EM-based identification of static errors-in-variables systems utilizing Gaussian Mixture models [★]

Angel L. Cedeño [∗] **Rafael Orellana** [∗,∗∗] **Rodrigo Carvajal** [∗]
Juan C. Agüero [∗]

[∗] *Electronic Engineering Department, Universidad Técnica Federico Santa María, Valparaíso, Chile.*

[∗∗] *School of Electrical Engineering, Universidad de Los Andes, Mérida, Venezuela.*

Abstract: In this paper we address the problem of identifying a static errors-in-variables system. Our proposal is based on the Expectation-Maximization algorithm, in which we consider that the distribution of the noise-free input is approximated by a finite Gaussian mixture. This approach allows us to estimate the static system parameters, the input and output noise variances, and the Gaussian mixture parameters. We show the benefits of our proposal via numerical simulations.

Keywords: Errors-in-variables, Maximum Likelihood, Expectation-Maximization, Gaussian Mixture, Estimation, Optimization.

1. INTRODUCTION

Errors-in-variables (EIV) models (Söderström, 2007) are fundamental research problems of systems identification, where both input and output are corrupted by additive errors, see e.g. (Deistler and Anderson, 1989; Tugnait, 1990; Thil et al., 2008; Fuller, 2009; Buonaccorsi, 2010; Söderström, 2018). EIV models arise in many applications, such as medical, agricultural, economical, among others, in which many variables can only be measured with errors. (Cheng and Van Ness, 1998). Initial studies of EIV systems go back to (Adcock, 1877, 1878) for the static case. In the literature, the identification of static EIV systems has been addressed by many authors utilizing different approaches, such as the Frisch Scheme (FS) (Guidorzi et al., 2008), Confirmatory Factor Analysis (CFA) (Bartholomew et al., 2011), Higher Order Moments (HOM) methods (Van Montfort et al., 1987), Maximum Likelihood (ML) method (Cheng and Van Ness, 1998), and Bias-Eliminated Least Square (BELS) method that can be interpreted as a form of weighted Instrumental Variable methods (Gilson and Van den Hof, 2001), among others. One of the difficulties in static EIV models is that the identifiability does not hold in general. For instance, it is well known that if all random variables are jointly normal, the *structural* model (where the noise-free input is a stochastic process) is not, in general, identifiable from second order properties, see e.g. (Agüero and Goodwin, 2008). On the other hand, (Reiersøl, 1950) shows that static EIV systems are identifiable when the noise-free input is non-Gaussian distributed.

Identification of static EIV systems with non-Gaussian distributed noise-free input has been solved using HOM methods in the FS approach, in which only the linear regression parameters are estimated (Van Montfort et al., 1987). The limitation of this approach is that it depends upon the availability of a large sample data set to obtain reliable estimates, see e.g. (Agüero and Goodwin, 2008). Another strategy to address the static EIV system identification problem has recently been presented in (Yao and Song, 2015), in which the distribution of the output noise is approximated by Gaussian Mixture Models (GMM) and the noise-free input distribution is estimated by a non-parametric deconvolution kernel method, assuming known distribution structures for the input and output errors.

On the other hand, for the ML estimator of the static EIV system one needs to consider prior knowledge about the noise-free input distribution, see e.g. (Cheng and Van Ness, 1998). For instance, in the time domain, it is usually assumed that the noise-free input is described as an Auto-Regressive Moving Average (ARMA) process (driven by a Gaussian noise) with known spectral density (Diversi et al., 2007; Söderström, 2007). In the frequency domain, it is assumed that the noise-free input is a periodic signal (Pintelon and Schoukens, 2007; Söderström and Soverini, 2017) or it has spectra different than zero for a sufficient number of frequencies (Carvajal et al., 2012). Nevertheless, if we do not consider any prior knowledge about the well-behaved but non-Gaussian noise-free input, we can approximate its unknown distribution by GMM. Notice that, this approach has been typically utilized in non-linear filtering (Arasaratnam et al., 2007), Bayesian inference (Carvajal et al., 2018; Orellana et al., 2018), dynamic systems estimation (Orellana et al., 2019a), Astronomy (Orellana et al., 2019b), among others.

In this paper we focus on the development of an ML estimator for the static EIV system problem. We approximate the noise-free input distribution by a GMM, and we estimate its *unknown distribution*. In our analysis we solve the EIV problem estimation with GMM, using an identifi-

cation technique based on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977).

2. MAXIMUM LIKELIHOOD ESTIMATION FOR STATIC EIV SYSTEMS USING GMM

2.1 System Model

For simplicity of the presentation, we assume that a straight line with an intercept equal to zero should be fitted to the available noise-corrupted data. We consider the setup depicted in Fig. 1, where the noise-free input and the undisturbed output, denoted by u_t^o and y_t^o respectively, are linked by:

$$y_t^o = K_s u_t^o, \tag{1}$$

where $K_s \in \mathbb{R}$ represents a constant slope. We consider that the observations are corrupted by additive measurement noise, i.e. the following holds:

$$\begin{aligned} u_t &= u_t^o + \tilde{u}_t, \\ y_t &= y_t^o + \tilde{y}_t, \end{aligned} \tag{2}$$

where $\tilde{u}_t$ and $\tilde{y}_t$ are additive zero-mean mutually uncorrelated Gaussian white noise with variance $\sigma_{\tilde{u}}^2$ and $\sigma_{\tilde{y}}^2$, respectively. We will focus on the structural problem, in which u_t^o is a stochastic process independent of the measurement noise sources $\tilde{u}_t$ and $\tilde{y}_t$. Then, the static EIV system can be rewritten as a system with a two-dimensional output vector and a three-dimensional mutually uncorrelated noise input vector (Söderström (2007)) as:

$$\begin{bmatrix} y_t \\ u_t \end{bmatrix} = \begin{bmatrix} K_s & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_t^o \\ \tilde{y}_t \\ \tilde{u}_t \end{bmatrix}. \tag{3}$$

Note that the static EIV system is cast into a multivariate system (3) with both u_t and y_t as outputs.

2.2 Gaussian Mixture Models

In many applications such as control, filtering, and estimation, it is generally necessary to estimate, from the contaminated measurement data, the corresponding noise distribution, see e.g. (Alspach and Sorenson, 1972; Zhuang et al., 1996; Mengersen et al., 2011). GMM is an approach widely used to approximate an unknown or mathematically difficult to handle probability density function (pdf). This approach is summarized as follows:

Lemma 1. Any pdf of an n-dimensional random variable u_t^o, $p(u_t^o)$, with compact support can be approximated as closely as desired in the space $L_1(\mathbb{R}^n)$ by a distribution of the form:

$$p(u_t^o) \approx \sum_{i=1}^{M} \alpha_i \phi(u_t^o; \mu_i, \Sigma_i), \tag{4}$$

where $\phi(u_t^o; \mu_i, \Sigma_i)$ represents an n-dimensional Gaussian distribution with mean μ_i, covariance matrix Σ_i, M is the number of elements in the sum, and $\alpha_i > 0$ is the i-th mixing weight subject to $\sum_{i=1}^{M} \alpha_i = 1$.
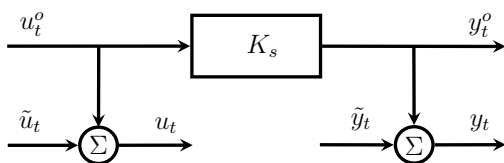

Fig. 1. Basic setup of static EIV problem.

Proof. See (Lo, 1972, Theorem 3). □

Remark 2. It is well known that GMM are not indentifiable due to the invariance of (4) under permutation of the indices (Mengersen et al., 2011, chapter 10), this is called Label switching problem. This means that for an arbitrary permutation in the ordering of the components, the result is a perfect symmetric distribution. Usually, this problem is solved by imposing an identifying ordering constraint on the parameters (Richardson and Green, 1997). ▽

2.3 Problem definition

Let us assume that the unknown distribution of the noise-free input, u_t^o, is approximated by a GMM as follows:

$$p(u_t^o) \approx \sum_{i=1}^{M} \alpha_i \phi(u_t^o; \mu_i, \sigma_i^2), \tag{5}$$

where α_i is the i-th mixing weight, $\phi(u_t^o; \mu_i, \sigma_i^2)$ represents a Gaussian distribution, with mean μ_i and variance σ_i^2, given by:

$$\phi(u_t^o; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ -\frac{(u_t^o - \mu_i)^2}{2\sigma_i^2} \right\}, \tag{6}$$

and M is the number of the mixture components required to approximate $p(u_t^o)$. Then, the problem under study can be formulated as follows:

From a set of noise-corrupted input and output signals, we estimate K_s, the variances $\sigma_{\tilde{u}}^2$ and $\sigma_{\tilde{y}}^2$, and the parameters that define the GMM for u_t^o utilizing ML. Thus, the vector of parameters of interest is given by:

$$\beta = \left[K_s, \sigma_{\tilde{u}}^2, \sigma_{\tilde{y}}^2, \alpha_1, \mu_1, \sigma_1^2, \dots, \alpha_M, \mu_M, \sigma_M^2 \right]^T. \tag{7}$$

Remark 3. Notice that our approach requires the existence of the parameter vector $\beta = \beta_0$ that defines the *true* system. We also consider that only u_t and y_t ($t = 1, \dots, N$) are available to be measured. ▽

2.4 Maximum Likelihood Estimation using GMM

In Maximum Likelihood methods we maximize the pdf of the data as a function of the unknown parameters, see e.g. (Söderström, 2018; Agüero et al., 2012). Let us consider that the sets $y_{1:N}$ and $u_{1:N}$ are collections of independent and identically distributed measurements. Then, utilizing the Bayes's theorem we obtain:

$$p(y_{1:N}, u_{1:N} | \beta) = \prod_{t=1}^{N} p(y_t, u_t | \beta), \tag{8}$$

where $p(y_t, u_t | \beta)$ is the joint pdf of the random variables y_t and u_t. Introducing $w_t = u_t^o$ as a latent variable and re-writing (3) in matrix form ($V = AX$) we obtain:

$$\underbrace{\begin{bmatrix} u_t^o \\ \tilde{y}_t \\ \tilde{u}_t \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -K_s \\ 0 & 1 & -1 \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} y_t \\ u_t \\ w_t \end{bmatrix}}_{V}. \tag{9}$$

Notice that due to the fact that $\tilde{u}_t$ and $\tilde{y}_t$ are assumed to be independent of u_t^o, the joint pdf, $p_X(x)$, of the noise input vector X is given by:

$$p_X(x) = \sum_{i=1}^{M} \alpha_i \phi(u_t^o; \mu_i, \sigma_i^2) \phi(\tilde{y}_t; 0, \sigma_{\tilde{y}}^2) \phi(\tilde{u}_t; 0, \sigma_{\tilde{u}}^2). \tag{10}$$

Utilizing the transformation of the random vectors theorem, see e.g. (Jazwinski, 1970, Theorem 2.7), $p_V(v)$ can be obtained as follows:

$$p_V(v) = \frac{1}{\det(A)} p_X(A^{-1}v), \qquad (11)$$

with $\det(A) = 1$. Then $p_V(v) = p_X(A^{-1}v)$. In order to obtain the log-likelihood function, we present the following result:

Lemma 4. The likelihood function for the available data is given by

$$\mathcal{L}_N(\beta) = \prod_{t=1}^{N} \sum_{i=1}^{M} \frac{\alpha_i}{2\pi\sqrt{a_i \varsigma_i}} \exp\left\{ \frac{b_{it}^2}{2a_i} - \frac{c_{it}}{2} \right\}, \qquad (12)$$

where:

$$\varsigma_i = \sigma_i^2 \sigma_{\tilde{y}}^2 \sigma_{\tilde{u}}^2, \qquad (13)$$

$$a_i = \left(\sigma_{\tilde{y}}^2 \sigma_{\tilde{u}}^2 + K_s^2 \sigma_i^2 \sigma_{\tilde{u}}^2 + \sigma_i^2 \sigma_{\tilde{y}}^2 \right) / \varsigma_i, \qquad (14)$$

$$b_{it} = \left(\mu_i \sigma_{\tilde{y}}^2 \sigma_{\tilde{u}}^2 + K_s y_t \sigma_i^2 \sigma_{\tilde{u}}^2 + u_t \sigma_i^2 \sigma_{\tilde{y}}^2 \right) / \varsigma_i, \qquad (15)$$

$$c_{it} = \left(\mu_i^2 \sigma_{\tilde{y}}^2 \sigma_{\tilde{u}}^2 + y_t^2 \sigma_i^2 \sigma_{\tilde{u}}^2 + u_t^2 \sigma_i^2 \sigma_{\tilde{y}}^2 \right) / \varsigma_i. \qquad (16)$$

Proof. See Appendix A. $\qquad \square$

Based on Lemma 4, the log-likelihood function is given by:

$$\ell_N(\beta) = \sum_{t=1}^{N} \log\left\{ \sum_{i=1}^{M} \frac{\alpha_i}{2\pi\sqrt{a_i \varsigma_i}} \exp\left\{ \frac{b_{it}^2}{2a_i} - \frac{c_{it}}{2} \right\} \right\}, \qquad (17)$$

where the vector of parameters β is defined in (7). Then the ML estimator is given by:

$$\hat{\beta} = \arg\max_{\beta} \ell_N(\beta), \quad \text{s.t.} \quad 0 \le \alpha_i \le 1, \quad \sum_{i=1}^{M} \alpha_i = 1. \quad (18)$$

3. AN EM-BASED ALGORITHM FOR STATIC EIV SYSTEMS USING GMM

3.1 Algorithm formulation

Let us define the following:

$$K(z_t, \beta_i) = \frac{\alpha_i}{2\pi\sqrt{a_i \varsigma_i}} \exp\left\{ \frac{b_{it}^2}{2a_i} - \frac{c_{it}}{2} \right\}, \qquad (19)$$

$$\mathcal{V}_t(\beta) = \sum_{i=1}^{M} K(z_t, \beta_i), \qquad (20)$$

where $z_t = \{y_t, u_t\}$ is the observed data, see e.g. (Carvajal et al., 2018). Then, the log-likelihood function in (17) can be expressed as:

$$\ell_N(\beta) = \sum_{t=1}^{N} \log\left[\mathcal{V}_t(\beta) \right]. \qquad (21)$$

In (Carvajal et al., 2018) an estimation algorithm for a class of problems with data augmentation was proposed. Following (Carvajal et al., 2018), we define $\mathcal{B}_t(\beta) = \log\left[\mathcal{V}_t(\beta) \right]$, obtaining:

$$\mathcal{B}_t(\beta) = \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\beta, \hat{\beta}^{(m)}), \qquad (22)$$

where the functions $\mathcal{Q}_t(\beta, \hat{\beta}^{(m)})$ and $\mathcal{H}_t(\beta, \hat{\beta}^{(m)})$ are given by

$$\mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{M} \log\left[K(z_t, \beta_i) \right] \frac{K(z_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}, \qquad (23)$$

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{M} \log\left[\frac{K(z_t, \beta_i)}{\mathcal{V}_t(\beta)} \right] \frac{K(z_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}. \qquad (24)$$

Lemma 5. The function $\mathcal{H}_t(\beta, \hat{\beta}^{(m)})$ is a decreasing function for any value of β and satisfies the following:

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\hat{\beta}^{(m)}, \hat{\beta}^{(m)}) \le 0. \qquad (25)$$

Proof. The result is directly obtained from Jensen's inequality, see e.g. (Carvajal et al. 2018) and the references therein). $\qquad \square$

From Lemma 5, we have that the log-likelihood function in (21) satisfies the following:

$$\ell_N(\hat{\beta}^{(m+1)}) \ge \ell_N(\hat{\beta}^{(m)}). \qquad (26)$$

Finally, we can formulate the following iterative algorithm:

$$\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}) = \sum_{t=1}^{N} \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}), \qquad (27)$$

$$\hat{\beta}^{(m+1)} = \arg\max_{\beta} \ \bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}), \qquad (28)$$

$$\text{s.t.} \qquad 0 \le \alpha_i \le 1, \qquad \sum_{i=1}^{M} \alpha_i = 1. \qquad (29)$$

Notice that (27) and (28) correspond to the E-step and M-step of the EM algorithm, respectively, see e.g. (Dempster et al., 1977). On the other hand, the proposed methodology in (Carvajal et al., 2018) can provide closed-form expressions for optimization problems, see e.g. (Orellana et al., 2018, 2019b). In addition, for solving the optimization problem in (28), the constraint $0 \le \alpha_i \le 1$ is not considered explicitly. For more details see Appendix B.

3.2 Optimization of the auxiliary function $\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)})$

We consider the coordinate descent algorithm (Wright (2015)) to optimize the auxiliary function $\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)})$ with respect to β. For the optimization of the auxiliary function in (27) we can obtain closed-form expressions for the estimates of the parameters $\{\alpha_i, \mu_i\}_{i=1}^{M}$. The optimization can be carried out as follows.

Lemma 6. The parameters $\{\hat{\alpha}_i, \hat{\mu}_i\}$ that optimize the auxiliary function $\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)})$ in (27) with respect to $\{\alpha_i, \mu_i\}$ are given by:

$$\hat{\alpha}_i^{(m+1)} = \left(\sum_{t=1}^{N} \mathcal{F}(z_t, \hat{\beta}_i^{(m)}) \right) \left(\sum_{t=1}^{N} \sum_{l=1}^{M} \mathcal{F}(z_t, \hat{\beta}_l^{(m)}) \right)^{-1} \qquad (30)$$

$$\hat{\mu}_i^{(m+1)} = \left(\sum_{t=1}^{N} \mathcal{M}(z_t, \hat{\beta}_i^{(m)}) \right) \left(\sum_{t=1}^{N} \mathcal{P}(z_t, \hat{\beta}_i^{(m)}) \right)^{-1} \qquad (31)$$

with

$$\mathcal{F}(z_t, \hat{\beta}_i^{(m)}) = \frac{K(z_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}, \qquad (32)$$

$$\mathcal{P}(z_t, \hat{\beta}_i^{(m)}) = \sum_{t=1}^{N} \left[\frac{\hat{\sigma}_{\tilde{y}}^{2(m)} \hat{\sigma}_{\tilde{u}}^{2(m)}}{\hat{\varsigma}_i^{(m)}} - \frac{\hat{\sigma}_{\tilde{y}}^{4(m)} \hat{\sigma}_{\tilde{u}}^{4(m)}}{(\hat{\varsigma}_i^{(m)})^2 \hat{a}_i^{(m)}} \right] \mathcal{F}(z_t, \hat{\beta}_i^{(m)}), \qquad (33)$$

and

$$\mathcal{M}(z_t, \hat{\beta}_i^{(m)}) = \sum_{t=1}^{N} \left[\left(y_t \hat{K}_s^{(m)} \hat{\sigma}_i^{2(m)} \hat{\sigma}_{\tilde{u}}^{2(m)} + u_t \hat{\sigma}_i^{2(m)} \hat{\sigma}_{\tilde{y}}^{2(m)} \right) \right.$$
$$\left. \left(\mathcal{F}(z_t, \hat{\beta}_i^{(m)}) \hat{\sigma}_{\tilde{y}}^{2(m)} \hat{\sigma}_{\tilde{u}}^{2(m)} / (\hat{\varsigma}_i^{(m)})^2 \hat{a}_i^{(m)} \right) \right]. \qquad (34)$$

Proof. See Appendix B. □

By defining $\theta = \{K_s, \sigma_{\tilde{u}}^2, \sigma_{\tilde{y}}^2, \{\sigma_i^2\}_{i=1}^M\}$, we obtain the estimates of the remaining parameters as follows:

$$\hat{\theta}^{(m+1)} = \arg\max_{\theta} \ \bar{\mathcal{Q}}(\theta, \hat{\theta}^{(m)}).$$
$$\text{s.t.} \quad \{\sigma_{\tilde{u}}^2, \sigma_{\tilde{y}}^2, \{\sigma_i^2\}_{i=1}^M\} > 0. \tag{35}$$

The optimization problem in (35) can be carried out by using Nonlinear Programming for constrained problems, see e.g. (Coleman and Li (1996) and the references therein). We summarize the proposed algorithm as follows:

(i) Fix the number components, M, of the GMM for u_t^o.
(ii) Choose an initial guess $\hat{\beta}^{(0)}$, and set $m = 0$.
(iii) Compute the GMM parameters $\{\hat{\alpha}_i^{(m+1)}, \hat{\mu}_i^{(m+1)}\}$ from (30) and (31) in Lemma 6.
(iv) Compute the estimates of $\hat{\theta}^{(m+1)}$ by solving (35).
(v) Set $m = m + 1$ and go back to step (iii) until a stopping criterion is satisfied.

4. NUMERICAL EXAMPLE

In this section we present two numerical examples to analyze the performance of our proposal. We solve the EIV problem in Fig. 1 considering two different noise-free input signals. We also consider that the measurements $\{u_{1:N}, y_{1:N}\}$ are generated from the system model in (1)-(2) with *true* values given by

$$K_s = 5, \quad \sigma_{\tilde{u}}^2 = 1, \quad \sigma_{\tilde{y}}^2 = 1, \tag{36}$$

with u_t^o non-Gaussian distributed. In order to have an adequate initialization of our algorithm, the initial guess will be obtained from the HOM in (Van Montfort et al., 1987). Thus, for comparison purposes, we compare both estimators in terms of mean value and variance.

The simulation setup is as follows:

(1) The initial value for the static gain K_s is given by HOM estimation.
(2) The initial guess for the GMM parameters is given by the sample variance of u_t for $\{\sigma_i^2\}_{i=1}^M$, and by $\alpha_i = 1/M$ for the mixing weights. The means of the mixture components, denoted by μ_i, are evenly spaced between the minimum and the maximum value of the input signal u_t.
(3) We consider 3 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 have been reached.

4.1 Example 1: Bimodal Noise-free input signal

In this example, we consider that the noise-free input signal u_t^o is drawn from a finite Gaussian mixture with two components given by:

$$p(u_t^o)^{(\text{True})} = \alpha_1 \phi(u_t^o; \mu_1, \sigma_1^2) + \alpha_2 \phi(u_t^o; \mu_2, \sigma_2^2), \tag{37}$$

with $\alpha_1 = \alpha_2 = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$, $\mu_1 = -5$ and $\mu_2 = 5$.

Fig. 2(a) shows the mean pdf of the 100 estimates, with $N = 5000$. The gray-shaded region represents the area in which all estimated pdf's lie. It can be observed that the

Table 1. Estimated parameters in Example 1.

N	Method	$\hat{K}_s$	$\hat{\sigma}_{\tilde{y}}^2$	$\hat{\sigma}_{\tilde{u}}^2$
1000	HOM	5.028 ± 0.256	-	-
	EM	5.000 ± 0.030	1.060 ± 0.757	0.999 ± 0.060
2000	HOM	5.040 ± 0.620	-	-
	EM	5.002 ± 0.023	1.013 ± 0.575	0.998 ± 0.041
5000	HOM	4.985 ± 0.156	-	-
	EM	4.998 ± 0.007	1.018 ± 0.416	1.000 ± 0.024

Table 2. Estimated parameters in Example 2.

N	Method	$\hat{K}_s$	$\hat{\sigma}_{\tilde{y}}^2$	$\hat{\sigma}_{\tilde{u}}^2$
1000	HOM	3.924 ± 2.884	-	-
	EM	5.046 ± 0.163	0.818 ± 0.642	1.001 ± 0.068
2000	HOM	4.197 ± 2.999	-	-
	EM	5.018 ± 0.128	0.959 ± 0.478	1.001 ± 0.036
5000	HOM	4.555 ± 3.469	-	-
	EM	4.996 ± 0.049	0.983 ± 0.255	1.001 ± 0.023

true Gaussian mixture pdf and the mean estimate pdf are very similar.

Table. 1 shows the mean and the corresponding standard deviations of the estimated parameters obtained with HOM and our proposal. We observed that the estimation is very accurate in this example for both HOM and our technique.

4.2 Example 2: Uniformly distributed noise-free input signal

In this example we consider that the noise-free input signal u_t^o is distributed as follows:

$$p(u_t^o)^{(\text{True})} = \frac{1}{l_b - l_a} I_{[l_a, l_b]}(u_t^o), \tag{38}$$

with $l_a = -2$, $l_b = 2$, and $I_{[l_a, l_b]}$ the indicator function given by

$$I_{[l_a, l_b]}(u_t^o) = \begin{cases} 1, & l_a \leq u_t^o \leq l_b & \text{(39a)} \\ 0, & \text{otherwise.} & \text{(39b)} \end{cases}$$

To approximate the noise-free input signal distribution, we consider $M = 7$ components for the GMM, see (Mengersen et al., 2011, pág 277). Fig. 2(b) shows the estimated average pdf for all MC realizations with $N = 5000$. As in the previous example, the gray-shaded region represents the area in which all estimated pdf's lie. We can observe that the estimated GMM pdf fits the uniform distribution.

In Table 2 we summarized the results of simulations, showing the mean and standard deviation for the estimated parameters. We observe that the estimated parameters are similar to the *true* value and that the standard deviation decreases while the data length N increases. In contrast, HOM method has a poor performance and requires a large data set to improve the estimations.

5. CONCLUSION

In this paper we proposed an identification algorithm for static EIV systems. We used the fact that any pdf can be approximated as closely as desired by a GMM, to approximate the unknown distribution of the noise-free input. We used the ML approach to estimate the static system parameters, input and output noise variances, and characteristic parameters of GMM. In order to deal with easier to handle expressions, we based our work
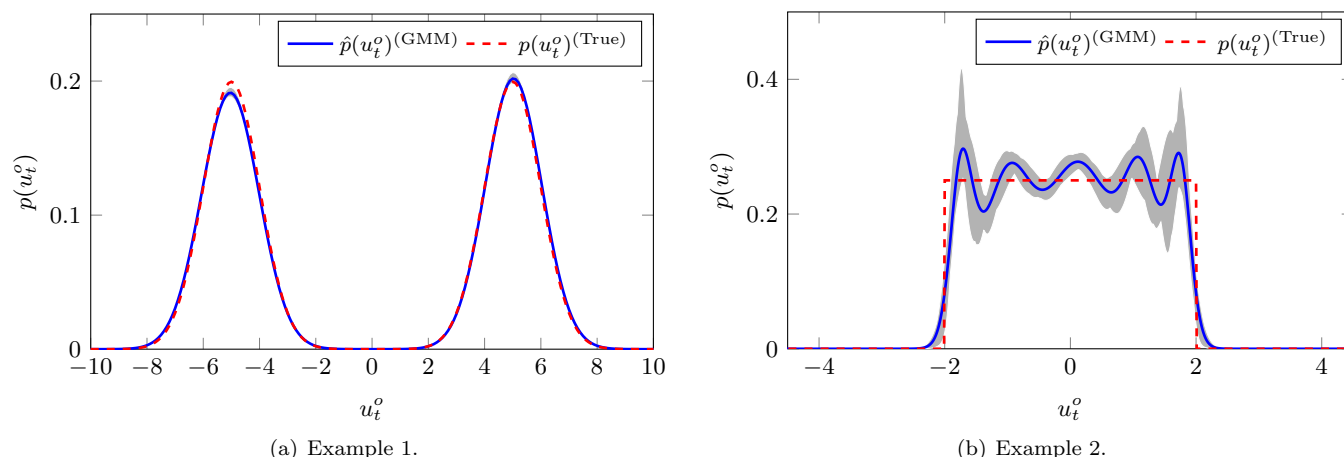
(a) Example 1.

(b) Example 2.

Fig. 2. Estimated noise-free input distribution $p(u_t^o)$ using GMM.

on the EM algorithm and the work in (Carvajal et al., 2018). We derived closed form expressions to estimate the weights and means of the GMM. From the simulations, we conclude that our proposal adequately handles non-Gaussian distributions, yielding more accurate estimates than HOM.

REFERENCES

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

Adcock, R.J. (1878). A problem in Least Squares. *The Analyst*, 5(2), 53–54.

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

Agüero, J.C., Tang, W., Yuz, J.I., Delgado, R., and Goodwin, G.C. (2012). Dual time–frequency domain system identification. *Automatica*, 48(12), 3031–3041.

Alspach, D. and Sorenson, H. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE transactions on automatic control*, 17(4), 439–448.

Arasaratnam, I., Haykin, S., and Elliott, R.J. (2007). Discrete-Time Nonlinear Filtering Algorithms Using Gauss-Hermite Quadrature. *Proceedings of the IEEE*, 95(5), 953–977.

Bartholomew, D.J., Knott, M., and Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*, volume 904. John Wiley & Sons.

Buonaccorsi, J.P. (2010). *Measurement error: models, methods, and applications*. Chapman and Hall/CRC.

Carvajal, R., Delgado, R., Agüero, J.C., and Goodwin, G.C. (2012). An identification method for errors-in-variables systems using incomplete data. *IFAC Proceedings Volumes*, 45(16), 1359–1364.

Carvajal, R., Orellana, R., Katselis, D., Escárate, P., and Agüero, J.C. (2018). A data augmentation approach for a class of statistical inference problems. *PLOS ONE*, 13(12), 1–24.

Cheng, C. and Van Ness, J. (1998). *Statistical Regression with Measurement Error*. Wiley.

Coleman, T. and Li, Y. (1996). An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIOPT*, 6(2), 418–445.

Deistler, M. and Anderson, B.D. (1989). Linear dynamic errors-in-variables models: Some structure theory. *Journal of Econometrics*, 41(1), 39–63.

Dempster, A.P., Laird, N.M., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Royal Statistical Society*, 39(1), 1–38.

Diversi, R., Guidorzi, R., and Soverini, U. (2007). Maximum likelihood identification of noisy input–output models. *Automatica*, 43(3), 464–472.

Fuller, W. (2009). *Measurement Error Models*. Wiley Series in Probability and Statistics. Wiley.

Gilson, M. and Van den Hof, P. (2001). On the relation between a bias-eliminated Least-Squares (BELS) and an IV estimator in closed-loop identification. *Automatica*, 37(10), 1593–1600.

Guidorzi, R., Diversi, R., and Soverini, U. (2008). The Frisch scheme in algebraic and dynamic identification problems. *Kybernetika*, 44(5), 585–616.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*. ISSN. Elsevier Science.

Lo, J. (1972). Finite-dimensional sensor orbits and optimal nonlinear filtering. *IEEE Transactions on Information Theory*, 18(5), 583–588.

Mengersen, K.L., Robert, C., and Titterington, M. (2011). *Mixtures: Estimation and Applications*, volume 896. John Wiley & Sons.

Orellana, R., Carvajal, R., and Agüero, J.C. (2018). Maximum Likelihood Infinite Mixture Distribution Estimation Utilizing Finite Gaussian Mixtures. *IFAC-PapersOnLine*, 51(15), 706 – 711.

Orellana, R., Carvajal, R., and Agüero, J.C. (2019a). Empirical Bayes estimation utilizing finite Gaussian Mixture Models. In *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 1–6. IEEE.

Orellana, R., Escárate, P., Curé, M., Christen, A., Carvajal, R., and Agüero, J.C. (2019b). A method to deconvolve stellar rotational velocities - III. The probability distribution function via maximum likelihood utilizing finite distribution mixtures. *A&A*, 623, A138.

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

Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica: Journal of the Econometric Society*, 375–389.

Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4), 731–792.

Söderström, T. (2018). *Errors-in-Variables Methods in System Identification.* Communications and Control Engineering. Springer International Publishing.

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

Söderström, T. and Soverini, U. (2017). Errors-in-variables identification using maximum likelihood estimation in the frequency domain. *Automatica*, 79, 131–143.

Thil, S., Hong, M., Söderström, T., Gilson, M., and Garnier, H. (2008). Statistical analysis of a third-order cumulants based algorithm for discrete-time errors-in-variables identification. *IFAC Proceedings Volumes*, 41(2), 420–425.

Tugnait, J.K. (1990). Stochastic system identification with noisy input using cumulant statistics. In *29th IEEE Conference on Decision and Control*, 1080–1085. IEEE.

Van Montfort, K., Mooijaart, A., and De Leeuw, J. (1987). Regression with errors in variables: estimators based on third order moments. *Stat. Neerl.*, 41(4), 223–238.

Wright, S. (2015). Coordinate descent algorithms. *Mathematical Programming*, 151(1), 3–34.

Yao, W. and Song, W. (2015). Mixtures of linear regression with measurement errors. *Communications in Statistics-Theory and Methods*, 44(8), 1602–1614.

Zhuang, X., Huang, Y., Palaniappan, K., and Zhao, Y. (1996). Gaussian mixture density modeling, decomposition, and applications. *IEEE Transactions on Image Processing*, 5(9), 1293–1302.

Appendix A. OBTAINING THE LIKELIHOOD FUNCTION

The joint pdf in (10) can be expressed as:

$$p_X(x) = \sum_{i=1}^{M} \frac{\alpha_i}{(2\pi)^{\frac{3}{2}}\sqrt{\varsigma_i}} \exp\left\{-\left(\frac{(u_t^o - \mu_i)^2}{2\sigma_i^2} + \frac{\tilde{y}^2}{2\sigma_{\tilde{y}}^2} + \frac{\tilde{u}^2}{2\sigma_{\tilde{u}}^2}\right)\right\}, \tag{A.1}$$

Evaluating (11) in (9), the joint pdf $p_V(v)$ is given by:

$$p_V(v) = \sum_{i=1}^{M} \frac{\alpha_i}{(2\pi)^{\frac{3}{2}}\sqrt{\varsigma_i}} \exp\left\{-\left(\frac{(w_t - \mu_i)^2}{2\sigma_i^2} + \frac{(y_t - K_s w_t)^2}{2\sigma_{\tilde{y}}^2} + \frac{(u_t - w_t)^2}{2\sigma_{\tilde{u}}^2}\right)\right\}. \tag{A.2}$$

Then, we can define:

$$p_V(v) = \sum_{i=1}^{M} \frac{\alpha_i}{(2\pi)^{\frac{3}{2}}\sqrt{\varsigma_i}} \exp\left\{-\frac{a_i w_t^2 - 2b_{it}w_t + c_{it}}{2}\right\}, \tag{A.3}$$

where ς_i, a_i, b_{it} and c_{it} are given by (13)-(16). Integrating (A.3) with respect to the latent variable w_t, we obtain:

$$p(y_t, u_t) = \sum_{i=1}^{M} \frac{\alpha_i}{2\pi\sqrt{a_i \varsigma_i}} \exp\left\{\frac{b_{it}^2}{2a_i} - \frac{c_{it}}{2}\right\}. \tag{A.4}$$

Finally, the likelihood function is given by:

$$\mathcal{L}_N(\beta) = \prod_{t=1}^{N}\sum_{i=1}^{M} \frac{\alpha_i}{2\pi\sqrt{a_i \varsigma_i}} \exp\left\{\frac{b_{it}^2}{2a_i} - \frac{c_{it}}{2}\right\}. \tag{A.5}$$

$\square$

Appendix B. COMPUTING THE PARAMETERS OF THE GMM

Using (32) we take the derivative of (27) with respect to μ_i and making it equal to zero yields:

$$\frac{\partial \bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)})}{\partial \mu_i} = \sum_{t=1}^{N}\left[\left(\hat{\mu}_i^{(m+1)}\hat{\sigma}_{\tilde{y}}^{2(m)}\hat{\sigma}_{\tilde{u}}^{2(m)} + y_t \hat{K}_s^{(m)}\hat{\sigma}_i^{2(m)}\hat{\sigma}_{\tilde{u}}^{2(m)}\right.\right.$$
$$\left. + u_t \hat{\sigma}_i^{2(m)}\hat{\sigma}_{\tilde{y}}^{2(m)}\right)\left(\hat{\sigma}_{\tilde{y}}^{2(m)}\hat{\sigma}_{\tilde{u}}^{2(m)}/(\hat{\varsigma}_i^{(m)})^2 \hat{a}_i^{(m)}\right) -$$
$$\left.\left(\hat{\mu}_i^{(m+1)}\hat{\sigma}_{\tilde{y}}^{2(m)}\hat{\sigma}_{\tilde{u}}^{2(m)}/\hat{\varsigma}_i^{(m)}\right)\right]\mathcal{F}(z_t, \hat{\beta}_i^{(m)}) = 0. \tag{B.1}$$

Using the definition in (33) and (34) we obtain:

$$\hat{\mu}_i^{(m+1)}\sum_{t=1}^{N}\left(\hat{\sigma}_{\tilde{y}}^{4(m)}\hat{\sigma}_{\tilde{u}}^{4(m)} - \hat{\sigma}_{\tilde{y}}^{2(m)}\hat{\sigma}_{\tilde{u}}^{2(m)}\hat{\varsigma}_i^{(m)}\hat{a}_i^{(m)}\right) \tag{B.2}$$

$$\left(\mathcal{F}(z_t, \hat{\beta}_i^{(m)})/(\hat{\varsigma}_i^{(m)})^2 \hat{a}_i^{(m)}\right) = \sum_{t=1}^{N}\mathcal{M}(z_t, \hat{\beta}_i^{(m)})$$

Then, we obtain:

$$\hat{\mu}_i^{(m+1)} = \left(\sum_{t=1}^{N}\mathcal{M}(z_t, \hat{\beta}_i^{(m)})\right)\left(\sum_{t=1}^{N}\mathcal{P}(z_t, \hat{\beta}_i^{(m)})\right)^{-1}.$$

For the parameter α_i we define $\mathcal{R}(\alpha_i)$ as follows:

$$\mathcal{R}(\alpha_i) = \sum_{t=1}^{N}\sum_{i=1}^{M}\log[\alpha_i]\left\{\frac{K(z_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}\right\}, \tag{B.3}$$

subject to $\sum_{i=1}^{M}\alpha_i = 1$. Notice that, we initially do not consider the constraint $0 \le \alpha_i \le 1$.

Then, using a Lagrange multiplier to deal with the constraint on α_i we define:

$$\mathcal{J}(\alpha_i, \gamma) = \sum_{t=1}^{N}\sum_{i=1}^{M}\log[\alpha_i]\mathcal{F}(z_t, \hat{\beta}_i^{(m)}) - \gamma\left(\sum_{i=1}^{M}\alpha_i - 1\right). \tag{B.4}$$

Taking the derivative of (B.4) with respect to α_i and γ, then equating to zero we obtain:

$$\frac{\partial \mathcal{J}(\alpha_i, \gamma)}{\partial \alpha_i} = \frac{1}{\hat{\alpha}_i^{(m+1)}}\sum_{t=1}^{N}\mathcal{F}(z_t, \hat{\beta}_i^{(m)}) - \gamma = 0. \tag{B.5}$$

$$\frac{\partial \mathcal{J}(\alpha_i, \gamma)}{\partial \gamma} = \sum_{i=1}^{M}\alpha_i - 1 = 0. \tag{B.6}$$

Then,

$$\hat{\alpha}_i^{(m+1)} = \sum_{t=1}^{N}\mathcal{F}(z_t, \hat{\beta}_i^{(m)})/\gamma. \tag{B.7}$$

Taking a summation over $i = 1, \ldots, M$ in (B.7) and use (B.6) we have:

$$\sum_{i=1}^{M}\hat{\alpha}_i^{(m+1)} = \sum_{i=1}^{M}\sum_{t=1}^{N}\mathcal{F}(z_t, \hat{\beta}_i^{(m)})/\gamma = 1, \tag{B.8}$$

$$\gamma = \sum_{i=1}^{M}\sum_{t=1}^{N}\mathcal{F}(z_t, \hat{\beta}_i^{(m)}). \tag{B.9}$$

Finally, we obtain:

$$\hat{\alpha}_i^{(m+1)} = \left(\sum_{t=1}^{N}\mathcal{F}(z_t, \hat{\beta}_i^{(m)})\right)\left(\sum_{t=1}^{N}\sum_{l=1}^{M}\mathcal{F}(z_t, \hat{\beta}_l^{(m)})\right)^{-1}$$

Notice that $0 \le \hat{\alpha}_i^{(m+1)} \le 1$ holds, even though we did not explicitly consider it in (B.4). $\square$