IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Finite Impulse Response Errors-in-Variables system identification utilizing Approximated Likelihood and Gaussian Mixture Models

ANGEL L. CEDEÑO[1], (Graduate Student Member, IEEE), RAFAEL ORELLANA[2],(Member, IEEE), RODRIGO CARVAJAL[3],(Member, IEEE), BORIS I. GODOY[4],(Member, IEEE), and JUAN C. AGÜERO[1], (Member, IEEE)

[1]Universidad Técnica Federico Santa María (UTFSM), Department of Electronic Engineering, Valparaíso, Chile.
[2]Universidad de Santiago de Chile, Department of Electrical Engineering, Santiago, Chile.
[3]Pontificia Universidad Católica de Valparaíso, School of Electrical Engineering, Valparaíso, Chile.
[4]Lund University, Department of Automatic Control, Lund, Sweden.

Corresponding author: Angel L. Cedeño (e-mail: angel.cedeno@sansano.usm.cl).

This work was supported in part by the Chilean National Agency for Research and Development (ANID) Scholarship Program/Doctorado Nacional/2020-21202410, in part by the PIIC program of DGP at Universidad Técnica Federico Santa María No 035/2021, and in part by grants ANID-Fondecyt 11201187, 3220403 and 1211630, ANID-Basal Project FB0008 (AC3E).

ABSTRACT In this paper a Maximum likelihood estimation algorithm for Finite Impulse Response Errors-in-Variables systems is developed. We consider that the noise-free input signal is Gaussian-mixture distributed. We propose an Expectation-Maximization-based algorithm to estimate the system model parameters, the input and output noise variances, and the Gaussian mixture noise-free input parameters. The benefits of our proposal are illustrated via numerical simulations.

INDEX TERMS Errors-in-Variables, Maximum Likelihood, Expectation-Maximization, Gaussian Mixture distribution, Estimation.

I. INTRODUCTION

Errors-in-Variables (EIV) models [1] refers to the representation of dynamic and static systems in which both input and output signals are corrupted by noise. EIV models play a fundamental role in many scientific areas, including a great number of applications, in which variables can only be measured with errors. Among others, medical variables–such as blood pressure, temperature– and agricultural variables–such as rainfall and soil nitrogen, are measured with errors, see e.g. [2]. On the other hand, economic and productive applications include forecasting in macroeconomics [3], electromagnetic mineral exploration [4], feedback control in 5G-wireless networks (where the quantization and delay of signals can be seen as measurements with errors [5]), modeling and tracking autonomous underwater vehicles, structural health monitoring and thermal models for many-core systems-on-chip, among others (see [1] for further details).

It is well known that classical methodologies are not suitable for the identification of EIV systems since they only consider measurement errors in the system output, thus resulting in non-consistent estimates when the system input is also measured with errors (see e.g. [6] and the references therein). In addition, the system identifiability does not hold in general, unless some additional assumptions are considered such as prior knowledge about the noise-free input, input and output noises [1], [2]. Moreover, if the measurement noises and the noise-free input are Gaussian distributed, it is not possible to uniquely identify static EIV systems [7], [8]. This implies that the joint distribution of the input and output signals can be modeled utilizing different sets of the system parameters and still obtain the same second order statistics. Then, it is not possible to obtain a consistent estimator when the data size goes to infinity because this limit is not uniquely defined (see e.g. [2]). However, it has been established that if the noise-free input is non-Gaussian distributed, then the system is generally identifiable [9]–[11].

A number of techniques have been developed to deal with EIV system identification, such as total least squares, instrumental variables, bias-compensation, covariance matching, high-order statistic (HOS), maximum likelihood (ML), among others (see e.g. [1], [2], [12]–[14]). Some assumptions can be used to achieve the EIV system identifiability. Then, depending upon such assumptions, we can have a particular scheme for EIV identification. For instance, total least squares methods are developed under the assumption that the ratio between the input noise variance and the output noise variance is known in order to obtain a consistent estimator [1]. HOS methods are developed under the assumption that the noise-free input is non-Gaussian distributed. These methods consider the information contained in the cumulants and moments of the third or fourth order, which are assumed to be nonzero [15]. However, HOS methods typically require a large number of samples to yield accurate estimators [15], [16]. In [17] an identification method was developed in which the non-Gaussian distribution is estimated using a non-parametric deconvolution kernel method, assuming that the distributions of both the noise input and output signals are known. In [18] an ML estimator for the identification of EIV systems is developed under the assumption that the noise-free input, the noise input and the noise output signals are modeled as Auto-regressive moving average (ARMA) processes. In this approach, ARMA processes are driven by Gaussian white noises, and the parameters of these three ARMA processes are jointly estimated with the system parameters.

On the other hand, ML is a method widely used for parameters estimation in dynamical systems, due to its properties of consistency and asymptotic efficiency [19]. The ML estimator maximizes the probability of the observed data conditioned to the system parameters. This optimization problem can be difficult to solve in presence of latent variables, thus the Expectation-Maximization algorithm (EM) can be used in this context, see e.g. [20]–[22]. Usually the ML is developed under the assumption of independent and identically distributed (i.i.d.) samples of the underlying random variables. However, when this assumption does not hold, the computation of the likelihood function of the available data can be intractable since the dependency upon signals on their past might lead to complex expressions. In this context, an alternative for obtaining an estimate of the system parameters is to consider an approximation of the likelihood function [23].

Composite likelihood is an approach used to approximate the likelihood function. This approach has been extensively used in many areas and applications (see e.g. [24]–[26] and the references therein). This approximated likelihood was introduced in [27] and [28] with the name pseudo-likelihood function in the context of spatially correlated data, and in [29] it was introduced with the name composite likelihood. In this paper we used a version of the composite likelihood where the main idea is to replace the conditioning on all previous observations by only the k most recent ones. This simplification is made by recognizing that practical systems have finite memory, i.e., distant data do not provide important information about the current observation (see e.g. [30], [31]). The consistency of this approximation was studied in [32], where the kth order likelihood function was defined.

In this work we are interested in the identification of FIR-EIV systems where both the input and output of a FIR model are corrupted by additive noises (where the usual i.i.d. assumption does not hold). In particular, we develop an ML estimator for an FIR-EIV system when the noise-free input probability density function (PDF) is given by a Gaussian Mixture Model (GMM). The motivations to consider the noise-free input PDF as GMM are i) to satisfy the condition of identifiability [9], [10] and ii) because the GMM approximates any PDF with compact support [33], which allows for identifying the FIR-EIV system with any noise-free input distribution. Furthermore, GMMs have been used in many applications such as filtering [34]–[37], static EIV system identification [38], Bayesian estimation [39], [40], linear dynamic systems estimation [41], [42], uncertainty modeling for FIR systems [43], and astronomy [44]. We use the approximated likelihood given in [32] in order to reduce the computational complexity that is produced by correlated data corresponding to the input and output measurements. The main contributions of our work are the following:

(i) We obtain an approximated likelihood function for an FIR-EIV system using GMMs. The composite likelihood is obtained based on the methodology proposed in [32]. Here, we consider that the noise-free input distribution is modeled as a GMM.
(ii) We propose an EM-based algorithm to solve the associated ML estimation problem with GMMs, obtaining the estimates of the FIR-EIV system model parameters, the noise-free input distribution parameters, and the noise input and output signals variances.

The EIV system identification algorithm presented in this paper considers that the PDF of the noise-free input corresponds to a GMM. Nevertheless, it can be used to approximate non-Gaussian distributions that are not Gaussian mixture distributions. The remainder of the paper is as follows: In Section II a motivation example is presented. In Section III the problem of interest is stated. In Section IV the estimation problem for the FIR-EIV systems is addressed using ML methodology with GMMs. In Section V an EM algorithm is presented to solve the corresponding ML estimation problem. Numerical examples are presented in Section VI, showing that the proposed method in this paper yields more accurate estimations than HOS methods. Finally, in Section VIII, we present conclusions.

II. MOTIVATION EXAMPLE

Consider the problem to estimate the constant slope K in the following static system:

$$y_t^0 = K u_t^0, \tag{1}$$

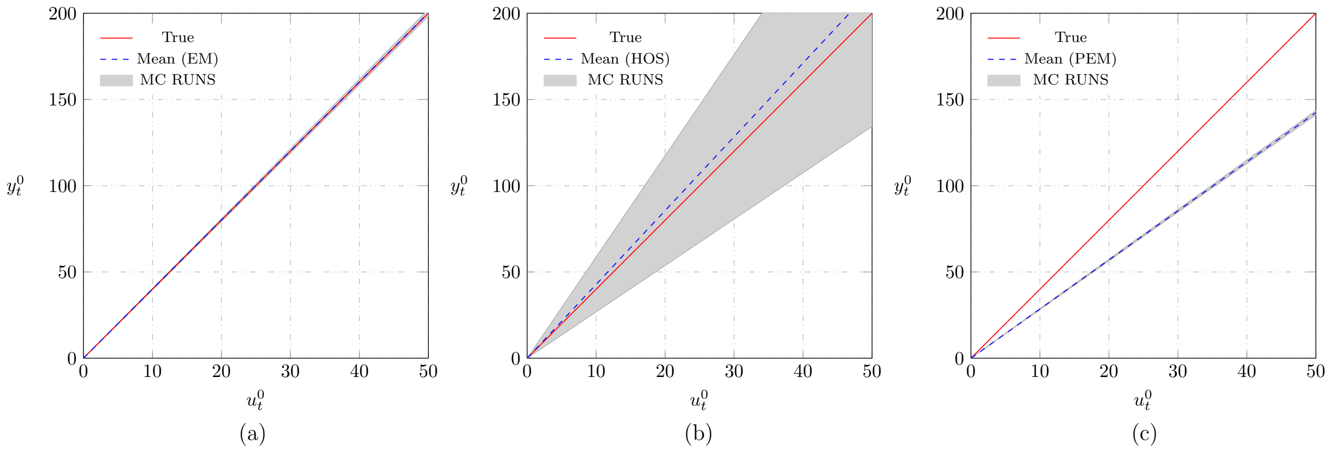

FIGURE 1. Estimation of the constant slope $K = 4$ with 5000 data points and a uniformly distributed noise-free input $\mathcal{U}[-3, 3]$: (a) using EM method: $\hat{K} = 4.004 \pm 0.039$, (b) using HOS method: $\hat{K} = 4.281 \pm 1.591$, and (C) using PEM method: $\hat{K} = 2.847 \pm 0.025$. The noises are both zero-mean Gaussian with unit variance.

where y_t^0 and u_t^0 are not available for the estimation process. Instead, they are measured with additive noises as follows:

$$y_t = y_t^0 + \tilde{y}_t, \tag{2}$$
$$u_t = u_t^0 + \tilde{u}_t, \tag{3}$$

where $\tilde{y}_t$ and $\tilde{u}_t$ are zero-men Gaussian white noises. Then, as it is shown in our previous work [38], an estimation of the parameter K is non-trivial, and some of the available methods in the literature produce inaccurate estimates, such as HOS method and Prediction Error Method (PEM) [19]. In addition, the HOS technique requires a large data set to improve the estimation, and the PEM approach only considers errors in the signal y_t. In Figure 1 we show the estimation results from 100 Monte Carlo simulations using the EM algorithm for static EIV systems proposed in [38], HOS method [45], and the classical PEM. In this example we consider that the true slope is $K = 4$, the noise-free input u_t^0 is a uniformly distributed signal in the interval $[-3, 3]$, and a set of observations containing 5000 data points. We observe that the estimator obtained using the EM method is more accurate than the estimator obtained using the HOS method, which produced estimates that show a large variance (gray-shaded region) and biased mean. Additionally, the PEM algorithm produced a biased estimation with a small variance. As we mentioned before, there are many applications that involve dynamical systems where both input and output variables can only be measured with errors, i.e. dynamic EIV systems. Thus, it is important to develop identification methodologies to obtain an accurate estimator of the system's parameters in the EIV framework. This paper presents a method to identify dynamic EIV systems that can be seen as an extension of our previous work [38] developed for static EIV models.

III. STATEMENT OF THE PROBLEM

A. SYSTEM DESCRIPTION

Consider the FIR-EIV system shown in Fig. 2. The noise-free input and output, denoted by $u_t^0 \in \mathbb{R}$ and $y_t^0 \in \mathbb{R}$ respectively, are related by:

$$y_t^0 = H(q^{-1}, \theta) u_t^0, \tag{4}$$

where $H(q^{-1}, \theta)$ is a FIR system given by:

$$H(q^{-1}, \theta) = h_0 + h_1 q^{-1} + \cdots + h_r q^{-r}, \tag{5}$$

q^{-1} is the backward shift operator ($q^{-1} u_t = u_{t-1}$) and $r \in \mathbb{N}$ is the FIR system order. In addition

$$\begin{aligned} y_t &= y_t^0 + \tilde{y}_t, \\ u_t &= u_t^0 + \tilde{u}_t, \end{aligned} \tag{6}$$

where $y_t \in \mathbb{R}$ is the noisy output signal, $u_t \in \mathbb{R}$ is the noisy input signal, and $\tilde{u}_t \in \mathbb{R}$ and $\tilde{y}_t \in \mathbb{R}$ are zero-mean mutually uncorrelated Gaussian white noise sequences with variance Γ_u and Γ_y, respectively.

The noise-free input, u_t^0, is Gaussian-mixture distributed as follows:

$$p(u_t^0) = \sum_{j=1}^{K} \rho_j \mathcal{N}(u_t^0; \eta_j, \Gamma_j), \tag{7}$$

where $K \geq 2$ ($K \in \mathbb{N}$), $\rho_j > 0$ is the jth mixing weight subject to $\sum_{j=1}^{K} \rho_j = 1$, $\mathcal{N}(u_t^0; \eta_j, \Gamma_j)$ represents a Gaussian PDF with mean value $\eta_j \in \mathbb{R}$ and variance $\Gamma_j \in \mathbb{R}^+$, given by:

$$\mathcal{N}(u_t^0; \eta_j, \Gamma_j) = \frac{1}{\sqrt{2\pi\Gamma_j}} \exp\left\{ -\frac{(u_t^0 - \eta_j)^2}{2\Gamma_j} \right\}. \tag{8}$$

The system (6) can be rewritten in matrix form as follows:

$$z_t = A_r x_t + v_t, \tag{9}$$

where v_t is a zero-mean white Gaussian noise with covariance matrix $\Phi = \text{diag}\{\Gamma_y, \Gamma_u\}$ and

$$z_t = [y_t, u_t]^\top, \qquad A_r = \begin{bmatrix} h_0 & h_1 & \dots & h_r \\ 1 & 0 & \dots & 0 \end{bmatrix}, \tag{10}$$

$$x_t = \left[u_t^0, u_{t-1}^0, \cdots, u_{t-r}^0\right]^\top, \qquad v_t = [\tilde{y}_t, \tilde{u}_t]^\top. \tag{11}$$

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3255827

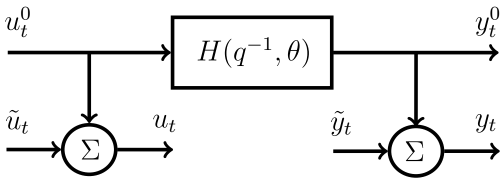

FIGURE 2. Basic setup of errors-in-variables problem.

The difficulty in the identification of the parameters of the system (9) is due to the fact that $x_1, x_2, \ldots, x_N$ is a latent correlated sequence, which results in $z_1, z_2, \ldots, z_N$ not being an i.i.d. sequence of z_t. This, in turn yields a likelihood function of the data that can not be written as the product of independent PDFs, difficulting the optimization procedure. The problem of interest can be stated as follows: From a set of noise-corrupted input and output signals[1], $z_{1:N} = \{y_{1:N}, u_{1:N}\}$, and the system model in (9), we develop an ML estimator for the system parameters defined as follows:

$$\hat{\beta}_{\text{ML}} = \arg\max_{\beta} \ \ell_N(z_{1:N}|\beta), \qquad \text{s.t. } \rho_j > 0, \ \sum_{i=1}^{M} \rho_j = 1, \tag{12}$$

where $\ell_N(z_{1:N}|\beta) = \log\{p(z_{1:N}|\beta)\}$ is the log-likelihood function, and β is defined as

$$\beta = \left[\theta^\top, \gamma^\top, \Gamma_y, \Gamma_u\right]^\top, \tag{13}$$

$$\theta = [h_0, h_1, \ldots, h_r]^\top, \tag{14}$$

$$\gamma = [\rho_1, \eta_1, \Gamma_1, \ldots, \rho_M, \eta_M, \Gamma_M]^\top. \tag{15}$$

B. STANDING ASSUMPTIONS

We consider that there exists a vector of parameters $\beta = \beta_0$ that defines the "true" system. In order to formulate the ML estimation algorithm, we introduce the following standing assumptions [18], [42], [46]:

Assumption 1. *The vector of parameters β_0, the input signal u_t, the noise-free input u_t^0, the noise $\tilde{u}_t$ and $\tilde{y}_t$ in (6) satisfy regularity conditions, guaranteeing that the ML estimate β_{ML} converges (in probability or a.s.) to the true solution β_0 as $N \to \infty$.*

Assumption 2. *The noise-free input u_t^0 is a stationary, independent, and identically distributed stochastic signal with a non-Gaussian distribution.*

Assumption 3. *u_t^0, $\tilde{u}_t$, and $\tilde{y}_t$ are jointly independent.*

Assumption 4. *The filter order r of the system* (4)*, and the number of Gaussian components K of the noise-free input distribution are known.*

[1]We use $x_{1:N}$ to denote the signal x_t for $t = 1, \ldots, N$.

Assumption 5. *System* (4) *is a Single Input Single Output (SISO), linear, and asymptotically stable system.*

Assumption 6. *The system* (4) *is operating in open loop, and only u_t and y_t with $t = 1, \ldots, N$ are available to be measured.*

Assumption 1 is necessary to develop an estimation algorithm that holds the asymptotic properties of the ML estimator. Assumptions 2 and 3 are necessary to compute the conditional PDF of the random variable z_t given its past, in order to obtain the likelihood function. Assumption 4 can be relaxed, for example, using an information criterion to determine the correct order r or by using a non-Gaussian noise-free input distribution that does not correspond to a GMM but can be approximated by one. We will address this case in Section VI. Assumption 5 is necessary to obtain an asymptotically unbiased ML estimator [47] and a system model that is controllable and observable [48].

The EIV framework utilized in this paper is inspired in Figure 3, as was stated in [1, pag 54]. The signal w_t can be considered as an accessible control variable and $\mathbf{F}$ describes an actuator with potentially complicated dynamics. This implies that the manipulated signal w_t is not a signal that directly excites the system. The signal entering the system is u_t^0 which is not accessible to be manipulated but only measured with errors. This situation is illustrated in the practical experiment setup addressed in section VII, where the control signal obtained in the workstation, does not directly excite the DC motor. This control signal first passes through an amplifier whose output is the signal that really excites the DC motor. Then, this signal is measured with a sensor that introduces errors in the measurements. Thus, we have an EIV problem.

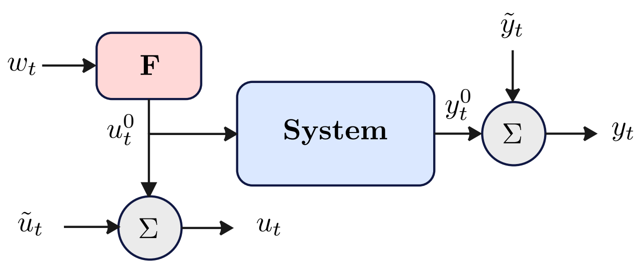

FIGURE 3. Basic setup of EIV problem, including input generation

IV. MAXIMUM LIKELIHOOD ESTIMATION FOR FIR-EIV SYSTEMS USING GMMS

In this section we develop an ML estimation algorithm for system (9) based on an approximation of the full likelihood function. We consider the composite likelihood approach in [49] to approximate the full likelihood.

In order to obtain the full and approximated log-likelihood functions for system (9), we define the following vectors:

$$\phi_{l:t} = \begin{bmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_l \end{bmatrix}, \quad \xi_{s:t} = \begin{bmatrix} u_t^0 \\ u_{t-1}^0 \\ \vdots \\ u_s^0 \end{bmatrix}, \quad \vartheta_{l:t} = \begin{bmatrix} v_t \\ v_{t-1} \\ \vdots \\ v_l \end{bmatrix}, \quad (16)$$

where the notation $\mathbb{X}_{a:b}$ describes the data of $\mathbb{X}$ in a discrete-time window from a to b. For instance, $\phi_{1:t}$, i.e. $l = 1$, is the vector of the observations $z_{1:t}$ that corresponds to the full likelihood function up to time t, and $\phi_{t-k:t}$, i.e. $l = t - k$, is the vector of the observations $z_{t-k:t}$ that corresponds to the likelihood function in the window of length $k + 1$ (approximated likelihood function). For a general description, using the observations $z_{l:t}$, the system (9) can be expressed as follows:

$$\phi_{l:t} = \mathcal{A}_n \xi_{s:t} + \vartheta_{l:t}, \quad (17)$$

where $n = t - l + 1$, $s = l - r$, $\phi_{l:t} \in \mathbb{R}^{2n}$, $\xi_{s:t} \in \mathbb{R}^{n+r}$, $\vartheta_{l:t} \in \mathbb{R}^{2n}$ and $\mathcal{A}_n \in \mathbb{R}^{2n \times n+r}$ is given by:

$$\mathcal{A}_n = \begin{bmatrix} h_0 & h_1 & \dots & h_r & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & h_0 & h_1 & \dots & h_r & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & \dots & h_r \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \quad (18)$$

where the index n in (18) denotes the number of times that the matrix A_r is repeated (see the gray shadow in $\mathcal{A}_n$).

A. COMPUTING THE FULL LIKELIHOOD FUNCTION

Utilizing Bayes's theorem we obtain the full likelihood function for the available data $z_{1:N}$ as follows:

$$\mathcal{L}_N(\beta) = \prod_{t=r+1}^{N} p(z_t | z_{1:t-1}, \beta) = \prod_{t=r+1}^{N} \frac{p(\phi_{1:t}|\beta)}{p(\phi_{1:t-1}|\beta)}, \quad (19)$$

where $p(\phi_{1:t}|\beta) = p(z_t, z_{t-1}, \dots, z_1|\beta)$ is the joint PDF of the observations $z_{1:t}$, and $p(\phi_{1:t-1}|\beta) = p(z_{t-1}, \dots, z_1|\beta)$ is the marginal PDF that corresponds to the observations $z_{1:t-1}$.

We will now present two Lemmas that will be used to find the joint and marginal PDFs $p(\phi_{1:t}|\beta)$ and $p(\phi_{1:t-1}|\beta)$. In the first Lemma, we present the PDF $p(\xi_m)$ of the random vector $\xi_m = [u_1^0, u_2^0, \dots, u_m^0]^\top$. Since $p(u_\kappa^0)$, with $\kappa = 1, 2, \dots, m$, is the GMM given in (7), the PDF $p(\xi_m)$ is also a GMM. The ith Gaussian component in $p(\xi_m)$ is obtained from each possible combination of the individual weights, means, and variances of the random variables u_κ^0. The indices of each combination can be thought as the ith experiment in a full factorial design with m factors and K levels by each factor (see e.g. [50], [51] for more details regarding a full factorial design). The weight, mean, and covariance of each Gaussian components in $p(\xi_m)$ are given in Figure 4. In order to compute each combination in the full factorial design we use the Matlab function `fullfact`.

In the second Lemma, we establish the PDF of the linear transformation $\phi_\tau = \mathcal{A}\xi_m + \vartheta_\tau$ where $\mathcal{A}$ is a matrix, ξ_m is a Gaussian-Mixture distributed random vector, and ϑ_τ is a Gaussian distributed random vector.

Lemma 1. *Consider the i.i.d. sequence of random variables $u_1^0, u_2^0, \dots, u_m^0$, where $u_\kappa^0 \in \mathbb{R}$ with $\kappa = 1, 2, \dots, m$, satisfying*

$$p(u_\kappa^0) = \sum_{j=1}^{K} \rho_j \mathcal{N}(u_\kappa^0; \eta_j, \Gamma_j), \quad (20)$$

where $\rho_j, \eta_j, \Gamma_j \in \mathbb{R}$ are the weight, mean, and variance of each Gaussian component, respectively. $K \in \mathbb{N}$ is the number of components of the mixture. Then, the PDF of the vector $\xi_m = [u_1^0, u_2^0, \dots, u_m^0]^\top$ is given by:

$$p(\xi_m) = \sum_{i=1}^{M} \alpha_i \mathcal{N}(\xi_m; \mu_i, \Sigma_i), \quad (21)$$

where $M = K^m$. The weight $\alpha_i \in \mathbb{R}$, the mean $\mu_i \in \mathbb{R}^m$ and the covariance matrix $\Sigma_i \in \mathbb{R}^{m\times m}$ are given in Figure 4.

Proof. See appendix B. □

Lemma 2. *Let $\xi_m \in \mathbb{R}^m$ be a random vector that satisfies*

$$p(\xi_m) = \sum_{i=1}^{M} \alpha_i \mathcal{N}(\xi_m; \mu_i, \Sigma_i), \quad (22)$$

where $\alpha_i \in \mathbb{R}$ is the weight of each component, $\mu_i \in \mathbb{R}^m$ is the vector of means, $\Sigma_i \in \mathbb{R}^{m\times m}$ is a positive definite covariance matrix, and $M \in \mathbb{N}$ is the number of components of the mixture. Consider $\phi_\tau = \mathcal{A}\xi_m + \vartheta_\tau$ where $\phi_\tau \in \mathbb{R}^\tau$, $\mathcal{A} \in \mathbb{R}^{\tau\times m}$, and $\vartheta_\tau \in \mathbb{R}^\tau$. ϑ_τ is independent of ξ_m and satisfies

$$p(\vartheta_\tau) = \mathcal{N}(\vartheta_\tau; \varepsilon, \Phi), \quad (23)$$

where $\varepsilon \in \mathbb{R}^\tau$ is the vector of means, $\Phi \in \mathbb{R}^{\tau\times\tau}$ is a positive definite covariance matrix. Then,

$$p(\phi_\tau) = \sum_{i=1}^{M} \alpha_i \mathcal{N}(\phi_\tau; \mathcal{A}\mu_i + \varepsilon, \mathcal{A}\Sigma_i\mathcal{A}^\top + \Phi). \quad (24)$$

Proof. See appendix C. □

Based on Lemmas 1 and 2, we first define the PDF of the observed data set $z_{l:t} = \{z_l, z_2, \dots, z_t\}$. Then, setting $l = 1$, the required PDFs to compute the full likelihood function are directly obtained. Consider the system in (17), then the PDF of the random variable $\phi_{l:t}$ is the parameterized PDF

$$f(\phi_{l:t}; \mathcal{A}_n) = \sum_{i=1}^{M} \alpha_i \mathcal{N}\left(\phi_{l:t}; \mathcal{A}_n\mu_i, \mathcal{A}_n\Sigma_i\mathcal{A}_n^\top + \Lambda_n\right), \quad (25)$$

where $M = K^{n+r}$. The weight $\alpha_i \in \mathbb{R}$, the mean $\mu_i \in \mathbb{R}^{n+r}$, and the covariance matrix $\Sigma_i \in \mathbb{R}^{n+r\times n+r}$ are

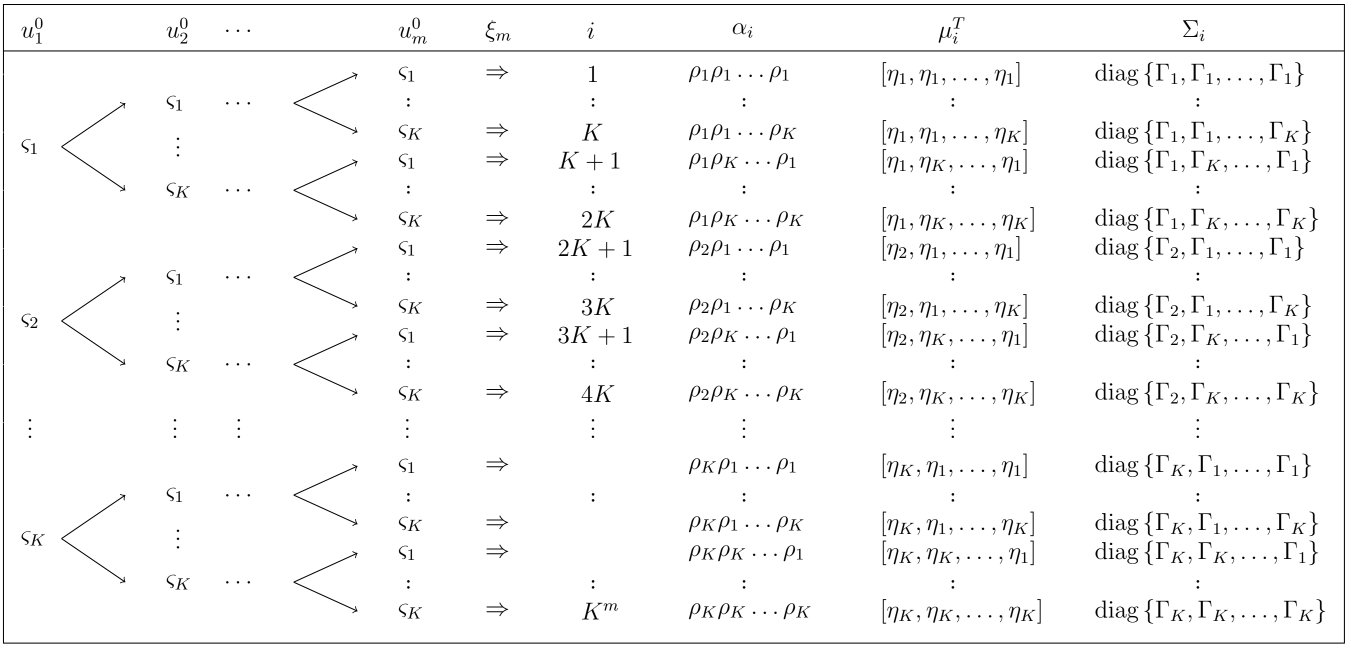

FIGURE 4. Full factorial design with m factors and K levels by each factor to obtain the the weights, means, and variances of the joint PDF $p(\xi_m)$ defined in Lemma 1. The variable ς represents α, μ, and Σ.

obtained following the indices in the ith experiment in the full factorial design with $n+r$ factors and K levels for each factor (see Figure 4). $\Lambda_n \in \mathbb{R}^{2n\times 2n}$ is a (block) diagonal matrix given by

$$\Lambda_n = \begin{bmatrix} \Phi & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Phi \end{bmatrix}. \tag{26}$$

In the following example we illustrate how to find the PDF $p(\phi_{l:t})$.

Example 1. *Suppose we have the system in* (17) *and we want to compute* $p(\phi_{7:8})$ *for an FIR-EIV system with* $r=1$*, then*

$$\begin{bmatrix} z_8 \\ \\ z_7 \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & 0 \\ 1 & 0 & 0 \\ 0 & h_0 & h_1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u_8^0 \\ u_7^0 \\ u_6^0 \end{bmatrix} + \begin{bmatrix} v_8 \\ \\ v_7 \end{bmatrix}. \tag{27}$$

Consider that each $u_\kappa^0 \in \mathbb{R}$ *with* $\kappa = 6,7,8$ *satisfies that*

$$p(u_\kappa^0) = \rho_1 \mathcal{N}(u_\kappa^0; \eta_1, \Gamma_1) + \rho_2 \mathcal{N}(u_\kappa^0; \eta_2, \Gamma_2). \tag{28}$$

$v_7 \in \mathbb{R}^2$ *and* $v_8 \in \mathbb{R}^2$ *are zero-mean Gaussian random variables with covariance matrix* $\Phi \in \mathbb{R}^{2\times 2}$*. Then, the PDF of the random variable* $\phi_{7:8}$ *is the parameterized PDF*

$$p(\phi_{7:8}) = f(\phi_{7:8}; \mathcal{A}_2), \tag{29}$$

where the weight $\alpha_i \in \mathbb{R}$*, the mean* $\mu_i \in \mathbb{R}^4$ *and the covariance matrix* $\Sigma_i \in \mathbb{R}^{4\times 4}$ *are given in Table 1.* $\Lambda \in \mathbb{R}^{4\times 4}$ *is the block diagonal matrix* $\Lambda_2 = diag\{\Phi, \Phi\}$*.* △

TABLE 1. Weights, means, and Covariance matrices for example 1.

i	Weight	Mean	Covariance
1	$\alpha_1 = \rho_1\rho_1\rho_1$	$\mu_1 = [\eta_1\ \ \eta_1\ \ \eta_1]^\top$	$\Sigma_1 = \mathrm{diag}\{\Gamma_1, \Gamma_1, \Gamma_1\}$
2	$\alpha_2 = \rho_1\rho_1\rho_2$	$\mu_2 = [\eta_1\ \ \eta_1\ \ \eta_2]^\top$	$\Sigma_2 = \mathrm{diag}\{\Gamma_1, \Gamma_1, \Gamma_2\}$
3	$\alpha_3 = \rho_1\rho_2\rho_1$	$\mu_3 = [\eta_1\ \ \eta_2\ \ \eta_1]^\top$	$\Sigma_3 = \mathrm{diag}\{\Gamma_1, \Gamma_2, \Gamma_1\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
8	$\alpha_8 = \rho_2\rho_2\rho_2$	$\mu_8 = [\eta_2\ \ \eta_2\ \ \eta_2]^\top$	$\Sigma_8 = \mathrm{diag}\{\Gamma_2, \Gamma_2, \Gamma_2\}$

Using the system model (17) with $l=1$ and equation (25), the PDFs to compute the full likelihood function, $p(\phi_{1:t}|\beta)$ and $p(\phi_{1:t-1}|\beta)$, are given by:

$$\begin{aligned} p(\phi_{1:t}|\beta) &= f(\phi_{1:t}; \mathcal{A}_t), \\ p(\phi_{1:t-1}|\beta) &= f(\phi_{1:t-1}; \mathcal{A}_{t-1}), \end{aligned} \tag{30}$$

where $f(x;W)$ represents a PDF corresponding to a Gaussian Mixture distribution in the random vector x parameterized by the matrices W as it is shown in equation (25).

B. APPROXIMATED LIKELIHOOD FUNCTION

From equation (19) we have that the full-log-likelihood function is given by

$$\mathcal{L}_N(\beta) = \sum_{t=r+1}^{N} \log\left\{\frac{p(\phi_{1:t}|\beta)}{p(\phi_{1:t-1}|\beta)}\right\}, \tag{31}$$

where both $p(\phi_{1:t}|\beta)$ and $p(\phi_{1:t-1}|\beta)$ in (30) are GMM. It is clear that in every instant t the number of Gaussian components of the full likelihood function increases and the dimension of each component increases as well with respect to the PDFs in the previous instant $t-1$. This results in the maximization of the full likelihood function to be computationally demanding when the data length N increases. In fact, it requires more than $\sum_{t=r+1}^{N} K^{t+r}$ operations for computing the full likelihood function. This large quantity of operations makes it complex to estimate $\hat{\beta}_{\text{ML}}$. In this situation, an approximated likelihood function can be considered in order to keep the computational complexity manageable.

Here, the composite likelihood approach to approximate the full likelihood function is considered. This approximation is made by omitting the distant data that is out of a window of length k. This kth-order log-likelihood function was defined in [32] as follows:

$$\ell_N^k(\beta) = \sum_{t=r+1}^{N} \log\left\{p(z_t|z_{t-1:t-k},\beta)\right\}, \qquad (32)$$

where $k \in \{1,\dots,N-1\}$. Thus the ML problem with the approximated log-likelihood function is as follows:

Theorem 1. *Consider the parameters to be estimated β given in* (13). *Then, under the standing assumptions, the ML estimator is given by:*

$$\hat{\beta}_{ML} = \arg\max_{\beta}\ \ell_N^k(\beta),\ \text{s.t. } \alpha_i > 0,\ \sum_{i=1}^{M} \alpha_i = 1, \qquad (33)$$

where the approximated log-likelihood function for the available data $z_{1:N}$ is given by

$$\ell_N^k(\beta) = \sum_{t=r+1}^{N} \log\left\{\frac{p(\phi_{t-k:t}|\beta)}{p(\phi_{t-k:t-1}|\beta)}\right\}, \qquad (34)$$

where $p(\phi_{t-k:t}|\beta)$ and $p(\phi_{t-k:t-1}|\beta)$ are GMM distributions obtained from (25) *as follows*

$$\begin{aligned} p(\phi_{t-k:t}|\beta) &= f(\phi_{t-k:t};\mathcal{A}_{k+1}), \\ p(\phi_{t-k:t-1}|\beta) &= f(\phi_{t-k:t-1};\mathcal{A}_k). \end{aligned} \qquad (35)$$

Proof. Directly applying (25) (that was derived using Lemmas 1 and 2) to the systems $\phi_{t-k:t}$ and $\phi_{t-k:t-1}$ given in (17). □

Remark 1. *Notice that when $H(q^{-1},\theta) = h_0$, and $k = 0$, we have the particular case of static EIV system treated in our previous work [38] where $p(\phi_{t:t-1}|\beta) = 1$ and*

$$p(\phi_{t:t}|\beta) = f(\phi_{t:t};\mathcal{A}_1). \qquad (36)$$

Clearly, in the case of static EIV systems we use the full log-likelihood function since $\{z_{1:N}\}$ is an i.i.d. sequence. △

We next focus on understanding how to choose k sufficiently large in order to obtain a good approximation of the likelihood function and keep the computational complexity low. Firstly, we observe that the covariance matrices in the GMMs $p(\phi_{t-k:t}|\beta)$ and $p(\phi_{t-k:t-1}|\beta)$ in (35) are diagonal-banded matrices, which implies that k should at least satisfy $k \geq r$. Secondly, we analyze the behavior of the log-likelihood function when the distant data increases i.e. increasing the number of past data points in the approximated log-likelihood function. For this purpose we consider several numerical simulations for FIR-EIV systems with orders $r \in \{1,2,3,4,5\}$. We test these systems, computing the approximated log-likelihood function, varying the parameter $k \in \{r, r+1, \dots\}$. In all cases, we obtain that the approximated log-likelihood function does not change significantly when k increases, thus we can consider that $r \leq k \leq 3r$. This results is similar to what was reported in [30], where, in the context of spacial correlated data, the authors concluded that $1 \leq k \leq 10$. In Figure 5 we show the approximated log-likelihood function for the system:

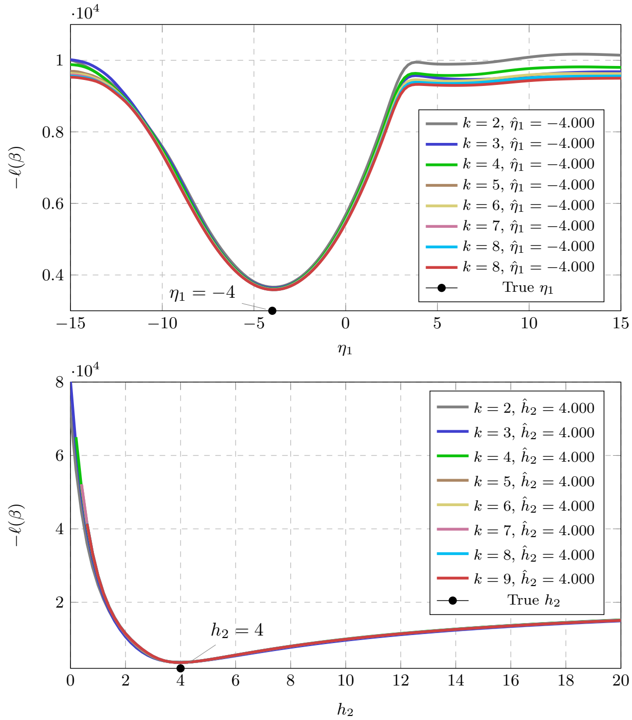

FIGURE 5. Log-likelihood function for the system in (37). All parameters are fixed in their true values except: η_1 (top) and h_2 (bottom).

$$H(q^{-1},\theta) = -1 + q^{-1} + 4q^{-2}, \qquad (37)$$

with $\Gamma_u = 1$ and $\Gamma_y = 2$. The *true* parameters of the PDF $p(u_t^0)$ are: $\rho_1 = 0.25$, $\rho_2 = 0.75$, $\eta_1 = -4$, $\eta_2 = 6$, $\Gamma_1 = 1$, and $\Gamma_2 = 2$. In this example all parameters are fixed in their true values except one of them. We observe that the optimal value of the parameters remains invariant when increasing of k.

V. AN ITERATIVE ALGORITHM FOR IDENTIFYING FIR-EIV SYSTEMS USING GMMS

In order to obtain the estimation $\hat{\beta}_{\text{ML}}$ of the parameter vector β, the approximated log-likelihood function $\ell_{\text{N}}^k(\beta)$ to be optimized is given by

$$\ell_{\text{N}}^k(\beta) = \sum_{t=r+1}^{N} \log\left\{\sum_{i=1}^{M} \frac{\alpha_i \mathcal{N}\left(\phi_{t-k:t};\psi_{k+1}^i,\Delta_{k+1}^i\right)}{p(\phi_{t-k:t-1}|\beta)}\right\}, \qquad (38)$$

with weights α_i, mean values $\psi_{k+1}^i = \mathcal{A}_{k+1}\mu_i$, and covariance matrices $\Delta_{k+1}^i = \mathcal{A}_{k+1}\Sigma_i\mathcal{A}_{k+1}^\top + \Lambda_{k+1}$. The PDF, $p(\phi_{t-k:t-1}|\beta)$, is given by:

$$p(\phi_{t-k:t-1}|\beta) = \sum_{\varrho=1}^{M} \alpha_\varrho \mathcal{N}\left(\phi_{t-k:t-1}; \psi_k^\varrho, \Delta_k^\varrho\right), \tag{39}$$

with weights α_ϱ, mean values $\psi_k^\varrho = \mathcal{A}_k\mu_\varrho$, and covariance matrices $\Delta_k^\varrho = \mathcal{A}_k\Sigma_\varrho\mathcal{A}_k^\top + \Lambda_k$. The maximization of the log-likelihood function in (38) can be intractable since, in general, it is a non-convex function of the data. Additionally, the log-sum structure in the log-likelihood function makes it difficult to obtain closed-form expressions for the estimators. In this case, the EM algorithm is typically used for the attainment of the ML estimator (see e.g. [6], [20], [22]).

The EM algorithm is a popular technique for identifying linear and non-linear dynamic systems in the time domain (see e.g. [52], [53]) and frequency domain [21]. The main idea behind the EM algorithm is that at each iteration, a simpler-to-optimize auxiliary function is handled, instead of the original log-likelihood function. The solution of the EM algorithm is obtained by iteratively alternating between two steps: (i) an Expectation step (E-step), where the auxiliary function is obtained, and (ii) a Maximization step (M-step), where the parameter estimates are found. A common strategy to develop an EM algorithm [20] with GMMs is to define the likelihood function with a data augmentation approach. That is, the likelihood function is defined from the observed data and with a *hidden* random variable that indicates from which GMM component an observation comes from [54]. On the other hand, in [22] a systematic procedure to develop an EM-based algorithm without explicitly defining a *hidden* variable was proposed, obtaining an iterative algorithm that holds the same properties of a traditional EM algorithms with GMMs. In this section, and inspired by [22], we will show how an EM-based estimation algorithm can be developed to solve the FIR-EIV system identification problem stated in (33).

A. EM-BASED ALGORITHM FORMULATION

Let us define the following:

$$\mathcal{K}_i(\phi_{t-k:t}, \beta) = \frac{\alpha_i \mathcal{N}\left(\phi_{t-k:t}; \psi_{k+1}^i, \Delta_{k+1}^i\right)}{p(\phi_{t-k:t-1})}, \tag{40}$$

$$\mathcal{V}_t(\beta) = \sum_{i=1}^{M} \mathcal{K}_i(\phi_{t-k:t}, \beta). \tag{41}$$

Then, the log-likelihood function in (34) can be expressed as

$$\ell_N^k(\beta) = \sum_{t=1}^{N} \log\left[\mathcal{V}_t(\beta)\right]. \tag{42}$$

The logarithm in (42) can be written as $\mathcal{B}_t(\beta) = \log\left[\mathcal{V}_t(\beta)\right]$, where:

$$\mathcal{B}_t(\beta) = \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\beta, \hat{\beta}^{(m)}), \tag{43}$$

with

$$\mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{M} \log\left\{\mathcal{K}_i(\phi_{t-k:t}, \beta)\right\} \frac{\mathcal{K}_i(\phi_{t-k:t}, \hat{\beta}^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}, \tag{44}$$

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{M} \log\left\{\frac{\mathcal{K}_i(\phi_{t-k:t}, \beta)}{\mathcal{V}_t(\beta)}\right\} \frac{\mathcal{K}_i(\phi_{t-k:t}, \hat{\beta}^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}. \tag{45}$$

Lemma 3. *The function $\mathcal{H}_t(\beta, \hat{\beta}^{(m)})$ is a decreasing function for any value of β and satisfies the following:*

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\hat{\beta}^{(m)}, \hat{\beta}^{(m)}) \leq 0. \tag{46}$$

Proof. The result is directly obtained from Jensen's inequality (see e.g. [22] and the references therein). □

From Lemma 3 we can formulate the following iterative algorithm:

$$\mathcal{Q}_N(\beta, \hat{\beta}^{(m)}) = \sum_{t=r+1}^{N} \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}), \tag{47}$$

$$\begin{aligned} \hat{\beta}^{(m+1)} &= \arg\max_{\beta}\ \mathcal{Q}_N(\beta, \hat{\beta}^{(m)}), \\ \text{s.t. } \alpha_i &> 0,\ \sum_{i=1}^{M} \alpha_i = 1. \end{aligned} \tag{48}$$

Notice that (47) and (48) correspond to the E-step and M-step of the EM algorithm, respectively [20].

B. OPTIMIZATION OF THE AUXILIARY FUNCTION

The optimization problem in (48) can be carried out by using Nonlinear Programming for constrained problems (see e.g. [55] and the references therein). There is a variety of methods and solvers for this type of optimization problem that involves a nonlinear cost function and linear constraints. The most common used techniques are the active set-, interior point-, augmented Lagrangian-, and Newton-based-methods, to mention a few (see for example [56]). Additionally, there are solvers that implement methods like branch-and-bound such as BARON [57], or primal-dual interior point such as the Matlab optimization toolbox [58] or SciPy package in Python [59]. In this paper we used the Matlab function *fmincon* with the interior point method. To solve the optimization problem, the following constraints are defined: $\Gamma_u, \Gamma_y, \Gamma_j > 0$, $\rho_j > 0$, and $\sum_{i=1}^{M} \rho_j = 1$. We summarize the proposed algorithm as follows:

1) Fix the number components, M, for the GMM of the noise-free input $p(u_t^0)$.
2) Choose an initial guess $\hat{\beta}^{(0)}$ and set $m = 1$.
3) Compute the estimates of $\hat{\beta}^{(m+1)}$ by solving (48).
4) Set $m = m + 1$ and go back to step 3 until a stopping criterion is satisfied.

Initialization methodologies for iterative algorithms with GMMs have been studied in order to obtain accurate estimates and to improve the rate of convergence [60], [61].

From system (4), we consider an initialization procedure as follows:

(1) The initial value of the FIR system parameters are obtained using HOS method.
(2) The initial guess of the mean values of the GMM parameters are evenly spaced between the minimum and the maximum value of the input signal u_t.
(3) The initial values of the variances $\{\Gamma_j\}_{j=1}^{K}$ are set equal to the sample variance of u_t.
(4) The mixing weight for each GMM component is given by $\rho_j = 1/K$.

VI. NUMERICAL EXAMPLES

In this section, we present three numerical examples, similar to [14], [62], to analyze the performance of our proposal. This framework of numerical simulations is typically used when the performance of new estimation algorithms are tested with problems for which an experimental setup cannot be planned or to reduce the costs of experiments [63], [64]. We compare the estimation accuracy of the results obtained using our proposal with the HOS method [15]. We solve the EIV problem in Figure 2 considering that the measurements $\{z_{1:N}\}$ are generated from the system model in (4)-(6) with $\Gamma_u = 1$ and $\Gamma_y = 1$.

In the first example, the system model in Figure 2 holds all the standing assumptions stated in *Section III-B*. In the second and third examples, our proposed EM-based algorithm is used to estimate the FIR-EIV system model relaxing Assumption 4. That is, the noise-free input distribution is not a GMM but can be approximated by one. This approximation framework is based on the Wiener approximation theorem that claims that any PDF with compact support can be, in general, approximated by a finite summation of Gaussian distributions (see Appendix A, Theorem 2). The true values of the system parameters and the noise-free input distribution for the three examples are summarized in Table 2. The simulation setup is as follows:

1) The initial value the FIR parameters are obtained using HOS method.
2) The initial guess for the GMM parameters is given by the sample variance of u_t for $\{\Gamma_j\}_{j=1}^{K}$, and by $\rho_j = 1/K$ for the mixing weights. The means of the mixture components, denoted by η_j, are evenly spaced between the minimum and the maximum value of the input signal u_t.
3) We consider different data lengths, namely $N = 1000$, $N = 2000$, and $N = 5000$.
4) The number of Monte Carlo (MC) simulations is 100.
5) The stopping criterion is chosen as

$$\left\|\hat{\beta}^{(m)} - \hat{\beta}^{(m-1)}\right\| / \left\|\hat{\beta}^{(m)}\right\| \leq 5 \times 10^{-6},$$

or when 1000 iterations of the EM algorithm have been reached. Here $\|\cdot\|$ denotes the Euclidean vector norm operator.

A. EXAMPLE 1: THE NOISE-FREE INPUT IS A GMM DISTRIBUTED SIGNAL

Consider the FIR system given in the first row of Table 2, where we define the true parameters of the system and the true noise-free input distribution. In this example the approximated log-likelihood function is defined using the joint PDF $p(\phi_{t-3:t}|\beta) = f(\phi_{t-3:t}; \mathcal{A}_4)$ and the marginal PDF $p(\phi_{t-3:t-1}|\beta) = f(\phi_{t-3:t-1}; \mathcal{A}_3)$. Notice that the number of past data points in the approximated log-likelihood function is $k = r = 3$. Figure 6, column 1, shows the mean PDF of the 100 estimates, for the data length $N \in \{1000, 2000, 5000\}$. The gray-shaded region represents the area in which all estimated PDFs lie. We observe that the average of all estimated PDFs fits the true PDF as N increases. Table 3 and 4 (first row) show the mean and the corresponding standard deviations of the estimated parameters obtained with HOS and our proposal, respectively. We observe that the estimation using our proposal, with a small data length (up to $N = 5000$), is more accurate than the HOS method, which requires a large data length (up to $N = 50000$) to yield a small estimation error.

B. EXAMPLE 2: THE NOISE-FREE INPUT IS A EXPONENTIAL DISTRIBUTED SIGNAL

Consider the FIR system given in the second row of Table 2, where we define the true parameters of the FIR system and the true noise-free input distribution. In this example, the approximated log-likelihood function can be calculated using the joint PDF $p(\phi_{t-2:t}|\beta) = f(\phi_{t-2:t}; \mathcal{A}_3)$ and the marginal PDF $p(\phi_{t-2:t-1}|\beta) = f(\phi_{t-2:t-1}; \mathcal{A}_2)$. The number of past data points in the approximated log-likelihood function is $k = r = 2$. Notice that, in this example the noise-free input signal follows an exponential distribution, and we do not have prior knowledge of the number of Gaussian components to approximate it. In this example we choose $K = 2$. In Figure 6, second column, we show the mean PDF of the 100 estimates, for the data length $N \in \{1000, 2000, 5000\}$. The gray-shaded region represents the area in which all estimated PDFs lie. We observe that the Gaussian Mixture approximation of the exponential distributions fits adequately enough the true distribution to obtain an accurate parameter estimation. Tables 3 and 4 (second row) show the mean and the corresponding standard deviations of the estimated parameters obtained with HOS and our proposal, respectively. We observe that the estimation using our proposal is more accurate than HOS method. Notice that the latter exhibits a poor performance when the data length is not too large.

C. EXAMPLE 3: THE NOISE-FREE INPUT IS UNIFORMLY DISTRIBUTED

Consider the FIR system given in the third row of Table 2, where we define the true parameters of the FIR system and the true noise-free input distribution. In this example the approximated log-likelihood function is defined using the joint PDF $p(\phi_{t-1:t}|\beta) = f(\phi_{t-1:t}; \mathcal{A}_2)$ and the marginal PDF $p(\phi_{t-1:t-1}|\beta) = f(\phi_{t-1:t-1}; \mathcal{A}_1)$. The number of past

TABLE 2. Distributions and true parameters for simulate the noise-free input distribution $p(u_t^o)$.

	FIR System	True values	Noise-free Distribution	True values
Example 1	$H(q^{-1},\theta) = h_0 + h_1q^{-1} + h_2q^{-2} + h_3q^{-3}$	$h_0 = -2.5$ $h_1 = 0.5$ $h_2 = -1.0$ $h_3 = 1.5$	$p(u_t^o) = \sum_{j=1}^{2} \rho_j \mathcal{N}(u_t^o; \eta_j, \Gamma_j)$	$\rho_1 = 0.3,\ \rho_2 = 0.7$ $\eta_1 = -1,\ \eta_2 = 3.0$ $\Gamma_1 = 1.0,\ \Gamma_2 = 1.0$
Example 2	$H(q^{-1},\theta) = h_0 + h_1q^{-1} + h_2q^{-2}$	$h_0 = 3.2$ $h_1 = 1.8$ $h_2 = -0.75$	$p(u_t^o) = \begin{cases} \lambda \exp\{-\lambda u_t^o\} & u_t^o \geq 0 \\ 0 & u_t^o < 0 \end{cases}$	$\lambda = 1,$ $K = 2$
Example 3	$H(q^{-1},\theta) = h_0 + h_1q^{-1}$	$h_0 = 4.25$ $h_1 = 1.25$	$p(u_t^o) = \begin{cases} (b-a)^{-1} & a \leq u_t^o \leq b \\ 0 & \text{otherwise} \end{cases}$	$a = -3.0,$ $b = 3.0,$ $K = 3$

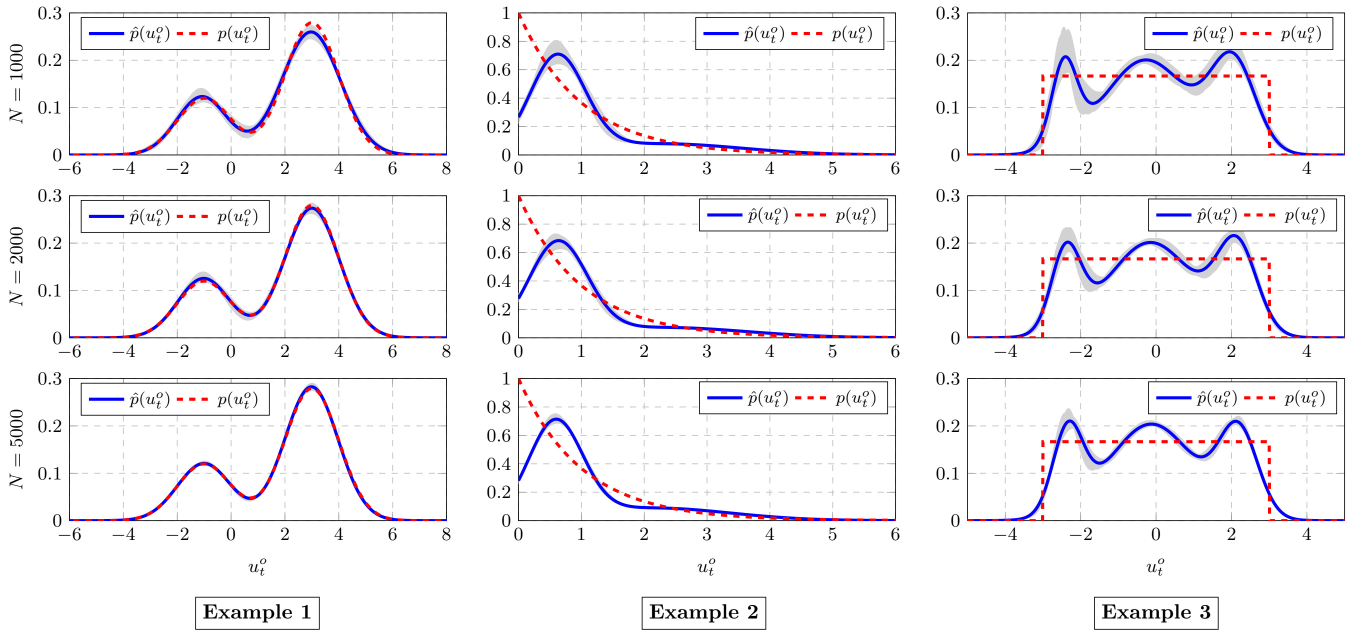

FIGURE 6. Estimated noise-free input distribution $p(u_t^0)$ using GMM for the data length $N \in \{1000, 2000, 5000\}$.

data points in the approximated log-likelihood function is $k = r = 1$. We choose $K = 3$ Gaussian components to approximate the noise-free input distribution. Figure 6, column 3, shows the estimated average PDF for all Monte Carlo realizations with $N \in \{1000, 2000, 5000\}$. As in the previous example, the gray-shaded region represents the area in which all estimated PDFs lie. We observe that the estimated GMM PDFs fits the uniform distribution making our proposal adequate for estimate the FIR-EIV system in presence of any unknown non-Gaussian distribution. In Tables 3 and 4 (third row) we summarized the results of simulations, showing the mean and standard deviation for the estimated parameters. We observe that the estimated parameters, using our proposal, are similar to the *true* value and that the standard deviation decreases while the data length N increases. In contrast, the HOS method has a poor performance and requires a large data set to improve the estimations.

TABLE 3. Estimated parameters for the numerical examples using EM-based method.

	True values	$N = 1000$	$N = 2000$	$N = 5000$
Example 1	$h_0 = -2.5$	-2.477 ± 0.046	-2.496 ± 0.032	-2.501 ± 0.020
	$h_1 = 0.5$	0.492 ± 0.032	0.504 ± 0.025	0.501 ± 0.015
	$h_2 = -1.0$	-1.026 ± 0.033	-1.000 ± 0.024	-1.000 ± 0.015
	$h_3 = 1.5$	1.515 ± 0.046	1.499 ± 0.029	1.500 ± 0.020
	$\Gamma_y = 1.0$	0.991 ± 0.424	1.022 ± 0.281	1.002 ± 0.166
	$\Gamma_u = 1.0$	1.000 ± 0.066	1.008 ± 0.052	1.000 ± 0.030
Example 2	$h_0 = 3.2$	3.209 ± 0.086	3.206 ± 0.063	3.204 ± 0.039
	$h_1 = 1.8$	1.789 ± 0.059	1.799 ± 0.041	1.801 ± 0.028
	$h_2 = -0.8$	-0.803 ± 0.062	-0.798 ± 0.047	-0.802 ± 0.028
	$\Gamma_y = 1.0$	0.967 ± 0.248	0.990 ± 0.148	1.005 ± 0.101
	$\Gamma_u = 1.0$	1.009 ± 0.052	1.010 ± 0.037	1.002 ± 0.020
Example 3	$h_0 = 4.2$	4.247 ± 0.095	4.226 ± 0.069	4.205 ± 0.045
	$h_1 = 1.2$	1.240 ± 0.060	1.222 ± 0.044	1.201 ± 0.024
	$\Gamma_y = 1.0$	0.977 ± 0.650	0.987 ± 0.504	0.999 ± 0.348
	$\Gamma_u = 1.0$	0.992 ± 0.060	1.007 ± 0.042	1.001 ± 0.030

Remark 2. *Notice that the GMMs obtained as approximations of the non-Gaussian distributions of the noise-free signals used in examples 2 and 3, are the best approximations that we can obtain using the reduced numbers of Gaussian*

TABLE 4. Estimated parameters for the numerical examples using HOS method.

	True values	$N = 1000$	$N = 2000$	$N = 5000$	$N = 10000$	$N = 30000$	$N = 50000$
Example 1	$h_0 = -2.5$	-1.942 ± 0.269	-2.087 ± 0.232	-2.137 ± 0.245	-2.205 ± 0.241	-2.329 ± 0.236	-2.423 ± 0.223
	$h_1 = 0.5$	0.333 ± 0.273	0.386 ± 0.264	0.534 ± 0.235	0.460 ± 0.231	0.522 ± 0.200	0.480 ± 0.151
	$h_2 = -1.0$	-0.808 ± 0.261	-0.813 ± 0.253	-0.799 ± 0.242	-0.815 ± 0.238	-0.820 ± 0.228	-0.866 ± 0.220
	$h_3 = 1.5$	1.442 ± 0.109	1.478 ± 0.085	1.481 ± 0.048	1.497 ± 0.023	1.501 ± 0.006	1.500 ± 0.003
Example 2	$h_0 = 3.2$	1.471 ± 0.425	1.725 ± 0.415	1.553 ± 0.351	1.753 ± 0.345	2.899 ± 0.312	2.996 ± 0.303
	$h_1 = 1.8$	0.759 ± 0.442	0.901 ± 0.337	0.990 ± 0.336	1.046 ± 0.321	1.074 ± 0.311	1.424 ± 0.300
	$h_2 = -0.8$	-0.684 ± 0.122	-0.714 ± 0.103	-0.726 ± 0.068	-0.739 ± 0.030	-0.744 ± 0.008	-0.746 ± 0.005
Example 3	$h_0 = 4.2$	3.221 ± 0.393	3.255 ± 0.370	3.308 ± 0.365	3.441 ± 0.341	3.507 ± 0.320	3.907 ± 0.308
	$h_1 = 1.2$	0.980 ± 0.358	0.990 ± 0.373	0.997 ± 0.375	1.028 ± 0.357	1.199 ± 0.336	1.202 ± 0.324

components of the examples. In particular, fitting a GMM with 3 components to the uniformly distributed noise-free signal in Example 3 yields: $\rho_1 = 0.1785$, $\eta_1 = -2.3343$, $\Gamma_1 = 0.1625$, $\rho_2 = 0.5851$, $\eta_2 = -0.1318$, $\Gamma_2 = 1.3089$, $\rho_3 = 0.2364$, $\eta_3 = 2.1796$, *and* $\Gamma_3 = 0.2697$, *which are the parameters obtained using the proposed method for EIV-FIR identification. An information criterion can be used to select the number of components of the GMM, e.g., Akaike information criterion [65]. However, a further analysis of this issue is out of the scope of the paper.*

VII. PRACTICAL EXPERIMENTAL RESULTS

In this section, we first utilize our methodology to approximate a sampled first-order continuous-time system with an FIR system model. Then, we consider an experimental setup, namely a Rotary Servo Unit from Quanser, to test both EIV and HOS methods utilizing FIR system models.

A. APPROXIMATION OF A FIRST-ORDER SYSTEM USING AN FIR MODEL

Consider the first-order continuous-time transfer function given by:

$$\frac{Y(s)}{U(s)} = \frac{K_c}{\tau s + 1}, \tag{49}$$

where $Y(s)$ is the Laplace transform of output $y(t)$, $U(s)$ is the Laplace transform of input $u(t)$, K_c is the steady-state gain, τ is the time constant, and s is the argument of the Laplace transform. The corresponding sampled-data transfer function with sampling period, Δ, and instantaneous sampling is given by [66]:

$$\frac{Y(z)}{U(z)} = K_d \frac{b_1}{z - a_1}, \tag{50}$$

where $K_d = K_c/\tau$, $b_1 = 1 - e^{-\tau\Delta}$, $a_1 = e^{-\tau\Delta}$, and z is the forward shift operator or the Z-transform variable. To illustrate the FIR system model approximation for the sampled-data model in (50), we consider a simulation setup with data length $N = 5000$, $K_c = 46.59$, $\tau = 31.81$, $\Delta = 50$ ms, and 50 Monte Carlo simulations. The PDF of the noise-free input signal is shown in Fig. 7(a). We also consider a 4th order FIR system model to approximate the sampled model in (50), and noise variances $\Gamma_u = \Gamma_y = 0.1$. Fig. 7(a) shows the estimated average of the noise-free input voltage PDF as a GMM for all Monte Carlo realizations (red line). The blue-shaded region represents the area in which all the estimated PDFs lie. The gray-shaded bars correspond to the histogram of the corresponding noise-free input voltage signal. We observe that the average of the estimated GMM fits the noise-free input distribution. Fig. 7(b) shows the impulse response corresponding to the average of all Monte Carlo simulations for the estimated FIR system model (red line). The blue line represents the impulse response of the sampled-data model in (50). We observe an accurate approximation of the first-order system using a FIR system model. Additionally, from the simulations, the estimated input and output noise variances are $\hat{\Gamma}_u = 0.12 \pm 0.011$ and $\hat{\Gamma}_y = 0.11 \pm 0.009$, respectively.

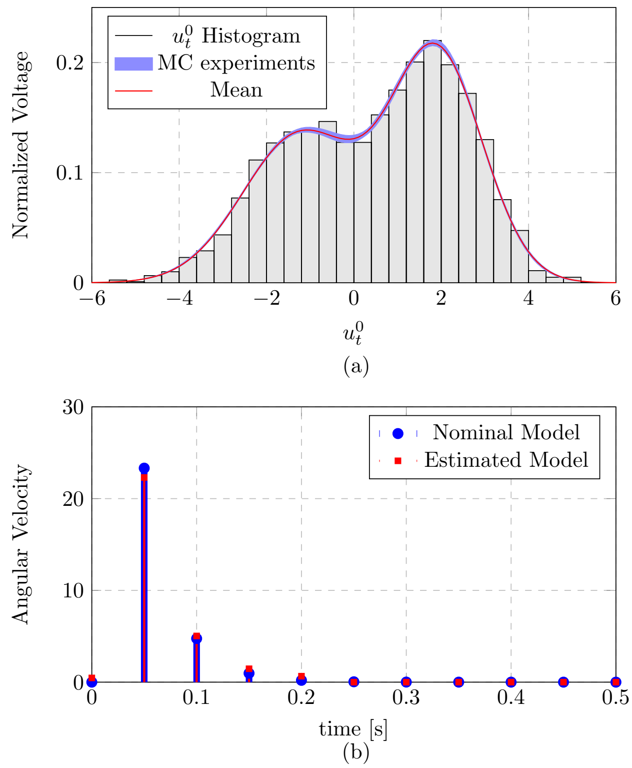

FIGURE 7. Simulations Results: (a) Noise-free input distribution, (b) Impulse response of the true and estimated model with $\Delta = 0.05$ [s].

B. APPLICATION TO QUANSER ROTARY SERVO UNIT

The angular speed of the Rotary Servo Base Unit load shaft with respect to the input motor voltage can be modeled with the first-order continuous-time transfer function in (49), see e.g., [67]. Here, $Y(s)$ is the Laplace transform of the load shaft speed, and $U(s)$ is the Laplace transform of the motor input voltage.

Fig. 8 shows the hardware configuration of the experimental setup, which consists of a SRV02 Rotary Servo Base Unit, a Q8-USB Data Acquisition (DAQ) device, a Quanser VoltPaq-X2 power amplifier, and a personal computer with MATLAB 2021a/Simulink software. This experimental setup has been used in several applications related to modeling and control [68], [69]. In this case we aim at obtaining an FIR system model for the Rotatory Servo Base unit. The angular speed measurements are obtained from the encoder signal utilizing a second order low pass filter [67]. The experimental setup is built using Matlab/Simulink and QUARC software to drive the experimental plant in real time. The sampling rate to collect the experimental data ($N = 5000$) is selected to 1 kHz. The noise-free input was the same as the simulation experiment, see Fig. 7(a). As in the simulation results, we consider a 4th order FIR system for modeling the sampled-data transfer function of the angular speed with respect to the input motor voltage utilizing both EIV and HOS methods. The estimation results of the input and output noise variances obtained with the proposed method are $\hat{\Gamma}_u = 0.0026$ and $\hat{\Gamma}_y = 0.0012$, respectively. Since the noise source variances are small, the FIR model of both the EIV and HOS methods are similar, as it is shown by the frequency response (magnitude and phase) in Fig. 9(a). Additionally, the EIV method yields an approximation of the PDF of the noise-free input motor voltage, as a Gaussian mixture model, see Fig. 9(b).

VIII. CONCLUSIONS

In this paper, we have proposed an identification algorithm for FIR-EIV systems. We have considered that the unknown distribution of the noise-free input $p(u_t^0)$ is a GMM. We have used the ML estimator to identify system parameters, input and output noise variances, and the parameters defining the GMM distribution. In our approach we have not assume any prior knowledge about the input and output noise variances. In order to deal with easier-to-handle expressions, we have used the EM algorithm and a composite likelihood function. Based on the Wiener approximation theorem, we have shown that our proposed algorithm adequately handles non-Gaussian distributions that are not Gaussian mixture distributions but can be approximated by a GMM, yielding more accurate estimates than the HOS method.

APPENDIX A TECHNICAL LEMMATA

Theorem 2. *Any PDF of an n-dimensional random variable u_t^0, $p(u_t^0)$, with compact support can be approximated as closely as desired in the space $L_1(\mathbb{R}^n)$ by a distribution of the form:*

$$p(u_t^0) \approx \sum_{j=1}^{K} \rho_j \mathcal{N}\left(u_t^0; \eta_j, \Gamma_j\right), \tag{51}$$

where $\mathcal{N}\left(u_t^0; \eta_j, \Gamma_j\right)$ represents an n-dimensional Gaussian distribution with mean $\eta_j \in \mathbb{R}^n$, covariance matrix $\Gamma_j \in \mathbb{R}^{n\times n}$, K is the number of elements in the sum, and $\rho_j > 0$ is the jth mixing weight subject to $\sum_{j=1}^{K} \rho_j = 1$.

Proof. See [33, Theorem 3]. □

APPENDIX B PROOF LEMMA 1

Without loss of generality we prove the case for $m = 2$ and $K = 2$, since for $m > 2$ and any number of Gaussian components the procedure is similar. Due to the fact that u_1^0 and u_2^0 are independent and identically distributed, the joint PDF of $\xi_2 = \left[u_1^0, u_2^0\right]^\top$ is given by

$$p(\xi_2) = \prod_{\kappa=1}^{2} \sum_{j=1}^{2} \rho_j \mathcal{N}(u_\kappa^0; \eta_j, \Gamma_j), \tag{52}$$

then, in extended form, we have

$$\begin{aligned} p(\xi_2) =& \rho_1\rho_1 \mathcal{N}\left(u_1^0; \eta_1, \Gamma_1\right) \mathcal{N}\left(u_2^0; \eta_1, \Gamma_1\right) + \\ & \rho_2\rho_1 \mathcal{N}\left(u_1^0; \eta_2, \Gamma_2\right) \mathcal{N}\left(u_2^0; \eta_1, \Gamma_1\right) + \\ & \rho_1\rho_2 \mathcal{N}\left(u_1^0; \eta_1, \Gamma_1\right) \mathcal{N}\left(u_2^0; \eta_2, \Gamma_2\right) + \\ & \rho_2\rho_2 \mathcal{N}\left(u_1^0; \eta_2, \Gamma_2\right) \mathcal{N}\left(u_2^0; \eta_2, \Gamma_2\right). \end{aligned} \tag{53}$$

Notice that, all of the elements in $p(\xi_2)$ are products of two Gaussian PDFs, and due to the fact u_1^0 and u_2^0 are independent, then each product is a joint PDF, thus

$$\begin{aligned} p(\xi_2) =& \rho_1\rho_1 \mathcal{N}\left(\begin{bmatrix}u_1^0\\u_2^0\end{bmatrix}; \begin{bmatrix}\eta_1\\\eta_1\end{bmatrix}, \begin{bmatrix}\Gamma_1 & 0\\0 & \Gamma_1\end{bmatrix}\right) + \\ & \rho_2\rho_1 \mathcal{N}\left(\begin{bmatrix}u_1^0\\u_2^0\end{bmatrix}; \begin{bmatrix}\eta_2\\\eta_1\end{bmatrix}, \begin{bmatrix}\Gamma_2 & 0\\0 & \Gamma_1\end{bmatrix}\right) + \\ & \rho_1\rho_2 \mathcal{N}\left(\begin{bmatrix}u_1^0\\u_2^0\end{bmatrix}; \begin{bmatrix}\eta_1\\\eta_2\end{bmatrix}, \begin{bmatrix}\Gamma_1 & 0\\0 & \Gamma_2\end{bmatrix}\right) + \\ & \rho_2\rho_2 \mathcal{N}\left(\begin{bmatrix}u_1^0\\u_2^0\end{bmatrix}; \begin{bmatrix}\eta_2\\\eta_2\end{bmatrix}, \begin{bmatrix}\Gamma_2 & 0\\0 & \Gamma_2\end{bmatrix}\right). \end{aligned} \tag{54}$$

Putting all in compact form we get

$$p(\xi_2) = \sum_{i=1}^{M} \alpha_i \mathcal{N}\left(\xi_2; \mu_i, \Sigma_i\right), \tag{55}$$

where $M = 4$. The weight $\alpha_i \in \mathbb{R}$, the mean $\mu_i \in \mathbb{R}^2$, and the covariance matrix $\Sigma_i \in \mathbb{R}^{2\times 2}$ are obtained using the indices of the ith experiment in the full factorial design with 2 factors and 2 levels for each factor, i.e. $\{\{1,1\},\{1,2\},\{2,1\},\{2,2\}\}$ as it is shown in Figure 4. Finally, we can conclude that the joint PDF in (21) holds for all m and K.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3255827

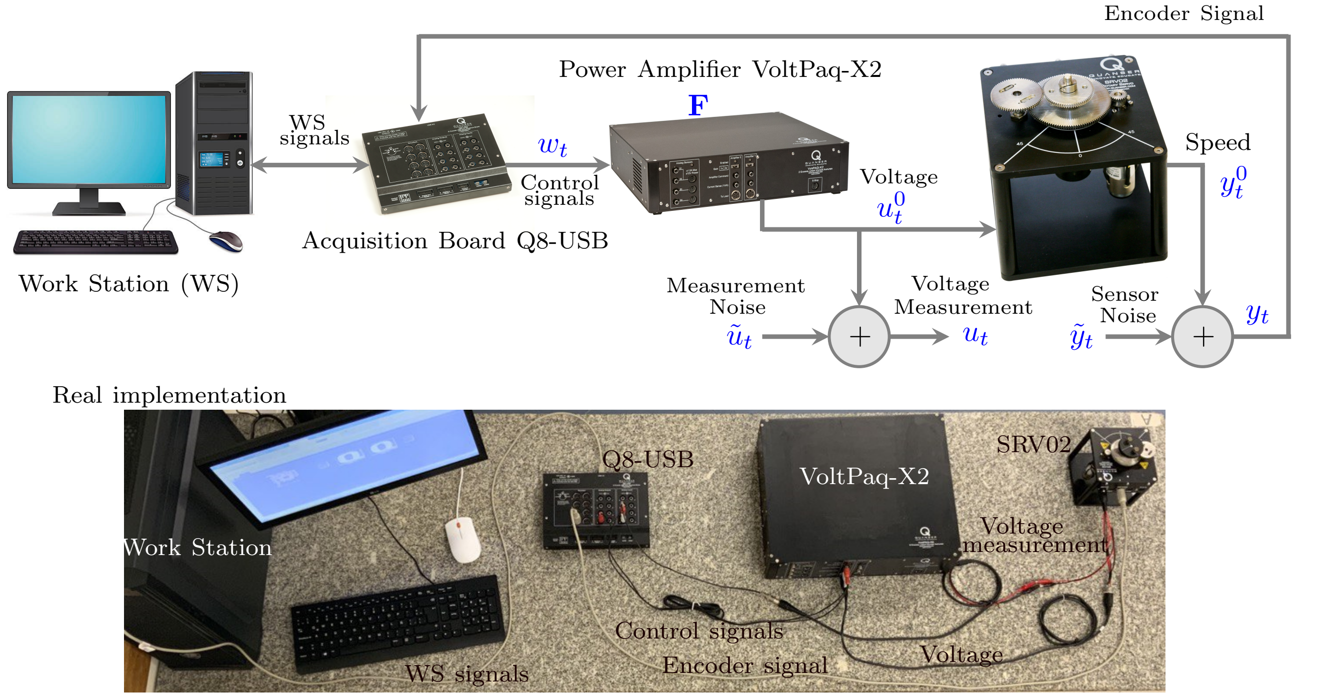

FIGURE 8. Experimental EIV setup using a Rotary Servo Base Quanser unit.

APPENDIX C PROOF LEMMA 2

Consider the characteristic function of the random vector $X \in \mathbb{R}^\tau$ given by

$$\Psi(X,t) = \mathbb{E}\left\{\exp\left\{jt^\top X\right\}\right\}, \tag{56}$$

where $t \in \mathbb{R}^\tau$, $j = \sqrt{-1}$, and $\mathbb{E}\{\cdot\}$ is the expectation operator. The characteristic function of the sum $\phi_\tau = \mathcal{A}\xi_m + \vartheta_\tau$ is given by:

$$\begin{aligned}\Psi(\phi_\tau,t) &= \mathbb{E}\left\{\exp\left\{jt^\top(\mathcal{A}\xi_m + \vartheta_\tau)\right\}\right\}\\ &= \mathbb{E}\left\{\exp\left\{j(\mathcal{A}^\top t)^\top \xi_m\right\}\exp\left\{jt^\top\vartheta_\tau\right\}\right\}.\end{aligned} \tag{57}$$

Due to the independence of ξ_m and ϑ_τ we have

$$\Psi(\phi_\tau,t) = \Psi(\xi_m,\mathcal{A}^\top t)\Psi(\vartheta_\tau,t). \tag{58}$$

We need to find the characteristic function of ξ_m and ϑ_τ, where ξ_m is Gaussian mixture distributed and ϑ_τ is Gaussian distributed. By definition

$$\begin{aligned}\Psi(\xi_m,\zeta) &= \mathbb{E}\left\{\exp\left\{j\zeta^\top\xi_m\right\}\right\}\\ &= \int \exp\left\{j\zeta^\top\xi_m\right\} p(\xi_m)d\xi_m\\ &= \int \exp\left\{j\zeta^\top\xi_m\right\}\sum_{i=1}^{M}\alpha_i\mathcal{N}(\xi_m;\mu_i,\Sigma_i)d\xi_m\\ &= \sum_{i=1}^{M}\alpha_i\int \exp\left\{j\zeta^\top\xi_m\right\}\mathcal{N}(\xi_m;\mu_i,\Sigma_i)d\xi_m\\ &= \sum_{i=1}^{M}\alpha_i\Psi_i(\xi_m,\zeta),\end{aligned} \tag{59}$$

where $\zeta = \mathcal{A}^\top t$ and $\Psi_i(\xi_m,\zeta)$ is the characteristic function of ith multivariate Gaussian component. Now, consider a general multivariate Gaussian PDF of $W \in \mathbb{R}^\tau$, given by $\mathcal{N}(w;\mu,\Sigma)$. Then, its characteristic function is obtained as:

$$\begin{aligned}\Psi(W,t) &= \int \exp\left\{jt^\top w\right\}\mathcal{N}(w;\mu,\Sigma)dw\\ &= \frac{1}{\sqrt{\det 2\pi\Sigma}}\int \Big[\exp\left\{jt^\top w\right\}\\ &\quad \exp\left\{-\frac{1}{2}\left[w-\mu\right]^\top\Sigma^{-1}\left[w-\mu\right]\right\}\Big]dw\\ &= \frac{1}{\sqrt{\det 2\pi\Sigma}}\int \exp\Big\{-\frac{1}{2}\left[w-\mu\right]^\top\\ &\qquad \Sigma^{-1}\left[w-\mu\right] + jt^\top w\Big\}dw.\end{aligned} \tag{60}$$

Completing the square in the exponential argument we get

$$\begin{aligned}\text{Arg} &= -\frac{1}{2}\left[w-\mu\right]^\top\Sigma^{-1}\left[w-\mu\right] + jt^\top w\\ &= -\frac{1}{2}\left[w-\mu\right]^\top\Sigma^{-1}\left[w-\mu\right] + jt^\top w\\ &+ jt^\top\mu - jt^\top\mu + \frac{1}{2}t^\top\Sigma t - \frac{1}{2}t^\top\Sigma t\\ &= -\frac{1}{2}\left[w-\mu\right]^\top\Sigma^{-1}\left[w-\mu\right] + jt^\top\Sigma\Sigma^{-1}(w-\mu)\\ &+ jt^\top\mu + \frac{1}{2}t^\top\Sigma\Sigma^{-1}\Sigma t - \frac{1}{2}t^\top\Sigma t\\ &= -\frac{1}{2}\left[w-b\right]^\top\Sigma^{-1}\left[w-b\right] + jt^\top\mu - \frac{1}{2}t^\top\Sigma t,\end{aligned} \tag{61}$$

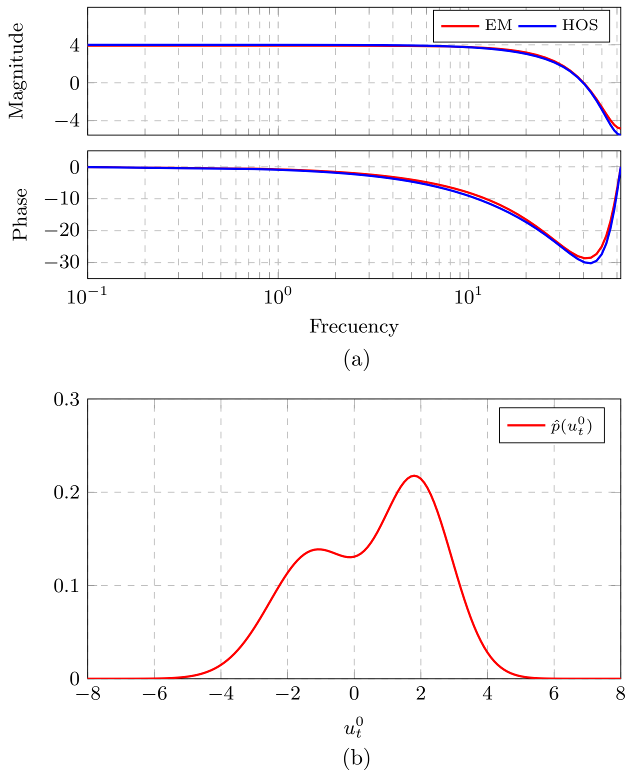

FIGURE 9. Experimental Results: (a) Frequency response of the experimental system, (b) Estimated PDF of noisy-free input signal as a GMM.

where $b = (\mu + j\Sigma t)$. Then

$$\Psi(W,t) = \frac{1}{\sqrt{\det 2\pi\Sigma}} \exp\left\{ jt^\top \mu - \frac{1}{2} t^\top \Sigma t \right\} \int \exp\left\{ -\frac{1}{2} [w-b]^\top \Sigma^{-1} [w-b] \right\} dw. \tag{62}$$

The last integral has closed form [70, pág 321] given by $\sqrt{\det 2\pi\Sigma}$, then

$$\Psi(W,t) = \exp\left\{ jt^\top \mu - \frac{1}{2} t^\top \Sigma t \right\}. \tag{63}$$

Thus, using the result derived in (63), the characteristic functions of ξ_m and ϑ_τ are given respectively by

$$\Psi(\xi_m, \mathcal{A}^\top t) = \sum_{i=1}^{M} \alpha_i \exp\left\{ jt^\top \mathcal{A}\mu_i - \frac{1}{2} t^\top \mathcal{A}\Sigma_i \mathcal{A}^\top t \right\} \tag{64}$$

$$\Psi(\vartheta_\tau, t) = \exp\left\{ jt^\top \varepsilon - \frac{1}{2} t^\top \Phi t \right\}. \tag{65}$$

Finally the characteristic function of ϕ_τ is as follows

$$\begin{aligned} \Psi(\phi_\tau, t) &= \Psi(\xi_m, \mathcal{A}^\top t) \Psi(\vartheta_\tau, t) \\ &= \sum_{i=1}^{M} \alpha_i \exp\left\{ jt^\top (\mathcal{A}\mu_i + \varepsilon) - \frac{1}{2} t^\top \left(\mathcal{A}\Sigma_i \mathcal{A}^\top + \Phi \right) t \right\}, \end{aligned} \tag{66}$$

where each element of the characteristic function of ϕ_τ corresponds to characteristic function of Gaussian distribution with mean $\mathcal{A}\mu_i + \varepsilon$ and covariance matrix $\mathcal{A}\Sigma_i \mathcal{A}^\top + \Phi$, then (24) holds.

REFERENCES

[1] T. Söderström, *Errors-in-Variables Methods in System Identification*, ser. Communications and Control Engineering. Springer International Publishing, 2018.

[2] C. L. Cheng and J. W. Van Ness, *Statistical Regression with Measurement Error*. Wiley, 1998.

[3] J. H. Stock and M. W. Watson, "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, vol. 20, no. 2, pp. 147–162, 2002.

[4] K. Lau, J. H. Braslavsky, J. C. Agüero, and G. C. Goodwin, "An errors-in-variables method for non-stationary data with application to mineral exploration," *Automatica*, vol. 45, no. 12, pp. 2971–2976, 2009.

[5] R. H. Middleton, T. Wigren, K. Lau, and R. A. Delgado, "Data flow delay equalization for feedback control applications using 5G wireless dual connectivity," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*. IEEE, 2017, pp. 1–7.

[6] R. Carvajal, R. Delgado, J. C. Agüero, and G. C. Goodwin, "An identification method for Errors-in-Variables systems using incomplete data," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 1359–1364, 2012.

[7] B. Anderson and M. Deistler, "Identifiability in dynamic errors-in-variables models." *Journal of Time Series Analysis*, vol. 5, no. 1, pp. 1–13, 1984.

[8] ——, "Dynamic errors-in-variables systems with three variables," *Automatica*, vol. 23, no. 5, pp. 611–616, 1987.

[9] O. Reiersøl, "Identifiability of a linear relation between variables which are subject to error," *Econometrica: Journal of the Econometric Society*, pp. 375–389, 1950.

[10] M. Deistler, "Linear dynamic errors-in-variables models," *Journal of Applied Probability*, pp. 23–39, 1986.

[11] J. C. Agüero and G. C. Goodwin, "Identifiability of errors in variables dynamic systems," *Automatica*, vol. 44, no. 2, pp. 371–382, 2008.

[12] W. A. Fuller, *Measurement Error Models*, ser. Wiley Series in Probability and Statistics. Wiley, 2009.

[13] J. P. Buonaccorsi, *Measurement error: models, methods, and applications*. Chapman and Hall/CRC, 2010.

[14] U. Soverini and T. Söderström, "Frequency domain identification of FIR models in the presence of additive input–output noise," *Automatica*, vol. 115, p. 108879, 2020.

[15] S. Thil, W. X. Zheng, M. Gilson, and H. Garnier, "Unifying some higher-order statistic-based methods for errors-in-variables model identification," *Automatica*, vol. 45, no. 8, pp. 1937–1942, 2009.

[16] J. K. Tugnait, "Stochastic system identification with noisy input using cumulant statistics," in *IEEE Transactions on Automatic Control*. IEEE, 1992, pp. 476–485.

[17] W. Yao and W. Song, "Mixtures of linear regression with measurement errors," *Communications in Statistics-Theory and Methods*, vol. 44, no. 8, pp. 1602–1614, 2015.

[18] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, no. 6, pp. 939–958, 2007.

[19] L. Ljung, "System Identification: Theory for the User," *PTR Prentice Hall, Upper Saddle River, NJ*, vol. 28, 1999.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[21] J. C. Agüero, W. Tang, J. I. Yuz, R. Delgado, and G. C. Goodwin, "Dual time–frequency domain system identification," *Automatica*, vol. 48, no. 12, pp. 3031–3041, 2012.

[22] R. Carvajal, R. Orellana, D. Katselis, P. Escárate, and J. C. Agüero, "A data augmentation approach for a class of statistical inference problems," *PloS one*, vol. 13, no. 12, 2018.

[23] J. J. Hanfelt and K. Y. Liang, "Approximate likelihoods for generalized linear errors-in-variables models," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 59, no. 3, pp. 627–637, 1997.

[24] X. Gao and P. X.-K. Song, "Composite likelihood EM algorithm with applications to multivariate hidden Markov model," *Statistica Sinica*, pp. 165–185, 2011.

[25] W. He and G. Y. Yi, "A pairwise likelihood method for correlated binary data with/without missing observations under generalized partially linear single-index models," *Statistica Sinica*, pp. 207–229, 2011.

[26] L. Naduvilezhath, L. E. Rose, and Metzler, "Jaatha: a fast composite-likelihood approach to estimate demographic parameters," *Molecular Ecology*, vol. 20, no. 13, pp. 2709–2723, 2011.
[27] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 192–225, 1974.
[28] ——, "Statistical analysis of non-lattice data," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 24, no. 3, pp. 179–195, 1975.
[29] B. G. Lindsay, "Composite likelihood methods," *Contemporary mathematics*, vol. 80, no. 1, pp. 221–239, 1988.
[30] A. V. Vecchia, "Estimation and model identification for continuous spatial processes," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 50, no. 2, pp. 297–312, 1988.
[31] G. C. Goodwin and J. C. Agüero, "Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 368–373.
[32] A. Azzalini, "Maximum likelihood estimation of order m for stationary stochastic processes," *Biometrika*, vol. 70, no. 2, pp. 381–387, 1983.
[33] J. Lo, "Finite-dimensional sensor orbits and optimal nonlinear filtering," *IEEE Transactions on information theory*, vol. 18, no. 5, pp. 583–588, 1972.
[34] D. L. Alspach and H. W. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.
[35] A. L. Cedeño, R. Albornoz, R. Carvajal, B. I. Godoy, and J. C. Agüero, "A Two-Filter Approach for State Estimation Utilizing Quantized Output Data," *Sensors*, vol. 21, no. 22, 2021.
[36] ——, "On Filtering Methods for State-Space Systems having Binary Output Measurements," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 815–820, 2021.
[37] A. L. Cedeño, R. Carvajal, and J. C. Agüero, "A Novel Filtering Method for Hammerstein-Wiener State-Space Systems," in *2021 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2021, pp. 1–7.
[38] A. L. Cedeño, R. Orellana, R. Carvajal, and J. C. Agüero, "EM-based identification of static errors-in-variables systems utilizing Gaussian Mixture models," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 863–868, 2020.
[39] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.
[40] R. Orellana, R. Carvajal, and J. C. Agüero, "Empirical Bayes estimation utilizing finite Gaussian Mixture Models," in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2019, pp. 1–6.
[41] G. Bittner, R. Orellana, R. Carvajal, and J. C. Agüero, "Maximum Likelihood identification for Linear Dynamic Systems with finite Gaussian mixture noise distribution," in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2019, pp. 1–7.
[42] R. Orellana, G. Bittner, R. Carvajal, and J. C. Agüero, "Maximum Likelihood estimation for non-minimum-phase noise transfer function with Gaussian mixture noise distribution," *Automatica*, vol. 135, p. 109937, 2022.
[43] R. Orellana, R. Carvajal, J. C. Agüero, and G. C. Goodwin, "Model Error Modelling using a Stochastic Embedding approach with Gaussian Mixture Models for FIR system," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 845–850, 2020, 21th IFAC World Congress.
[44] R. Orellana, P. Escárate, M. Curé, A. Christen, R. Carvajal, and J. C. Agüero, "A method to deconvolve stellar rotational velocities-iii. the probability distribution function via maximum likelihood utilizing finite distribution mixtures," *Astronomy & Astrophysics*, vol. 623, p. A138, 2019.
[45] K. Van Montfort, A. Mooijaart, and J. De Leeuw, "Regression with errors in variables: estimators based on third order moments," *Stat. Neerl.*, vol. 41, no. 4, pp. 223–238, 1987.
[46] R. Orellana, R. Carvajal, P. Escárate, and J. C. Agüero, "On the Uncertainty Identification for Linear Dynamic Systems Using Stochastic Embedding Approach with Gaussian Mixture Models," *Sensors*, vol. 21, no. 11, p. 3837, 2021.
[47] C. Yu and M. Verhaegen, "Data-driven fault estimation of non-minimum phase LTI systems," *Automatica*, vol. 92, pp. 181–187, 2018.
[48] M. C. Campi, "The problem of pole-zero cancellation in transfer function identification and application to adaptive stabilization," *Automatica*, vol. 32, no. 6, pp. 849–857, 1996.
[49] C. Varin and P. Vidoni, "A note on composite likelihood inference and model selection," *Biometrika*, vol. 92, no. 3, pp. 519–528, 2005.
[50] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, Å. Nyström, J. Pettersen, and R. Bergman, "Experimental design and optimization," *Chemometrics and Intelligent Laboratory Systems*, vol. 42, no. 1, pp. 3–40, 1998.
[51] G. E. Dieter and L. C. Schmidt, *Engineering design*. McGraw-Hill Higher Education Boston, 2009.
[52] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667 – 1682, 2005.
[53] R. B. Gopaluni, "A particle filter approach to identification of nonlinear processes under missing observations," *The Canadian Journal of Chemical Engineering*, vol. 86, no. 6, pp. 1081–1092, 2008.
[54] S. Frühwirth-Schnatter, G. Celeux, and C. Robert, *Handbook of Mixture Analysis*. New York: Chapman and Hall/CRC, 2018.
[55] T. Coleman and Y. Li, "An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds," *SIOPT*, vol. 6, no. 2, pp. 418–445, 1996.
[56] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*. Springer, 2016, vol. 4.
[57] N. V. Sahinidis, "Baron: A general purpose global optimization software package," *Journal of global optimization*, vol. 8, no. 2, pp. 201–205, 1996.
[58] "Matlab optimization toolbox," 2021, the MathWorks, Natick, MA, USA.
[59] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
[60] L. Xu and M. Jordan, "On Convergence properties of the EM Algorithm for Gaussian Mixtures," *Neural Computation*, vol. 8, no. 1, pp. 129–151, 1996.
[61] R. Zhao, Y. Li, and Y. Sun, "Statistical convergence of the EM algorithm on Gaussian mixture models," *Electron. J. Statist.*, vol. 14, no. 1, pp. 632–660, 2020.
[62] D. Feng and W. Xing Zheng, "Bilinear equation method for unbiased identification of linear fir systems in the presence of input and output noises," *Signal Processing*, vol. 87, no. 5, pp. 1147–1155, 2007.
[63] P. Lichota, "Inclusion of the D-optimality in multisine manoeuvre design for aircraft parameter estimation," *Journal of Theoretical and Applied Mechanics*, vol. 54, no. 1, 2016.
[64] A. Pawełek and P. Lichota, "Arrival air traffic separations assessment using Maximum Likelihood Estimation and Fisher Information Matrix," in *20th International Carpathian Control Conference*, 2019, pp. 1–6.
[65] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
[66] G. Goodwin, J. C. Agüero, M. E. Cea-Garrido, M. Salgado, and J. Yuz, "Sampling and Sampled-Data Models: The Interface Between the Continuous World and Digital Algorithms," *IEEE Control Systems Magazine*, vol. 33, no. 5, pp. 34–53, 2013.
[67] Quanser, "Rotary servo base unit." https://www.quanser.com/products/rotary-servo-base-unit/, [Online; accessed 20-Oct.-2022].
[68] S. R. Adarsh and S. Selvakumar, "Model identification and position control of quanser qube servo dc motor," in *Advances in Electrical and Computer Technologies*, T. Sengodan, M. Murugappan, and S. Misra, Eds. Singapore: Springer Nature Singapore, 2021, pp. 1321–1335.
[69] S. Balamurugan, P. Venkatesh, and M. Varatharajan, "Fuzzy sliding-mode control with low pass filter to reduce chattering effect: an experimental validation on Quanser SRIP," *Sādhanā*, vol. 42, no. 10, pp. 1693–1703, 2017.
[70] D. A. Harville, "Matrix algebra from a statistician's perspective," 1998.

ANGEL L. CEDEÑO (LWJG) was born in Manabí, Ecuador. He received the undergraduate degree of Ingeniero en Electrónica y Control from Escuela Politécnica Nacional, Quito, Ecuador. He is currently a student of the electronic engineering Ph.D. program with Universidad Técnica Federico Santa María, Valparaíso, Chile. His research interests include system identification, control, and optimization-based predictive control for embedded systems.

RAFAEL ORELLANA was born in San Cristóbal, Venezuela. He obtained the professional title of Electronic Engineer from the Universidad Nacional Experimental del Táchira (Venezuela). He also obtained the M.Sc. degree in Automation and Instrumentation from the Universidad de Los Andes (Venezuela). He received the M.Sc. degree and a Ph.D. in Electronic Engineering from the Universidad Técnica Federico Santa María (Chile) in 2020 and 2021, respectively. He is currently an Academic of the Department of Electrical Engineering at Universidad de Santiago de Chile. His research interests are System Identification, Statistical Signal Processing, Control and Instrumentation.

RODRIGO CARVAJAL was born in Valparaíso, Chile. He received the Ingeniero Civil Electrónico degree and the M.Sc. degree from the Pontificia Universidad Católica de Valparaíso in 2003 and 2007, respectively, and a Ph.D. from The University of Newcastle (Australia) in 2013. Dr. Carvajal received the Postgraduate Research Excellence Prize for Electrical, Computer, and Telecommunication Engineering in 2011. He is currently with the School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso (Chile). His research interests include Communication Systems, Statistical Signal Processing, and Systems Identification.

BORIS I. GODOY was born in Valparaíso, Chile, in 1975. He received his professional title of Ingeniero Civil Electrónico and the M.Sc. degree in electronics engineering from the Universidad Técnica Federico Santa María in (Chile) in 2001, and his Ph.D in electrical engineering from The University of Newcastle (Australia) in 2008. Since then, he has held several academic and industry positions in Australia. Currently he is a Researcher at Lund University, Lund, Sweden. His main research interests are in system identification, time-varying systems, process control, and dynamical systems in general.

JUAN C. AGÜERO (Member, IEEE) was born in Osorno, Chile. He obtained the professional title of Ingeniero Civil Electrónico and a Master of Engineering from Universidad Técnica Federico Santa María (Chile) in 2000, and a Ph.D. from The University of Newcastle (Australia) in 2006. He gained industrial experience from 1997 to 1998 in the copper mining industry at El Teniente, Codelco (Chile). He was a Research Academic at The University of Newcastle (Australia) until 2014 and is currently an Associate Professor at Universidad Técnica Federico Santa María (Chile). He is an Associate editor of Automatica. His research interests include system identification, control and statistical signal processing.