C. Agüero, J. S. Welsh, J.I. Yuz, G. J. Adams,
and C. R. Rojas, Robust identification of process models from
plant data, *Journal of Process Control*, 18:810–820, 2008.

E. J Hannan and P. M. Robinson, Lagged Regression with Unknown
Lags, *Journal of the Royal Statistical Society. Series B (Method-
ological)*, 35(2):252–267, 1973.

H. Hjalmarsson, J. Mårtensson , C. R. Rojas and T. Söder-
ström, On the accuracy in errors-in-variables identification
compared to prediction-error identification Automatica, 2011.
doi:10.1016/j.automatica.2011.009.002.

T. McKelvey and L. Ljung, Frequency domain maximum likelihood
identification *11th IFAC symposium on system identification,*
Fukuoka, Japan, 1997.

S. Mittnik, Computation of theoretical autocovariance matrices for
multivariate autoregressive moving average time series, *J. R. Stat.
Soc. B* 52(1): 151–155, 1990.

A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time
signal processing,* 2nd ed., Prentice Hall, New Jersey, 1999.

R. Pintelon and J. Schoukens, *System Identification: A frequency
domain approach* IEEE Press, 2001.

R. Pintelon and J. Schoukens, Frequency domain maximum like-
lihood estimation of linear dynamic errors-in-variables models,
Automatica, 43(4):621–630, 2007

G. C. Reinsel, *Elements of multivariate time series analysis,* 2nd
ed., New York, Springer, 1997.

P. M. Robinson, Stochastic Difference Equations with Non-Integral
Differences, *Advances in Applied Probability*, 6(3):524–545, 1974.

P. M. Robinson, On the Errors-in-Variables problem for time series,
Journal of Multivariate Analysis, 19:240–250, 1986.

J. Schoukens, Y. Rolain, and R. Pintelon, On the use of parametric
and non-parametric noise-models in time- and frequency domain
system identification, *Proc. of the 49th IEEE Conference on
Decision and Control, CDC,* 2010.

T. Söderström, Errors-in-variables methods in system identification
Automatica, 43(6):939–958, 2007.

V. Solo, Identifiability of time series models with errors in variables,
Journal of Applied Probability, Essays in Time Series and Allied
Processes, 23:63–71, 1986.

A. Stuart, K. Ord, and S. Arnold, *Kendall's advanced theory of
statistics,* 6th ed., vol. 2A, London, Arnold, 1999.

S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem:
Computational aspects and analysis*, SIAM Frontiers Applied
Math, 1991.

J. I. Yuz and G. C. Goodwin, Robust Identification of Continuous-
Time Systems from Sampled Data, in *Robust Identification of
Continuous-Time Systems from Sampled Data*, H. Garnier and
L. Wang Ed., Springer, 2008.