Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Model Error Modelling using a Stochastic Embedding approach with Gaussian Mixture Models for FIR systems ⋆

Rafael Orellana [∗,∗∗] **Rodrigo Carvajal** [∗] **Juan C. Agüero** [∗]
Graham C. Goodwin [∗∗∗]

[∗] *Electronic Engineering Department, Universidad Técnica Federico Santa María, Valparaíso, Chile.*
[∗∗] *School of Electrical Engineering, Universidad de Los Andes, Mérida, Venezuela.*
[∗∗∗] *School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, Australia.*

Abstract: In this paper a Maximum Likelihood estimation algorithm for error-model modelling using a *stochastic embedding* approach is developed. The error-model distribution is approximated by a finite Gaussian mixture. An Expectation-Maximization based algorithm is proposed to estimate the nominal model and the distribution of the parameters of the error-model by using the data from independent experiments. The benefits of our proposal are illustrated via numerical simulations.

Keywords: Model errors, Stochastic Embedding, Maximum Likelihood, Gaussian Mixture, Expectation-Maximization, Estimation.

1. INTRODUCTION

Most System Identification techniques available in the literature assume that the system lies in the model set when estimating the corresponding vector of parameters (see e.g. Goodwin and Payne (1977); Söderström and Stoica (1988); Ljung (1999)). However real systems have arbitrary complexity. However, using a complex model structure can lead to large variance estimation errors (Ljung et al. (2014)).

An alternative view of Modelling and System Identification of dynamic systems combines a *nominal model* with an *error-model*, i.e., uncertainty modelling is part of the structure. Modelling uncertainty has been addressed in different frameworks, such as Set Membership (Milanese and Vicino (1991)), characterization of model bias (Hakvoort and Van den Hof (1997); Wahlberg and Ljung (1986)), Model Error Modelling (Reinelt et al. (2002); Ljung et al. (2015)), and Stochastic Embedding (Goodwin and Salgado (1989); Goodwin et al. (1992)).

Stochastic Embedding (SE) describes model uncertainty in a stochastic framework. The key idea is to think of the model as a realization drawn from an underlying probability space, where the parameters that define the error-model are characterized by a probability density function (pdf) (see e.g. Goodwin and Salgado (1989); Goodwin et al. (1992)). In this approach the uncertainty can be quantified by using Maximum Likelihood (ML) estimation to obtain the parameters of the error-model. In Delgado et al. (2012); Ljung et al. (2014) an interpretation of SE has been proposed, where the parameters that define the error-model are considered as *latent* variables (also known as *hidden* variables). In this approach, the use of Expectation-Maximization (EM) (Dempster et al. (1977)) based algorithms is proposed. It is also possible to adopt a Bayesian perspective by utilizing a prior distribution for the parameters of both the nominal and error-model (Ljung et al. (2014)). In Ljung et al. (2015) a Gaussian distribution for the vector that defines the error-model is assumed. An explicit expression for the likelihood function is obtained by marginalising over the error-model as a linear regression. There are works closely related with this framework (see e.g. Pillonetto and Nicolao (2010); Ljung et al. (2020)).

In this paper, we focus on the development of an ML estimator for model error modelling for linear dynamic systems using the SE approach. We approximate the probability distribution of the vector of parameters that define the error-model by a Gaussian Mixture Model (GMM). We then propose an estimation algorithm to estimate the vector of parameters that define both the nominal model and the probability density function of the GMM based on the EM algorithm.

GMMs have been typically utilised in non-linear filtering (Anderson and Moore (1979); Söderström (2002)) (see also Wills et al. (2017)). Recently, GMMs have also been used to develop algorithms to identify dynamic systems in the Maximum Likelihood framework (Bittner et al. (2019)), and a Bayesian framework (Sorenson and Alspach (1971); Dahlin et al. (2018); Orellana et al. (2019)). Static systems have also been addressed using finite mixtures models, such as in the deconvolution of stellar rotational velocities (Orellana et al. (2018, 2019)). To the best of our knowledge, utilizing a GMM approximation for unknown or non-Gaussian distributions, has not previously been used for uncertainty modelling in an SE framework.

⋆ This work was partially supported by FONDECYT trough grant No 1181158, PIIC program 05/2019 of DGP at Universidad Técnica Federico Santa María and the Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008, ANID, Chile.

The remainder of the paper is as follows: In Section 2 the problem of interest is stated as a linear regression using the SE approach. In Section 3 the uncertainty modelling problem is addressed using ML estimation with GMM. In Section 4 an EM-based identification algorithm is presented. A numerical simulation example is presented in Section 5. Finally, in Section 6, we present conclusions.

2. STOCHASTIC EMBEDDING ERROR MODEL AS A LINEAR REGRESSION

Consider the following set of linear dynamic systems:

$$y_t^{[r]} = G_T^{[r]}(z)u_t + \omega_t^{[r]}, \tag{1}$$

where $r = 1, \ldots, M$ denotes the r-th realization of the system, M corresponds the number of independent experiments or batches, $G_T^{[r]}(z)$ denotes the *true* system, z denotes the forward shift operator or z-transform variable, $y_t^{[r]}$ denotes the output signal, u_t is the input signal and $\omega_t^{[r]}$ is a white noise sequence with zero mean and variance σ_ω^2. Notice that $\omega_t^{[r]}$ describes a different noise realizations for each independent experiment.

We assume that $G_T(z)^{[r]}$ can be described as follows (see e.g. Goodwin and Salgado (1989); Ljung et al. (2015)):

$$G_T(z)^{[r]} = \begin{cases} G_o(z,\theta) + G_\epsilon(z, \eta(\gamma)^{[r]}), & \text{(2a)} \\ G_o(z,\theta)(1 + G_\Delta(z, \eta(\gamma)^{[r]})), & \text{(2b)} \end{cases}$$

where $G_o(z,\theta)$ is the nominal model parametrized by θ, $G_\epsilon(z, \eta(\gamma)^{[r]})$ is an additive error-model (2a), $G_\Delta(z, \eta(\gamma)^{[r]})$ is a multiplicative error-model (2b) parametrized by $\eta(\gamma)^{[r]}$. Here γ is the vector of parameters that define a given pdf that models η. We consider that all transfer functions are FIR systems. In this particular case, there is a clear connection with Bayesian approaches, since the corresponding parameters of both the nominal and error-model can be considered in a unified model. However, the approach presented in this paper can be extended for more complex model structures.

Remark 1. We assume that the the nominal model $G_o(z,\theta)$ does not change between experiments, whilst the error-model $G_\Delta(z, \eta(\gamma)^{[r]})$ (or $G_\epsilon(z, \eta(\gamma)^{[r]})$) may change for each experiment. In addition, all the realizations of η are drawn from the same pdf parametrized by γ. ▽

We assume that the observed data $Y^{[r]} = [y_1^{[r]} \ \ldots \ y_N^{[r]}]^T$ is a collection of measurements [1] for each experiment. We then obtain a regression model combining (1) and (2a) (or (2b)) as follows:

$$Y^{[r]} = \Phi^{[r]}\theta + \Psi^{[r]}\eta^{[r]} + W^{[r]}, \tag{3}$$

where $Y^{[r]}, W^{[r]} \in \mathbb{R}^{N\times 1}$, $\theta \in \mathbb{R}^{n_o\times 1}$, $\eta^{[r]} \in \mathbb{R}^{n_\Delta \times 1}$, $\Phi^{[r]} \in \mathbb{R}^{N\times n_o}$, $\Psi^{[r]} \in \mathbb{R}^{N\times n_\Delta}$. The term $\Phi^{[r]}\theta$ represents the output response corresponding to the nominal model structure, $\Psi^{[r]}\eta^{[r]}$ corresponds to the output signal due to the error-model related to the structures defined in (2a) or (2b) and $W^{[r]} \in \mathcal{N}(0, \sigma_\omega^2 I_N)$ (is Gaussian white noise with zero mean and covariance matrix $\sigma_w^2 I_N$) [2].

[1] We use capital letters to denote the vector of signals, u_t or ω_t for $t = 1, \ldots N$

[2] I_x represents the identity matrix with dimension given by x.

3. MAXIMUM LIKELIHOOD ESTIMATION FOR MODEL ERROR MODELLING USING GMM

3.1 Using a Gaussian Mixture Model to approximate the error-model distribution

In this paper we approximate the error-model distribution as a GMM. The GMM can be tailored to approximate non-Gaussian distributions, (see e.g. Mengersen et al. (2011) and the references therein). Based on the Wiener approximation theorem, it is known that any pdf with compact support can be approximated by a finite sum of Gaussian distributions (Lo (1972); Achieser (1992)). For completeness of the presentation, the Gaussian sum approximation approach is summarized as follows (See (Lo, 1972, Theorem 3)):

Lemma 2. Any probability density function, $p(\eta|\gamma)$, of an n dimensional random variable η with compact support can be approximated as closely as desired in the space $L_1(\mathbb{R}^n)$ by a distribution of the form

$$p(\eta|\gamma) \approx \sum_{j=1}^{\kappa} \lambda_j \phi\left(\eta; \mu_j, \Gamma_j\right), \tag{4}$$

where $\lambda_j > 0$, $\sum_{j=1}^{\kappa} \lambda_j = 1$ and $\phi\left(\eta; \mu_j, \Gamma_j\right)$ represents an n dimensional Gaussian distribution with mean μ_j and covariance matrix Γ_j.

3.2 Likelihood function for GMM

For the system in (3), we define the vector of parameters to be estimated as $\beta = [\theta^T \ \gamma^T \ \sigma_\omega^2]^T$, where

$$\gamma = [\underbrace{\lambda_1 \ \mu_1 \ \Gamma_1}_{\gamma_1} \ \ldots \ \underbrace{\lambda_\kappa \ \mu_\kappa \ \Gamma_\kappa}_{\gamma_\kappa}]^T.$$

We let β_0 be the *true* vector of parameters. For the model (3) using the GMM in (4), the likelihood function, $\mathcal{L}(\beta)$, can be obtained by marginalizing the *hidden* variable, $[\eta^{[1]} \ \ldots \ \eta^{[M]}]$, as follows [3]:

$$\mathcal{L}(\beta) = p(Y^{[1]}, \ldots, Y^{[M]}|\beta) \tag{5}$$

$$= \prod_{r=1}^{M} \int_{-\infty}^{\infty} p(Y^{[r]}|\eta^{[r]}, \beta)p(\eta^{[r]}|\beta)d\eta^{[r]}, \tag{6}$$

where

$$p(Y^{[r]}|\eta^{[r]}, \beta) = \phi(Y^{[r]}; \Phi^{[r]}\theta + \Psi^{[r]}\eta^{[r]}, \sigma_\omega^2 I_N). \tag{7}$$

Considering that $p(\eta^{[r]}|\beta)$ in (6) is a GMM, then the log-likelihood function is given by

$$\ell(\beta) = \sum_{r=1}^{M} \log\left\{ \sum_{j=1}^{\kappa} \int_{-\infty}^{\infty} \lambda_j p(Y^{[r]}|\eta^{[r]}, \beta)\phi(\eta^{[r]}; \mu_j, \Gamma_j)d\eta^{[r]} \right\}. \tag{8}$$

The ML estimator is then given by

$$\hat{\beta}_{\mathrm{ML}} = \arg\max_{\beta} \ \ell(\beta) \quad \text{s.t.} \quad 0 \leq \lambda_j \leq 1, \ \sum_{j=1}^{\kappa} \lambda_j = 1. \tag{9}$$

Remark 3. We assume that the vector of parameters β_0, the input u_t and the noise ω_t satisfy regularity conditions, guaranteeing that the solution $\hat{\beta}_{\mathrm{ML}}$ of the optimization problem in (9) converges (in probability or a.s.) to the

[3] Typical methods of ML estimation for GMM do not consider the presence of *hidden* variables.

true solution β_0 as $N \to \infty$. $\triangledown$

The solution of the optimization problem in (9) may be difficult to obtain when the number of components in the GMM increases, due to the fact that $\ell(\beta)$ may exhibit several local maxima and the size of the parameter space tends to be large [4] (see e.g. Jin et al. (2016)). For simplicity, we focus on the case where the number of components is known.

4. AN ALGORITHM FOR MODEL ERROR MODELLING USING GMM

The EM algorithm is a popular tool for identifying linear and non-linear dynamic systems in the time domain (see e.g. Gibson et al. (2005); Gopaluni (2008)) and frequency domain (Agüero et al. (2012)). In this section we will show how an EM-based estimation algorithm can be developed to solve the problem of interest.

4.1 EM based algorithm formulation

From (7) and (8), we define the following:

$$K(\beta,\eta^{[r]}) = \lambda_j \phi(Y^{[r]};\Phi^{[r]}\theta + \Psi^{[r]}\eta^{[r]}, \sigma_\omega^2 I_N)\phi(\eta^{[r]};\mu_j,\Gamma_j). \tag{10}$$

Then, the log-likelihood function in (8) can be expressed as

$$\ell(\beta) = \sum_{r=1}^{M} \log\left[\mathcal{V}^{[r]}(\beta)\right], \tag{11}$$

with

$$\mathcal{V}^{[r]}(\beta) = \sum_{j=1}^{\kappa}\int_{-\infty}^{\infty} K(\beta,\eta^{[r]})d\eta^{[r]}. \tag{12}$$

The expression in (12) can be written as $\log\big(\mathcal{V}^{[r]}(\beta)\big) = \mathcal{Q}^{[r]}(\beta,\hat{\beta}^{(m)}) - \mathcal{H}^{[r]}(\beta,\hat{\beta}^{(m)})$, where:

$$\mathcal{Q}^{[r]}(\beta,\hat{\beta}^{(m)}) = \sum_{j=1}^{\kappa}\int_{-\infty}^{\infty}\log[K(\beta,\eta^{[r]})]\frac{K(\hat{\beta}^{(m)},\eta^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})}d\eta^{[r]}. \tag{13}$$

$$\mathcal{H}^{[r]}(\beta,\hat{\beta}^{(m)}) = \sum_{j=1}^{\kappa}\int_{-\infty}^{\infty}\log\left[\frac{K(\beta,\eta^{[r]})}{\mathcal{V}^{[r]}(\beta)}\right]\frac{K(\hat{\beta}^{(m)},\eta^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})}d\eta^{[r]}. \tag{14}$$

The function $\mathcal{H}^{[r]}(\beta,\hat{\beta}^{(m)})$ is a decreasing function for any value of β (see e.g. Carvajal et al. (2018)).

In order to develop the estimation algorithm, we first obtain the following result. This will be used to compute the integral in (13):

Lemma 4. The expression in (10) evaluated at the m-th estimate, $\hat{\beta}^{(m)}$, can be rewritten as follows:

$$K(\hat{\beta}^{(m)},\eta^{[r]}) = \phi(Y^{[r]};m_Y^{(j)},\Sigma_Y^{(j)})\phi(\eta^{[r]};m_{\eta^{[r]}}^{(j)},\Sigma_{\eta^{[r]}}^{(j)}), \tag{15}$$

with

$$m_Y^{(j)} = \Psi^{[r]}\hat{\mu}_j^{(m)} + \Phi^{[r]}\hat{\theta}^{(m)}, \tag{16}$$

$$\Sigma_Y^{(j)} = [\hat{\sigma}_\omega^2]^{(m)} I_N + \Psi^{[r]}\hat{\Gamma}_j^{(m)}\Psi^{[r]^T}, \tag{17}$$

$$Z^{(j)} = \hat{\Gamma}_j^{(m)}\Psi^{[r]^T}\left(\Sigma_Y^{(j)}\right)^{-1}, \tag{18}$$

$$m_{\eta^{[r]}}^{(j)} = \hat{\mu}_j^{(m)} + Z^{(j)}\left(Y^{[r]} - m_Y^{(j)}\right), \tag{19}$$

$$\Sigma_{\eta^{[r]}}^{(j)} = \left(I_{n_\Delta} - Z^{(j)}\Psi^{[r]}\right)\hat{\Gamma}_j^{(m)}. \tag{20}$$

Proof. The result is directly obtained from (10) by using the following identities:

$$\begin{bmatrix} A & B\\ C & D\end{bmatrix} = \begin{bmatrix} I & 0\\ CA^{-1} & I\end{bmatrix}\begin{bmatrix} A & 0\\ 0 & D-CA^{-1}B\end{bmatrix}\begin{bmatrix} I & A^{-1}B\\ 0 & I\end{bmatrix}, \tag{21}$$

$$(I+BD^{-1}C)^{-1} = I - B(D+CB)^{-1}C, \tag{22}$$

$$\det(A)\det\big(D-CA^{-1}B\big) = \det(D)\det\big(A-BD^{-1}C\big). \tag{23}$$

Using this result, we can express $\log[K(\beta,\eta^{[r]})]$ as follows:

$$\begin{aligned}\log[K(\beta,\eta^{[r]})] &= \log[\lambda_j] - \frac{N}{2}\log[\sigma_\omega^2] - \frac{n_\Delta}{2}\log[\det(\Gamma_j)] - \\ &\frac{1}{2\sigma_\omega^2}\Big[(Y^{[r]} - \Phi^{[r]}\theta - \Psi^{[r]}m_{\eta^{[r]}}^{(j)})^T(Y^{[r]} - \Phi^{[r]}\theta - \Psi^{[r]}m_{\eta^{[r]}}^{(j)}) \\ &+ \mathrm{tr}\Big(\Psi^{[r]}\Sigma_{\eta^{[r]}}^{(j)}\Psi^{[r]^T}\Big)\Big] - \frac{1}{2}\Big[\mathrm{tr}\Big(\Gamma_j^{-1}\Sigma_{\eta^{[r]}}^{(j)}\Big) + \\ &(m_{\eta^{[r]}}^{(j)} - \mu_j)^T\Gamma_j^{-1}(m_{\eta^{[r]}}^{(j)} - \mu_j)\Big]. \end{aligned} \tag{24}$$

$\square$

From Lemma 4, the auxiliary function in (13) can be expressed as

$$\mathcal{Q}^{[r]}(\beta,\hat{\beta}^{(m)}) = \sum_{j=1}^{\kappa}\log[K(\beta,\eta^{[r]})]\frac{\mathcal{F}_j(Y^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})}. \tag{25}$$

with

$$\mathcal{F}_j(Y^{[r]}) = \hat{\lambda}_j^{(m)}\phi(Y^{[r]};m_Y^{(j)},\Sigma_Y^{(j)}). \tag{26}$$

Finally, we can formulate the following iterative algorithm:

$$\bar{\mathcal{Q}}(\beta,\hat{\beta}^{(m)}) = \sum_{r=1}^{M}\mathcal{Q}^{[r]}(\beta,\hat{\beta}^{(m)}), \tag{27}$$

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

where (28) is solved subject to $\sum_{j=1}^{\kappa}\lambda_j = 1$, $0 \le \lambda_j \le 1$. Notice that (27) and (28) are closely related to the E-step and M-step of the EM algorithm, respectively.

4.2 Optimization of the auxiliary function

For the optimization of the auxiliary function $\bar{\mathcal{Q}}(\beta,\hat{\beta}^{(m)})$ in (27), we can obtain closed-form expressions for the estimate of β. Specifically, the optimization with respect to β can be carried out as described below.

Lemma 5. The vector of parameters $\hat{\beta}$ that optimizes the auxiliary function $\bar{\mathcal{Q}}(\beta,\hat{\beta}^{(m)})$ in (27) with respect to β is given by:

$$\hat{\mu}_j^{(m+1)} = \mathcal{M}_j(Y,\hat{\beta}^{(m)})/\mathcal{P}_j(Y,\hat{\beta}^{(m)}), \tag{29}$$

$$\hat{\Gamma}_j^{(m+1)} = \mathcal{S}_j(Y,\hat{\beta}^{(m)})/n_\Delta\mathcal{P}_j(Y,\hat{\beta}^{(m)}), \tag{30}$$

$$\hat{\lambda}_j^{(m+1)} = \mathcal{P}_j(Y,\hat{\beta}^{(m)})/\sum_{l=1}^{\kappa}\mathcal{P}_l(Y,\hat{\beta}^{(m)}), \tag{31}$$

$$\begin{aligned}\hat{\theta}^{(m+1)} = &\left[\sum_{r=1}^{M}\sum_{j=1}^{\kappa}\left(\Phi^{[r]^T}\Phi^{[r]}\right)^{-1}\Phi^{[r]^T}\mathcal{F}_j(Y)\left(Y^{[r]} - \right.\right. \\ &\left.\left.\Psi^{[r]}m_{\eta^{[r]}}^{(j)}\right)/\mathcal{V}^{[r]}(\hat{\beta}^{(m)})\right]\Big/\left[\sum_{l=1}^{\kappa}\mathcal{P}_l(Y,\hat{\beta}^{(m)})\right],\end{aligned} \tag{32}$$

[4] Each additional component of the GMM increases the search space by $1 + n_\Delta + n_\Delta(n_\Delta - 1)/2$.

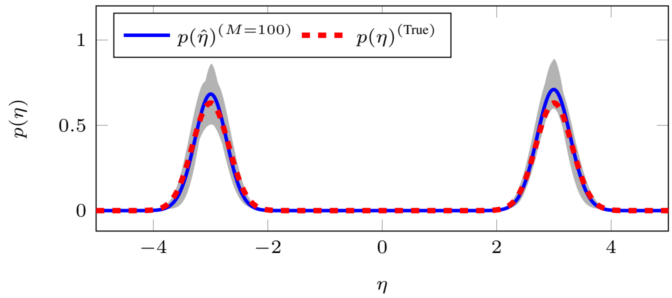

Fig. 1. Estimation of the error-model distribution $p(\eta)$ for $M = 100$ experiments.

$$[\hat{\sigma}^2_\omega]^{(m+1)} = \frac{\sum_{r=1}^{M}\sum_{j=1}^{\kappa}\mathcal{F}_j(Y^{[r]})\mathcal{B}_j(Y^{[r]})/\mathcal{V}^{[r]}(\hat{\beta}^{(m)})}{N\sum_{l=1}^{\kappa}\mathcal{P}_l(Y,\hat{\beta}^{(m)})}, \tag{33}$$

with

$$\mathcal{P}_j(Y,\hat{\beta}^{(m)}) = \sum_{r=1}^{M}\mathcal{F}_j(Y^{[r]})/\mathcal{V}^{[r]}(\hat{\beta}^{(m)}), \tag{34}$$

$$\mathcal{M}_j(Y,\hat{\beta}^{(m)}) = \sum_{r=1}^{M} m^{(j)}_{\eta^{[r]}}\mathcal{F}_j(Y^{[r]})/\mathcal{V}^{[r]}(\hat{\beta}^{(m)}), \tag{35}$$

$$\mathcal{S}_j(Y,\hat{\beta}^{(m)}) = \sum_{r=1}^{M}\Big[(m^{(j)}_{\eta^{[r]}} - \hat{\mu}^{(m)}_j)(m^{(j)}_{\eta^{[r]}} - \hat{\mu}^{(m)}_j)^T + \Sigma^{(j)}_{\eta^{[r]}}\Big]\mathcal{F}_j(Y^{[r]})/\mathcal{V}^{[r]}(\hat{\beta}^{(m)}), \tag{36}$$

$$\mathcal{B}_j(Y^{[r]}) = (Y^{[r]} - \Phi^{[r]}\hat{\theta}^{(m)} - \Psi^{[r]}m^{(j)}_{\eta^{[r]}})^T (Y^{[r]} - \Phi^{[r]}\hat{\theta}^{(m)} - \Psi^{[r]}m^{(j)}_{\eta^{[r]}}) + \mathrm{tr}\Big(\Psi^{[r]}\Sigma^{(j)}_{\eta^{[r]}}\Psi^{[r]^T}\Big). \tag{37}$$

Proof. See Appendix A. □

We summarize our algorithm as follows:

(i) Fix the number of experiments, M, and components κ for the GMM.
(ii) Choose an initial guess $\hat{\beta}^{(0)}$, and set $m = 0$.
(iii) Estimate the GMM parameters from (29)-(31).
(iv) Compute the nominal model parameters and the noise variance using (32) and (33), respectively.
(v) Set $m = m + 1$ and go back to step (iii) until a stopping criterion is satisfied.

5. NUMERICAL EXAMPLE

In this section we present a numerical example to illustrate the performance of the proposed algorithm. This example is a variant of the example used in Ljung et al. (2015) for a simple FIR model. Consider (1) with relative error-model (2b) as follows:

$$G_o(z,\theta) = g_0 + g_1 z^{-1}, \tag{38}$$

$$G_\Delta(z,\eta) = \eta_0 - \eta_0 z^{-1}, \tag{39}$$

where the *true* (but unknown) value of θ is $\theta_0 = [1 \;\; 0.5]^T$, $u_t \sim \mathcal{N}(0,\sigma_u^2)$, $\sigma_u^2 = 10$, $\sigma_\omega^2 = 0.1$ and η_0 is a *hidden* variable. In this example, we consider that in each experiment, $\eta = \eta_0$ is drawn from a finite Gaussian mixture distribution given by [5] :

[5] We have run several numerical examples with different error-model distributions. However, for illustrations purposes we focus on a simple case.

Table 1. Nominal model parameter and noise variance estimates

Param/	$M=10$		$M=100$	
Experiments	$\kappa=1$	$\kappa=2$	$\kappa=1$	$\kappa=2$
g_0	1.087 ± 0.161	1.001 ± 0.024	1.039 ± 0.119	0.999 ± 0.002
g_1	0.461 ± 0.166	0.514 ± 0.030	0.483 ± 0.123	0.500 ± 0.002
σ_w^2	17.301 ± 19.584	0.322 ± 0.549	0.101 ± 0.030	0.101 ± 0.002

$$p(\eta)^{(\mathrm{True})} = \lambda_1\phi(\eta;\mu_1,\Gamma_1) + \lambda_2\phi(\eta;\mu_2,\Gamma_2), \tag{40}$$

with *true* (but unknown) parameters $\lambda_1 = \lambda_2 = 0.5$, $\Gamma_1 = \Gamma_2 = 0.1$, $\mu_1 = -3$ and $\mu_2 = 3$.

The simulation setup is as follows:

(1) The data length is $N = 100$.
(2) We consider M different independent experiments where $M = \{10, 100\}$.
(3) The number of Monte Carlo (MC) [6] simulation is 100.
(4) The stopping criterion is chosen as:
$$\left\|\hat{\beta}^{(m)} - \hat{\beta}^{(m-1)}\right\| / \left\|\hat{\beta}^{(m)}\right\| < 10^{-6},$$
or when 500 iterations have been reached.

Table 1 shows the estimation results of the nominal model parameters for different number of experiments ($M = \{10, 100\}$). Fig. 1 shows the results of the estimation of the error-model distribution. The gray-shaded area corresponds to the region in which the corresponding pdf of all the estimates from the MC simulations lie. The blue line corresponds to the average of all the estimates with $M = 100$. It is clear that the estimated pdf is similar to the *true* pdf when the number of experiments is high ($M = 100$). We compare our results, with the estimation obtained utilizing the SE approach proposed in Ljung et al. (2015). To that end, we marginalize the relative error-model in (2b) and assume a single Gaussian distribution ($\kappa = 1$) for the error-model vector η. Fig. 2 shows the magnitude and phase of the frequency response corresponding to the average of all MC simulations for the estimated nominal model with $M = 10$. The gray-shaded region represents the area in which all the estimated nominal models lie. We observe a small difference between the estimated frequency response using our proposed algorithm and the nominal model $G_o(z)$. These results confirm the benefits of obtaining an accurate estimation of the error-model distribution.

6. CONCLUSION

In this paper we have addressed the model error modelling problem by combining the SE approach with GMMs. We proposed an identification algorithm to estimate the nominal model and the error-model distribution as a GMM by using Maximum Likelihood. The proposed algorithm utilizes closed-form expressions to estimate parameters of the nominal model, noise variance and GMM. Our proposal exhibits good accuracy, specially when the number of experiments is high.

REFERENCES

Achieser, N.I. (1992). *Theory of Approximation.* Dover Publications.

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.

[6] Each MC simulation corresponds to an estimation obtained from M independent experiments.

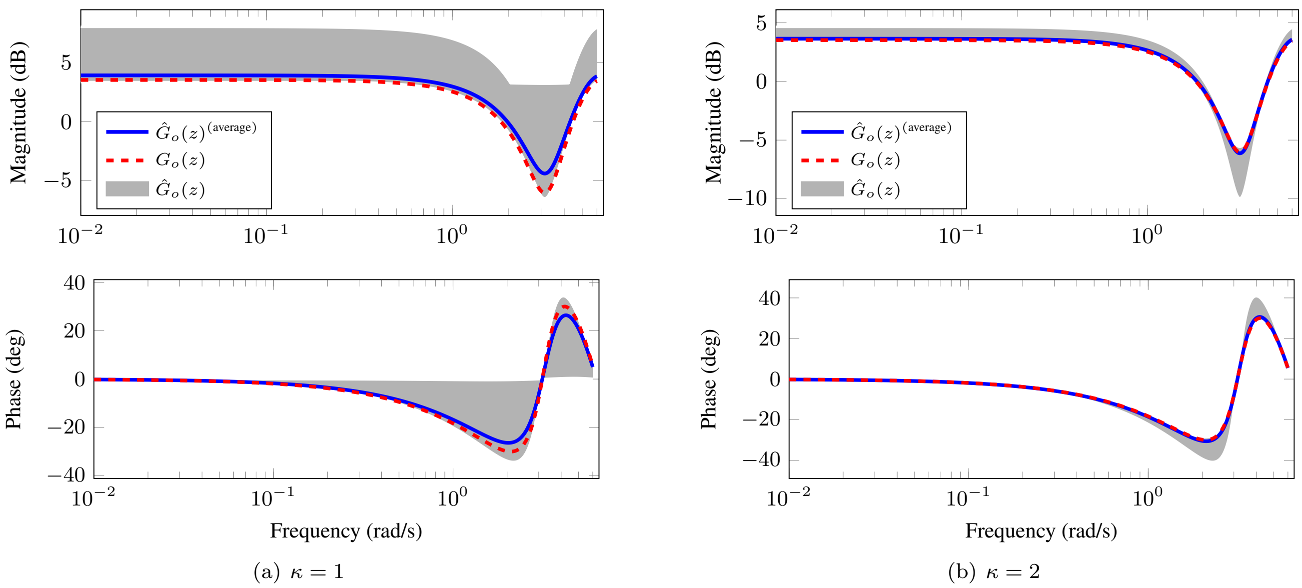

Fig. 2. Bode plots for the estimated nominal system using $p(\hat{\eta})$ obtained with $M = 10$.

Anderson, B. and Moore, J. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

Bittner, G., Orellana, R., Carvajal, R., and Agüero, J.C. (2019). 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)*, 1–7.

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.

Dahlin, J., Wills, A., and Ninness, B. (2018). Sparse Bayesian ARX models with flexible noise distributions. *IFAC-PapersOnLine*, 51(15), 25 – 30. 18th IFAC Symposium on System Identification SYSID.

Delgado, R.A., Goodwin, G.C., Carvajal, R., and Agüero, J.C. (2012). A novel approach to model error modelling using the expectation-maximization algorithm. In *51st IEEE Conference on Decision and Control*, 7327–7332.

Dempster, A.P., Laird, N.M., and Rubin, D. (1977). Royal Statistical Society. *Royal Statistical Society*, 39(1), 1–38.

Gibson, S., Wills, A., and Ninness, B. (2005). Maximum-likelihood parameter estimation of bilinear systems. *IEEE Transactions on Automatic Control*, 50(10), 1581–1596.

Goodwin, G.C., Gevers, M., and Ninness, B. (1992). Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Transactions on Automatic Control*, 37(7), 913–928.

Goodwin, G.C. and Payne, R.L. (1977). *Dynamic system identification : experiment design and data analysis*. Academic Press New York.

Goodwin, G.C. and Salgado, M.E. (1989). A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. *International Journal of Adaptive Control and Signal Processing*, 3(4), 333–356.

Gopaluni, R.B. (2008). A particle filter approach to identification of nonlinear processes under missing observations. *The Canadian Journal of Chemical Engineering*, 86(6), 1081–1092.

Hakvoort, R. and Van den Hof, M. (1997). Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Transactions on Automatic Control*, 42(11), 1516–1528.

Jin, C., Zhang, Y., Balakrishnan, S., Wainwright, M.J., and Jordan, M.I. (2016). Local maxima in the likelihood of gaussian mixture models: Structural results and algorithmic consequences. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4123–4131.

Ljung, L., Goodwin, G., Agüero, J., and Chen, T. (2015). Model Error Modeling and Stochastic Embedding. *IFAC-PapersOnLine*, 48(28), 75 – 79. 17th IFAC Symposium on System Identification.

Ljung, L., Goodwin, G.C., and Agüero, J.C. (2014). Stochastic Embedding revisited: A modern interpretation. In *53rd IEEE Conference on Decision and Control*, 3340–3345.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

Ljung, L., Chen, T., and Mu, B. (2020). A shift in paradigm for system identification. *International Journal of Control*, 93(2), 173–180.

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

Mengersen, K., Robert, C., and Titterington, M. (2011). *Mixtures: Estimation and Applications*. Wiley.

Milanese, M. and Vicino, A. (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6), 997 – 1009.

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

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. 18th IFAC Symposium on System Identification.

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

Pillonetto, G. and Nicolao, G.D. (2010). A new kernel-based approach for linear system identification. *Automatica*, 46(1), 81–93.

Reinelt, W., Garulli, A., and Ljung, L. (2002). Comparing Different Approaches to Model Error Modeling in Robust Identification. *Automatica*, 38(5), 787–803.

Söderström, T. and Stoica, P. (eds.) (1988). *System Identification*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Söderström, T. (2002). *Discrete-Time Stochastic Systems: Estimation and Control*. Springer-Verlag, Berlin, Heidelberg, 2nd edition.

Sorenson, H.W. and Alspach, D.L. (1971). Recursive bayesian estimation using gaussian sums. *Automatica*, 7(4), 465–479.

Wahlberg, B. and Ljung, L. (1986). Design variables for bias distribution in transfer function estimation. *IEEE Transactions on Automatic Control*, 31(2), 134–144.

Wills, A.G., Hendriks, J., Renton, C., and Ninness, B. (2017). A Bayesian Filtering Algorithm for Gaussian Mixture Models. *ArXiv e-prints*.

Appendix A. COMPUTING THE PARAMETERS OF THE GMM

Taking the derivative of (27) with respect to μ_j and equating to zero yields:

$$\frac{\partial \bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)})}{\partial \mu_j} = \sum_{r=1}^{M} \left[\Gamma_j^{-1} (m_{\eta^{[r]}}^{(j)} - \hat{\mu}_j^{(m+1)}) \right] \frac{\mathcal{F}_j(Y^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} = 0 \tag{A.1}$$

Using (34) and (35) we obtain:

$$\sum_{r=1}^{M} m_{\eta^{[r]}}^{(j)} \frac{\mathcal{F}_j(Y^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} = \hat{\mu}_j^{(m+1)} \sum_{r=1}^{M} \frac{\mathcal{F}_j(Y^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})}, \tag{A.2}$$

$$\hat{\mu}_j^{(m+1)} = \mathcal{M}_j(Y, \hat{\beta}^{(m)}) / \mathcal{P}_j(Y, \hat{\beta}^{(m)}). \tag{A.3}$$

Then, taking the derivative of (27) with respect to Γ_j^{-1} and equating to zero:

$$\frac{\partial \bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)})}{\partial \Gamma_j^{-1}} = \sum_{r=1}^{M} \left(\left[(m_{\eta^{[r]}}^{(j)} - \hat{\mu}_j^{(m)})(m_{\eta^{[r]}}^{(j)} - \hat{\mu}_j^{(m)})^T + \Sigma_{\eta^{[r]}}^{(j)} \right] + n_\Delta \hat{\Gamma}_j^{(m+1)} \right) \frac{\mathcal{F}_j(Y^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} = 0. \tag{A.4}$$

Using (34) and (36) we obtain:

$$\hat{\Gamma}_j^{(m+1)} = \mathcal{S}_j(Y, \hat{\beta}^{(m)}) / n_\Delta \mathcal{P}_j(Y, \hat{\beta}^{(m)}). \tag{A.5}$$

Similarly, using (26) and taking the derivative of (27) with respect to θ and equating to zero yields:

$$\sum_{r=1}^{M} \sum_{j=1}^{\kappa} \mathcal{F}_j(Y^{[r]}) \Phi^{[r]^T} (Y^{[r]} - \Psi^{[r]} m_{\eta^{[r]}}^{(j)}) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}) = \sum_{r=1}^{M} \sum_{j=1}^{\kappa} \mathcal{F}_j(Y^{[r]}) \Phi^{[r]^T} \Phi^{[r]} \hat{\theta}^{(m+1)} / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}). \tag{A.6}$$

Utilizing (34) we obtain:

$$\hat{\theta}^{(m+1)} = \left[\sum_{r=1}^{M} \sum_{j=1}^{\kappa} \left(\Phi^{[r]^T} \Phi^{[r]} \right)^{-1} \Phi^{[r]^T} \mathcal{F}_j(Y) \left(Y^{[r]} - \Psi^{[r]} m_{\eta^{[r]}}^{(j)} \right) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}) \right] / \left[\sum_{l=1}^{\kappa} \mathcal{P}_l(Y, \hat{\beta}^{(m)}) \right]. \tag{A.7}$$

Then, using (26), (37) and taking the derivative of (27) with respect to $\alpha = 1/\sigma_\omega^2$ and equating to zero:

$$\sum_{r=1}^{M} \sum_{j=1}^{\kappa} N \mathcal{F}_j(Y^{[r]}) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}) \hat{\alpha}^{(m+1)} = \sum_{r=1}^{M} \sum_{j=1}^{\kappa} \mathcal{F}_j(Y^{[r]}) \mathcal{B}_j(Y^{[r]}) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}). \tag{A.8}$$

Thus, we obtain:

$$[\hat{\sigma}_\omega^2]^{(m+1)} = \frac{\sum_{r=1}^{M} \sum_{j=1}^{\kappa} \mathcal{F}_j(Y^{[r]}) \mathcal{B}_j(Y^{[r]}) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)})}{N \sum_{l=1}^{\kappa} \mathcal{P}_l(Y, \hat{\beta}^{(m)})}. \tag{A.9}$$

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

$$\mathcal{R}(\lambda_j) = \sum_{r=1}^{M} \sum_{j=1}^{\kappa} \log[\lambda_j] \left\{ \mathcal{F}_j(Y^{[r]}) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}) \right\}, \tag{A.10}$$

subject to $\sum_{j=1}^{\kappa} \lambda_j = 1$. Notice that, we initially do not consider the constraint $0 \leq \lambda_j \leq 1$. Then, using a Lagrange multiplier to deal with the constraint on λ_j we define:

$$\mathcal{G}(\lambda_j, \zeta) = \sum_{r=1}^{M} \sum_{j=1}^{\kappa} \log[\lambda_j] \left\{ \mathcal{F}_j(Y^{[r]}) / \mathcal{V}^{[r]}(\hat{\beta}^{(m)}) \right\} - \zeta(\sum_{j=1}^{\kappa} \lambda_j - 1). \tag{A.11}$$

Using (34) and taking the derivative of (A.11) with respect to λ_j and ζ and equating to zero we obtain:

$$\frac{\partial \mathcal{G}(\lambda_j, \zeta)}{\partial \lambda_j} = \left(\mathcal{P}_j(Y, \hat{\beta}^{(m)}) / \hat{\lambda}_j^{(m+1)} \right) - \zeta = 0. \tag{A.12}$$

$$\frac{\partial \mathcal{G}(\lambda_j, \zeta)}{\partial \zeta} = \sum_{j=1}^{\kappa} \lambda_j - 1 = 0. \tag{A.13}$$

Then, $\hat{\lambda}_j^{(m+1)} = \mathcal{P}_j(Y, \hat{\beta}^{(m)}) / \zeta$. Taking a summation over $j = 1 \ldots \kappa$ and using (A.13) we have:

$$\sum_{j=1}^{\kappa} \hat{\lambda}_j^{(m+1)} = \sum_{j=1}^{\kappa} \mathcal{P}_j(Y, \hat{\beta}^{(m)}) / \zeta = 1. \tag{A.14}$$

Finally, we obtain:

$$\hat{\lambda}_j^{(m+1)} = \mathcal{P}_j(Y, \hat{\beta}^{(m)}) / \sum_{l=1}^{\kappa} \mathcal{P}_l(Y, \hat{\beta}^{(m)}). \tag{A.15}$$

Notice that $0 \leq \hat{\lambda}_j^{(m+1)} \leq 1$ holds, even though we did not explicitly consider it in (A.11). □