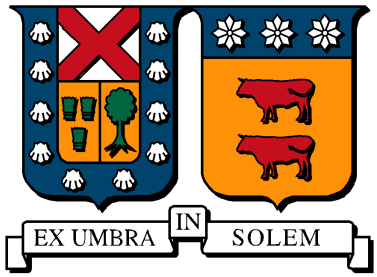
2021

A STOCHASTIC EMBEDDING APPROACH FOR LINEAR SYSTEM UNCERTAINTY MODELING UTILIZING FINITE MIXTURE DISTRIBUTIONS

ORELLANA, RAFAEL



A Stochastic Embedding Approach for Linear System Uncertainty Modeling Utilizing Finite Mixture Distributions

by

Rafael Orellana, BSc. MSc.

A thesis submitted in fulfillment of the

requirements for the degree of

Doctorate in Electronics Engineering

Supervised by
A/Prof. Juan C. Agüero

Electronic Engineering Department
Universidad Técnica Federico Santa María

Valparaíso, Chile

2021

To my wife Maria, and our child Jorge Alejandro.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Prof. Juan C. Agüero for the continuous support during my Ph.D. studies and research, for his patience, motivation, enthusiasm, and really immense knowledge. His guidance helped me through my research and writing of this thesis. I could have not imagined having a better supervisor and mentor for my Ph.D. studies. Besides my supervisor, I would like to thank to Dr. Rodrigo Carvajal for his essential support, academic and personal, since my arrival from Venezuela to Chile. Without his constant help this dissertation would have not been possible. I would also like to thank Dr. Pedro Escárate for giving me the possibility to apply this work through the area of astronomy and astrophysics, I have learned a lot from you.

I acknowledge Universidad Técnica Federico Santa Maria for financial support of my studies and research through the fellowship Beca Doctorado USM. I also acknowledge the Advanced Center of Electrical and Electronic Engineering, the DPP and the Office of International Affairs (OAI), for your support and help during my postgraduate studies.

I would also like to thank my friends and postgraduate students with whom I shared Lab Fourier at USM. I want to thank Luis Mora, Angel Cedeño, Claudia Sánchez, Gustavo Bittner, José Rojas, Marco Gordon, Javier Gonzalez and my beloved wife Maria Coronel. They all contributed in their own way to achieving this goal. Thank you guys.

To our friends in Chile, Maryori, Emiro, Lunita, Diana and Steffy. Thank you for your true friendship and all the fantastic moments we shared with you.

This thesis is dedicated to my parents Nidia and Rafael who taught me that only with dedication, hard work and faith we can construct our future. Thank you for your advice and love since my childhood, I love you. I would like to thank my brother Juan Carlos. Thank you for his support and company during the time that you lived in Chile.

Last but not least, I would like to thank the most important people to me. My beloved wife Maria. We started this adventure together and without your support I could have never achieved all of this, I love you. To my favorite person, my little boy Jorge Alejandro. Thank you for coming into my life, you are all that I need to start a new day every day, I love you Jorgito.

Abstract

This thesis is concerned with modeling and identification of linear dynamic systems with uncertainties. In order to adequately represent the behavior of real processes, identified models must include a quantification of their uncertainties. This thesis presents identification methods where the uncertainty model is assumed to be a linear filter and its corresponding parameters are random variables. This modeling framework is known in the literature as *Stochastic Embedding* and it has been traditionally developed under the assumption that the parameters that define the uncertainty are Gaussian distributed latent variables. In this thesis we extend this framework in order to consider that the distribution of the parameters that define the uncertainty is an unknown non-Gaussian distribution, which is approximated by a Gaussian mixture. Thus, in the first part of this thesis, we develop identification methodologies for systems that consider non-Gaussian latent variables. In the second part of this thesis, we develop identification methodologies driven by non-Gaussian noise. Based on these results, we develop identification methodologies in a *Stochastic Embedding* framework in the third part of this thesis.

The first part of the thesis addresses an estimation problem where latent variables are modeled as a Gaussian Mixture Model, and Maximum Likelihood estimation algorithms with data augmentation approach are developed. The associated estimation problem is solved iteratively based on the Expectation-Maximization algorithm. To address the uncertainty modeling estimation problem, we firstly present a Maximum Likelihood estimation algorithm expressing the likelihood function in terms of an infinite mixture utilizing a Gaussian Mixture Model. A systematic procedure to construct an auxiliary function with finite Gaussian mixtures is developed in order to solve the expectation step of the iterative algorithm. This approach provides closed form expressions for the Gaussian mixture estimators in the maximization step. In addition, the proposed estimation procedure is used to address the problem of stellar rotational velocities estimation with a finite mixture of Maxwellian distributions. This approach provides an accurate estimation of the mixture distribution and closed form expressions for the estimators of the proposed iterative algorithm.

The second part of the thesis addresses systems in which the measurements are subject to non-Gaussian noise. We focus on linear dynamic systems with non-minimum-phase noise transfer function and a Gaussian mixture noise distribution. An iterative algorithm with Gaussian Mixture Models is developed based on the corresponding prediction error. The prediction error is obtained using anti-causal and causal filtering techniques. The proposed iterative algorithm also provides closed form expressions for the Gaussian mixture estimators of the noise distribution. In general, the simulation results show accurate estimations of the system model parameters when the stochastic description is appropriately modeled as a finite Gaussian mixture, even if the non-Gaussian distribution does not correspond to a finite Gaussian mixture but can be approximated by one.

Finally, in order to consider structural and parametric uncertainty quantification, a *Stochastic Embedding* framework is used. In this approach, a nominal model and a Gaussian mixture distributed error-model are considered. The error-model parameters are selected as

hidden variables and the likelihood function is obtained by marginalizing the *hidden* variables. We jointly use the measurements of independent experiments to solve the expectation step of the iterative algorithm. The proposed algorithm results in closed form expressions for the Gaussian mixture estimators of the error-model distribution. This framework provides an accurate estimation of both the nominal model and the error-model distribution when the number of experiments is high. Moreover, this identification approach shows similar performance for modeling error-model in continuous-time linear systems utilizing sampled data.

Key words: Maximum Likelihood, Stochastic Embedding, Gaussian Mixture Models, Non-minimum-phase noise transfer function, Expectation-Maximization.

Resumen

En esta tesis se aborda el problema de modelado e identificación de sistemas dinámicos lineales con incertidumbres. Con el propósito de representar adecuadamente el comportamiento de procesos reales, es necesario que los modelos estimados incorporen un modelo de la incertidumbre. En este trabajo se desarrollan metodologías de identificación en donde el modelo de la incertidumbre es considerado como un filtro lineal cuyos parámetros son variables estocásticas. Este enfoque de modelado es conocido en la literatura como *Stochastic Embedding*, donde típicamente ha sido usado considerando que los parámetros que definen el modelo de la incertidumbre son variables latentes con distribución de probabilidad gaussiana. En esta tesis se hace una extensión de este paradigma de modelado, considerando que la distribución de los parámetros que definen la incertidumbre es desconocida y puede ser aproximada por una suma finita de distribuciones gaussianas.

En la primera parte de esta tesis se considera un problema de estimación en donde las variables latentes son modeladas como una suma de distribuciones gaussianas, desarrollando algoritmos de estimación por máxima verosimilitud con el enfoque de datos aumentados. El problema de estimación se resuelve de forma iterativa usando un algoritmo basado en el algoritmo que maximiza la esperanza. Con el objetivo de poder abordar el problema de estimación de modelos con incertidumbre, se desarrolla un algoritmo de estimación por máxima verosimilitud en donde la función de verosimilitud se expresa como una mezcla infinita de distribuciones utilizando suma de distribuciones gaussianas. Se desarrolla un procedimiento sistemático para construir una función auxiliar con una suma de distribuciones gaussianas con el objetivo de desarrollar un algoritmo iterativo basado en maximizar la esperanza. Bajo este enfoque, al maximizar la función auxiliar, se puede obtener expresiones cerradas para los estimadores de la suma de distribuciones gaussianas. Adicionalmente, este procedimiento es usado para resolver el problema de estimación de velocidades de rotación estelar usando una suma finita de distribuciones maxwellianas. Este enfoque permite obtener una buena exactitud en la estimación de la suma de distribuciones, y además expresiones cerradas para los estimadores del algoritmo propuesto.

En la segunda parte de esta tesis se consideran sistemas dinámicos donde las mediciones están sujetas a fuentes de ruido con distribución no-gaussiana, en particular, sistemas dinámicos lineales con funciones de transferencia de ruido de fase no-mínima y distribución de probabilidad dada por la suma de distribuciones gaussianas para el ruido. Se desarrolla un algoritmo iterativo con suma de distribuciones gaussianas basado en el error de predicción. El error de predicción se obtiene usando técnicas de filtrado causal y anti-causal. El algoritmo iterativo propuesto permite obtener expresiones cerradas para los estimadores de la suma de distribuciones gaussianas que definen la distribución de probabilidad del ruido. En general, los resultados de simulación muestran buena exactitud en la estimación de los parámetros del sistema cuando el comportamiento estocástico del ruido es modelado o aproximado de forma apropiada con una suma de distribuciones gaussianas.

Finalmente, el enfoque de *Stochastic Embedding* es usado para cuantificar incertidumbres paramétricas y estructurales en un sistema dinámico lineal. Bajo esta idea, se considera que el comportamiento del sistema viene definido por un modelo nominal y un modelo de la incertidumbre con una distribución de probabilidad dada por una suma de distribuciones

gaussianas. Los parámetros del modelo de la incertidumbre se escogen como variables latentes y la función de verosimilitud se obtiene con la marginalización de dichas variables. Se considera el manejo simultáneo de las mediciones de experimentos independientes para formular un algoritmo iterativo. Para el algoritmo propuesto, se obtienen expresiones en forma cerrada para cada uno de los estimadores de los parámetros que definen la suma de distribuciones gaussianas. Bajo este paradigma se obtiene una gran exactitud en la estimación del modelo nominal y del modelo del error cuando el número de experimentos es alto. Además, este enfoque de identificación muestra un desempeño similar para modelar la incertidumbre en sistemas lineales en tiempo continuo utilizando datos que se obtienen del muestreo.

Palabras clave: Máxima verosimilitud, Stochastic Embedding, Suma de distribuciones gaussianas, Función de transferencia de fase no-mínima, Maximizar la esperanza.

Contents

Nomenclature

The notation and abbreviations used in this thesis are the following:

Notation:

$\mathbb{R}$:	Real number set.
$\mathbb{N}$:	Natural number set.
N:	denotes the number of sampled data
M:	denotes the number of experiments.
κ:	denotes the number of a mixture model distribution components
x_t:	denotes the measurement t of a signal x.
ε_t:	prediction error sequence.
$\varepsilon_t(\theta)$:	prediction error sequence parameterized by θ.
$x_{1:N}$:	denotes de signal x_t for $t = 1, ..., N$.
$\mathcal{N}(x; \mu_x, \Sigma_x)$:	denotes a normal distribution of a random variable x with mean value μ_x and covariance matrix Σ_x.
$\phi_{\mathrm{M}}(x; \sigma)$:	Maxwellian probability density function with dispersion parameter σ.
β:	vector of parameters to be estimated.
$\hat{\beta}$:	estimate of a vector of parameters β.
z^{-1}:	denotes the backward shift operator or the z-transform variable.
s:	denotes the time derivative operator $\frac{d}{dt}$ or the Laplace transform variable.
$\mathbb{E}\{a\|b\}$:	expected value of the random variable a given the random variable b.
$\mathcal{L}(\beta)$:	likelihood function with the vector of parameters β.
$\ell(\beta)$:	log-likelihood function with the vector of parameters β.
$\mathbf{tr}(\cdot)$:	denotes the trace operator.
$\|\cdot\|$:	Euclidean norm.
$\|\cdot\|$:	denotes the determinant operator.
$p(a\|b)$:	probability density function of the random variable a given the random variable b.
$p(a, b)$:	joint probability density function of random variables a and b.

Abbreviations and Acronyms:

AIC:	Akaike information criterion.
ARMA:	Auto-regressive moving average.
BF:	Basis function.
CT:	Continuous-time.
DT:	Discrete-time.
EM:	Expectation-Maximization.

FIR:	Finite impulse response.
FSS:	Feasible system set.
GMM:	Gaussian mixture model.
HOM:	High order moments.
i.i.d.:	Independent and identically distributed.
MA:	Moving average.
MC:	Monte Carlo.
MEM:	Model error modeling.
ML:	Maximum likelihood.
MSA:	Maxwellian mixture model.
MSE:	Mean square error.
PEM:	Prediction error method.
PDF:	Probability density function.
SE:	Stochastic embedding.
SM:	Set membership.
s.t.:	Subject to.
TRM:	Tikhonov regularization method.

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Motivation

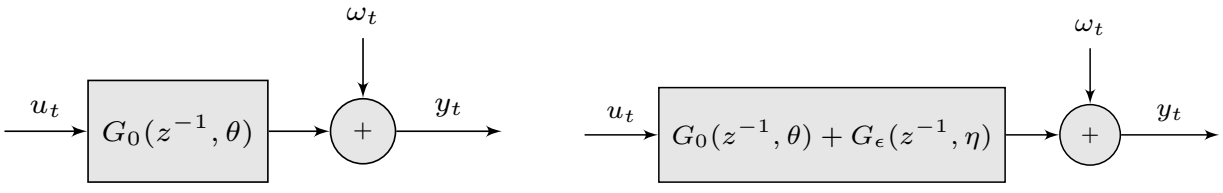
System identification is a scientific area that provides methodologies to model and estimate dynamic systems from a set of experimental data, i.e., a model is fitted to the measured data by estimating its corresponding system model parameters [1, 2]. The formulation of identification algorithms is basically a statistical problem, i.e., we consider the disturbances and measurements errors as realizations of stochastic processes. Then, the accuracy of estimated system models depends of the random behavior of these processes.

Because of its valuable statistical properties [1–3], many identification algorithms in the literature involve the Maximum Likelihood (ML) principle with particular scenarios in the field of system modeling, such as dynamic systems [4–6], static systems [7], systems with quantized output data [8–11], and communications [12, 13], to mention a few. In particular, the ML estimators are asymptotically *unbiased*, that is, the estimates approach the *true* value when the number of measurements is large. However, if the number of measurements is small, then the estimates can be far from the *true* value due to the measurement noise variance [1, 2]. On the other hand, there are scenarios where the number of sampled data is large and the estimated system model can considerably change from one estimation to another, i.e, there are uncertainties in the system model that cannot be accounted for by the noise measurements and a large variance.

To illustrate this behavior, we consider the system models shown in Figure 1.1, where y_t is the output signal, u_t is the input signal, z^{-1} denotes the backward shift operator ($z^{-1}u_t = u_{t-1}$), ω_t is a zero-mean Gaussian white noise with variance σ^2. The main goal is to obtain an ML estimation of $\beta = [\theta^T \;\; \sigma^2]^T$, where $G_0(z^{-1}, \theta)$ is parameterized by θ. The estimation is performed utilizing the corresponding set of measurements $\{u_{1:N}, y_{1:N}\}$ for both *Model A* and *Model B*, where N is the data length. For simplicity of the presentation, we adopt finite impulse response (FIR) system models for all transfer functions in Figure 1.1 as follows:

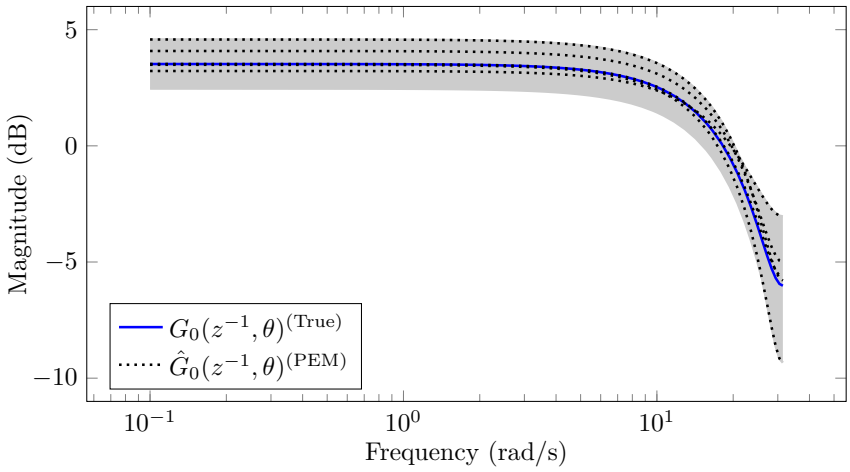
$$G_0(z^{-1}, \theta) = g_0 + g_1 z^{-1}, \quad G_\epsilon(z^{-1}, \eta) = \eta_0 + \eta_1 z^{-1}, \tag{1.1}$$

where $\theta = [g_0 \;\; g_1]^T$ is a deterministic vector of parameters, and $\eta = [\eta_0 \;\; \eta_1]^T$ is a stochastic vector of parameters with a known probability density function (PDF). For *Model A*, an

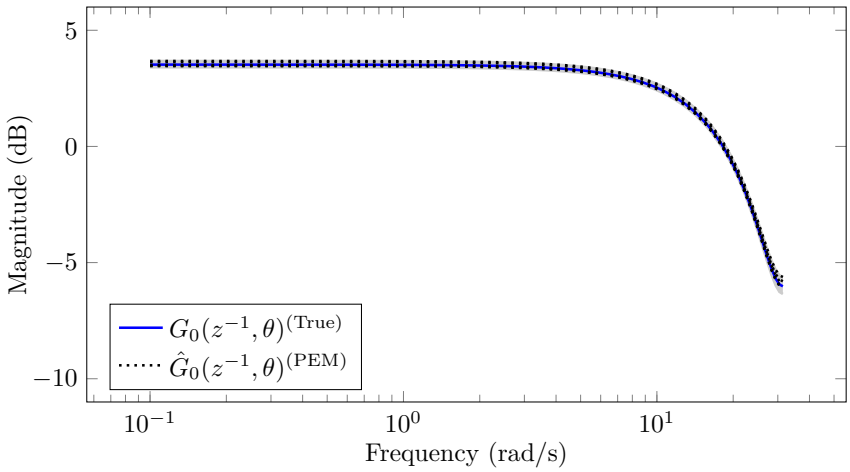


(a) *Model A* (b) *Model B*

Figure 1.1: Models for linear dynamic system identification.



(a) $N = 250$



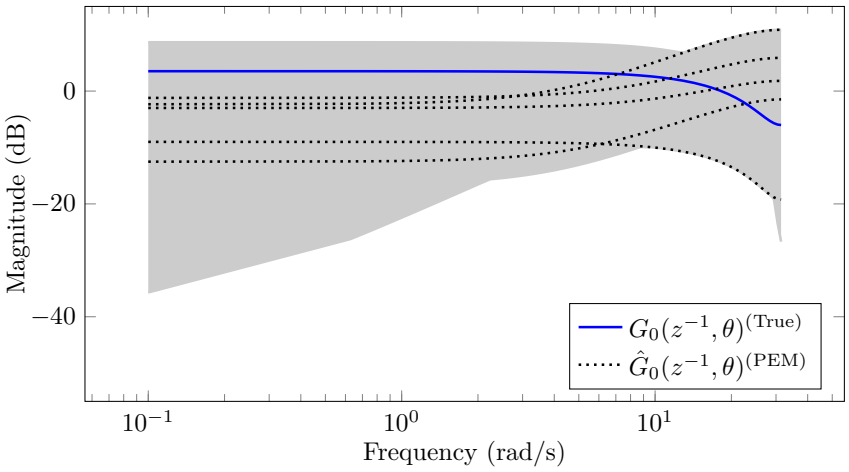
(b) $N = 10^4$

Figure 1.2: Frequency response of the magnitude of the estimated system for *Model A* using PEM.

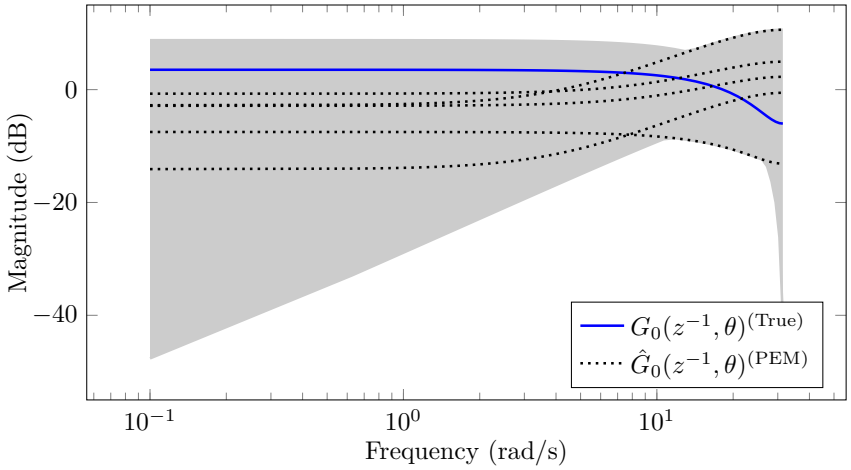
additive noise sequence ω_t is considered. Similarly, for *Model B*, structural and parametric uncertainties are incorporated using an additive error-model, $G_\epsilon(z^{-1},\eta)$, parameterized by a random vector η.

We perform 50 simulations to obtain a data set $\{u_{1:N}, y_{1:N}\}$ considering that u_t is a zero-mean Gaussian distributed input signal with variance $\sigma_u^2 = 10$, and ω_t is zero-mean Gaussian distributed with variance $\sigma^2 = 8$ for each scenario. In addition, for *Model B*, η is Gaussian distributed with mean value $\mu_\eta = [0 \ \ -1]$ and covariance matrix $\Sigma_\eta = \text{diag}[0.2]$. We then estimate a system model $G_0(z^{-1},\theta)$, for *Model A* and *Model B*, utilizing the data set of each simulation and the classical prediction error method (PEM) [2][1]. In Figure 1.2 and Figure 1.3 the gray shaded regions represent the area in which the magnitude of the frequency response of 50 estimations of the system lie, for both *Model A* and *Model B*. In Figure 1.2, we observe biased estimations (dotted lines) since the number of measurements used to obtain the ML estimations is small ($N = 250$). The bias in the estimated models is reduced when the number of measurements is increased ($N = 10^4$). In contrast, in Figure 1.3, since structural and

[1] In this case the PEM estimations are computed using `pem()` function of Matlab®.



(a) $N = 250$



(b) $N = 10^4$

Figure 1.3: Frequency response of the magnitude of the estimated system for *Model B* using PEM.

parametric uncertainties are added by $G_\epsilon(z^{-1},\eta)$, biased estimated models (dotted lines) are obtained utilizing both small and large data measurements set. This means that a different system is estimated for each data set. Hence, it is desirable that from system identification techniques provide a quantification of the estimation uncertainty, that is, modeling the error-model that comes from structural and/or parametric uncertainties [14–16].

An alternative to modeling structural and/or parametric uncertainties is to combine a *nominal model* with an *error-model*, i.e., the uncertainty modeling is part of the estimation algorithm. This idea can be used for robust control design, where optimization methods are used in order to minimize the expected variation (variance) of the *true* system in the control performance [17], or to obtain a probabilistic solution for robust control design [18, 19]. This approach of uncertainty modeling has been addressed in different frameworks, such as Set Membership [20], Model Error Modeling [21], and Stochastic Embedding [14, 22]. In [20] the Set Membership (SM) approach was used to deal with the problem of uncertainty modeling in dynamic systems. The authors of [20] considered the error-model estimation in a deterministic framework in which it is unknown-but-bounded, and obtained a set of possible

solutions. However, this method does not guarantee a small set of solutions to properly describe the system model uncertainty. In [21] classical prediction error methods (PEMs) [2] were used to obtain an estimation of the nominal model. Then, the unmodeled dynamics were estimated from the residuals dynamics using PEM and obtaining the corresponding model error modeling (MEM). This methodology assumes that the nominal model is available.

In [14, 22] a Stochastic Embedding (SE) approach was used to describe the model uncertainty by considering the model as a realization drawn from an underlying probability space, where the parameters that define the error-model are characterized by a PDF. In this approach, a flexible error-model distribution is needed to obtain a correct description of the system model uncertainty. The uncertainty can be quantified by using ML estimation considering the parameters of the error-model as *hidden* variables [23,24] and solving the associated estimation problem using the Expectation-Maximization (EM) [25] algorithm under Gaussian assumptions for the error-model distribution. However, when the error-model distribution is not Gaussian, these techniques perform poorly under Gaussian assumptions. In [24,26] a Bayesian perspective was adopted, where the parameters that define both the nominal model and the error-model can be modeled as realizations of random variables with certain prior distributions, and the posterior densities of the parameters can be estimated. There are works closely related with this framework [27, 28] based on the kernel approach for system identification. Under this Bayesian framework, it is possible to obtain a unified system model of both the nominal model and error-model if all system models are FIR systems. However, this approach does not allow for obtaining the nominal model and the error-model separately when more complex model structures are involved in the *true* system.

On the other hand, classical ML algorithms have been tailored for estimating linear dynamic systems where measurements are subject to noise. In particular, Gaussian distributed noise sources are considered to attain the corresponding PEM estimators to obtain models for linear systems [1, 2]. Based on this noise distribution, the estimation is improved when the number of sample data is increased. However, when a Gaussian distribution for the noise is assumed and the actual distribution is not Gaussian, the estimates can be far from the *true* value, since the number of samples required to achieve consistency and efficiency tends to be large.

In contrast, in [29–31] non-Gaussian assumptions have been used to obtain identification methodologies for dynamic systems. Particularly, the identification of purely stochastic non-minimum-phase systems under non-Gaussian assumptions is a problem that arises in many applications such as communications [32], signal processing [33], and deconvolution problems [34]. Typically, the identification of non-minimum-phase noise transfer functions requires the use of high-order statistics [35] due to the fact that second-order statistics of the output observations do not include sufficient information to differentiate minimum-phase zeros from non-minimum-phase zeros [36,37]. In [37,38] an ML estimation algorithm for purely stochastic non-minimum-phase linear dynamic systems was developed. In [36, 39, 40] an ML estimation algorithm using an approximation of the likelihood function was considered. This approximated likelihood function is obtained by using a truncation in the representation of the innovations in terms of the observed data, assuming that i) the noise sequence distribution is known, and ii) a specific system model structure is given (moving average (MA) models [36] or auto-regressive moving average (ARMA) models [39]). Under this approach,

the estimation accuracy is improved when the number of samples is large.

In [41] a traditional EM algorithm to obtain an exact ML estimator for a non-minimum-phase MA system with non-Gaussian noise was developed. The initial conditions of the system are considered as *hidden* (unobserved) variables. The authors of [41] showed that, when the initial conditions are included as parameters to be estimated, the estimation accuracy is better than when the initial conditions are fixed to zero (conditional ML estimation). On the other hand, non-Gaussian assumptions have been used in the literature, specially for error-in-variables systems (see e.g. [42,43] and the references therein). Moreover, most of the techniques found in the literature for the identification of systems with non-minimum-phase zeros utilize the Method of Moments [35,44–46], which is known to require a large number of samples to yield accurate estimators [47]. In addition, these high-order estimation techniques are restricted for non-Gaussian noise distributions such that the fourth-order cumulant is non-zero [44].

In this thesis, we develop an ML estimation methodology for modeling the error-model in linear dynamic systems under non-Gaussian assumptions. For structural and parametric uncertainties, we consider the *Stochastic Embedding* approach, obtaining an estimation of both a nominal model and a non-Gaussian distributed error-model. We also established an ML estimation algorithm for linear dynamic systems where the output signal measurements are subject to non-Gaussian noise, specifically, system models with non-minimum-phase noise transfer functions driven by an exogenous input signal. For both scenarios, the non-Gaussian distributions are modeled utilizing finite mixture distributions, specially, Gaussian Mixture Models (GMMs). GMMs have been utilized in filtering [48, 49], tracking [50, 51], communications [52, 53], Bayesian estimation [54, 55], linear dynamic systems [56], and estimation of stellar rotational velocities [57], to mention a few. Moreover, some of these authors use GMMs to approximate non-Gaussian-sum distributions based on the Wiener approximation theorem, which establishes that any PDF with compact support can be approximated by a finite sum of Gaussian distributions [58, Theorem 3].

The main contributions of this thesis can be summarized as follows:

Main contributions

(i) We develop a Maximum Likelihood methodology for modeling structural and parametric uncertainties in linear dynamic systems utilizing a *Stochastic Embedding* approach with GMMs. We obtain a system model that combines the nominal model and an error-model distribution defined by a GMM.

(ii) We develop a Maximum Likelihood method for estimating linear dynamic systems with a Gaussian mixture noise distribution. We focus on the identification of the parameters of linear systems with non-minimum-phase noise transfer functions and the parameters of Gaussian mixture noise distributions.

(iii) We propose Expectation-Maximization algorithms to solve the associated Maximum Likelihood estimation problems with GMMs, obtaining the system model parameters and closed form expressions for the GMM parameter estimators that describe the non-Gaussian distributions.

1.2 Thesis organization

This thesis is organized in chapters as follows:

In Chapter 2 the basic concepts of Finite mixture distributions and the approximation properties of the GMMs are presented. We also show the ML formulation for estimation problems in which the likelihood function can be expressed as infinite mixtures with GMMs. An iterative EM-based algorithm to solve the corresponding ML estimation problem with GMMs is developed. We also use the proposed ML methodology with GMMs to address the stellar rotational distribution estimation problem. In this problem, the *true* distribution of the rotational velocities is modeled as a finite mixture of known distributions, specifically, a finite mixture of Maxwellian distributions. These results are the baseline to develop the ML estimation algorithms for uncertainty modeling in linear dynamic systems.

In Chapter 3 the identification of linear dynamic systems with non-Gaussian noise distribution is addressed, namely, dynamic systems with non-minimum-phase noise transfer functions driven by an exogenous input signal. The noise sequence distribution is modeled as a GMM. An ML estimation algorithm with GMMs to obtain both the system model and the GMM estimators is developed. An EM algorithm with GMMs is developed to solve the associated ML estimation problem.

In Chapter 4 the problem of model error modeling for linear dynamic systems is addressed. We focus on the *Stochastic Embedding* approach using GMMs for modeling the error-model distribution for dynamic systems with structural and/or parametric uncertainties. The estimation problem is addressed using the ML method with GMMs. An iterative EM-based algorithm to estimate the nominal model and the error-model distribution is proposed. In addition, the *Stochastic Embedding* approach using GMMs is extended to address the problem of uncertainty modeling for continuous-time linear dynamic systems using sampled data.

Finally, in Chapter 5, we present the conclusions and future research directions.

1.3 Associated publications

This thesis is supported by a set of publications in which the candidate is a joint author. Related publications made during the research period are also included.

Journal papers

J.1 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, e0208499, 2018. doi: 10.1371/journal.pone.0208499

J.2 **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, A138, 2019. doi: 10.1051/0004-6361/201833455

J.3 **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, 3837, 2021. doi: 10.3390/s21113837

J.4 **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*, Accepted for publication, 2021.

Conference papers

C.1 **R. Orellana**, R. Carvajal and J. C. Agüero. "Maximum Likelihood Infinite Mixture Distribution Estimation Utilizing Finite Gaussian Mixtures", in *18th IFAC Symposium on System Identification, (SYSID)*, Sweden. IFAC-PapersOnline, vol. 51, no. 15, pp. 706-711, 2018. doi: 10.1016/j.ifacol.2018.09.200

C.2 **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 systems", in *21th IFAC World Congress,* Berlin, Germany. IFAC-PapersOnline, vol. 53, no. 2, pp. 845-850, 2020. doi: 10.1016/j.ifacol.2020.12.841

C.3 **R. Orellana**, M. Coronel, R. Carvajal, R. Delgado, P. Escárate and J. C Agüero. "On the Uncertainty Modelling for Linear Continuous-Time Systems Utilising Sampled Data and Gaussian Mixture Models", in *19th IFAC Symposium on System Identification, (SYSID)*, Italy, 2021.

Other conference and journal publications

O.1 **R. Orellana**, R. Carvajal and J. C. Agüero. "Empirical Bayes estimation utilizing finite Gaussian Mixture Models", in *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON),* Valparaíso, Chile, 2019, pp. 1-6. doi: 10.1109/CHILECON47746.2019.8987584.

O.2 G. Bittner, **R. Orellana**, R. Carvajal and J. C. Agüero. "Maximum Likelihood identification for Linear Dynamic Systems with finite Gaussian mixture noise distribution", in *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON),* Valparaíso, Chile, 2019, pp. 1-6. doi: 10.1109/CHILECON47746.2019.8987642.

O.3 M. Coronel, **R. Orellana**, L. Mora, R. Rojas and J.C. Agüero. A Sliding Mode Control Strategy for Cascade Systems. *IEEE Latin America Transactions*, vol. 17, no. 9, pp. 1410-1417, 2019. doi: 10.1109/TLA.2019.8931133.

O.4 A. Cedeño, **R. Orellana**, R. Carvajal and J. C. Agüero. "EM-based identification of static errors-in-variables systems utilizing Gaussian Mixture models", in *21th IFAC World Congress,* Berlín, Germany. IFAC-PapersOnline, vol. 53, no. 2, pp. 863-868, 2020. doi: 10.1016/j.ifacol.2020.12.844.

In conference paper **O.1**, we use the proposed methodology for error-model estimation to address the *prior* distribution estimation problem in the classical Bayesian inference statement. An Empirical Bayes estimation approach with GMMs was used with the measurements of independent experiments. In **O.2**, the ML estimation of linear dynamic systems with Gaussian mixture noise distribution is addressed. A general class of linear system with minimum-phase noise transfer functions was considered, solving the associated ML estimation problem utilizing a global optimization technique. In **O.3** a variable structure control with sliding modes is proposed for cascade systems with dead time. However, these results are

out of the scope of this thesis. Finally, in **O.4**, the ML estimation of static Error-in-Variables systems is addressed. GMMs are used for modeling the noise-free input distribution, and also to approximate non-Gaussian-sum noise-free input distributions.

The diagram in Figure 1.4 summarizes the thesis organization.

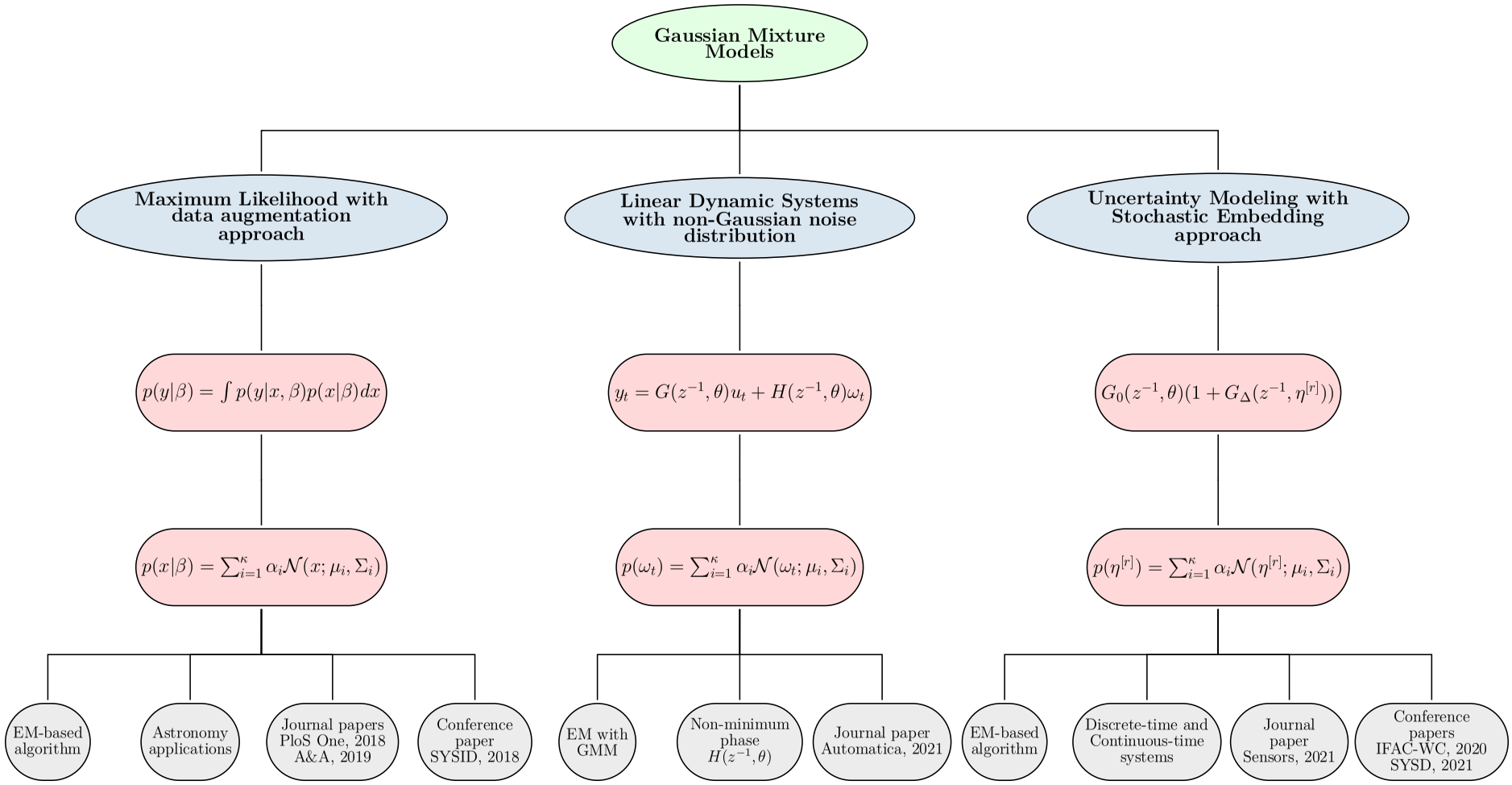


Figure 1.4: Thesis organization diagram

Chapter 2
Maximum Likelihood Estimation of Mixture Distributions with Data Augmentation Approach

Problems in statistics and system identification often involve variables for which measurements are not available. In the ML framework this difficulty is traditionally avoided by the utilization of iterative algorithms, where an auxiliary function that includes both the available data and the non-available data as random variables is used. However, this approach typically involves a cost function in terms of an infinite mixture distribution that comes from marginalizing the measurements with respect to the non-available data. In this chapter the ML parameter estimation is discussed for problems where the likelihood function can be expressed as an infinite mixture distribution that includes a finite Gaussian Mixture Model. The parameter estimation technique is based on the EM algorithm, which provides an iterative framework to solve the ML estimation problem. The benefits of the proposed algorithm are illustrated via simulation examples. In addition, we show how this approach can be used in Astronomy problems, specifically for estimating stellar rotational velocities with real data from three sets of measurements. This chapter summarizes the results presented in the journal publications **J.1** and **J.2**, and also in the conference paper **C.1**.

Contribution

We propose a methodology to solve the Maximum Likelihood estimation problem for infinite mixture distributions utilizing Gaussian Mixture Models. We develop a systematic procedure to generate an auxiliary function and to solve the estimation problem based on an iterative structure of an EM algorithm. This methodology provides closed form expressions for the finite Gaussian mixture estimators.

2.1 Motivation

Data augmentation algorithms in inference problems are based on the construction of the *augmented data* and its many-to-one mapping from the sample space of the complete data (observed and unobserved data) to the sample space of observed data. This *augmented data* is assumed to describe a model from which the observed data set, y, is obtained via marginalization [59]. That is, a system with a likelihood function, $p(y|\beta)$, can be understood to arise from

$$p(y|\beta) = \int_{-\infty}^{\infty} p(y|x,\beta)p(x|\beta)dx, \tag{2.1}$$

where β is the vector of parameters that defines the marginal distributions, $p(y|x,\beta)$ and $p(x|\beta)$, the *augmented data* corresponds to (y,x), and x is the *latent* data. Notice that (2.1) is a common representation of an infinite mixture distribution used in a plethora of

engineering and statistical inverse problems. For example, the likelihood function in (2.1) is closely related with the ML formulation where the y corresponds to the measurements and x may correspond to the internal state (in state-space representation). Another common mixture distribution can be found in Empirical Bayes method [60], in which the goal is the attainment of the ML estimator of the hyperparameters that define the PDF of the parameters that in turn define the likelihood function. The mixture distribution in (2.1) can also be interpreted as a kernel function in a Bayesian framework [61, 62].

On the other hand, this kind of inverse problems appears in several practical problems. One of the many problems in Astronomy deals with is the estimation of rotational velocities of stars. This particular problem is of great importance, since it allows astronomers to describe and model the stars formation, their internal structure and evolution, as well as how they interact with other stars, see e.g. [57, 63, 64]. Modern telescopes allow for the measurement of the rotational velocities from the telescope point of view, that is, a projection of the *true* rotational velocity. Then, (2.1) can be used to deconvolve stellar rotational velocities where $p(x|\beta)$ is the *true* rotational velocity PDF to be estimated (for more details see e.g. [63]). Similar approaches have been used in channel modeling in wireless communications [65] and neutrino mass search in particle physics [66]. In particular, the authors of [24] used the data augmentation approach in (2.1) to obtain a description of the uncertainty in linear dynamic systems.

In this chapter, we propose an identification algorithm to obtain estimates of the parameters that define $p(x|\beta)$ in (2.1). Here, we consider the ML principle to develop an estimation algorithm where $p(x|\beta)$ corresponds to a finite mixture Gaussian distribution. We propose a systematic procedure to build a surrogate function in order to obtain an iterative algorithm for solving the associated ML estimation problem. This procedure is the baseline to develop the estimation algorithms presented in Chapters 3 and 4 for estimating linear dynamic systems with GMMs.

2.2 Finite mixture distributions

2.2.1 General description

Many statistical models involve finite mixture distributions in order to determine subgroups in a population when individual observations of each subgroup are not available [67, 68]. In Figure 2.1 a simple finite mixture distribution with three components is shown. It is assumed that the measurements of an n_y dimensional random variable y are drawn from a density, $p(y|\beta)$, modeled by a linear combination of probability density functions, $p_i(y|\gamma_i)$, as follows:

$$p(y|\beta) = \sum_{i=1}^{\kappa} \alpha_i p_i(y|\gamma_i), \tag{2.2}$$

where κ is the number of components of the mixture model, $p_i(y|\gamma_i)$ is the i-th density function parameterized by γ_i, α_i is the mixing weight subject to $\sum_{i=1}^{\kappa} \alpha_i = 1$, $0 \leq \alpha_i \leq 1$, and the vector of parameters β is given by

$$\beta = [\underbrace{\alpha_1 \ \gamma_1}_{\beta_1} \ \cdots \ \underbrace{\alpha_\kappa \ \gamma_\kappa}_{\beta_\kappa}]. \tag{2.3}$$

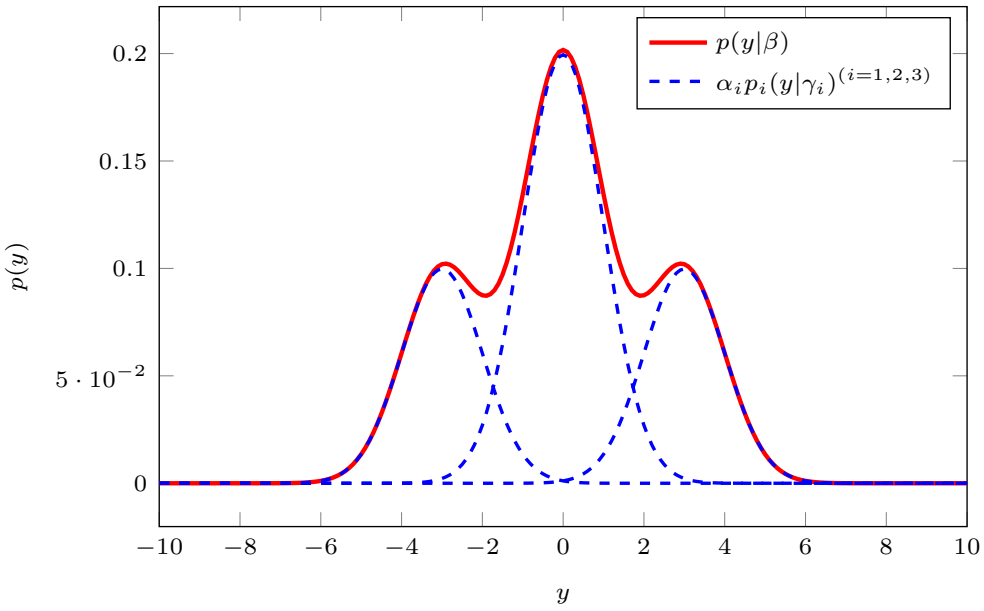


Figure 2.1: Example of a finite mixture of three Gaussian distributions.

In general, it is considered that all the mixture components $p_i(y|\gamma_i)$ arise from the same parametric family of distributions [67, 68]. However, there are some applications in which a linear combination of different parametric distribution family is considered (see e.g. [69]).

2.2.2 Gaussian mixture models (GMMs) approximation

There is a particular case in which the mixture component $p_i(y|\gamma_i)$ in (2.2) is given by a normal distribution, that is a Gaussian Mixture Model (GMM). For the mixture distribution in (2.2), the GMM is defined as follows:

$$p(y|\beta) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(y; \mu_i, \Sigma_i), \tag{2.4}$$

where $\kappa \in \mathbb{N}$ ($\kappa \geq 2$), $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^{\kappa} \alpha_i = 1$, $\mathcal{N}(y; \mu_i, \Sigma_i)$ corresponds to an n_y-dimensional normal (Gaussian) distribution given by:

$$\mathcal{N}(y; \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{n_y/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(y-\mu_i)^T \Sigma_i^{-1}(y-\mu_i)\right\}, \tag{2.5}$$

where $\mu_i \in \mathbb{R}^{n_y \times 1}$ is the mean value, $\Sigma_i \in \mathbb{R}^{n_y \times n_y}$ is the covariance matrix, and $|\cdot|$ corresponds to the determinant operator.

GMMs have been utilized to approximate non-Gaussian distributions (see e.g. [70] and the references therein) in a variety of problems such as tracking [71], classification [72, 73], and estimation of rotational velocities of stars [57], to mention a few. 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 [58, 74]. For completeness of the presentation, the Gaussian sum approximation approach is summarized as follows (See [58, Theorem 3]):

Lemma 1. *Any probability density function, $p(x|\beta)$, of an n-dimensional random variable x with compact support can be approximated as closely as desired in the space $L_1(\mathbb{R}^n)$ by a distribution of the form*

$$p(x|\beta) \approx \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(x; \mu_i, \Sigma_i), \tag{2.6}$$

where $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^{\kappa} \alpha_i = 1$, and $\mathcal{N}(x; \mu_i, \Sigma_i)$ represents an n-dimensional normal distribution with mean value μ_i and covariance matrix Σ_i. ▽

Remark 1. *Based on the Tauberian theorems, the result of Lemma 1 can be extended using a linear combination of translated density functions [74, 75]. Then, the Wiener problem shows that for a certain set $\mathcal{M}$ in the space L of all measurable functions in the interval $(-\infty, \infty)$, every function of the form*

$$\sum_{i,k} c_{i,k} f_i(x + \xi_{i,k}) \tag{2.7}$$

where $c_{i,k}$ are any real numbers, $\xi_{i,k}$ are any real numbers and $f_i \in \mathcal{M}$, lies in L and these functions generate a linear manifold in the space L. ▽

2.3 Maximum Likelihood estimation for infinite mixtures using GMMs

2.3.1 The Maximum Likelihood (ML) method

Maximum Likelihood (ML) is a common method to estimate parameters of a system using the information provided by observations that can be described as realizations of stochastic variables [2]. We can think that the observations (data) are more likely under some members of an arbitrary family of probability distributions than under others [3]. Then, in the ML framework, the probability distribution that makes more likely the given data in the sense of maximizing the *likelihood function* is chosen.

Suppose that the measurements are represented by a random variable $y_{1:N} = [y_1 \ \cdots \ y_N]$, then the *likelihood function* corresponds to the PDF given the vector of parameters β, denoted by $p(y_{1:N}|\beta)$. Hence, the ML estimator is given by:

$$\beta_{\text{ML}} = \arg \max_{\beta} \ p(y_{1:N}|\beta). \tag{2.8}$$

Under mild conditions, the ML estimators have the following properties [1, 2]:

(a) *Consistency:* ML estimators are consistent. That is, they tend to the *true* values in a probabilistic or almost sure sense as the number of measurements N tends to infinity, i.e., $N \to \infty$.

(b) *Efficiency:* Efficiency is concerned with the variance of the estimator. Then, ML estimators are efficient, since their covariance is equal to the Cramér-Rao lower bound (CRLB) [76].

(c) *Invariance:* If there exists a transformation of the parameters, defined as a function, such as $\beta = f(\tilde{\theta})$, the principle of invariance states that the ML estimator of β is given by $\beta_{\text{ML}} = f(\theta_{\text{ML}})$.

(d) *Bias:* The ML estimator is in general biased, i.e., the expected value is not equal to the *true* value of the parameters. Given the property (a), ML estimators are asymptotically unbiased.

For more details on the ML principle, we refer to the reader the classical literature on statistics and system identification, see e.g. [1–3, 77].

2.3.2 Likelihood function from infinite and finite mixtures with augmented data

Consider the infinite mixture distribution in (2.1). Here, $p(y|\beta)$, is a function of a variable accessible to observation, y, and $p(x|\beta)$ is the PDF to be estimated. The conditional PDF $p(y|x)$ is known. Then, the problem to solve is to obtain an estimation of $p(x|\beta)$ from measurements of y using the GMM in (2.6) as follows:

$$p(x|\beta) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(x; \mu_i, \Sigma_i), \tag{2.9}$$

with $\sum_{i=1}^{\kappa} \alpha_i = 1$, $0 \leq \alpha_i \leq 1$. Typically, (2.9) is expressed as an approximation (see e.g. [78]). Nevertheless, the problem of interest is formulated with the equality assumption. Thus, the vector of parameters to be estimated, β, is given by

$$\beta = [\ \underbrace{\alpha_1\ \mu_1\ \Sigma_1}_{\beta_1}\ \cdots\ \underbrace{\alpha_\kappa\ \mu_\kappa\ \Sigma_\kappa}_{\beta_\kappa}\ ]^T. \tag{2.10}$$

On the other hand, if we assume that the available data $y_{1:N}$ and also $x_{1:N}$ are independent and identically distributed (i.i.d) random variables, we obtain:

$$p(y_{1:N}|\beta) = \prod_{t=1}^{N} p(y_t|\beta), \tag{2.11}$$

$$p(x_{1:N}|\beta) = \prod_{t=1}^{N} p(x_t|\beta), \tag{2.12}$$

with

$$p(y_t|\beta) = \sum_{i=1}^{\kappa} \alpha_i \int_{-\infty}^{\infty} p(y_t|x_t)\mathcal{N}(x_t; \mu_i, \Sigma_i)dx_t. \tag{2.13}$$

The likelihood function, $\mathcal{L}(\beta)$, can then be expressed as:

$$\begin{aligned} \mathcal{L}(\beta) &= \prod_{t=1}^{N} p(y_t|\beta), \\ &= \prod_{t=1}^{N} \sum_{i=1}^{\kappa} \alpha_i \int_{-\infty}^{\infty} p(y_t|x_t)\mathcal{N}(x_t; \mu_i, \Sigma_i)dx_t. \end{aligned} \tag{2.14}$$

Obtaining the ML estimator in (2.8) using the likelihood function in (2.14) is usually a difficult task due to the likelihood function nature and it can be even more difficult when the number of unknown parameters is high. A way to reduce the complexity is by considering

a strictly increasing function that, when applied to the likelihood function, yields the same maximum value, such as the logarithmic function. This nonlinear function often simplifies the computation of the maximum of the likelihood function, by turning products into sums and eliminating exponential terms and functions [77]. Then, the log-likelihood function, $\ell(\beta)$, is given by:

$$\ell(\beta) = \sum_{t=1}^{N} \log \left\{ \sum_{i=1}^{\kappa} \alpha_i \int_{-\infty}^{\infty} p(y_t|x_t)\mathcal{N}(x_t; \mu_i, \Sigma_i) dx_t \right\}. \tag{2.15}$$

Finally, using (2.15), the ML estimator is given by the following constrained optimization problem:

$$\beta_{\text{ML}} = \arg \max_{\beta} \ \ell(\beta) \ \text{ s.t. } \ \sum_{i=1}^{\kappa} \alpha_i = 1, \ 0 \leq \alpha_i \leq 1. \tag{2.16}$$

Notice that the ML estimation problem with GMMs in (2.16) involves an *incomplete data* problem, i.e., measurements of the random variable x are not available (see e.g. [79]). This fact can hinder the computation of the log-likelihood function. In addition, the likelihood function in (2.15) involves a logarithm of a sum that depends on the number of components κ of the GMM, and it may be difficult to solve when the number of components in the GMM increases [80]. An EM algorithm provides an elegant solution to the ML estimation problem with *incomplete data* and GMMs.

2.4 Expectation-Maximization (EM) methods for finite mixture distributions

2.4.1 The EM algorithm description

The iterative EM algorithm is a popular tool for identifying linear and non-linear dynamic systems in the time domain (see e.g. [81,82]) and the frequency domain [83]. The key idea of the EM algorithm is to obtain the ML estimation in (2.16) utilizing the data augmentation approach or complete data [25], that is, the information given by the observed data y and the *hidden* variable x. The iterative algorithm consists in obtaining a succession of estimates $\hat{\beta}^{(m)}$, $m = 1, 2, ...$, of the parameter β, alternating between computing an auxiliary function, $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$, with the current estimate $\hat{\beta}^{(m)}$ (*E-step*), and obtaining a new estimate $\hat{\beta}^{(m+1)}$ by maximizing the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ (*M-step*) [25,79]. Figure 2.2 depicts the behavior of the EM algorithm in terms of the auxiliary function. Here, at the m-th iteration, an auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ is obtained (blue dashed line). After that, it is maximized with respect to β, yielding a new estimate, $\hat{\beta}^{(m+1)}$. This procedure continues until a convergence criterion is achieved or a number of EM iterations is reached.

2.4.2 Classical formulation of EM algorithms with GMMs

The EM method has been successfully tailored for finite Gaussian mixtures (see e.g. [67,68] and the references therein). In this approach, the observed data, $y_{1:N} = \{y_1, ..., y_N\}$, are drawn from a density modeled as a convex combination of Gaussian density functions, i.e.,

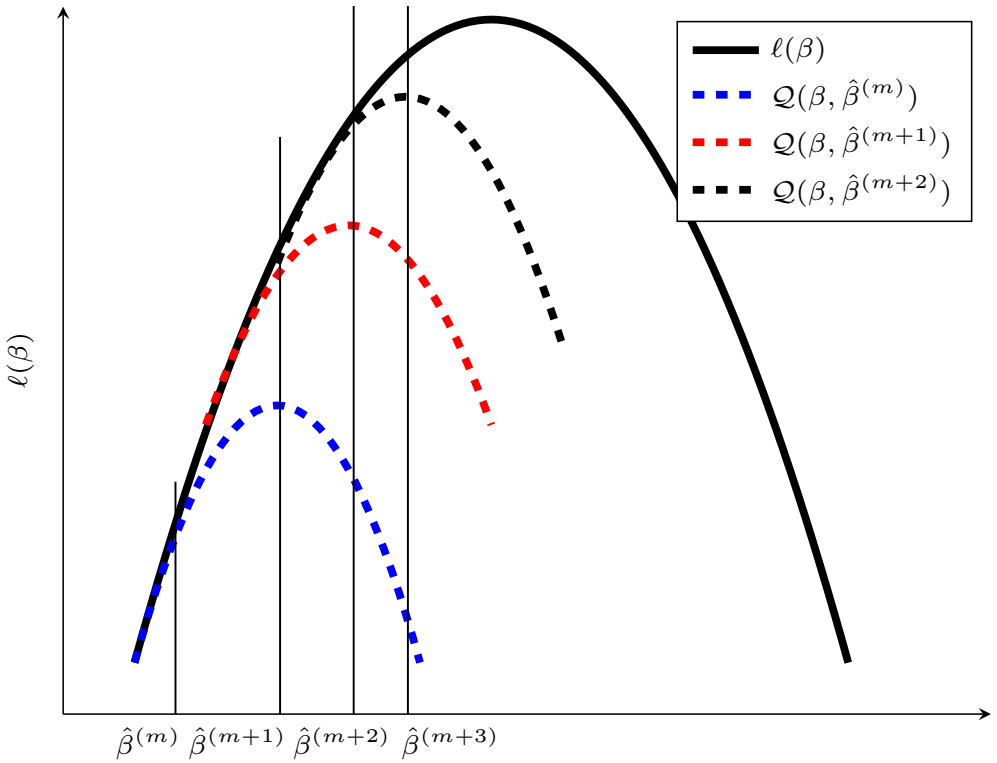


Figure 2.2: Iterative procedure of the EM algorithm.

the log-likelihood function has the form [67]:

$$\ell(\beta) = \sum_{t=1}^{N} \log \left\{ \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(y_t; \mu_i, \Sigma_i) \right\}. \tag{2.17}$$

In this context, the key idea of the EM algorithm is to define a label as a *hidden* discrete random variable, $\zeta_t \in \{1, ..., \kappa\}$, and then maximizing at each iteration the conditional expectation of the complete log-likelihood function, $\ell_c(\beta, \zeta) = p(y_{1:N}, \zeta_{1:N})$,

$$\ell_c(\beta, \zeta) = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \zeta_{ti} \log(\alpha_i \mathcal{N}(y_t; \mu_i, \Sigma_i)), \tag{2.18}$$

given the observed measurements and the current value of the vector of parameters to be estimated [68]. This *hidden* random variable ζ_t is an indicator that determines if an observation y_t arises from the ith component of the GMM. The formulation of the EM algorithm with GMMs takes advantage of the relationship between the observed and complete likelihood functions as follows [25]:

$$\ell(\beta) = \ell_c(\beta, \zeta) - \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \zeta_{ti} \log[p(y_t|\zeta_t = i)], \tag{2.19}$$

where $p(y_t|\zeta_t = i)$ is the conditional probability that y_t arises from the i-th Gaussian mixture component given the parameter value β. Then, the log-likelihood function in (2.19) can be

rewritten in the following form [25, 68]:

$$\ell(\beta) = \bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}) - \bar{\mathcal{H}}(\beta, \hat{\beta}^{(m)}), \tag{2.20}$$

where

$$\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}) = \mathbb{E}\left\{\ell_c(\beta, \zeta)|y_{1:N}, \hat{\beta}^{(m)}\right\}, \tag{2.21}$$

$$\bar{\mathcal{H}}(\beta, \hat{\beta}^{(m)}) = \mathbb{E}\left\{\sum_{t=1}^{N}\sum_{i=1}^{\kappa}\zeta_{ti}\log[p(y_t|\zeta_t = i)]|y_{1:N}, \hat{\beta}^{(m)}\right\}, \tag{2.22}$$

$\mathbb{E}\{\cdot|\cdot\}$ denotes the conditional expectation, and $\hat{\beta}^{(m)}$ is the parameter value at the current iteration. From Jensen's inequality [84], the function $\bar{\mathcal{H}}(\beta, \hat{\beta}^{(m)})$ in (2.22) is a decreasing function, i.e., $\bar{\mathcal{H}}(\beta, \hat{\beta}^{(m)}) \leq \bar{\mathcal{H}}(\hat{\beta}^{(m)}, \hat{\beta}^{(m)})$. Thus, if $\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}) \geq \bar{\mathcal{Q}}(\hat{\beta}^{(m)}, \hat{\beta}^{(m)})$ then $\ell(\beta) \geq \ell(\hat{\beta}^{(m)})$ and the likelihood value increases [79].

Therefore, the EM algorithm for GMMs can be summarized as follows[1]:

$$\bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}) = \mathbb{E}\left\{\ell_c(\beta, \zeta)|y_{1:N}, \hat{\beta}^{(m)}\right\}, \tag{2.23}$$

$$\hat{\beta}^{(m+1)} = \arg\max_{\beta}\ \bar{\mathcal{Q}}(\beta, \hat{\beta}^{(m)}), \quad \text{s.t} \quad \sum_{i=1}^{\kappa}\alpha_i = 1,\ 0 \leq \alpha_i \leq 1, \tag{2.24}$$

where (2.23) and (2.24) correspond to the *E-step* and the *M-step* for the EM algorithm, respectively [25].

The solution of the optimization problem in (2.24) provides closed form expressions for the estimators of the GMM. The *M-step* solution is given by the following expressions [68, 79]:

$$\hat{\alpha}_i^{(m+1)} = \frac{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}{N}, \tag{2.25}$$

$$\hat{\mu}_i^{(m+1)} = \frac{\sum_{t=1}^{N} y_t\hat{\zeta}_{ti}^{(m)}}{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}, \tag{2.26}$$

$$\hat{\Sigma}_i^{(m+1)} = \frac{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}(y_t - \hat{\mu}_i^{(m+1)})^2}{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}, \tag{2.27}$$

where

$$\hat{\zeta}_{ti}^{(m)} = \frac{\hat{\alpha}_i^{(m)}\mathcal{N}(y_t; \hat{\mu}_i^{(m)}, \hat{\Sigma}_i^{(m)})}{\sum_{l=1}^{\kappa}\hat{\alpha}_l^{(m)}\mathcal{N}(y_t; \hat{\mu}_l^{(m)}, \hat{\Sigma}_l^{(m)})}. \tag{2.28}$$

Notice that the *M-step* resulted in closed form expressions. However, for the problem of interest in (2.16), the likelihood function is an infinite mixture and the auxiliary function is difficult to compute. That is the case, for instance, of categorical variables in classification problems or system identification with quantized data, where the integral of the infinite mixture should be computed and optimized at every EM iteration. In the next section, we will show a systematic procedure to obtain an EM-based algorithm with GMMs when the

[1] $\mathbb{E}\{a|b\}$ denotes the expected value of the random variable a given the random variable b.

likelihood function is given by an infinite mixture distribution as (2.15).

2.5 An EM-based algorithm for ML estimation of infinite mixture of GMMs

2.5.1 Constructing the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

In order to formulate an iterative algorithm to solve the ML optimization problem in (2.16), we construct an auxiliary function, $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$, from the log-likelihood function in (2.15) using the following expressions:

$$\mathcal{K}(x_t, \beta_i) = \alpha_i \mathcal{N}(x_t; \mu_i, \Sigma_i), \tag{2.29}$$

$$d\mu(x_t) = p(y_t|x_t)dx_t. \tag{2.30}$$

Then, the log-likelihood function in (2.15) can be expressed as:

$$\ell(\beta) = \sum_{t=1}^{N} \log[\mathcal{V}_t(\beta)], \tag{2.31}$$

$$\mathcal{V}_t(\beta) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \mathcal{K}(x_t, \beta_i) d\mu(x_t). \tag{2.32}$$

Finally, the ML estimator is obtained from:

$$\beta_{\text{ML}} = \arg \max_{\beta} \sum_{t=1}^{N} \log[\mathcal{V}_t(\beta)], \quad \text{s.t.} \quad \sum_{i=1}^{\kappa} \alpha_i = 1, \ 0 \leq \alpha_i \leq 1. \tag{2.33}$$

From the optimization problem in (2.33), the term $\log[\mathcal{V}_t(\beta)]$ can be expressed as follows [85]:

$$\log[\mathcal{V}_t(\beta)] = \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\beta, \hat{\beta}^{(m)}), \tag{2.34}$$

where

$$\mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log[\mathcal{K}(x_t, \beta_i)] \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t), \tag{2.35}$$

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log\left[\frac{\mathcal{K}(x_t, \beta_i)}{\mathcal{V}_t(\beta)}\right] \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t), \tag{2.36}$$

are auxiliary functions. As in the EM algorithm, we have the following result in order to obtain an iterative EM-based algorithm:

Lemma 2. *The function $\mathcal{H}_t(\beta, \hat{\beta}^{(m)})$ in (2.36) is a decreasing function for any value of β and satisfies the following:*

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\hat{\beta}^{(m)}, \hat{\beta}^{(m)}) \leq 0. \tag{2.37}$$

Proof. Using Jensen's inequality [84, pp.24–25] we have:

$$\begin{aligned}\mathcal{H}_t(\beta,\hat{\beta}^{(m)})-\mathcal{H}_t(\hat{\beta}^{(m)},\hat{\beta}^{(m)}) &= \sum_{i=1}^{\kappa}\int_{-\infty}^{\infty}\log\left[\frac{\mathcal{K}(x_t,\beta_i)}{\mathcal{V}_t(\beta)}\right]\frac{\mathcal{K}(x_t,\hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}d\mu(x_t)- \\ &\sum_{i=1}^{\kappa}\int_{-\infty}^{\infty}\log\left[\frac{\mathcal{K}(x_t,\hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}\right]\frac{\mathcal{K}(x_t,\hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}d\mu(x_t), \\ &= \sum_{i=1}^{\kappa}\int_{-\infty}^{\infty}\log\left[\frac{\mathcal{K}(x_t,\beta_i)\mathcal{V}_t(\hat{\beta}^{(m)})}{\mathcal{V}_t(\beta)\mathcal{K}(x_t,\hat{\beta}_i^{(m)})}\right]\frac{\mathcal{K}(x_t,\hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}d\mu(x_t), \\ &\leq \sum_{i=1}^{\kappa}\log\int_{-\infty}^{\infty}\log\left[\frac{\mathcal{K}(x_t,\beta_i)}{\mathcal{V}_t(\beta)}\right]d\mu(x_t), \\ &= 0.\end{aligned} \tag{2.38}$$

Hence, for any value of β, the function $\mathcal{H}_t(\beta,\hat{\beta}^{(m)})$ in (2.36) is a decreasing function. □

From Lemma 2 and inspired in the EM algorithm, we can formulate the following iterative algorithm:

$$\mathcal{Q}(\beta,\hat{\beta}^{(m)}) = \sum_{t=1}^{N}\mathcal{Q}_t(\beta,\hat{\beta}^{(m)}), \tag{2.39}$$

$$\hat{\beta}^{(m+1)} = \arg\max_{\beta}\ \mathcal{Q}(\beta,\hat{\beta}^{(m)}),\quad \text{s.t.}\quad \sum_{i=1}^{\kappa}\alpha_i = 1,\quad 0\leq\alpha_i\leq 1. \tag{2.40}$$

Notice that (2.39) and (2.40) correspond to the *E-step* and the *M-step* of the EM algorithm, respectively [25].

Remark 2. *If we consider a finite mixture of the form*

$$p(y|\beta) = \prod_{t=1}^{N}\sum_{i=1}^{\kappa}\alpha_i\mathcal{N}(y_t;\mu_i,\Sigma_i), \tag{2.41}$$

we can utilize the same approach described here to solve the problem of estimating parameters in (2.39) *and* (2.40)*. In this case, we define* $\mathcal{K}(y_t,\beta_i)$ *as:*

$$\mathcal{K}(y_t,\beta_i) = \alpha_i\mathcal{N}(y_t;\mu_i,\Sigma_i). \tag{2.42}$$

Utilizing the expression derived in (2.35) *we obtain the following auxiliary function:*

$$\mathcal{Q}_t(\beta,\hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa}\log[\mathcal{K}(y_t,\beta_i)]\frac{\mathcal{K}(y_t,\hat{\beta}_i^{(m)})}{\sum_{l=1}^{\kappa}\mathcal{K}(y_t,\hat{\beta}_l^{(m)})}, \tag{2.43}$$

and the auxiliary function $\mathcal{Q}(\beta,\hat{\beta}^{(m)})$ *can be computed substituting* (2.43) *in* (2.39)*. The ML estimator for the finite mixture in* (2.41) *can be then locally obtained from the iterative algorithm* (2.39) *and* (2.40)*.* ▽

2.5.2 Optimizing the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

In order to obtain a new estimate $\hat{\beta}^{(m+1)}$, we solve the optimization problem in (2.40). For the optimization of the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ in (2.39), we can obtain closed form expressions for the estimate of β. Specifically, the optimization with respect to β can be carried out as follows:

Lemma 3. *The vector of parameters $\hat{\beta}$ that optimizes the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ in (2.39) with respect to β subject to $\sum_{i=1}^{\kappa} \alpha_i = 1,\ 0 \leq \alpha_i \leq 1$ is given by:*

$$\hat{\alpha}_i^{(m+1)} = \frac{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}{\sum_{l=1}^{\kappa} \mathcal{P}(x_t, \hat{\beta}_l^{(m)})}, \tag{2.44}$$

$$\hat{\mu}_i^{(m+1)} = \frac{\mathcal{M}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}, \tag{2.45}$$

$$\hat{\Sigma}_i^{(m+1)} = \frac{\mathcal{S}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})} - [\hat{\mu}_i^{(m+1)}]^2, \tag{2.46}$$

with

$$\mathcal{P}(x_t, \hat{\beta}_i^{(m)}) = \sum_{t=1}^{N} \int_{-\infty}^{\infty} \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t), \tag{2.47}$$

$$\mathcal{M}(x_t, \hat{\beta}_i^{(m)}) = \sum_{t=1}^{N} \int_{-\infty}^{\infty} x_t \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t), \tag{2.48}$$

$$\mathcal{S}(x_t, \hat{\beta}_i^{(m)}) = \sum_{t=1}^{N} \int_{-\infty}^{\infty} x_t^2 \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t), \tag{2.49}$$

where $\mathcal{V}_t(\hat{\beta}^{(m)})$ corresponds to (2.32) evaluated at the current value $\hat{\beta}^{(m)}$.

Proof. See Appendix 2.A □

The integrals of the proposed algorithm can be carried out with numerical integral approximations using global adaptive quadrature, e.g., Gauss-Kronrod quadrature. This method uses algebraic transformations and high-order global adaptive quadrature to solve problems on infinite intervals, see e.g. [86] and the references therein.

Finally, the iterative estimation procedure is summarized in *Algorithm 2.1*.

2.6 Numerical simulations

In this section we show three numerical examples of weighting PDF estimation utilizing the proposed algorithm. For the examples, let us consider the following system:

$$y_t = x_t + \omega_t, \tag{2.50}$$

where $\omega_t \sim \mathcal{N}(0, \sigma_\omega^2)$, $\sigma_\omega^2 = 1$ and $p(y_t|x_t) \sim \mathcal{N}(x_t, \sigma_\omega^2)$. We consider the weighting distribution $p(x_t|\beta)$ is modeled as a GMM parameterized by β. The vector of parameters to be

Algorithm 2.1 *(Iterative algorithm for infinite mixtures with GMMs)*

Inputs $y_{1:N}$, κ, $\hat{\beta}^{(0)}$.
Outputs $\hat{\beta}$.

1: $m \leftarrow 0$.
2: Compute $\mathcal{V}_t(\hat{\beta}^{(m)})$ from (2.29), (2.30) and (2.32).
3: Estimate $\hat{\alpha}_i^{(m+1)}$ from (2.44) for $i = 1, ..., \kappa$.
4: Estimate $\hat{\mu}_i^{(m+1)}$ from (2.45) for $i = 1, ..., \kappa$.
5: Estimate $\hat{\Sigma}_i^{(m+1)}$ from (2.46) for $i = 1, ..., \kappa$.
6: **if** stopping criterion is not satisfied **then**
7: $\quad m \leftarrow m + 1$,
8: $\quad$ **return** to 2
9: **else**
10: $\quad \hat{\alpha}^{(m+1)} \leftarrow \hat{\alpha}_i^{(m+1)}$, $\hat{\mu}^{(m+1)} \leftarrow \hat{\mu}_i^{(m+1)}$, $\hat{\Sigma}^{(m+1)} \leftarrow \hat{\Sigma}_i^{(m+1)}$ for $i = 1, ..., \kappa$.
11: $\quad \hat{\beta} \leftarrow [\hat{\alpha}^{(m+1)}\ \hat{\mu}^{(m+1)}\ \hat{\Sigma}^{(m+1)}]$.
12: **end if**
13: End

estimated is $\beta = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1}^{\kappa}$. In a first example, the *true* weighting distribution (but unknown) is given by a two component overlapped GMM. In a second example, a bimodal GMM is considered. Finally, a third example is considered where the *true* weighting distribution does not correspond to a GMM but can be approximated by one.

The simulation setup is as follows:

(1) The initial value $\hat{\beta}^{(0)}$ is given by the average of y_t for each mean value $\hat{\mu}_i^{(0)}$, sampling variance of y_t for each $\hat{\Sigma}_i^{(0)}$, and equal mixing weights $\hat{\alpha}_i^{(0)} = 1/\kappa$.
(2) The data length is $N = 5000$.
(3) The number of Monte Carlo (MC) simulations is 100.
(4) The stopping criterion is satisfied when:

$$\frac{\left\|\hat{\beta}^{(m)} - \hat{\beta}^{(m-1)}\right\|}{\left\|\hat{\beta}^{(m)}\right\|} < 10^{-6},$$

or when the maximum number of EM iterations of 80 has been reached.

Remark 3. *Notice that this kind of problem can be solved in a simple manner when the conditional PDF of the mixture, $p(y_t|x_t)$, is Gaussian (see e.g. [70, 87, 88]). However, this would not be the case for more complex distributions or functions that arise in problems such as system identification with quantized data (see e.g. [9, 10]) and estimation of rotational velocities of stars [57]. In those cases, the approach proposed can be directly utilized, yielding simple and closed form expressions that are, in general, not difficult to compute.* ▽

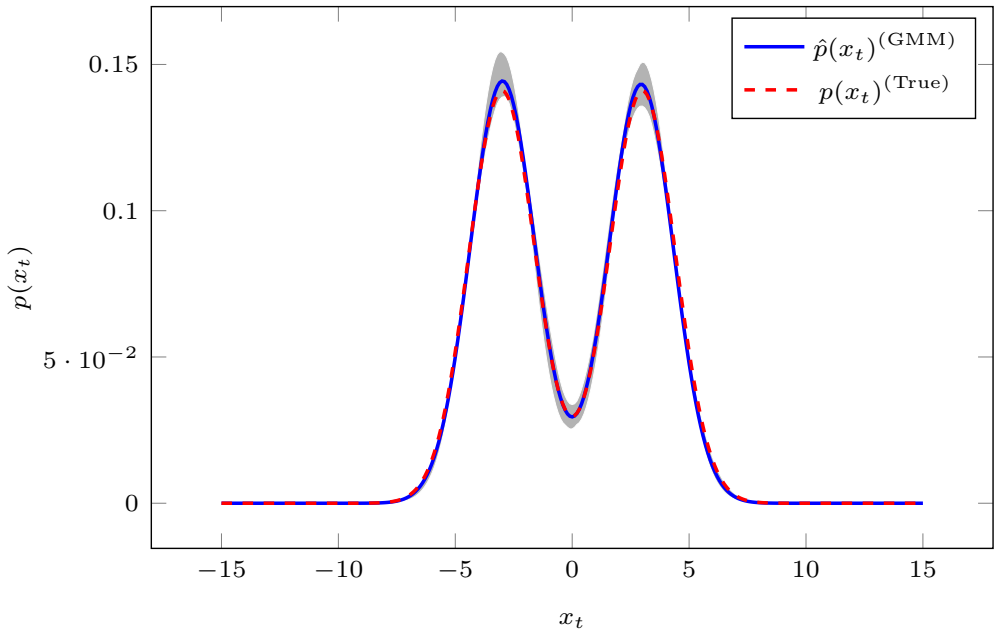


Figure 2.3: Estimation of bimodal distribution $p(x_t|\beta)$ using two Gaussian mixture components.

Table 2.1: Estimated Gaussian mixture parameters of MC simulations for the Example 1

GMM component	α_i	μ_i	Σ_i
$i = 1$	$0.503 \pm 5.20 \times 10^{-3}$	$-2.976 \pm 3.72 \times 10^{-2}$	$1.934 \pm 1.02 \times 10^{-1}$
$i = 2$	$0.497 \pm 5.20 \times 10^{-3}$	$2.949 \pm 3.37 \times 10^{-2}$	$1.919 \pm 1.03 \times 10^{-1}$

2.6.1 Example 1: Estimation of an overlapping Gaussian mixture distribution

In this example, the unknown random variable x_t is drawn from a finite Gaussian mixture distribution using the *Slice Sampler* [89], where:

$$p(x_t)^{(\mathrm{True})} = \alpha_1 \mathcal{N}(x_t; \mu_1, \Sigma_1) + \alpha_2 \mathcal{N}(x_t; \mu_2, \Sigma_2), \tag{2.51}$$

with $\alpha_1 = \alpha_2 = 0.5$, $\mu_1 = -3$, $\mu_2 = 3$, $\Sigma_1 = \Sigma_2 = 2$.

The estimation results are shown in Figure 2.3. The gray-shaded region corresponds to the area in which all the estimated GMMs lie. The blue line represents the average GMM for all MC realizations. We observe that the average of the estimated GMM is very similar to the *true* distribution. In Table 2.1 we show the mean value and standard deviation of the estimated parameters. We observed that, in general, the estimated parameters are "close to" the *true* values. However, the estimated variances of the corresponding Gaussian distributions, Σ_i, exhibit a greater dispersion with respect to the other parameters.

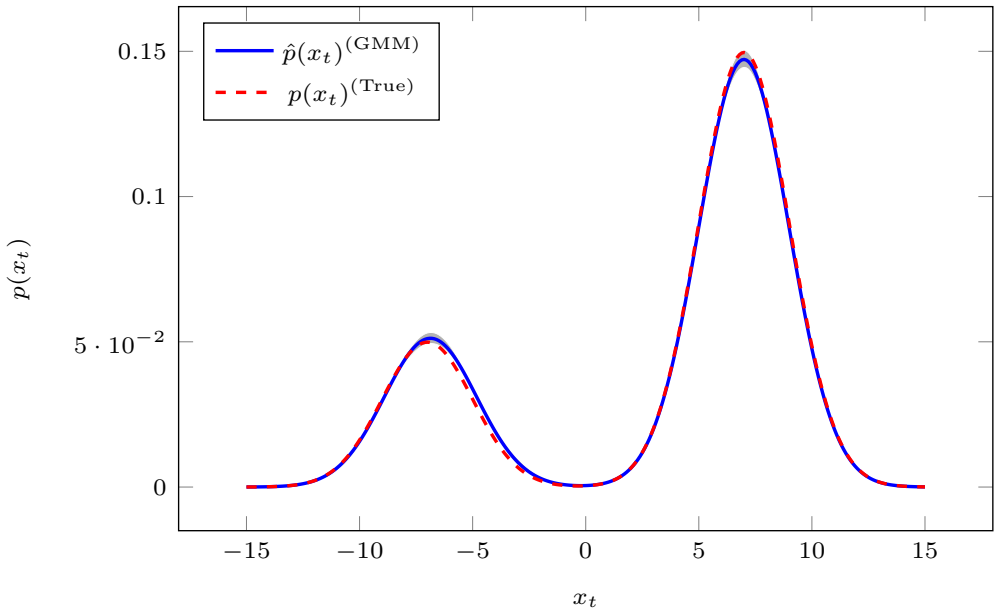


Figure 2.4: Estimation of overlapped distribution $p(x_t|\beta)$ using two Gaussian mixture components.

Table 2.2: Estimated Gaussian mixture parameters of MC simulations for the Example 2

GMM component	α_i	μ_i	Σ_i
$i = 1$	$0.262 \pm 3.00 \times 10^{-4}$	$-6.86 \pm 2.68 \times 10^{-2}$	$4.17 \pm 1.17 \times 10^{-1}$
$i = 2$	$0.738 \pm 3.01 \times 10^{-4}$	$6.99 \pm 1.55 \times 10^{-2}$	$4.01 \pm 6.29 \times 10^{-2}$

2.6.2 Example 2: Estimation of a non-overlapping Gaussian mixture distribution

In this example, the unknown data x_t in (2.51) is generated from a bimodal Gaussian mixture distribution, also using the *Slice sampler* [89], with parameters $\alpha_1 = 0.25$, $\alpha_2 = 0.75$, $\mu_1 = -7$, $\mu_2 = 7$, $\Sigma_1 = \Sigma_2 = 4$. In Figure 2.4 the average GMM estimated for all MC realizations is shown (blue line). The gray-shaded region represents the area in which all the estimated GMMs lie. We observe that the difference between the *true* bimodal distribution and the average of the estimated PDFs is negligible.

In Table 2.2, we show the mean value and standard deviation of the estimated parameters. We observed, in general, accurate estimations for the GMM parameters. However, the estimated variances of the corresponding Gaussian distributions, Σ_i, exhibit a greater dispersion with respect to the other parameters.

2.6.3 Example 3: Approximation of a Maxwellian distribution using GMMs

In this example, we consider that x_t in (2.51) is drawn from a Maxwellian distribution given by:

$$p(x_t)^{(\mathrm{True})} = \sqrt{\frac{2}{\pi}} \frac{x_t^2}{\sigma^3} \exp\left\{-\frac{x_t^2}{2\sigma^2}\right\}, \tag{2.52}$$

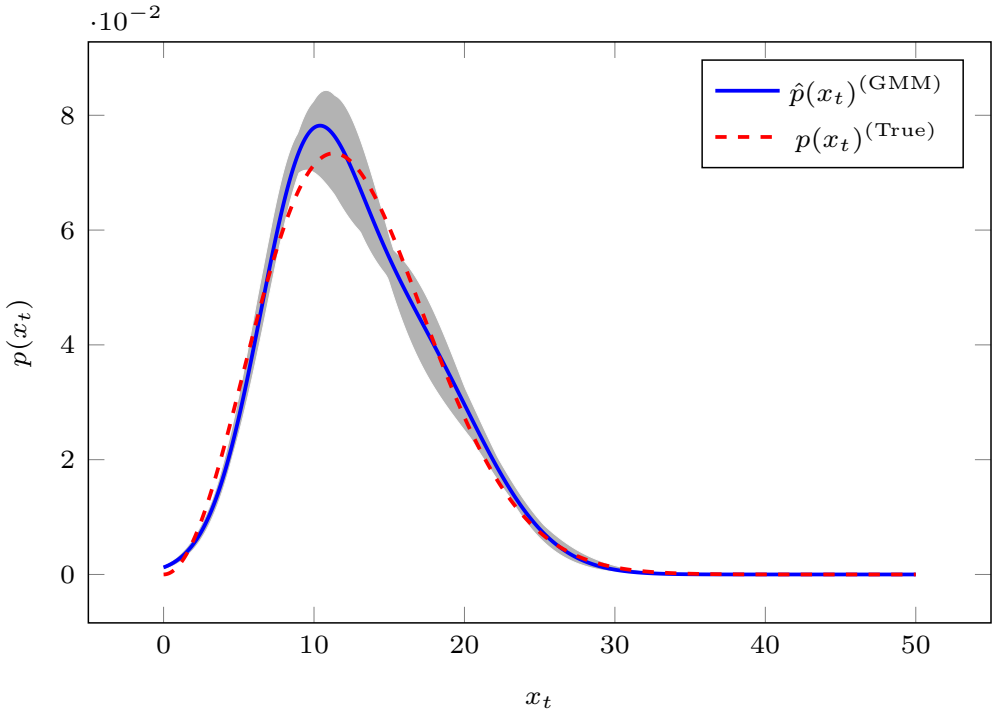


Figure 2.5: Approximation of a Maxwellian distribution $p(x_t|\beta)$ using two Gaussian mixture components.

Table 2.3: Estimated Gaussian mixture parameters of MC simulations for the Example 3

GMM component	α_i	μ_i	Σ_i
$i = 1$	$0.476 \pm 6.55 \times 10^{-2}$	$9.44 \pm 4.94 \times 10^{-1}$	10.74 ± 1.38
$i = 2$	$0.524 \pm 6.55 \times 10^{-2}$	$15.83 \pm 3.93 \times 10^{-1}$	25.73 ± 1.38

where $x_t > 0$, and σ is the dispersion parameter that defines the Maxwellian distribution. This distribution has been used to address estimation problems related with stellar rotational velocities in astronomy and astrophysics (see e.g. [57,63,64]). We consider a GMM in (2.9) with $\kappa = 2$ to approximate the PDF in (2.52).

Figure 2.5 shows the approximation of the Maxwellian distribution using a GMM. The blue line represents the estimated average GMM for all MC simulations. The gray-shaded region corresponds to the area in which all the estimated GMMs lie. We observe that the average of all the estimated GMMs fits the *true* probability distribution. In Table 2.3 it is shown the mean value and standard deviation of the estimated GMM parameters. As in the previous examples, we observed accurate estimations for the GMM parameters. However, the estimated variances, Σ_i, exhibit a greater dispersion with respect to the other parameters.

In the next section, we will show how the proposed methodology can be used to obtain an iterative algorithm to address the stellar rotational distribution estimation problem when the *true* distribution of the rotational velocities is modeled as a finite mixture of known distributions. Specifically, we consider a sum of Maxwellian distributions.

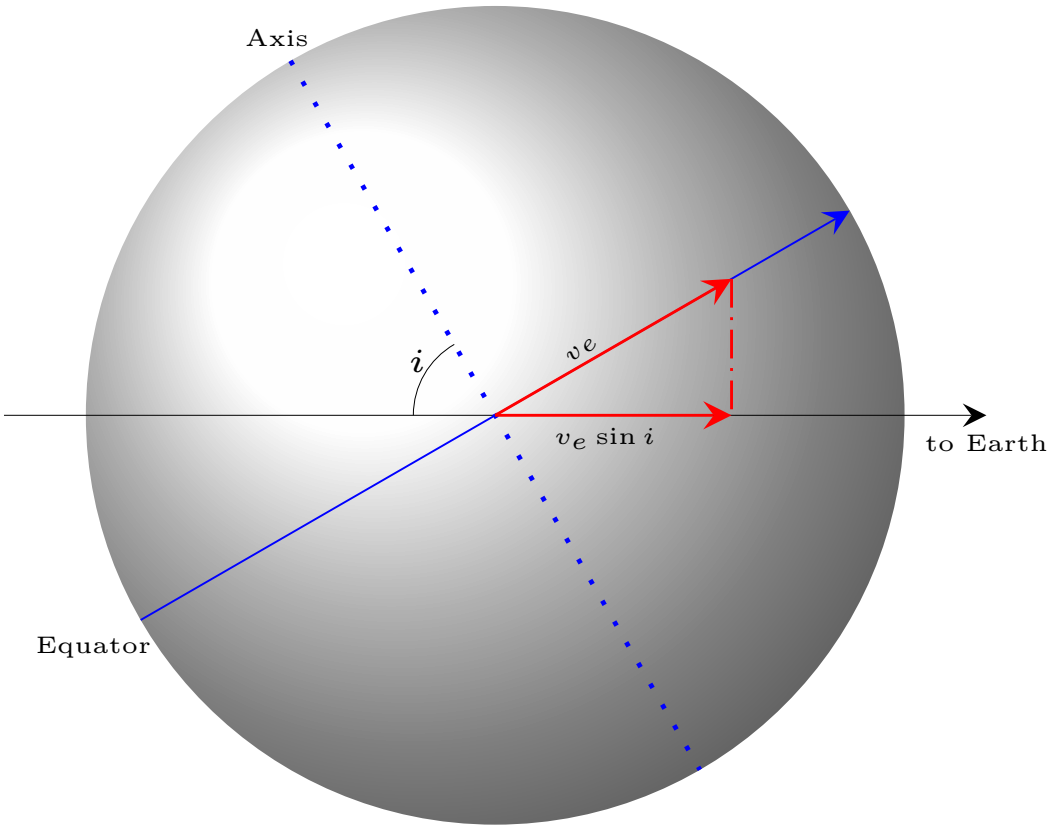


Figure 2.6: Stellar rotational velocities description.

2.7 Deconvolving stellar rotational velocities with data augmentation approach

The estimation of the probability distribution of rotational velocities of stars is essential to describe and model many aspects of stellar evolution. Figure 2.6 shows a description of the stellar rotational problem. From observations, y, it is only possible to obtain the projected velocity, $(v_e \sin i)$, where i is the inclination angle with respect to the line of sight and v_e is the *true* (non-projected) rotational velocity. These measurements are assumed to be realizations of a random variable drawn from a PDF, $p(y|\beta)$, that satisfies (2.1). Notice that the expression in (2.1) corresponds to the Fredholm integral of the first kind.

Many inverse problems in physics and astronomy are given in terms of the Fredholm integral of the first kind [90, 91]. In [63] a method to deconvolve the inverse problem given by (2.1) was developed, obtaining the cumulative distribution function (CDF) for stellar rotational velocities extending the work of [57]. Assuming a uniform distribution of stellar axes over the unit sphere, this integral equation reads as follows (see [63] for more details):

$$p(y|\beta) = \int_y^\infty \underbrace{\frac{y}{x\sqrt{x^2-y^2}}}_{p(y|x)} p(x|\beta)dx, \tag{2.53}$$

where $x = v_e$, is the *true* rotational speed, $y = x \sin i$, is the projected rotational speed, and i, is the (unknown) inclination angle. Furthermore, $p(y|\beta)$ represents the PDF of projected

rotational velocities and $p(x|\beta)$ is the density of true rotational velocities. The function $p(y|x)$ in this integral is related to the distribution of projected angles [63].

The distribution of rotational velocities has been studied in the literature, providing strong evidence for the occurrence of Gaussian and Maxwellian distributions in astrophysical systems. In [92,93] it was proved (analytically) that a Maxwellian distribution corresponds to the rotational speed distribution when the distribution of stellar axes is uniformly distributed over the unit sphere (random axes orientations). For non-Gaussian statistics, [94] proved that the Tsallis distribution (with the k parameter) corresponds to the distribution of rotational velocities, and in the limit case when $k \to 0$, the Maxwellian distribution is recovered. In this section, we focus on showing how the ML estimator in (2.16) can be used to obtain the *true* rotational velocity PDF in (2.53) as a finite mixture of Maxwellian distributions (MSA).

2.7.1 An EM-based algorithm with Maxwellian mixture distributions

The unknown *true* rotational velocities distribution function in (2.53) can be expressed as the following finite mixture model:

$$p(x|\beta) = \sum_{i=1}^{\kappa} \alpha_i \phi_{\mathcal{M}}(x;\sigma_i), \tag{2.54}$$

$$\phi_{\mathcal{M}}(x;\sigma_i) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\sigma_i^3} \exp\left\{-\frac{x^2}{2\sigma_i^2}\right\}, \tag{2.55}$$

where $x > 0$, $\phi_{\mathcal{M}}(x;\sigma_i)$ represents a Maxwellian PDF, and $\sigma_i > 0$ is the dispersion parameter that defines the Maxwellian PDF. The vector of parameters to be estimated is defined as follows:

$$\beta = [\underbrace{\alpha_1 \ \sigma_1}_{\beta_1} \ \cdots \ \underbrace{\alpha_\kappa \ \sigma_\kappa}_{\beta_\kappa}]. \tag{2.56}$$

From (2.15), the ML estimator is given by

$$\beta_{\text{ML}} = \arg\max_{\beta} \ \ell(\beta), \quad \text{s.t.} \quad \sum_{i=1}^{\kappa} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1, \tag{2.57}$$

with

$$\ell(\beta) = \sum_{t=1}^{N} \log \left[\sum_{i=1}^{\kappa} \alpha_i \left\{ \int_{y_t}^{\infty} \frac{y_t}{x\sqrt{x_t^2 - y_t^2}} \phi_{\mathcal{M}}(x_t;\sigma_i) dx_t \right\} \right]. \tag{2.58}$$

Finally, from the iterative algorithm in (2.39) and (2.40), the estimators are given by

$$\hat{\alpha}_i^{(m+1)} = \frac{\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{\sum_{t=1}^{N} \sum_{l=1}^{\kappa} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_l^{(m)})}, \tag{2.59}$$

$$\hat{\sigma}_i{}^{(m+1)} = \left[\frac{\sum_{t=1}^{N} \mathcal{S}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{3\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})} \right]^{\frac{1}{2}}, \tag{2.60}$$

where

$$\mathcal{K}(x_t, \beta_i) = \alpha_i \phi_{\mathcal{M}}(x_t; \sigma_i), \tag{2.61}$$

$$d\mu(x_t) = \frac{y_t}{x\sqrt{x_t^2 - y_t^2}} dx_t, \tag{2.62}$$

$$\mathcal{V}_t(\beta) = \sum_{i=1}^{\kappa} \int_{y_t}^{\infty} \mathcal{K}(x_t, \beta_i) d\mu(x_t), \tag{2.63}$$

$$\mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)}) = \int_{y_t}^{\infty} \frac{\mathcal{K}\left(x_t, \hat{\beta}_i^{(m)}\right)}{\mathcal{V}_t\left(\hat{\beta}^{(m)}\right)} d\mu(x_t), \tag{2.64}$$

$$\mathcal{S}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)}) = \int_{y_t}^{\infty} x_t^2 \frac{\mathcal{K}\left(x_t, \hat{\beta}_i^{(m)}\right)}{\mathcal{V}_t\left(\hat{\beta}^{(m)}\right)} d\mu(x_t). \tag{2.65}$$

More details of the EM-based formulation with a mixture of Maxwellian distributions are presented in Appendix 2.B.

Remark 4. *Notice that an MSA does not approximate every PDF unless a translation term is included in the Maxwellian distribution. Nevertheless, as shown in the following sections, an MSA, even without translation, may provide an adequate fit to the PDF of the stellar rotational velocities.* ▽

2.7.2 Deconvolving real samples

Here, we consider: i) to apply the proposed iterative algorithm to a sample of measured ($v \sin i$) data of stars in order to estimate the PDF of the *true* rotational velocities; and ii) to compare the performance of this proposed algorithm with the Tikhonov regularization method (TRM) proposed by [64] (see Appendix 2.C). We consider real data from three sets of measurements: Coma Berenice, Tarantula, and Geneva.

From the catalog [95], we select the Coma Berenice cluster (Melotte 111) data which has $N = 60$ values of $v \sin i > 0$ from 0 km/s up to 50 km/s for F-K dwarf stars. The estimated parameter using MSA algorithm with one Maxwellian distribution is $\sigma_1 = 7.5675$. In addition, the estimated parameters using two Maxwellian distributions ($\kappa = 2$) are $\sigma_1 = 2.7903$, $\sigma_2 = 11.9939$, $\alpha_1 = 0.6364$, and $\alpha_2 = 0.3636$. Similarly, we obtain the results considering a three Maxwellian mixture distribution ($\kappa = 3$), where the estimated parameters are $\sigma_1 = 2.7902$, $\sigma_2 = 11.9934$, $\sigma_3 = 11.9954$, $\alpha_1 = 0.6364$, $\alpha_2 = 0.2810$ and $\alpha_3 = 0.0827$. We can observe that the mixing weight of the third component (α_3) of Maxwellian mixture distribution is close to zero. In this sense, for simplicity and clarity on the presentation, we analyze the case for $\kappa = 2$.

We also select the Tarantula sample for single O-type stars from the VLT Flames Tarantula Survey presented in [96], where the authors deconvolved the rotational velocity distribution using the method in [90] and TRM. This sample contains 216 stars with $v \sin i$ data from 40 km/s up to 610 km/s. The estimated parameter using the MSA algorithm with one Maxwellian distribution ($\kappa = 1$) is $\sigma_1 = 131.39$. The parameters estimated for the two Maxwellian mixture distribution ($\kappa = 2$) are $\sigma_1 = 64.54$, $\sigma_2 = 185.67$, $\alpha_1 = 0.568$ and

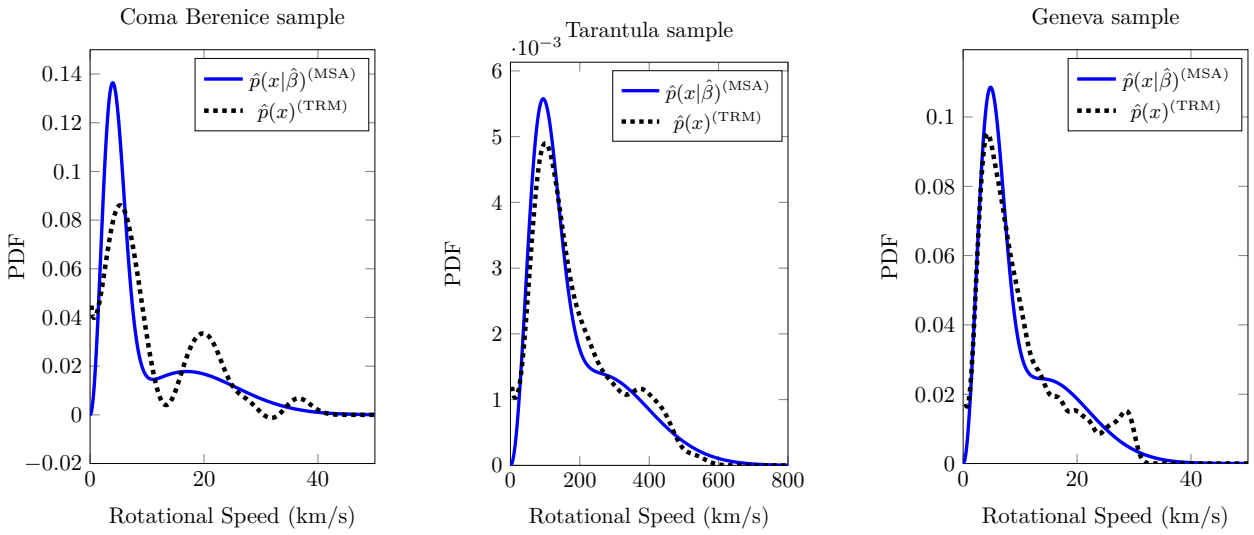


Figure 2.7: Estimated PDF of rotational velocities for real samples cases using $\kappa = 2$ (Coma Berenice and Tarantula samples) and $\kappa = 3$ (Geneva sample) for MSA algorithm.

$\alpha_2 = 0.432$. The estimation results using three Maxwellian mixture distribution ($\kappa = 3$) are $\sigma_1 = 64.54$, $\sigma_2 = 185.67$, $\sigma_3 = 64.54$, $\alpha_1 = 0.568$, $\alpha_2 = 0.432$ and $\alpha_3 = 8.03 \times 10^{-17}$. It is evident that the third component is not relevant in the Maxwellian mixture distribution model since α_3 is almost equal to zero. We also focus on the results we obtained with $\kappa = 2$.

For the third data set, we selected a large sample data of measured $v \sin i$ data of the Geneva-Copenhagen survey of the solar neighborhood [97, 98], which contains information about 16.500 F and G main-sequence field stars. We observed that this data sample presents important uncertainties showing velocities of 0 km/s, in consequence, we only selected stars with $0 < v \sin i \leq 30$ km/s, obtaining a sample of 11685 stars. The estimated parameter using MSA algorithm with one Maxwellian distribution is $\sigma_1 = 7.13$. Additionally, the estimated parameters for $\kappa = 2$ are $\sigma_1 = 3.32$, $\sigma_2 = 10.26$, $\alpha_1 = 0.578$ and $\alpha_2 = 0.422$. The estimation results using three Maxwellian mixture distribution are $\sigma_1 = 4.21$, $\sigma_2 = 10.66$, $\sigma_3 = 2.41$, $\alpha_1 = 0.40$, $\alpha_2 = 0.37$ and $\alpha_3 = 0.23$. In this case, we focus on the estimation obtained with three mixture components.

2.7.3 Analysis of the estimation results

For the analysis of the estimation of the rotational velocities for the three data sets, we consider the following: from the estimated PDF of the rotational velocity, using MSA and TRM, we obtain an estimation of the projected rotational velocity PDF, $\hat{p}(y)$, by solving the integral in (2.53), that is,

$$\hat{p}(y) = \int_y^\infty \frac{y}{x\sqrt{x^2 - y^2}} \hat{p}(x) dx, \tag{2.66}$$

where $\hat{p}(x) = \hat{p}(x|\hat{\beta})$ for MSA and $\hat{p}(x)$ is the estimated PDF using TRM.

In Figure 2.7 from a large number of samples (Tarantula and Geneva samples) both estimation methods show similar results, except for very low rotational speeds, where TRM

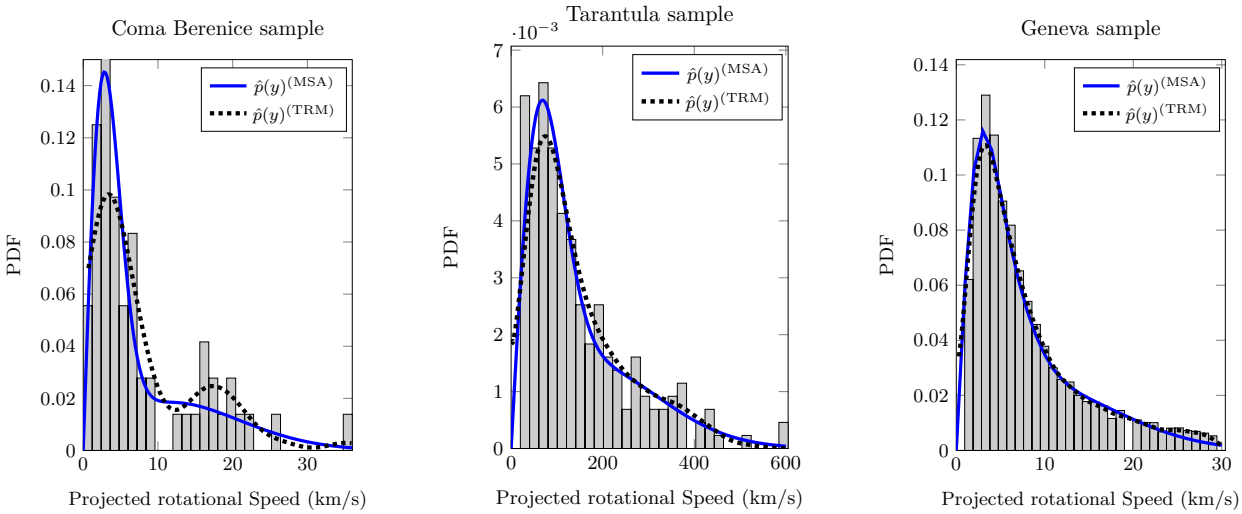


Figure 2.8: Contrasting projected and observed rotational velocities for the real samples using $\kappa = 2$ (Coma Berenice and Tarantula samples) and $\kappa = 3$ (Geneva sample) for MSA algorithm.

yields wrong estimation with non-zero probability for nought rotational velocities, that is, $\hat{p}(x = 0) > 0$. In contrast, our proposed method correctly provides a zero probability for nought rotational velocities, that is, $\hat{p}(x = 0) = 0$. In addition, the estimated projected rotational velocities are expected to provide a zero probability to nought projected rotational velocities. Our method provides the correct estimation as shown in Figure 2.8. On the other hand, similarly to the rotational velocities estimation, the result provided by TRM is incorrect for zero projected rotational velocities. For a small number of samples (i.e., Coma Berenice sample), the rotational velocity (Figure 2.7) and the projected rotational velocity (Figure 2.8) PDFs differ. However, the estimated projected rotational velocities we obtained with MSA closely resembles the histogram from collected data (see Figure 2.8). These results suggest that our method provide better estimate of the rotational velocity.

2.8 Conclusions

In this chapter, we developed a system identification algorithm for the estimation of a probability density function in terms of a Gaussian sum approximation using the ML approach. We formulated an EM-based algorithm with GMMs to solve the corresponding ML estimation problem. We obtained an auxiliary function for the problem of interest, which allowed us the attainment of closed form expressions to estimate parameters of the GMM. We illustrated the performance of the proposed identification algorithm via numerical simulations, where we used numerical approximations to compute the integrals on infinite intervals, as proposed in [86]. The proposed algorithm was also used to approximate a non-Gaussian-sum distribution with good accuracy.

In addition, this estimation methodology was utilized to obtain the estimated PDF of rotational stellar velocities. The advantage of this algorithm is that we were able to use the sample data from the experiments to obtain an estimated PDF of the projected rotational velocities $p(y)$. This algorithm allowed us to use the sample data directly, without intermediate

steps, unlike the estimation using TRM that requires the utilization of kernel density estimator, a non-parametric estimation technique. Finally, the iterative estimation algorithm was tested in a set of real observed data from the Coma Berenice, Tarantula clusters and Geneva samples. The results yielded an adequate statistical description of the projected rotational velocities.

Appendix

2.A Proof of Lemma 3

Taking the derivative of $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ in (2.35) with respect to μ_i and making it equal to zero yields:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \mu_i} = \sum_{t=1}^{N} \int_{-\infty}^{\infty} x_t \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t) - \int_{-\infty}^{\infty} \hat{\mu}_i^{(m+1)} \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t) = 0. \tag{2.67}$$

Using the definition of (2.47) and (2.48) we obtain

$$\sum_{t=1}^{N} \int_{-\infty}^{\infty} x_t \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t) = \hat{\mu}_i^{(m+1)} \mathcal{P}(x_t, \hat{\beta}_i^{(m)}), \tag{2.68}$$

$$\hat{\mu}_i^{(m+1)} = \frac{\mathcal{M}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}. \tag{2.69}$$

Then, taking the derivative of $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ in (2.35) with respect to $\rho = \Sigma_i^{-1}$ and equating to zero:

$$\begin{aligned} \frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \rho} = [\hat{\rho}^{(m+1)}]^{-1} \sum_{t=1}^{N} \int_{-\infty}^{\infty} \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t) - \\ \int_{-\infty}^{\infty} (x_t - \hat{\mu}_i^{(m+1)})^2 \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t) = 0. \end{aligned} \tag{2.70}$$

Expanding the quadratic term $(x_t - \hat{\mu}_i^{(m+1)})^2$ and using (2.47) and (2.49) we obtain:

$$\hat{\Sigma}_i^{(m+1)} = \frac{1}{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})} \sum_{t=1}^{N} \int_{-\infty}^{\infty} x_t^2 \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t) - \left[\hat{\mu}_i^{(m+1)}\right]^2, \tag{2.71}$$

$$\hat{\Sigma}_i^{(m+1)} = \frac{\mathcal{S}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})} - [\hat{\mu}_i^{(m+1)}]^2. \tag{2.72}$$

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

$$\mathcal{R}(\alpha_i) = \sum_{i=1}^{\kappa} \log[\alpha_i] \left\{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})\right\}, \tag{2.73}$$

subject to

$$\sum_{i=1}^{\kappa} \alpha_i = 1. \tag{2.74}$$

Using a Lagrange Multiplier and optimizing with respect to α_i, we then obtain:

$$\mathcal{J}(\alpha_i, \epsilon) = \sum_{i=1}^{\kappa} \log[\alpha_i] \left\{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})\right\} - \epsilon \left(\sum_{i=1}^{\kappa} \alpha_i - 1\right). \tag{2.75}$$

Taking the derivative of $\mathcal{J}(\alpha_i, \epsilon)$ with respect to α_i and ϵ, then equating to zero we obtain:

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \alpha_i} = \frac{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}{\hat{\alpha}_i^{(m+1)}} - \epsilon = 0, \tag{2.76}$$

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \epsilon} = \sum_{i=1}^{\kappa} \alpha_i - 1 = 0. \tag{2.77}$$

Then,

$$\hat{\alpha}_i^{(m+1)} = \frac{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}{\epsilon}. \tag{2.78}$$

If we sum over κ in (2.78) and use (2.77) we have:

$$\sum_{i=1}^{\kappa} \hat{\alpha}_i^{(m+1)} = \sum_{i=1}^{\kappa} \frac{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}{\epsilon} = 1, \tag{2.79}$$

$$\epsilon = \sum_{i=1}^{\kappa} \mathcal{P}(x_t, \hat{\beta}_i^{(m)}). \tag{2.80}$$

Finally we obtain:

$$\hat{\alpha}_i^{(m+1)} = \frac{\mathcal{P}(x_t, \hat{\beta}_i^{(m)})}{\sum_{l=1}^{\kappa} \mathcal{P}(x_t, \hat{\beta}_l^{(m)})}. \tag{2.81}$$

This completes the proof.

2.B EM-based algorithm formulation with Maxwellian mixture distributions

2.B.1 Computing the Auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

From (2.61) and (2.62) we obtain the likelihood function in (2.15) as follows:

$$\ell(\beta) = \sum_{t=1}^{N} \log\left[\mathcal{V}_t(\beta)\right], \tag{2.82}$$

with

$$\mathcal{V}_t(\beta) = \sum_{i=1}^{\kappa} \int_{y_t}^{\infty} \mathcal{K}(x_t, \beta_i) d\mu(x_t). \tag{2.83}$$

By defining $\mathcal{B}_t(\beta) = \log\left[\mathcal{V}_t(\beta)\right]$, we can follow a similar analysis as in *Section 2.5*, obtaining

$$\mathcal{B}_t(\beta) = \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\beta, \hat{\beta}^{(m)}), \tag{2.84}$$

where

$$\mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{y_t}^{\infty} \log\left[\mathcal{K}(x_t, \beta_i)\right] \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t), \tag{2.85}$$

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{y_t}^{\infty} \log\left[\frac{\mathcal{K}(x_t, \beta_i)}{\mathcal{V}_t(\beta)}\right] \frac{\mathcal{K}(x_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})} d\mu(x_t). \tag{2.86}$$

From Lemma 2, the function $\mathcal{H}_t(\beta, \hat{\beta}^{(m)})$ is a decreasing function for any value of β and satisfies the following:

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\hat{\beta}^{(m)}, \hat{\beta}^{(m)}) \leq 0. \tag{2.87}$$

From this inequality inspired by the EM algorithm, we can formulate the following iterative algorithm:

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{t=1}^{N} \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}), \tag{2.88}$$

$$\hat{\beta}^{(m+1)} = \arg\max_{\beta} \ \mathcal{Q}(\beta, \hat{\beta}^{(m)}), \quad \text{s.t.} \quad \sum_{i=1}^{\kappa} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1. \tag{2.89}$$

Notice that (2.88) and (2.89) correspond to the *E-step* and *M-step* of the EM algorithm, respectively.

2.B.2 Optimizing the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

Taking derivative of $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ with respect to $\rho = 1/\sigma_i^2$ and equating to zero we obtain

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \rho} = \frac{3}{2\hat{\rho}^{(m+1)}} \sum_{t=1}^{N} \int_{y_t}^{\infty} \frac{\mathcal{K}\left(x_t, \hat{\beta}_i^{(m)}\right)}{\mathcal{V}_t\left(\hat{\beta}^{(m)}\right)} d\mu(x_t) - \frac{1}{2} \int_{y_t}^{\infty} x_t^2 \frac{\mathcal{K}\left(x_t, \hat{\beta}_i^{(m)}\right)}{\mathcal{V}_t\left(\hat{\beta}^{(m)}\right)} d\mu(x_t) = 0. \tag{2.90}$$

Using $\mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})$ and $\mathcal{S}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})$ from (2.64) and (2.65) in (2.90), we have:

$$\hat{\sigma}_i^{2^{(m+1)}} = \frac{1}{3\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})} \sum_{t=1}^{N} \int_{y_t}^{\infty} x_t^2 \frac{\mathcal{K}\left(x_t, \hat{\beta}_i^{(m)}\right)}{\mathcal{V}_t\left(\hat{\beta}^{(m)}\right)} d\mu(x_t), \tag{2.91}$$

$$\hat{\sigma}_i^{(m+1)} = \left[\frac{\sum_{t=1}^{N} \mathcal{S}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{3\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}\right]^{\frac{1}{2}}. \tag{2.92}$$

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

$$\mathcal{R}(\alpha_i) = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \log\left[\alpha_i\right] \left\{\mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})\right\}, \tag{2.93}$$

subject to

$$\sum_{i=1}^{\kappa} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1. \tag{2.94}$$

Then, using a Lagrange multiplier we define:

$$\mathcal{J}(\alpha_i, \epsilon) = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \log\left[\alpha_i\right] \left\{\mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})\right\} - \epsilon \left(\sum_{i=1}^{\kappa} \alpha_i - 1\right). \tag{2.95}$$

Algorithm 2.2 *(Estimation algorithm for stellar velocities using MSA)*

Inputs $y_{1:N}$, κ, $\hat{\beta}^{(0)}$.
Outputs $\hat{\beta}$.

1: $i \leftarrow 1$
2: **while** $(i \leq \kappa)$ **do**
3: $m \leftarrow 0$
4: Compute $\hat{\alpha}_i^{(m+1)}$ from (2.59).
5: Compute $\hat{\sigma}_i{}^{(m+1)}$ from (2.60).
6: **if** stopping criterion is not satisfied **then**
7: $m \leftarrow m + 1$,
8: **return** to 4.
9: **else**
10: $\hat{\alpha}_i = \hat{\alpha}_i^{(m+1)}$, $\hat{\sigma}_i = \hat{\sigma}_i^{(m+1)}$.
11: **end if**
12: $\hat{\beta}_i = [\hat{\alpha}_i \ \hat{\sigma}_i]$.
13: $i \leftarrow i + 1$
14: **end while**
15: $\hat{\beta} \leftarrow [\hat{\beta}_1 \ \cdots \ \hat{\beta}_\kappa]$.
16: End

Taking the derivative of $\mathcal{J}(\alpha_i, \epsilon)$ with respect to α_i and ϵ, then equating to zero we obtain

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \alpha_i} = \frac{\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{\hat{\alpha}_i^{(m+1)}} - \epsilon = 0, \tag{2.96}$$

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \epsilon} = \sum_{i=1}^{\kappa} \alpha_i - 1 = 0. \tag{2.97}$$

Then,

$$\hat{\alpha}_i^{(m+1)} = \frac{\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{\epsilon}. \tag{2.98}$$

If we sum over κ in (2.98) and use (2.97) we have

$$\sum_{i=1}^{\kappa} \hat{\alpha}_i^{(m+1)} = \sum_{i=1}^{\kappa} \frac{\sum_{t=1}^{N} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{\epsilon} = 1, \tag{2.99}$$

$$\epsilon = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)}). \tag{2.100}$$

Finally, substituting (2.100) in (2.98), we obtain

$$\hat{\alpha}_i^{(m+1)} = \frac{\mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_i^{(m)})}{\sum_{t=1}^{N} \sum_{l=1}^{\kappa} \mathcal{P}_{\mathcal{M}}(x_t, \hat{\beta}_l^{(m)})}. \tag{2.101}$$

Notice that $0 \leq \hat{\alpha}_i^{(m+1)} \leq 1$ holds, even though we did not explicitly consider it in (2.95).

Finally, the proposed estimation procedure is summarized in *Algorithm 2.2*.

2.C Estimation with Tikhonov regularization method

In this section, we show the TRM proposed in [64] to estimate the PDF of the *true* stellar rotational velocities directly from the Fredholm integral (2.53). The integral equation (2.53) can be expressed in matrix form as follows:

$$Y = AX, \tag{2.102}$$

where A is a matrix that represents the kernel $p(y|x)$, Y is a vector representing the density of projected rotational velocities $p(y)$, and X is the unknown vector representing the density of *true* rotational velocities $p(x)$. A kernel density estimator is used to obtain an estimation of $p(y)$ [99], that is, it is used to obtain a non-parametric representation of the PDF of projected velocities. This estimation is sensitive to the smoothing of the function and the bandwidth value since these features control the smoothness of the resulting density distribution.

A standard method to solve (2.102) is to apply ordinary Least Squares (OLS), that is, $\min\{\|AX - Y\|^2\}$, where $\|\cdot\|$ represents the euclidean norm. However, for ill-posed problems, this method fails in the sense that it can produce unstable estimators. In order to avoid this problem the TRM imposes a regularization term to be included in the minimization problem as follows:

$$\min_X \|AX - Y\|^2 + \lambda^2 \|L(X - X_0)\|^2, \tag{2.103}$$

where λ is the Tikhonov factor. The standard definition for the L matrix is $L = I_X$, where I_X is the identity matrix and X_0 is an initial estimation, setting $X_0 = 0$ when there is no previous information. There exist different quantitative approaches to obtain the Tikhonov factor such as Generalized Cross-Validation (GCV), L-curve Method, Discrepancy Principle, and Restricted Maximum likelihood (see e.g. [91, 100, 101] and the references therein). In particular, in [64] an iterative procedure to obtain the value of the Tikhonov factor was proposed, which provided a Tikhonov estimator unbiased and consistent in the case of smooth solutions.

Remark 5. *Since the observed data y in (2.53) are measured with error, the matrix form (2.102) is an example of a discrete ill-conditioned problem, that is, small errors in the measured data can produce large variations in the recovered function which makes the solution unstable (see e.g. [102] and the references therein).* ▽

Chapter 3
Maximum Likelihood Estimation of Linear Dynamic Systems with Gaussian Mixture Noise Distribution

In this chapter, a Maximum Likelihood estimation algorithm for a linear dynamic system driven by an exogenous input signal, with non-minimum-phase noise transfer function and a Gaussian mixture noise is developed. We propose a flexible identification technique to estimate the system model parameters and the Gaussian mixture parameters based on the EM algorithm with GMMs. The benefits of the proposed algorithm are illustrated via numerical simulations. This chapter summarizes the results of the journal paper **J.4**.

Contribution

We obtain the likelihood function using the prediction error for a general class of linear system driven by an exogenous input signal, with a non-minimum-phase noise transfer function and a Gaussian mixture noise. The prediction error is computed by using causal and anti-causal filtering techniques considering its corresponding initial conditions as deterministic parameters to be estimated. We propose an EM algorithm to solve the associated ML estimation problem with GMMs, obtaining the estimates of the system model parameters and closed form expressions for the GMM estimators.

3.1 System of interest

The system of interest is as follows:

$$y_t = G(z^{-1}, \theta)u_t + H(z^{-1}, \theta)\omega_t, \tag{3.1}$$

where θ is the vector that parameterized the system model, z^{-1} should be understood as the backward-shift operator ($z^{-1}u_t = u_{t-1}$) or the z-transform variable, $u_t \in \mathbb{R}$ is a deterministic input signal, $y_t \in \mathbb{R}$ is the output signal of the system, and $\omega_t \in \mathbb{R}$ is an independent and identically distributed noise sequence with a Gaussian mixture distribution given by:

$$p(\omega_t) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(\omega_t; \mu_i, \Sigma_i), \tag{3.2}$$

where $\kappa \geq 2$ ($\kappa \in \mathbb{N}$) is assumed in order to have non-Gaussian noise ω_t, $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^{\kappa} \alpha_i = 1$ and $\mathcal{N}(\omega_t; \mu_i, \Sigma_i)$ represents a Gaussian PDF with mean value μ_i and covariance matrix $\Sigma_i > 0$. The system transfer function in (3.1) is given by:

$$G(z^{-1}, \theta) = \frac{B(z^{-1}, \theta)}{A(z^{-1}, \theta)}, \tag{3.3}$$

where

$$A(z^{-1},\theta) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}, \tag{3.4}$$
$$B(z^{-1},\theta) = b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}. \tag{3.5}$$

Similarly, the noise transfer function in (3.1) is given by:

$$H(z^{-1},\theta) = H_s(z^{-1},\theta) H_u(z^{-1},\theta), \tag{3.6}$$

where $H_s(z^{-1},\theta)$ and $H_u(z^{-1},\theta)$ are the minimum-phase and non-minimum-phase transfer functions of $H(z^{-1},\theta)$ respectively, and they are given by:

$$H_s(z^{-1},\theta) = \frac{C_s(z^{-1},\theta)}{D(z^{-1},\theta)}, \tag{3.7}$$
$$H_u(z^{-1},\theta) = C_u(z^{-1},\theta), \tag{3.8}$$

with

$$C_s(z^{-1},\theta) = 1 + c_{s_1} z^{-1} + \cdots + c_{s_r} z^{-r}, \tag{3.9}$$
$$C_u(z^{-1},\theta) = 1 + c_{u_1} z^{-1} + \cdots + c_{u_p} z^{-p}, \tag{3.10}$$
$$D(z^{-1},\theta) = 1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d}, \tag{3.11}$$

$n_c = r + p$, all zeros corresponding to $C_s(z^{-1},\theta)$ are inside the unit circle, and all zeros corresponding to $C_u(z^{-1},\theta)$ lie outside the unit circle.

3.2 Estimation approaches for linear dynamic systems

The problem of interest is to estimate the vector of parameters β that defines the parameters of the system model and the GMM in (3.1). In addition, we consider that β_0 is the *true* vector of parameters that defines the *true* model. In order to formulate the ML estimation algorithm with GMMs, we introduce the following standing assumptions:

- **A1** The general system in (3.1) is operating in open loop and the input signal u_t is an exogenous deterministic signal.
- **A2** The vector of parameters β_0, the input u_t and the noise ω_t in (3.1) satisfy regularity conditions, guaranteeing that the ML estimate $\hat{\beta}_{\text{ML}}$ converges (in probability or a.s.) to the *true* solution β_0 as $N \to \infty$.
- **A3** The orders n_a, n_b, r, p, and n_d of the polynomials of system (3.1), and the number of components κ of the noise sequence distribution (3.2) are known.
- **A4** System (3.1) is asymptotically stable, its transfer functions $G(z^{-1},\theta)$ and $H(z^{-1},\theta)$ have no poles-zeros on the unit circle and have no pole-zero cancellations.

Assumption **A2** is necessary to develop an estimation algorithm that holds the asymptotic properties of the ML estimator. The non-under-modeling restriction (Assumption **A3**) can be relaxed using an information criteria to determine the correct number of both minimum-phase zeros and non-minimum-phase zeros for the noise transfer function in (3.1). In this

approach, if the number of zeros $n_c = r + p$ of the noise transfer function is given, the number of combinations of minimum-phase zeros and non-minimum-phase zeros that must be tested is $n_c + 1$. Assumption **A4** is necessary to obtain an asymptotically unbiased ML estimator [103] and a system model that is controllable, for example, using adaptive control techniques for non-minimum-phase systems [104].

In addition, we consider the following estimation approaches for the system of interest in (3.1):

3.2.1 Prediction Error Method (PEM)

In general, most systems in nature are stochastic, that is, the output of the system at time t cannot be determined exactly from the measurements available at time $t-1$ [1–3]. In many applications, the model is used for prediction. Then, it makes sense that to estimate the vector of parameters θ of a model such that the prediction error, $\varepsilon_t(\theta)$, is small [1]. The prediction error can be defined as follows:

$$\varepsilon_t(\theta) = y_t - \hat{y}_{t|t-1}, \tag{3.12}$$

where $\hat{y}_{t|t-1}$ denotes the prediction of the output signal y_t given the data up to and including the time $t-1$ (e.g. for a system driven by exogenous input signal u_t, $\hat{y}_{t|t-1}$ corresponds to $\{y_{t-1}, u_{t-1}, y_{t-2}, u_{t-2}, ...\}$). Notice that the predictor $\hat{y}_{t|t-1}$ also depends on the vector of parameters θ.

For system (3.1) a general linear predictor is given by [1, 2]:

$$\hat{y}_{t|t-1} = L_1(z^{-1}, \theta)y_t + L_2(z^{-1}, \theta)u_t, \tag{3.13}$$

where the predictor $\hat{y}_{t|t-1}$ is a function of the past measurements only if the predictors filters, $L_1(z^{-1}, \theta)$ and $L_2(z^{-1}, \theta)$, satisfy the following:

$$L_1(0, \theta) = L_2(0, \theta) = 0. \tag{3.14}$$

The predictor $\hat{y}_{t|t-1}$ in (3.13) can be obtained based on the system model structure of (3.1). Then, the predictor filters $L_1(z^{-1}, \theta)$ and $L_2(z^{-1}, \theta)$ can be chosen such that (3.13) is the optimal mean-square predictor [1]. For system (3.1) the prediction error $\varepsilon_t(\theta)$ is given by [1, 2]:

$$\varepsilon_t(\theta) = \frac{1}{H(z^{-1}, \theta)}y_t + \frac{G(z^{-1}, \theta)}{H(z^{-1}, \theta)}u_t. \tag{3.15}$$

Finally, the criterion which maps the sequence of prediction errors, $\{\varepsilon_1, \varepsilon_2, ...\varepsilon_N\}$, into a scalar-valued function can be chosen as follows [2]:

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_t(\theta)^2. \tag{3.16}$$

The PEM estimator can be then obtained as follows:

$$\hat{\theta}^{(\text{PEM})} = \arg \min_{\theta} V_N(\theta). \tag{3.17}$$

Algorithm 3.1 *(PEM algorithm for linear dynamic systems)*

Inputs $y_{1:N}$, $u_{1:N}$.
Outputs $\hat{\theta}^{(\mathrm{PEM})}$.

1: Compute the prediction error $\varepsilon_t(\theta)$ in (3.15).
2: Compute the cost function $V_N(\theta)$ in (3.16).
3: Solve the optimization problem in (3.17) to obtain $\hat{\theta}^{(\mathrm{PEM})}$.
4: End

In most cases the estimation problem in (3.17) cannot be solved analytically. Then, the minimization problem in (3.17) can be performed by using classical optimization algorithms such as Gauss-Newton algorithm or Newton-Raphson algorithm, to mention a few. The PEM estimation is summarized in *Algorithm 3.1*.

Remark 6. *Notice that the predictor $\hat{y}_{t|t-1}$ is a mean square optimal predictor when the noise sequence ω_t in (3.1) is zero-mean Gaussian distributed [1, 2], i.e., the PEM estimator (3.17) is asymptotically statistically efficient. In addition, the prediction error $\varepsilon_t(\theta)$ needs to be stable, i.e., the noise transfer function $H(z^{-1}, \theta)$ in (3.1) should only have minimum-phase zeros ($H(z^{-1}, \theta) = H_s(z^{-1}, \theta)$, with $H_u(z^{-1}, \theta) = 1$).* ▽

3.2.2 High Order Moments (HOM) method

A particular class of the system of interest corresponds to ARMA models. In order to analyze the benefits of our proposal, here we consider the High Order Moment (HOM) method proposed in [44] for comparison purposes. The authors in [44] consider ARMA system models with non-minimum-phase zeros since they have been used in many applications (see e.g. [32, 33]). The key idea is to estimate the system model parameters which yields the best least-squares match between the theoretical cumulant function and the sampled cumulant function obtained from the measurements. In addition, the authors in [44], assume that the model order is known, and a non-minimum-phase purely stochastic system is considered as follows:

$$y_t = H(z^{-1}, \theta)\omega_t + v_t, \tag{3.18}$$

where $y_t \in \mathbb{R}$ is the system output, $H(z^{-1}, \theta)$ is a non-minimum-phase transfer function (see (3.6)), $\omega_t \in \mathbb{R}$ is a zero-mean, independent and identically distributed (i.i.d.) non-Gaussian sequence with variance Σ_ω, and $v_t \in \mathbb{R}$ is a zero-mean i.i.d. Gaussian sequence with variance Σ_v. For the problem of interest, the distribution of ω_t is given by a GMM (3.2). The noise transfer function, $H(z^{-1}, \theta)$, can be expressed as follows:

$$H(z^{-1}, \theta) = \frac{1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}}{1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d}}. \tag{3.19}$$

The HOM estimation procedure is carried out in two steps as follows [44]:

(1) *Estimation of an spectrally equivalent model*: Although one may use any method that yields consistent estimates of an spectrally equivalent model, a PEM estimation, $\hat{\theta}^{(\mathrm{PEM})}$, is used to obtain an spectrally equivalent model with all the poles inside the unit circle

from the data measurements $y_{1:N}$. The power spectral density (PSD) of the noise-free system in (3.18) is given by:

$$\bar{\Phi}(z) = \Sigma_\omega(\hat{\theta}^{(\text{PEM})})H(z,\hat{\theta}^{(\text{PEM})})H(z^{-1},\hat{\theta}^{(\text{PEM})}). \tag{3.20}$$

Therefore, there are at most 2^{n_c} possible combinations of the zeros which yield, up to a constant, the same PSD in (3.20). Let Θ denote the set of all possible combinations of the vector of parameters derived from $\hat{\theta}^{(\text{PEM})}$ by considering all possible locations of the zeros for the spectrally equivalent model. Notice that the elements of Θ differ in the coefficients c_i in (3.19) and Σ_ω since the estimation of Σ_ω depends on the estimation of the coefficients c_i in order to obtain an identical PSD in (3.20).
Summarizing, a finite set Θ containing at most 2^{n_c} elements is obtained. It can be defined as follows:

$$\Theta = \{\theta : \Phi(z,\theta) = \Phi(z,\hat{\theta}^{(\text{PEM})})\}, \tag{3.21}$$

where

$$\Phi(z,\theta) = \Sigma_\omega(\theta)H(z,\theta)H(z^{-1},\theta) + \Sigma_v(\theta). \tag{3.22}$$

(2) *Determination of the correct zeros*: Let λ denotes an arbitrary element of Θ. All λ's lead to the same set of poles and measurement noise variance Σ_v. The key idea is to obtain a λ from the set Θ which yields the best least square error between the theoretical cumulant function and the sampled cumulant function obtained from the measurements $y_{1:N}$. In [105] is presented estimators to compute the high-order cumulant functions of sampling distributions. These estimators of the cumulant functions are asymptotically unbiased and consistent, i.e., a large number of measurements are needed in order to obtain accurate estimations [44, 56]. Based on this fact, the authors in [44] consider the fourth-order cumulant, ν_4, to formulate a least square minimization problem as follows:

$$\{\hat{\lambda},\hat{\nu}_4\} = \min_{\lambda,\nu_4} J(\lambda,\nu_4), \tag{3.23}$$

with

$$J(\lambda,\nu) = \sum_{t_1=-L}^{0}\sum_{t_2=-L}^{0}\sum_{t_3=-L}^{0}\left[\mathfrak{C}_4(t_1,t_2,t_3|\lambda,\nu_4) - \hat{\mathfrak{C}}_4(t_1,t_2,t_3)\right]^2, \tag{3.24}$$

$$\begin{aligned}\hat{\mathfrak{C}}_4(t_1,t_2,t_3) = \frac{1}{N}\left[\sum_{t=n}^{N} y_t y_{t+t_1} y_{t+t_2} y_{t+t_3}\right] - \hat{\mathfrak{R}}_2(t_1)\hat{\mathfrak{R}}_2(t_3-t_2) - \\ \hat{\mathfrak{R}}_2(t_2)\hat{\mathfrak{R}}_2(t_3-t_1) - \hat{\mathfrak{R}}_2(t_3)\hat{\mathfrak{R}}_2(t_2-t_1),\end{aligned} \tag{3.25}$$

where $L >> 2n_c$, ν_4 is fourth order cumulant, $\mathfrak{C}_4(t_1,t_2,t_3|\lambda,\nu_4)$ is the theoretical fourth order cumulant function, and $\hat{\mathfrak{C}}_4(t_1,t_2,t_3)$ is the fourth order cumulant function estimated as follows where $n = \max(0,-t_1,-t_2,-t_3)$, $t_1,t_2,t_3 \leq 0$, and

$$\hat{\mathfrak{R}}_2(k) = \frac{1}{N}\sum_{t=-k}^{N} y_t y_{t+k}, \quad k \leq 0. \tag{3.26}$$

The theoretical fourth-order cumulant, $\mathfrak{C}_4(t_1,t_2,t_3|\lambda,\nu_4)$, in (3.24) can be obtained

from the impulse response function of the system model (3.18) (see e.g. [106,107]). Then, for the fourth-order cumulant, $\hat{\nu}_4$, we take the derivative in (3.24) with respect to ν_4, with λ fixed, and equating to zero a closed form expression for $\hat{\nu}_4(\lambda)$ is obtained as follows:

$$\hat{\nu}_4(\lambda) = \frac{\sum_{t_1=0}^{L}\sum_{t_2=0}^{L}\sum_{t_3=0}^{L}\hat{\mathfrak{C}}_4(t_1,t_2,t_3)\bar{\mathfrak{C}}_4(t_1,t_2,t_3|\lambda)}{\sum_{t_1=0}^{L}\sum_{t_2=0}^{L}\sum_{t_3=0}^{L}[\bar{\mathfrak{C}}_4(t_1,t_2,t_3|\lambda)]^2}, \tag{3.27}$$

where

$$\bar{\mathfrak{C}}_4(t_1,t_2,t_3|\lambda) = \frac{\mathfrak{C}_4(t_1,t_2,t_3|\lambda,\nu_4)}{\nu_4}. \tag{3.28}$$

Then, substituting (3.27) in (3.24) and concentrating the cost function on the parameter λ, we can obtain an estimate $\hat{\lambda}$ solving the estimation problem in (3.23).

Remark 7. *An extension of the HOM for a general class of system with an exogenous input signal as (3.1) was addressed in [56]. First, an spectrally equivalent model using PEM is obtained. Then, the location of the zeros is obtained utilizing the system* $y_t^{(aux)} = y_t - G(z^{-1},\hat{\theta}^{(PEM)})u_t$ *with the procedure described in the second step of the HOM.* ▽

3.2.3 Maximum Likelihood method with GMMs

In this section, we develop an ML estimation algorithm for system (3.1) which includes the computation of its corresponding prediction error [1, 2]. A procedure to compute the prediction error for purely stochastic non-minimum-phase systems was presented in [37]. Here, we extend this result in order to consider an exogenous input signal. The prediction error, $\varepsilon_t(\theta)$, is obtained using causal recursive filtering followed by anti-causal recursive filtering that involves time-reversing operations, that is, the time-reversed sequence $x_{1:N}^R$ of a sequence $x_{1:N}$ is obtained by flipping $x_{1:N}$ [108]. That is, $x_{1:N}^R = \{x_N, x_{N-1}, ..., x_1\}$.

In order to compute the prediction error, the system (3.1) is expressed as follows [37]:

$$y_t = G(z^{-1},\theta)u_t + H_s(z^{-1},\theta)\bar{H}_u(z,\theta)v_t, \tag{3.29}$$

where

$$\bar{H}_u(z,\theta) = 1 + \frac{c_{u_{p-1}}}{c_{u_p}}z + \cdots + \frac{1}{c_{u_p}}z^p, \tag{3.30}$$

and the PDF of the noise sequence $v_t = c_{u_p}\omega_{t-p}$ is given by a GMM in (3.2) with mixing weight α_i, mean $\tilde{\mu}_i$, and covariance matrix $\tilde{\Sigma}_i$ given by

$$\tilde{\mu}_i = c_{u_p}\mu_i, \qquad \tilde{\Sigma}_i = c_{u_p}^2\Sigma_i. \tag{3.31}$$

For system (3.29), it is well known that the prediction error $\varepsilon_k(\theta)$ is given by [37]:

$$\varepsilon_t(\theta) = \underbrace{\left[\frac{1}{\bar{H}_u(z,\theta)}\right]}_{\text{anti-causal filter}}\underbrace{\left[\frac{1}{H_s(z^{-1},\theta)}\right]}_{\text{causal filter}}\left[y_t - G(z^{-1},\theta)u_t\right], \tag{3.32}$$

where the anti-causal filtering can be computed using the following steps:

(i) Time-reverse the output response of the causal filter in (3.32) with $[y_t - G(z^{-1},\theta)u_t]$,

obtaining $\bar{\varepsilon}_t^R(\theta)$.

(ii) Filter the signal $\bar{\varepsilon}_t^R(\theta)$ with $[\bar{H}_u(z^{-1},\theta)]^{-1}$, obtaining $\varepsilon_t^R(\theta)$.

(iii) Time-reverse the output response $\varepsilon_t^R(\theta)$ to obtain $\varepsilon_t(\theta)$.

This procedure is based on the fact that if $[\bar{H}_u(z,\theta)]^{-1}$ in (3.32) is an anti-causal filter, then $[\bar{H}_u(z^{-1},\theta)]^{-1}$ is a causal filter, that is, $\bar{H}_u(z^{-1},\theta)$ is a monic polynomial with zeros equal to the zeros in (3.30) reflected into the unit disc (see e.g. [108, Chapter 4], [37]). In addition, time-reversing operations are used based on the fact that if the Z-transform of a real sequence x_t is $X(z)$, then the Z-transform of the time-reversed sequence x_{-t} is $X(z^{-1})$.

In our formulation, and inspired in [41], we include the initial conditions of the recursive filters in (3.32) as deterministic parameters to be estimated. We define the vector of parameters to be estimated as $\beta = [\theta^T\ \gamma^T\ \eta^T]^T$ in order to formulate the ML estimator for the system (3.1):

$$\theta = [a_1\ b_1\ c_{s_1}\ c_{u_1}\ d_1 \dots\ a_{n_a}\ b_{n_b}\ c_{s_r}\ c_{u_p}\ d_{n_d}]^T, \tag{3.33}$$

$$\gamma = [\alpha_1\ \mu_1\ \Sigma_1\ \dots\ \alpha_\kappa\ \mu_\kappa\ \Sigma_\kappa]^T, \tag{3.34}$$

$$\eta = [\eta^{g^T}\ \eta^{s^T}\ \eta^{u^T}]^T, \tag{3.35}$$

where θ is the vector of parameters of the system model, γ is the vector of parameters of the Gaussian mixture distribution of the noise sequence ω_k, and η defines the initial conditions of the corresponding prediction error recursive filters in (3.32) with $\eta^g \in \mathbb{R}^{\max(n_b,n_a)\times 1}$, $\eta^s \in \mathbb{R}^{\max(r,n_d)\times 1}$ and $\eta^u \in \mathbb{R}^{p\times 1}$. Initial conditions η are computed from the state-space representation $\{\mathcal{A}_s,\mathcal{C}_s,\mathcal{K}_s,\mathcal{D}_s\}$, $\{\mathcal{A}_g,\mathcal{C}_g,\mathcal{K}_g,\mathcal{D}_g\}$ and $\{\mathcal{A}_u,\mathcal{C}_u\mathcal{K}_u,\mathcal{D}_u\}$ of $[H_s(z^{-1},\theta)]^{-1}$, $G(z^{-1},\theta)$ and $[\bar{H}_u(z^{-1},\theta)]^{-1}$, respectively. That is, $[H_s(z^{-1},\theta)]^{-1} = \mathcal{D}_s + \mathcal{C}_s(zI-\mathcal{A}_s)^{-1}\mathcal{K}_s$, $G(z^{-1},\theta) = \mathcal{D}_g + \mathcal{C}_g(zI-\mathcal{A}_g)^{-1}\mathcal{K}_g$ and $[\bar{H}_u(z^{-1},\theta)]^{-1} = \mathcal{D}_u + \mathcal{C}_u(zI-\mathcal{A}_u)^{-1}\mathcal{K}_u$, where I denotes the identity matrix with appropriate dimensions. Given the above, the corresponding ML estimation algorithm using GMMs is obtained as follows:

Lemma 4. *Consider the vector of parameters to be estimated as $\beta = [\theta^T\ \gamma^T\ \eta^T]^T$ in (3.33)–(3.35). Under the standing assumptions, the ML estimator for system (3.1) is given by*

$$\hat{\beta}_{ML} = \arg\max_{\beta}\ \ell(\beta), \quad s.t.\ \ 0 \le \alpha_i \le 1,\ \sum_{i=1}^{\kappa}\alpha_i = 1, \tag{3.36}$$

where the log-likelihood function is given by

$$\ell(\beta) = \sum_{t=1}^{N}\log\left\{\sum_{i=1}^{\kappa}\alpha_i\mathcal{N}(\varepsilon_t(\theta,\eta);\tilde{\mu}_i,\tilde{\Sigma}_i)\right\}, \tag{3.37}$$

and $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ are defined in (3.31).

The prediction error, $\varepsilon_t(\theta,\eta)$, in (3.37) is obtained using the following steps:

(i) Compute a causal filter using $G(z^{-1},\theta)$ and $H_s(z^{-1},\theta)$:

$$f_t^g(\theta) = \mathcal{C}_g\mathcal{A}_g^t, \quad f_t^s(\theta) = \mathcal{C}_s\mathcal{A}_s^t, \tag{3.38}$$

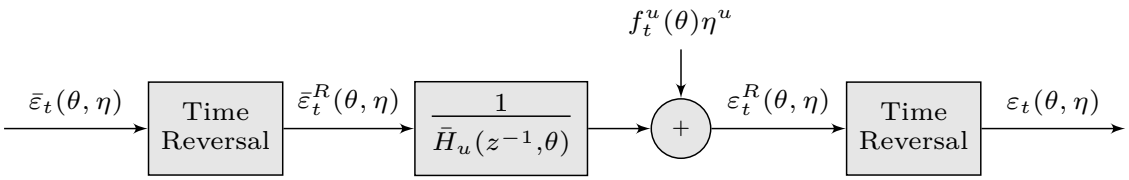


Figure 3.1: Processing scheme to address the anti-causal filtering to obtain $\varepsilon_t(\theta,\eta)$.

$$\bar{y}_t = G(z^{-1},\theta)u_t + f_t^g(\theta)\eta^g, \tag{3.39}$$

$$\bar{\varepsilon}_t(\theta,\eta) = [H_s(z^{-1},\theta)]^{-1}(y_t - \bar{y}_t) + f_t^s(\theta)\eta^s, \tag{3.40}$$

where $\{\mathcal{A}_g, \mathcal{C}_g\}$ *and* $\{\mathcal{A}_s, \mathcal{C}_s\}$ *are obtained from the state-space representation of* $G(z^{-1},\theta)$ *and* $[H_s(z^{-1},\theta)]^{-1}$*, respectively.*

(ii) Time-reverse the sequence $\bar{\varepsilon}_t(\theta,\eta)$ *in* (3.40) *and solving an anti-causal filter using* $\bar{H}_u(z^{-1},\theta)$*:*

$$f_t^u(\theta) = \mathcal{C}_u\mathcal{A}_u^t, \tag{3.41}$$

$$\varepsilon_t^R(\theta,\eta) = [\bar{H}_u(z^{-1},\theta)]^{-1}\bar{\varepsilon}_t^R(\theta,\eta) + f_t^u(\theta)\eta^u, \tag{3.42}$$

where $\{\mathcal{A}_u, \mathcal{C}_u\}$ *is obtained from the state-space representation of* $[\bar{H}_u(z^{-1},\theta)]^{-1}$*.*

(iii) Time-reverse the sequence $\varepsilon_t^R(\theta,\eta)$ *in* (3.42) *to obtain the prediction error* $\varepsilon_t(\theta,\eta)$*.*

Proof. See Appendix 3.A. □

Remark 8. *We can obtain the prediction error for noise transfer functions with only minimum-phase zeros considering* $\bar{H}_u(z^{-1},\theta) = 1$ *in* (3.42)*. Then, the prediction error is obtained using* (3.40) *straightforwardly with causal recursive filtering [1, 2], i.e., we have* $\varepsilon_t(\theta,\eta) = \bar{\varepsilon}_t(\theta,\eta)$*.*▽

The result obtained in Lemma 4 differs from the results shown in [37], where ARMA system models are considered with initial conditions fixed to zero and the ML estimation algorithm is developed considering a non-Gaussian noise distribution that is not modeled as a GMM. In contrast, from Lemma 4, the terms $f_t^g(\theta)$, $f_t^s(\theta)$, and $f_t^u(\theta)$ introduce the effect of initial conditions on the output filter responses in (3.39), (3.40), and (3.42), respectively. Figure 3.1 shows a block diagram of the anti-causal filtering utilized to obtain $\varepsilon_t(\theta,\eta)$, where the sequence $\bar{\varepsilon}_t(\theta,\eta)$ from (3.40) is time-reversed to obtain $\bar{\varepsilon}_t^R(\theta,\eta)$, and then filtered by using $[\bar{H}_u(z^{-1},\theta)]^{-1}$ to obtain $\varepsilon_t^R(\theta,\eta)$. Thus, the prediction error, $\varepsilon_t(\theta,\eta)$, is computed by time-reversing the sequence $\varepsilon_t^R(\theta,\eta)$.

As in the previous chapter, the optimization problem in (3.36) may be difficult to solve using gradient-based methods when the number of components in the GMM increases [80]. The EM algorithm [25] can provide a solution to overcome this difficulty due to the fact that the EM algorithm for GMMs typically provides closed form estimators for the Gaussian mixture parameters [67, 68].

3.3 An EM estimation algorithm with GMMs for linear dynamic systems

3.3.1 EM algorithm formulation

In this section, we consider the classical EM formulation with GMMs presented in Section 2.4.2 in order to solve the ML estimation problem in (3.36). To solve the estimation problem in (3.36), and inspired in [68], an EM algorithm [25, 70] with GMMs is developed from the definition of the likelihood function using the observed data $y_{1:N}$ and a *hidden* discrete random variable, $\zeta_{1:N}$, that is, defining the likelihood function using the *complete data*[1] [68]. Hence, the EM algorithm is given by (see e.g. [25, 68]):

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \mathbb{E}\left\{\log[p(y_{1:N}, \zeta_{1:N}|\beta)]|y_{1:N}, \hat{\beta}^{(m)}\right\}, \tag{3.43}$$

$$\hat{\beta}^{(m+1)} = \arg\max_{\beta} \mathcal{Q}(\beta, \hat{\beta}^{(m)}), \quad \text{s.t.} \quad \sum_{i=1}^{\kappa} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1, \tag{3.44}$$

where $\hat{\beta}^{(m)}$ is the current estimate, $p(y_{1:N}, \zeta_{1:N}|\beta)$ is the joint PDF of $y_{1:N}$ and $\zeta_{1:N}$, and $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ is the auxiliary function of the EM algorithm.

In order to develop the EM algorithm in (3.43) and (3.44), we first obtain the following result. This will be used to compute the auxiliary function in (3.43):

Lemma 5. *Consider the vector of parameters to be estimated* $\beta = [\theta^T \ \gamma^T \ \eta^T]^T$ *defined in* (3.33)–(3.35) *for system* (3.1). *The joint log-PDF,* $\log[p(y_{1:N}, \zeta_{1:N}|\beta)]$, *in* (3.43) *is given by*

$$\log[p(y_{1:N}, \zeta_{1:N}|\beta)] = \sum_{t=1}^{N}\sum_{i=1}^{\kappa} \log[p(\zeta_t = i|\beta)] + \log[\mathcal{N}(\varepsilon_t(\theta, \eta); \tilde{\mu}_i, \tilde{\Sigma}_i)], \tag{3.45}$$

where $\tilde{\mu}_i$, $\tilde{\Sigma}_i$ *and* $\varepsilon_t(\theta, \eta)$ *are given by* (3.31) *and* (3.42) *respectively.*

Proof. See Appendix 3.B □

In order to obtain a new estimate in (3.44), we solve the optimization problem using coordinate descent algorithm [109]. First, we fix the parameters $\{\theta, \eta\}$ at its value from the current iteration $\{\hat{\theta}^{(m)}, \hat{\eta}^{(m)}\}$ to optimize (3.43) with respect to the GMM parameters γ. Then, we fix the GMM parameters at its value from the iteration $\hat{\gamma}^{(m+1)}$ and to solve the optimization problem in (3.44) to obtain $\{\hat{\theta}^{(m+1)}, \hat{\eta}^{(m+1)}\}$. From Lemma 5, the proposed EM algorithm can be computed using the following:

Theorem 1. *Consider the vector of parameters to be estimated* $\beta = [\theta^T \ \gamma^T \ \eta^T]^T$ *defined from* (3.33)–(3.35) *for the system in* (3.1). *Substituting* (3.45) *in* (3.43), *the E-step in the EM algorithm is given by*

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{t=1}^{N}\sum_{i=1}^{M} \hat{\zeta}_{ti}^{(m)} \log\left[\alpha_i \mathcal{N}(\varepsilon_t(\theta, \eta); \tilde{\mu}_i, \tilde{\Sigma}_i)\right], \tag{3.46}$$

[1] The complete data $\{y_{1:N}, \zeta_{1:N}\}$ corresponds to the set defined by the observed data and the unobserved data. The discrete random variable $\zeta_{1:N}$ is defined as $\zeta_t \in \{1, \ldots, M\}$ for $t = 1, \ldots, N$.

where $\tilde{\mu}_i$ *and* $\tilde{\Sigma}_i$ *are given by* (3.31)*, and*

$$\hat{\zeta}_{ti}^{(m)} = \frac{\hat{\alpha}_i^{(m)} \mathcal{N}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}); \hat{\tilde{\mu}}_i^{(m)}, \hat{\tilde{\Sigma}}_i^{(m)})}{\sum_{l=1}^{M} \hat{\alpha}_l^{(m)} \mathcal{N}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}); \hat{\tilde{\mu}}_l^{(m)}, \hat{\tilde{\Sigma}}_l^{(m)})}. \tag{3.47}$$

Consider the maximization problem stated in (3.44) *using* (3.46)*. The M-step in the EM algorithm is carried out using the following steps:*

(i) Solving (3.44) *using* (3.46) *with* $\theta = \hat{\theta}^{(m)}$ *and* $\eta = \hat{\eta}^{(m)}$*:*

$$\hat{\alpha}_i^{(m+1)} = \frac{\sum_{t=1}^{N} \hat{\zeta}_{ti}^{(m)}}{N}, \tag{3.48}$$

$$\hat{\mu}_i^{(m+1)} = \frac{\sum_{t=1}^{N} \hat{\zeta}_{ti}^{(m)} \varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)})}{\hat{c}_{u_p}^{(m)} \sum_{t=1}^{N} \hat{\zeta}_{ti}^{(m)}}, \tag{3.49}$$

$$\hat{\Sigma}_i^{(m+1)} = \frac{\sum_{t=1}^{N} \hat{\zeta}_{ti}^{(m)} (\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}) - \hat{c}_{u_p}^{(m)} \hat{\mu}_i^{(m)})^2}{\left[\hat{c}_{u_p}^{(m)}\right]^2 \sum_{t=1}^{N} \hat{\zeta}_{ti}^{(m)}}. \tag{3.50}$$

(ii) Solving (3.44) *using* (3.46) *with* $\alpha_i = \hat{\alpha}_i^{(m+1)}$*,* $\mu_i = \hat{\mu}_i^{(m+1)}$ *and* $\Sigma_i = \hat{\Sigma}_i^{(m+1)}$*:*

$$\{\hat{\theta}^{(m+1)}, \hat{\eta}^{(m+1)}\} = \arg\min_{\theta,\eta} \ \mathcal{B}(\theta, \eta), \tag{3.51}$$

where

$$\mathcal{B}(\theta, \eta) = \sum_{t=1}^{N} \sum_{i=1}^{M} \frac{\hat{\zeta}_{ti}^{(m)} \left(\varepsilon_t(\theta, \eta) - \hat{c}_{u_p}^{(m)} \hat{\mu}_i^{(m+1)}\right)^2}{\left[\hat{c}_{u_p}^{(m)}\right]^2 \hat{\Sigma}_i^{(m+1)}}. \tag{3.52}$$

Proof. See Appendix 3.C □

Remark 9. *The iterative algorithm in Theorem 1 is developed from defining* $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ *in terms of the complete log-likelihood function in* (3.45)*. That differs from the procedure proposed in Section 2.5, where an auxiliary function is built without explicitly defining a hidden variable. However, the auxiliary function obtained in Section 2.5 is equal to* (3.46) *and the estimators are obtained from to* (3.48)–(3.51) *(see Appendix 3.D).* ▽

Finally, the estimation procedure is summarized in *Algorithm 3.2*.

3.3.2 An initialization procedure for the EM algorithm with GMMs

Initialization techniques for EM algorithms with GMMs have been studied in many research works (see e.g. [110–113]), showing that a careful initialization of the GMM parameters is important to obtain accurate estimates and to improve the rate of convergence. However, these approaches consider that a GMM describes the probabilistic model of the observed data. The problem we address here is different since we consider that a GMM describes the noise sequence distribution. From system (3.1), the prediction error in (3.42) provides statistical information of the Gaussian mixture noise distribution that can be used to obtain an initialization of the GMM parameters for the EM algorithm.

Algorithm 3.2 *(EM algorithm with GMMs)*

Inputs $y_{1:N}$, $u_{1:N}$, κ, $\hat{\theta}^{(0)}$, $\hat{\gamma}^{(0)}$ and $\hat{\eta}^{(0)}$.
Outputs $\hat{\theta}$, $\hat{\gamma}$ and $\hat{\eta}$.

1: $m \leftarrow 0$
2: **E-step:**
3: Compute $\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)})$ from (3.42).
4: **M-step:**
5: Estimate $\hat{\gamma}^{(m+1)}$ from (3.47)–(3.50).
6: Estimate $\{\hat{\theta}^{(m+1)}, \hat{\eta}^{(m+1)}\}$ by solving (3.51).
7: **if** stopping criterion is not satisfied **then**
8: $\quad m \leftarrow m + 1$,
9: $\quad$ **return** to 2
10: **else**
11: $\quad \hat{\theta} \leftarrow \hat{\theta}^{(m+1)}$, $\hat{\gamma} \leftarrow \hat{\gamma}^{(m+1)}$, $\hat{\eta} \leftarrow \hat{\eta}^{(m+1)}$
12: **end if**
13: End

To illustrate our initialization procedure, we consider a simple MA system in (3.1) with $H_s(z^{-1}, \theta) = 1$ and $H_u(z^{-1}, \theta) = 1 - 1.5z^{-1}$. The distribution of ω_k is given by a GMM with two components with $\alpha_1 = \alpha_2 = 0.5$, variances $\Sigma_1 = \Sigma_2 = 1$ and mean values $\mu_1 = -2$ and $\mu_2 = 2$. Then, the corresponding GMM parameters $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ are given by $\tilde{\Sigma}_1 = \tilde{\Sigma}_2 = 2.25$, $\tilde{\mu}_1 = -3$ and $\tilde{\mu}_2 = 3$. The PEM [1, 2] is used to estimate the system model parameters, obtaining $\hat{\theta}^{(\text{PEM})}$. In Figure 3.2 the gray bars represent the histogram of the prediction error obtained from (3.42) with $\theta = \hat{\theta}^{(\text{PEM})}$ and fixing initial conditions to zero. The red line is the *true* Gaussian mixture noise PDF and the black dashed line describes a Gaussian mixture PDF with parameters computed as follows:

(1) The initial value for the system model parameters, $\hat{\theta}^{(0)}$, is obtained using PEM.
(2) The initial values for the initial conditions, $\hat{\eta}^{(0)}$, are fixed to zero.
(3) The initial guess of mean values $\hat{\mu}_i^{(0)}$ are evenly spaced between the maximum and minimum value for the prediction error $\varepsilon_t(\hat{\theta}^{(\text{PEM})})$.
(4) The variances $\hat{\Sigma}_i^{(0)}$ for each component of the GMM are equal to the sample variance of $\varepsilon_t(\hat{\theta}^{(\text{PEM})})$.
(5) The mixing weight $\hat{\alpha}_i^{(0)}$ for each component are set all equal as $\hat{\alpha}_i^{(0)} = 1/M$.

Using the prediction error histogram (see Figure 3.2), one could (intuitively) initialize the parameters of the Gaussian mixture noise model. It can be obtained, for example, using a classical EM algorithm with finite mixture models [68]. Our experience shows that this procedure provides an estimation of the system (3.1) with less EM iterations than using the initialization procedure previously described. However, this alternative initialization requires extra computation. The accuracy of the estimates of system (3.1) is very similar using both initialization procedures (see Appendix 3.E).

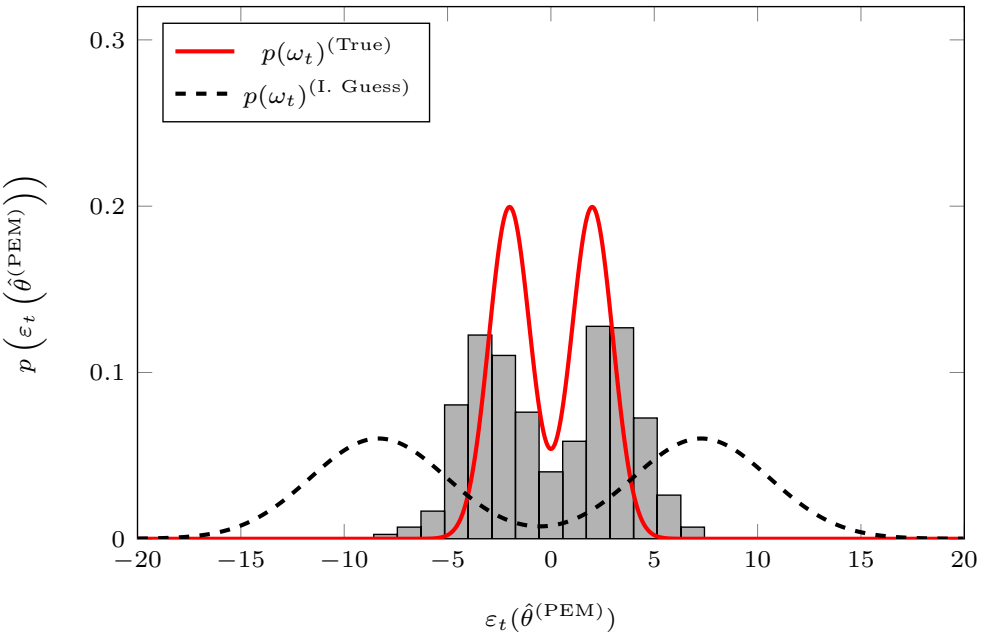


Figure 3.2: Histogram of the prediction error $\varepsilon_t(\hat{\theta}^{(\text{PEM})})$ obtained by using PEM.

3.4 Akaike information criterion for model structure determination

In the area of system identification both the model structure determination and model validation are important aspects, due to an over-parameterized model structure can lead unnecessarily computations in order to obtain the parameter estimates or well the system model obtained can be very inaccurate [1, 2]. In the following sections, we show how an information criterion is used to determine the correct number of both minimum-phase zeros and non-minimum-phase zeros for the noise transfer function in (3.1).

Akaike information criterion (AIC) [114] is a popular technique for evaluating how well a model fits the data it was generated from. In particular, if the ML method is used to estimate the system model parameters from a set of candidate system models and the models are non-nested or nested, AIC can provide a system model that explains the data measurements with good accuracy [115, Chapter 2], [116]. AIC can be utilized for the estimation problem in (3.36) as follows:

$$\{\hat{\beta}, \hat{\mathcal{M}}\} = \arg\min_{\beta,\mathcal{M}} \frac{1}{N}\left[-\ell(\beta) + d_{\mathcal{M}}\right], \tag{3.53}$$

where $\mathcal{M}$ is the model index, and $d_{\mathcal{M}}$ is the dimension of the vector of parameters β in the model $\mathcal{M}$.

In particular, the proposed algorithm can be used to estimate (3.1) with a non-minimum-phase noise transfer function when the standing assumptions are relaxed as follows: i) the noise sequence distribution does not correspond to a GMM, but can be approximated by one with known κ, and ii) the polynomials orders r and p in (3.1) are unknown, but their sum $n_c = r + p$ is known. We define $n_c + 1$ candidate system models that correspond to the number of all possible combinations of minimum-phase zeros and non-minimum-phase zeros of the noise transfer function, and we select the model according to AIC. Despite AIC is

not asymptotically consistent, it can be used to select a system model from a set of nested or non-nested system models in which the *true* system does not lie [115, 116]. This fact represents an advantage to obtain an estimation of the system (3.1) when a GMM is used to approximate a non-Gaussian noise distribution.

3.5 Numerical examples

Consider the following system:

$$y_t = \frac{b_1^o z^{-1}}{1 + a_1^o z^{-1}} u_t + \frac{1 + c_1^o z^{-1}}{1 + d_1^o z^{-1}} \omega_t, \tag{3.54}$$

where the *true* parameters are $b_1^o = 1$, $a_1^o = -0.5$, $d_1^o = 0.8$, $u_t \sim \mathcal{N}(0, \sigma_u^2)$, $\sigma_u^2 = 10$ and the noise signal ω_t is Gaussian mixture distributed.

We focus on two scenarios for (3.54) considering: i) a minimum-phase noise transfer function with $c_1^o = 0.1$, and ii) a non-minimum-phase noise transfer function with $c_1^o = 1.5$. The vector of parameters to estimate is given by $\beta = [\theta^T \ \gamma^T \ \eta^T]^T$, with $\theta = \{b_1, a_1, c_1, d_1\}$, $\gamma = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1}^{\kappa}$ and $\eta = \{\eta^g \ \eta^s \ \eta^u\}$.

The simulation setup is as follows:

(1) The data length is $N = 5000$.
(2) The number of Monte Carlo (MC) simulations are 100.
(3) The stopping criterion is satisfied when:

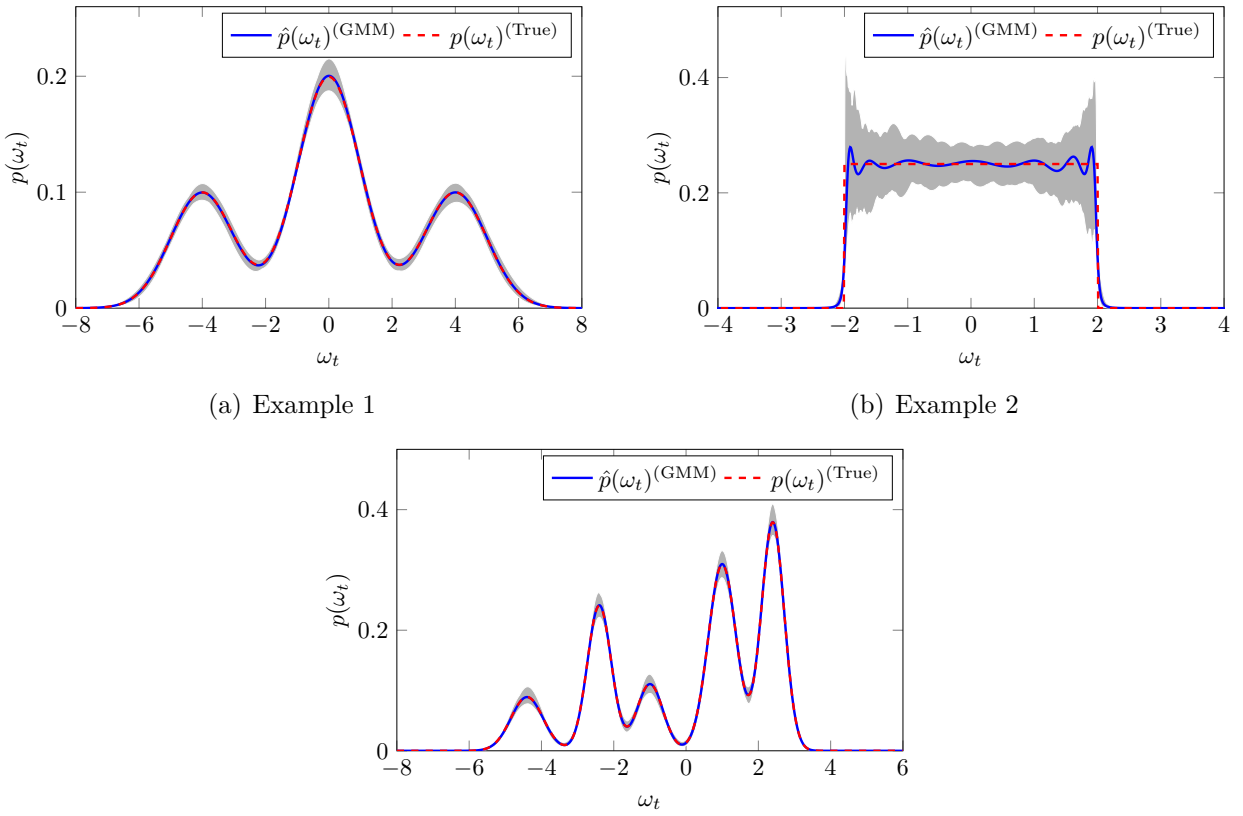
$$\frac{\left\| \hat{\beta}^{(m)} - \hat{\beta}^{(m-1)} \right\|}{\left\| \hat{\beta}^{(m)} \right\|} < 10^{-6},$$

or when 2000 iterations of the EM algorithm have been reached.

In order to show the benefits of the proposed algorithm, we consider three examples with different approaches to increment the difficulty to estimate the system (3.54). In a first example, a non-minimum-phase noise transfer function is considered in (3.54) holding all standing assumptions. We consider a three components Gaussian mixture noise with mean values well separated (non-overlapping Gaussian distributions). For comparison purposes, we also estimate the system model parameters using the methodology proposed for non-minimum-phase systems in [44] based on the HOM method.

In a second example, we show how the proposed EM algorithm can be used to estimate (3.54) with a non-minimum-phase noise transfer function when Assumption **A3** is relaxed as follows: i) the noise sequence distribution does not correspond to a GMM, but can be approximated by a GMM, and ii) the polynomials orders r and p in (3.54) are unknown, but their sum $n_c = r + p$ is known. We compare our estimation results with the estimation obtained using HOM. For our proposed algorithm, AIC [114] is used to obtain the correct number of both minimum-phase zeros and non-minimum-phase zeros of the noise transfer function.

Finally, a third example shows how our proposal can be used to estimate (3.54) with a minimum-phase noise transfer function and a Gaussian mixture noise. We consider a five



(a) Example 1

(b) Example 2

(c) Example 3

Figure 3.3: Estimated noise distribution using Gaussian mixture models.

components overlapped Gaussian mixture noise. Assumption **A3** is also relaxed considering that the polynomials orders r and p in (3.54) are unknown, but their sum $n_c = r + p$ is known. AIC is used to obtain the correct number of minimum-phase zeros of the noise transfer function in (3.54). For comparison analysis, we compare the outcome of the proposed method with PEM estimation [2].

3.5.1 Example 1: Non-minimum-phase dynamic system with a non-overlapped Gaussian mixture noise distribution

In this example we consider a non-minimum-phase noise transfer function in (3.54) holding all standing assumptions. The noise sequence ω_t is drawn from a GMM in (3.2) with $M = 3$, mean values $\mu_1 = -4$, $\mu_2 = 0$, $\mu_3 = 4$, variances $\Sigma_1 = \Sigma_2 = \Sigma_3 = 1$ and mixing weights $\alpha_1 = \alpha_3 = 0.25$, $\alpha_2 = 0.5$. Figure 3.3(a) shows the estimated average PDF for all MC realizations. The gray shaded region represents the area in which all the estimated probability density functions lie. We can observe that the average of the estimated GMM probability density functions fits the *true* noise signal distribution. Figure 3.4 shows the estimation results for the estimated system model. The large red cross indicates the *true* value for θ. We observe that the estimation using our proposal, namely EM–GMM, is better than using HOM algorithm. Table 3.1 shows the estimation results of the system model parameters for both EM–GMM and HOM algorithms. We observe that our proposal exhibits better

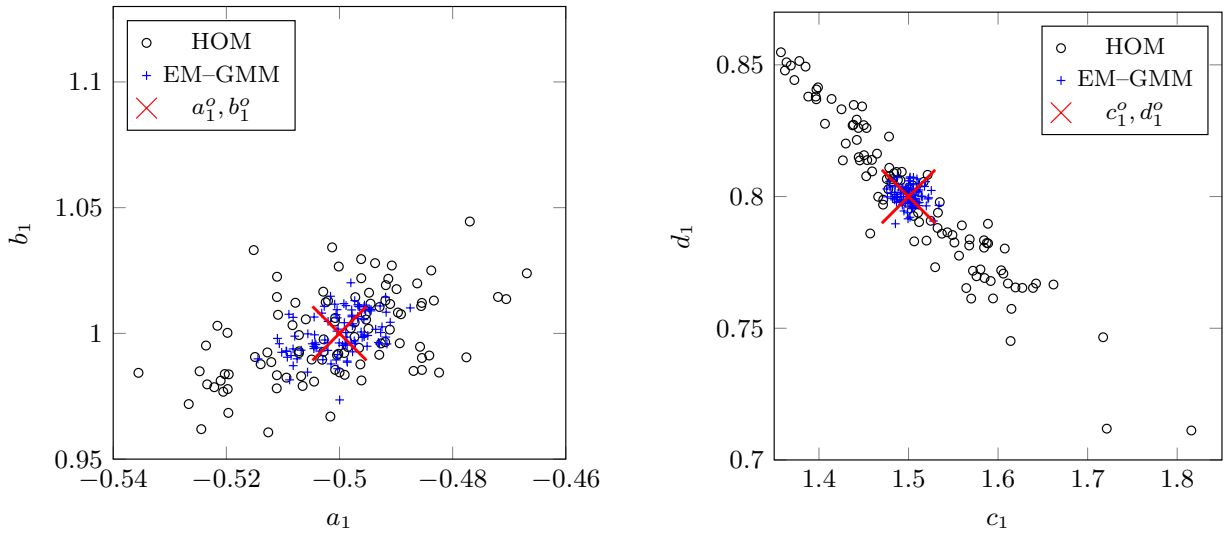


Figure 3.4: System model parameters estimated for the Example 1.

Table 3.1: Estimated system model parameters for numerical Example 1

Method	Parameters			
	$\hat{a}_1$	$\hat{b}_1$	$\hat{c}_1$	$\hat{d}_1$
HOM	$-0.5009 \pm 1.4 \times 10^{-2}$	$0.9992 \pm 1.7 \times 10^{-2}$	$1.5081 \pm 8.6 \times 10^{-2}$	$0.7994 \pm 2.9 \times 10^{-2}$
EM–GMM	$-0.5001 \pm 5.4 \times 10^{-3}$	$1.0001 \pm 8.5 \times 10^{-3}$	$1.5001 \pm 1.1 \times 10^{-2}$	$0.8002 \pm 3.6 \times 10^{-3}$

accuracy than HOM algorithm.

3.5.2 Example 2: Non-minimum-phase dynamic system with uniform distributed noise signal

In this example we show how our proposal can be used to estimate system (3.54) with a non-minimum-phase noise transfer function when the noise sequence distribution does not correspond to a GMM, but can be approximated by a GMM. We also consider that the polynomials orders r and p in (3.54) are unknown, but their sum $n_c = r + p$ is known. The noise input signal ω_t is distributed as follows:

$$p(\omega_t)^{(\text{True})} = \frac{1}{l_b - l_a} I_{[l_a, l_b]}(\omega_t), \tag{3.55}$$

where $l_a = -2$, $l_b = 2$, and $I_{[l_a,l_b]}$ is an indicator function given by

$$I_{[l_a,l_b]}(\omega_t) = \begin{cases} 1, & l_a \leq \omega_t \leq l_b, \\ 0, & \text{otherwise.} \end{cases}$$

To approximate the noise sequence distribution, we consider $\kappa = 7$ components for the GMM. Table 3.2 shows the AIC results for different orders of $C_s(z^{-1}, \theta)$ and $C_u(z^{-1}, \theta)$. We observe that the correct orders of the polynomials correspond to $r = 0$ and $p = 1$, i.e., one non-minimum-phase zero.

Figure 3.3(b) shows the estimated average PDF for all MC realizations. As in the previous

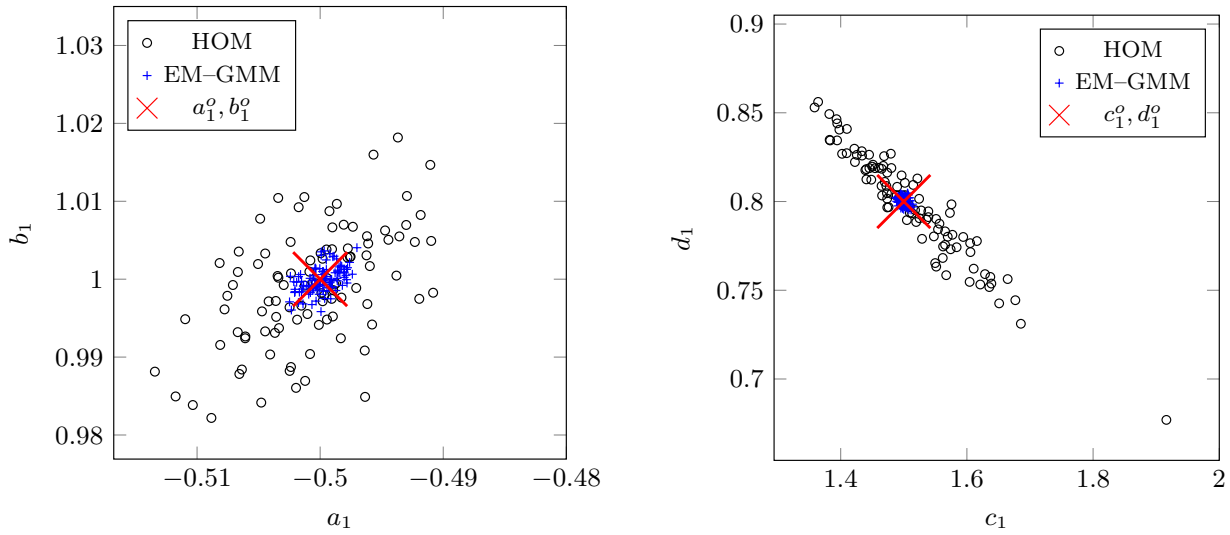


Figure 3.5: System model parameters estimated for the Example 2.

Table 3.2: AIC values for the numerical Examples 2 and 3

Order	AIC	
	Example 2	Example 3
$r = 1, p = 0$	1.9474	1.7993
$r = 0, p = 1$	1.7893	1.8751

example, the gray shaded region represents the area in which all the estimated probability density functions lie. We can observe that the average of the estimated GMM probability density functions fits the *true* noise signal distribution.

Figure 3.5 shows the estimation results for the estimated system model. The large red cross indicates the *true* value for θ. We observe that the estimation using EM–GMM exhibits better accuracy than using HOM. Table 3.3 shows the estimation results, confirming the benefits of our proposal in the estimation accuracy for the model parameters with respect to HOM algorithm. These results confirm that our proposed technique performs well even when the noise PDF is not a GMM, but can be approximated by one.

Remark 10. *The convergence properties of the EM algorithm have been discussed in [117] and for finite mixture models in [67, 110, 111]. The proposed iterative algorithm is developed from the classical definition of an EM algorithm with GMMs and then it holds the same convergence properties if the noise sequence distribution corresponds to a GMM. However, it can converge to good estimates if a GMM with an appropriate number of mixture components*

Table 3.3: Estimated system model parameters for numerical Example 2

Method	Parameters			
	$\hat{a}_1$	$\hat{b}_1$	$\hat{c}_1$	$\hat{d}_1$
HOM	$-0.5007 \pm 4.9 \times 10^{-3}$	$0.9990 \pm 7.3 \times 10^{-3}$	$1.5114 \pm 8.6 \times 10^{-2}$	$0.7982 \pm 2.9 \times 10^{-2}$
EM–GMM	$-0.4998 \pm 1.3 \times 10^{-3}$	$0.9999 \pm 1.7 \times 10^{-3}$	$1.5000 \pm 5.9 \times 10^{-3}$	$0.8000 \pm 2.3 \times 10^{-3}$

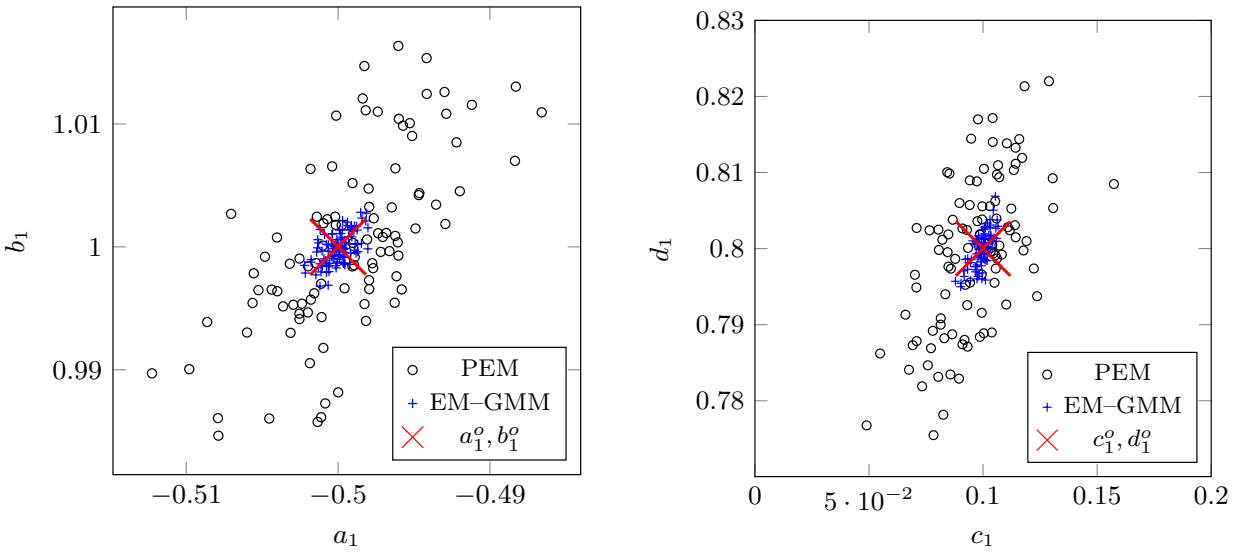


Figure 3.6: System model parameters estimated for the Example 3.

is used to approximate a non-Gaussian-sum noise sequence distribution. ▽

3.5.3 Example 3: Minimum-phase dynamic system with an overlapped Gaussian mixture noise distribution

In this example we consider a minimum-phase noise transfer function in (3.54) relaxing the Assumption **A3**. We assume that the polynomials orders r and p in (3.54) are unknown, but their sum $n_c = r + p$ is known. The noise sequence ω_t is drawn from a GMM in (3.2) with $\kappa = 5$, mean values $\mu_1 = -4.4$, $\mu_2 = -2.4$, $\mu_3 = -1$, $\mu_4 = 1$, $\mu_5 = 2.4$, variances $\Sigma_1 = 0.2$, $\Sigma_2 = 0.11$, $\Sigma_3 = 0.13$, $\Sigma_4 = 0.15$, $\Sigma_5 = 0.1$, and mixing weights $\alpha_1 = \alpha_3 = 0.1$, $\alpha_2 = 0.2$, $\alpha_4 = \alpha_5 = 0.3$. To approximate the noise signal distribution, we consider $\kappa = 5$ components for the GMM. Table 3.2 shows the AIC results for different orders of $C_s(z^{-1}, \theta)$ and $C_u(z^{-1}, \theta)$. We observe that using AIC the correct orders $r = 1$ and $p = 0$ are obtained, i.e. one minimum-phase zero. Figure 3.3(c) shows the results of the estimation of the noise distribution. The gray shaded region 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 as a GMM. It is clear that the average of all the estimated probability density functions is "close to" the *true* PDF. The results of the estimation of the system model parameters are shown in Figure 3.6. In both figures, the large red cross indicates the *true* value for θ. We observe that the estimation using the proposed method is more accurate than using PEM algorithm. Table 3.4 shows the mean and the corresponding standard deviations of all the estimated system model parameters obtained with both PEM and our proposed method. We observe an accurate estimation of the parameters using our proposal.

Table 3.4: Estimated system model parameters for numerical Example 3

Method	Parameters			
	$\hat{a}_1$	$\hat{b}_1$	$\hat{c}_1$	$\hat{d}_1$
PEM	$-0.4990 \pm 4.6 \times 10^{-3}$	$1.0011 \pm 7.2 \times 10^{-3}$	$0.0960 \pm 1.8 \times 10^{-2}$	$0.7980 \pm 1.1 \times 10^{-2}$
EM–GMM	$-0.4999 \pm 9.3 \times 10^{-4}$	$0.9999 \pm 1.2 \times 10^{-3}$	$0.0994 \pm 3.1 \times 10^{-3}$	$0.7999 \pm 2.1 \times 10^{-3}$

3.6 Conclusions

In this chapter, we have addressed the estimation of a general class of linear dynamic system driven by an exogenous input signal, with non-minimum-phase noise transfer function and a Gaussian mixture noise. We proposed an identification algorithm using ML to estimate the system model parameters and the GMM parameters of the noise distribution. The ML algorithm was formulated using the prediction error for systems with non-minimum-phase noise transfer functions that is computed using causal filtering and anti-causal filtering. We included the initial conditions of the filters of the prediction error as deterministic parameters to be estimated. The proposed EM algorithm utilizes closed form expressions to estimate the parameters of the GMM. In the simulation examples, we considered two scenarios: i) a non-minimum-phase noise transfer function holding all standing assumptions, and ii) relaxing the standing assumptions for a minimum-phase and a non-minimum-phase transfer functions with Gaussian mixture distributed noise signal and an uniform distributed noise signal, respectively. We obtained the correct number of minimum-phase zeros and non-minimum-phase zeros using AIC for all possible combinations of the orders r and p of the noise transfer function. For all cases, our proposal exhibited an accurate estimation of the system model parameters, specially in the estimation of the parameters of the noise transfer function. We also showed that GMMs can be used to approximate a non-Gaussian-sum noise distribution, obtaining accurate estimations of the system model parameters. Finally, the proposed EM algorithm and HOM method performed similarly when the signal to noise ratio is high, whilst the proposed EM algorithm exhibited better accuracy of the estimations than HOM when the signal to noise ratio is low.

Appendix

3.A Proof of Lemma 4

The prediction error in (3.32) can be obtained using causal recursive filtering followed by anti-causal recursive filtering [37] including the effect of initial conditions. From (3.32), the causal recursive filtering is obtained as follows:

$$\bar{y}_t = G(z^{-1}, \theta)u_t + f_t^g(\theta)\eta^g, \tag{3.57}$$

where η^g is the initial state of the state-space representation $\{\mathcal{A}_g, \mathcal{K}_g, \mathcal{C}_g, \mathcal{D}_g\}$ of $G(z^{-1}, \theta)$, and the term $f_t^g(\theta)$ in (3.57) is given by $f_t^g(\theta) = \mathcal{C}_g\mathcal{A}_g^t$.

Using (3.32) and (3.57), we obtain $\bar{\varepsilon}_t(\theta, \eta)$ as follows:

$$\bar{\varepsilon}_t(\theta, \eta) = \left[\frac{1}{H_s(z^{-1}, \theta)}\right](y_t - \bar{y}_t) + f_t^s(\theta)\eta^s, \tag{3.58}$$

where η^s is the initial state of the state-space representation $\{\mathcal{A}_s, \mathcal{K}_s, \mathcal{C}_s, \mathcal{D}_s\}$ of $[H_s(z^{-1}, \theta)]^{-1}$, and the term $f_t^s(\theta)$ in (3.58) is computed as $f_t^s(\theta) = \mathcal{C}_s\mathcal{A}_s^t$.

Then, the anti-causal recursive filtering is obtained using the time-reversed sequence of (3.58), $\bar{\varepsilon}_t^R(\theta, \eta)$, as follows:

$$\varepsilon_t^R(\theta, \eta) = \left[\frac{1}{\bar{H}_u(z^{-1}, \theta)}\right]\bar{\varepsilon}_t^R(\theta, \eta) + f_t^u(\theta)\eta^u, \tag{3.59}$$

where η^u is the initial state of the state-space representation $\{\mathcal{A}_u, \mathcal{K}_u, \mathcal{C}_u, \mathcal{D}_u\}$ of $[\bar{H}_u(z^{-1}, \theta)]^{-1}$, and the term $f_t^u(\theta)$ in (3.59) is obtained as $f_t^u(\theta) = \mathcal{C}_u\mathcal{A}_u^t$. Then, the prediction error $\varepsilon_t(\theta, \eta)$ is obtained by reversing the sequence $\varepsilon_t^R(\theta, \eta)$ in (3.59).

Assuming that ω_t in (3.1) is an independent and identically distributed noise sequence and using (3.33)–(3.35) to define the vector of parameters $\beta = [\theta^T \ \gamma^T \ \eta^T]^T$, the likelihood function $\mathcal{L}(\beta)$ is obtained from (3.29) as follows:

$$\begin{aligned}\mathcal{L}(\beta) &= p(y_{1:N}|\beta, u_{1:N}), \\ &= \prod_{t=1}^{N}\sum_{i=1}^{\kappa}\alpha_i\mathcal{N}(\varepsilon_t(\theta, \eta); \tilde{\mu}_i, \tilde{\Sigma}_i),\end{aligned} \tag{3.60}$$

where $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ are given by (3.31). Therefore, the log-likelihood function is given by

$$\ell(\beta) = \sum_{t=1}^{N}\log\left\{\sum_{i=1}^{\kappa}\alpha_i\mathcal{N}(\varepsilon_t(\theta, \eta); \tilde{\mu}_i, \tilde{\Sigma}_i)\right\}. \tag{3.61}$$

3.B Proof of Lemma 5

From the likelihood function in (3.60) we have:

$$p(y_t|\beta) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(\varepsilon_t(\theta,\eta); \tilde{\mu}_i, \tilde{\Sigma}_i), \tag{3.62}$$

where $\tilde{\mu}_i$, $\tilde{\Sigma}_i$ and $\varepsilon_k(\theta,\eta)$ are given by (3.31) and (3.42) respectively. Let consider

$$p(\zeta_t = i) = \alpha_i, \quad i = 1, \ldots, \kappa, \tag{3.63}$$

where $\zeta_{1:N}$ is a discrete random variable (*hidden* variable) such that ζ_t is an indicator that determines if an observation y_t arises from the ith component of the GMM, that is

$$\zeta_t \in \{1, \ldots, \kappa\}, \quad t = 1, \ldots, N. \tag{3.64}$$

Then, (3.62) can be obtained by marginalizing the *hidden* variable, ζ_t, as follows:

$$p(y_t|\beta) = \sum_{i=1}^{\kappa} p(y_t|\zeta_t = i, \beta) p(\zeta_t = i|\beta), \tag{3.65}$$

where

$$p(y_t|\zeta_t = i, \beta) = \mathcal{N}(\varepsilon_t(\theta,\eta); \tilde{\mu}_i, \tilde{\Sigma}_i). \tag{3.66}$$

Defining the *complete data* set as $\{y_{1:N}, \zeta_{1:N}\}$ and utilizing Bayes' rule, the complete likelihood function is obtained as follows:

$$p(y_{1:N}, \zeta_{1:N}|\beta) = \prod_{t=1}^{N} p(y_t|\zeta_t, \beta) p(\zeta_t|\beta). \tag{3.67}$$

The complete log-likelihood function is given by:

$$\log[p(y_{1:N}, \zeta_{1:N}|\beta)] = \sum_{t=1}^{N} \log[p(\zeta_t|\beta)] + \log[p(y_t|\zeta_t, \beta)]. \tag{3.68}$$

Using (3.63) and (3.65) in (3.68) we obtain:

$$\log[p(y_{1:N}, \zeta_{1:N}|\beta)] = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \log[p(\zeta_t = i|\beta)] + \log[\mathcal{N}(\varepsilon_t(\theta,\eta); \tilde{\mu}_i, \tilde{\Sigma}_i)]. \tag{3.69}$$

3.C Proof of Theorem 1

Computing the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

Using (3.69) in (3.43) we obtain:

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \mathbb{E}\left\{\log[p(\zeta_t = i|\beta)]|y_{1:N}, \hat{\beta}^{(m)}\right\} + \mathbb{E}\left\{\log[\mathcal{N}(\varepsilon_t(\theta,\eta); \tilde{\mu}_i, \tilde{\Sigma}_i)]|y_{1:N}, \hat{\beta}^{(m)}\right\}. \tag{3.70}$$

Solving the expected value in (3.70) and substituting (3.63) into (3.70) we have:

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{t=1}^{N}\sum_{i=1}^{\kappa} \log[\alpha_i]\hat{\zeta}_{ti}^{(m)} + \log[\mathcal{N}(\varepsilon_t(\theta,\eta); \tilde{\mu}_i, \tilde{\Sigma}_i)]\hat{\zeta}_{ti}^{(m)}, \tag{3.71}$$

where

$$\hat{\zeta}_{ti}^{(m)} = p(\zeta_t = i | y_{1:N}, \hat{\beta}^{(m)}). \tag{3.72}$$

Using the Bayes' Theorem, (3.62), (3.63) and (3.65) to compute (3.72) we obtain:

$$\begin{aligned}\hat{\zeta}_{ti}^{(m)} &= \frac{p(\zeta_t = i|\hat{\beta}^{(m)})p(y_t|\zeta_t = i, \hat{\beta}^{(m)})}{p(y_t|\hat{\beta}^{(m)})}, \\ &= \frac{\hat{\alpha}_i^{(m)}\mathcal{N}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}); \hat{\tilde{\mu}}_i^{(m)}, \hat{\tilde{\Sigma}}_i^{(m)})}{\sum_{l=1}^{\kappa}\hat{\alpha}_l^{(m)}\mathcal{N}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}); \hat{\tilde{\mu}}_l^{(m)}, \hat{\tilde{\Sigma}}_l^{(m)})}.\end{aligned} \tag{3.73}$$

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

Taking the derivative of (3.71) with respect to $\tilde{\mu}_i$ and equating to zero we obtain:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \tilde{\mu}_i} = \sum_{t=1}^{N}\left(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}) - \hat{\tilde{\mu}}_i^{(m+1)}\right)\hat{\zeta}_{ti}^{(m)} = 0. \tag{3.74}$$

Substituting (3.31) in (3.74), we obtain:

$$\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}) = \hat{c}_{u_p}^{(m)}\hat{\mu}_i^{(m+1)}\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}, \tag{3.75}$$

$$\hat{\mu}_i^{(m+1)} = \frac{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)})}{\hat{c}_{u_p}^{(m)}\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}. \tag{3.76}$$

Taking the derivative of (3.71) with respect to $\rho = \tilde{\Sigma}_i^{-1}$ and equating to zero yields:

$$\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}\left(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}) - \hat{\tilde{\mu}}_i^{(m)}\right)^2 = \hat{\tilde{\Sigma}}_i^{(m+1)}\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}. \tag{3.77}$$

Substituting (3.31) in (3.77) we obtain:

$$\hat{\Sigma}_i^{(m+1)} = \frac{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}) - \hat{c}_{u_p}^{(m)}\hat{\mu}_i^{(m)})^2}{\left[\hat{c}_{u_p}^{(m)}\right]^2\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}. \tag{3.78}$$

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

$$\mathcal{R}(\alpha_i) = \sum_{t=1}^{N}\sum_{i=1}^{\kappa}\log[\alpha_i]\hat{\zeta}_{ti}^{(m)}, \tag{3.79}$$

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

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

$$\mathcal{J}(\alpha_i, \epsilon) = \sum_{t=1}^{N}\sum_{i=1}^{\kappa} \log[\alpha_i]\hat{\zeta}_{ti}^{(m)} - \epsilon\left(\sum_{i=1}^{\kappa}\alpha_i - 1\right). \tag{3.80}$$

Taking the derivative of (3.80) with respect to α_i and ϵ and equating to zero we obtain:

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \alpha_i} = \frac{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}{\hat{\alpha}_i^{(m+1)}} - \epsilon = 0, \tag{3.81}$$

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \epsilon} = \sum_{i=1}^{\kappa}\alpha_i - 1 = 0. \tag{3.82}$$

Then, $\hat{\alpha}_i^{(m+1)} = \sum_{t=1}^{N}\hat{\zeta}_{ki}^{(m)}/\epsilon$. Taking a summation over $i = 1 \dots \kappa$ and using (3.82) we have:

$$\sum_{i=1}^{\kappa}\hat{\alpha}_i^{(m+1)} = \sum_{i=1}^{\kappa}\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}/\epsilon = 1, \tag{3.83}$$

$$\epsilon = \sum_{i=1}^{\kappa}\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}. \tag{3.84}$$

Notice that $\sum_{i=1}^{\kappa}\hat{\zeta}_{ti}^{(m)} = 1$, then we have $\epsilon = N$. Finally, we obtain:

$$\hat{\alpha}_i^{(m+1)} = \frac{\sum_{t=1}^{N}\hat{\zeta}_{ti}^{(m)}}{N}. \tag{3.85}$$

Notice that $0 \leq \hat{\alpha}_i^{(m+1)} \leq 1$ holds, even though we did not explicitly consider it in (3.80).

3.D An alternative EM formulation with GMMs for linear dynamic systems

Let us define the following:

$$\mathcal{K}(y_t, \beta_i) = \alpha_i \mathcal{N}(\varepsilon_t(\theta, \eta); \tilde{\mu}_i, \tilde{\Sigma}_i), \tag{3.86}$$

$$\mathcal{V}_t(\beta) = \sum_{i=1}^{\kappa}\mathcal{K}(y_t, \beta_i). \tag{3.87}$$

Then, the log-likelihood function in (3.37) can be expressed as:

$$\ell(\beta) = \sum_{t=1}^{N}\log\left[\mathcal{V}_t(\beta)\right]. \tag{3.88}$$

Defining $\mathcal{B}_t(\beta) = \log\left[\mathcal{V}_t(\beta)\right]$, we can obtain:

$$\mathcal{B}_t(\beta) = \mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) - \mathcal{H}_t(\beta, \hat{\beta}^{(m)}), \tag{3.89}$$

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

$$\mathcal{Q}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \log\left[\mathcal{K}(y_t, \beta_i)\right] \frac{\mathcal{K}(y_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}, \tag{3.90}$$

$$\mathcal{H}_t(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \log\left[\frac{\mathcal{K}(y_t, \beta_i)}{\mathcal{V}_t(\beta)}\right] \frac{\mathcal{K}(y_t, \hat{\beta}_i^{(m)})}{\mathcal{V}_t(\hat{\beta}^{(m)})}. \tag{3.91}$$

The function $\mathcal{H}_t(\beta, \hat{\beta}^{(m)})$ is a decreasing function for any value of β. This result is directly obtained from Jensen's inequality (see Lemma 2). Then, the log-likelihood function in (3.88) satisfies the following:

$$\ell(\hat{\beta}^{(m+1)}) \geq \ell(\hat{\beta}^{(m)}). \tag{3.92}$$

Substituting (3.89) in (3.88) and using (3.86) and (3.87), the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ is obtained as follows (*E-step*):

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{t=1}^{N} \sum_{i=1}^{\kappa} \log\left[\alpha_i \mathcal{N}(\varepsilon_k(\theta, \eta); \tilde{\mu}_i, \tilde{\Sigma}_i)\right] \hat{\zeta}_{ti}^{(m)}, \tag{3.93}$$

where

$$\hat{\zeta}_{ti}^{(m)} = \frac{\hat{\alpha}_i^{(m)} \mathcal{N}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}); \hat{\tilde{\mu}}_i^{(m)}, \hat{\tilde{\Sigma}}_i^{(m)})}{\sum_{l=1}^{\kappa} \hat{\alpha}_l^{(m)} \mathcal{N}(\varepsilon_t(\hat{\theta}^{(m)}, \hat{\eta}^{(m)}); \hat{\tilde{\mu}}_l^{(m)}, \hat{\tilde{\Sigma}}_l^{(m)})}. \tag{3.94}$$

The *M-step* is then given by

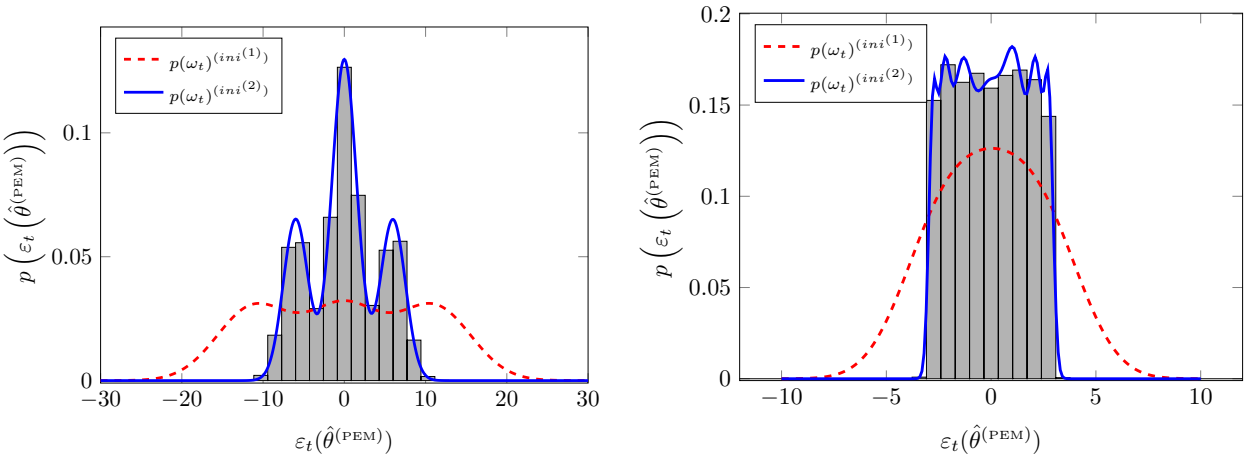
$$\hat{\beta}^{(m+1)} = \arg\max_{\beta} \mathcal{Q}(\beta, \hat{\beta}^{(m)}), \quad \text{s.t} \quad 0 \leq \alpha_i \leq 1, \quad \sum_{i=1}^{M} \alpha_i = 1. \tag{3.95}$$

Finally, the auxiliary function, $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$, in (3.93) is equal to (3.46) and the estimators are obtained from to (3.48)–(3.51).

3.E Initialization procedures for the EM algorithm with GMMs

For comparison purposes, we consider the Example 1 and Example 2 using our proposed algorithm (EM–GMM) with $N = 5000$, 50 Monte Carlo (MC) simulations, our initialization procedure for the GMM components ($ini^{(1)}$), and an initialization based on the classical EM algorithm with GMMs [68] (namely $ini^{(2)}$) using the prediction error from the Prediction Error Method (PEM), $\varepsilon_t(\hat{\theta}^{(\text{PEM})})$. We run the simulations using Matlab® with an Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz, RAM 12GB.

In Figure 3.7 we show the histogram of the prediction error and the proposed initialization for the noise distribution. In Table 3.5 we compare the estimation results using both initialization procedures considering the number of EM–GMM iterations and the Mean Square Error (MSE) of the estimates of the vector of parameters. We also included the average running time for each MC simulation for both Example 1 and Example 2. That is, the time



(a) Example 1.

(b) Example 2.

Figure 3.7: Histogram of the prediction error $\varepsilon_t(\hat{\theta}^{(\mathrm{PEM})})$ obtained by using PEM for the Example 1 and Example 2.

involved in the GMM initialization plus the time for the EM estimation (Total time). We clearly observe that $ini^{(2)}$ corresponds to a better fit of the prediction error histogram than the one corresponding to $ini^{(1)}$. In addition, we also observe that the EM–GMM yields an optimal solution with less iterations when we use the $ini^{(2)}$ than using $ini^{(1)}$. Our experience shows that using $ini^{(2)}$ spends around 40% less total time than $ini^{(1)}$ to yield the estimates for each MC estimation. However, for both initialization procedures, the accuracy of the estimates is very similar.

Table 3.5: MSE and EM–GMM iterations for Examples 1 and 2

Initialization	Example 1			Example 2		
procedure	Iterations	MSE	Total time [s]	Iterations	MSE	Total time [s]
$ini^{(1)}$	68	2.5243×10^{-4}	18.61 ± 2.49	1477	1.2564×10^{-5}	103.22 ± 12.24
$ini^{(2)}$	39	2.5239×10^{-4}	8.30 ± 1.77	1057	1.2453×10^{-5}	56.68 ± 15.38

Chapter 4
Model Error Modeling Using Stochastic Embedding Approach with GMMs

Most System Identification techniques available in the literature are developed under the assumption that the system lies in the model set when estimating the corresponding vector of parameters (see e.g. [1–3]). However, real systems have arbitrary complexity and using a complex model structure can lead to large estimation variance errors [24]. An alternative view of Modeling and System Identification of dynamic systems combines a nominal model with an error-model. Among popular alternatives for uncertainty modeling we find *Stochastic Embedding* and Set Membership. In this chapter, an ML estimator for uncertainty modeling in linear dynamic systems with *Stochastic Embedding* approach is developed. The PDF of the vector of parameters that define the error-model is given by a GMM. An estimation algorithm to estimate the vector of parameters that define both the nominal model and the PDF of the GMM based on the EM algorithm is proposed. In this chapter we summarize the results presented in the journal paper **J.3**, and the conference papers **C.2** and **C.3**.

Contribution

We obtain the likelihood function for a linear dynamic system modeled with *Stochastic Embedding* approach and a finite Gaussian mixture distribution. The likelihood function is computed by marginalizing the vector of parameters of the error-model as a *hidden* variable. We propose an EM-based algorithm to solve the associated ML estimation problem with GMMs, obtaining the estimates of the vector of parameters that define the nominal model and closed form expressions for the GMM estimators of the error-model distribution.

4.1 Motivation

In control and monitoring of manufacturing processes, it is key to understand model uncertainty in order to achieve the required levels of consistency, quality, and economy, among others. In aerospace applications, models need to be very precise and able to describe the entire dynamics of an aircraft [118]. In addition, the complexity of modern real systems has turned deterministic models impractical, since they cannot adequately represent the behavior of disturbances in sensors and actuators, and tool and machine wear, to name a few. Thus, it is necessary to deal with model uncertainties in the dynamics of the plant by incorporating a stochastic behavior.

In Figure 4.1, a typical representation of a real process and measured experimental data is shown, where u_t is the input measurements, y_t denotes the output measurements, and ω_t

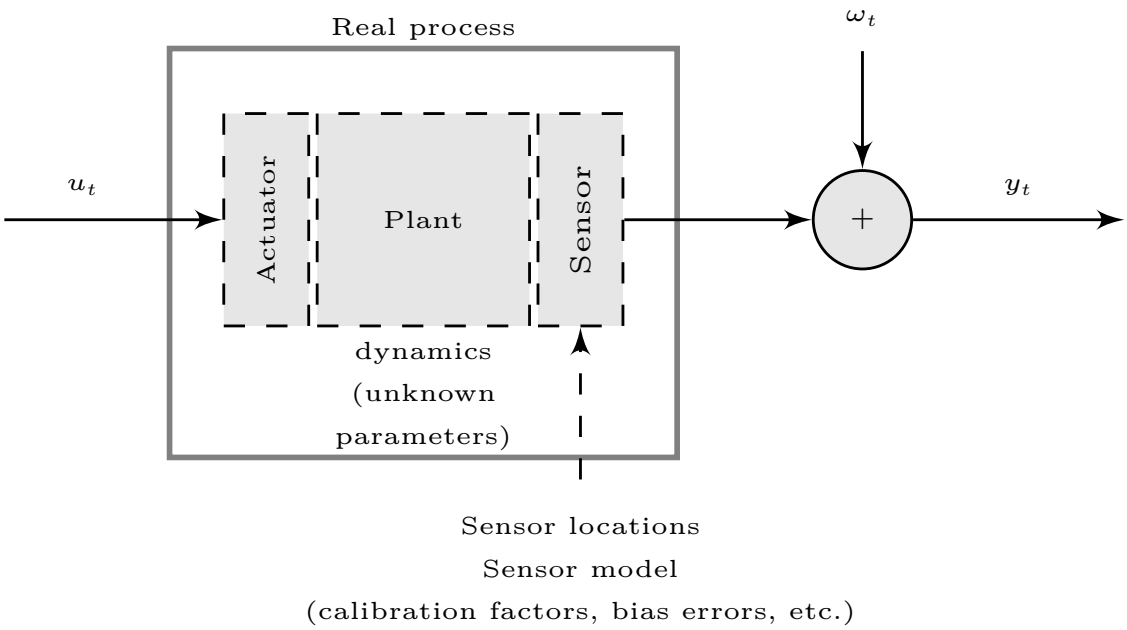


Figure 4.1: Representation of a real dynamic system with data measurements for system identification.

denotes a measurement noise. The real process (*true* system) is represented by the interaction of three components: (i) an actuator, (ii) a plant, and (iii) a sensor. This representation has been used to develop system identification methodologies for flight vehicle development in order to obtain accurate models and to validate mathematical models of the flight vehicle [119]. These system models are required, for example, to perform fault-diagnosis and adaptive control, to analyze handling qualities specification compliance, to develop high-fidelity aerodynamic databases for flight simulators, among others [120]. However, the quality of these results can be affected by the uncertainties incorporated by the (unknown) dynamics that are not modeled, the instrumentation, model simplifications, and measurement noise and sensor errors [121]. For these reasons, a suitable formulation of a model that incorporates uncertainties is an important aspect to consider in system identification methods, in order to obtain system models that represent as closely as possible the actual process behavior [1, 2, 119]. Determining suitable dynamic system models of real processes is essential to obtain effective process control strategies and accurate fault detection and diagnosis methodologies that deliver good performance [122].

In the next sections, we will show an identification methodology to address the uncertainty modeling problem in a general class linear dynamic system using SE approach. That is, we aim at obtaining an estimation of a nominal model with an error-model in which its vector of parameters are Gaussian mixture distributed.

4.2 General system description

Consider the linear dynamic system given by:

$$y_t^{[r]} = G^{[r]}(z^{-1})u_t^{[r]} + H(z^{-1}, \theta)\omega_t^{[r]}, \tag{4.1}$$

where $r = 1, ..., M$ denotes the r-th realization of the system, M corresponds the number of independent experiments or batches, z^{-1} denotes the backward shift operator ($z^{-1}x_t = x_{t-1}$), $y_t^{[r]} \in \mathbb{R}$ is the output signal, $u_t^{[r]} \in \mathbb{R}$ denotes the input signal, $\omega_t^{[r]} \in \mathbb{R}$ is a zero-mean white Gaussian noise with variance σ^2, and $H(z^{-1}, \theta)$ is the noise model parameterized by θ. Notice that $u_t^{[r]}$ and $\omega_t^{[r]}$ describe, respectively, a different input signal and noise realizations for each independent experiment. We consider that the system $G^{[r]}(z^{-1})$ can be described as follows (see e.g. [14, 26]):

$$G^{[r]}(z^{-1}) = G_0(z^{-1}, \theta)(1 + G_\Delta(z^{-1}, \eta^{[r]})), \tag{4.2}$$

where $G_0(z^{-1}, \theta)$ is the nominal model parameterized by θ, $G_\Delta(z^{-1}, \eta^{[r]})$ is a multiplicative error-model (4.2) parameterized by $\eta^{[r]}$. Here, we also consider that the PDF of the vector of parameters of the error-model, $\eta^{[r]}$, is given by a finite Gaussian mixture distribution as follows:

$$p(\eta^{[r]}|\gamma) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i), \tag{4.3}$$

$$\gamma = [\underbrace{\alpha_1 \ \mu_1 \ \Gamma_1}_{\gamma_1} \ \cdots \ \underbrace{\alpha_\kappa \ \mu_\kappa \ \Gamma_\kappa}_{\gamma_\kappa}]^T, \tag{4.4}$$

where $\alpha_i > 0$, $\sum_{i=1}^{\kappa} \alpha_i = 1$ and $\mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i)$ represents a Gaussian PDF with mean value μ_i and covariance matrix Γ_i. Note that we consider that $H(z^{-1}, \theta)$ in (4.1) is part of the nominal model[1], i.e., it is only parameterized by θ.

The nominal system model, $G_0(z^{-1}, \theta)$, for the system model (4.1) is given by:

$$G_0(z^{-1}, \theta) = \frac{B(z^{-1}, \theta)}{A(z^{-1}, \theta)}, \tag{4.5}$$

where

$$B(z^{-1}, \theta) = b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}, \tag{4.6}$$

$$A(z^{-1}, \theta) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}. \tag{4.7}$$

Similarly, the noise system model in (4.1) is given by

$$H(z^{-1}, \theta) = \frac{C(z^{-1}, \theta)}{D(z^{-1}, \theta)}, \tag{4.8}$$

with

$$C(z^{-1}, \theta) = 1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}, \tag{4.9}$$

$$D(z^{-1}, \theta) = 1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d}. \tag{4.10}$$

We also consider that the error-model $G_\Delta(z^{-1}, \eta^{[r]})$ in (4.2) is a linear regression as follows:

$$G_\Delta(z^{-1}, \eta^{[r]}) = \eta_0^{[r]} + \eta_1^{[r]} z^{-1} + \cdots + \eta_{n_\Delta}^{[r]} z^{-n_\Delta}, \tag{4.11}$$

[1] We use the term nominal model to refer to the system model $G_0(z^{-1}, \theta)$ and the noise system model $H(z^{-1}, \theta)$.

where $\eta^{[r]} \in \mathbb{R}^{n_\Delta \times 1}$.

4.2.1 System model as a linear regression

We consider that the observed data $Y^{[r]} = [y_1^{[r]} \ \cdots \ y_N^{[r]}]$ is a collection of measurements for each experiment. Then, the system (4.1) can be described as follows[2]:

$$Y^{[r]} = \mathcal{G}(\theta)U^{[r]} + \Psi^{[r]}(\theta)\eta^{[r]} + W^{[r]}, \tag{4.12}$$

where $Y^{[r]}, U^{[r]}, W^{[r]} \in \mathbb{R}^{N\times 1}$, $\theta \in \mathbb{R}^{n_0 \times 1}$, $\eta^{[r]} \in \mathbb{R}^{n_\Delta \times 1}$, $\mathcal{G}(\theta)U^{[r]} \in \mathbb{R}^{N\times 1}$, $\Psi^{[r]}(\theta) \in \mathbb{R}^{N\times n_\Delta}$. The term $\mathcal{G}(\theta)U^{[r]}$ corresponds to the output response corresponding to the nominal model structure $\mathcal{G}(\theta)$, $\Psi^{[r]}(\theta)\eta^{[r]}$ corresponds to the output signal due to the error-model structure in (4.2), and $W^{[r]} \sim \mathcal{N}(0, \sigma^2 I_N)$ ($I_N \in \mathbb{R}^{N\times N}$ is the identity matrix). Notice that $\mathcal{G}(\theta)$ involves the nominal model structure utilized for $G_0(z^{-1}, \theta)$ and $H(z^{-1}, \theta)$ in (4.1). The term $\Psi^{[r]}(\theta)$ describes the model structure given by the product $G_0(z^{-1}, \theta)G_\Delta\left(z^{-1}, \eta^{[r]}\right)$ in (4.2).

Remark 11. *The system model in (4.12) is constructed under the assumption that the error-model structure $\Psi^{[r]}(\theta)\eta^{[r]}$ is a linear regression and the nominal model structure $\mathcal{G}(\theta)$ can have an arbitrary structure [26]. The fact of considering the error-model as a linear regression can provide flexibility, lower computational complexity and contain FIR model of arbitrary orders, Laguerre, and Kautz models (see e.g. [123]).* ▽

4.2.2 Standing assumptions

The problem of interest is to estimate the vector of parameters, $\beta = [\ \theta^T \ \gamma^T \ \sigma^2\ ]^T$, that defines the parameters of the nominal model, error-model and the noise variance. In addition, we consider that β_0 is the *true* vector of parameters that defines the *true* model. In order to formulate the ML estimation algorithm, we introduce the following standing assumptions:

- **A1** The system in (4.1) is operating in open loop and the input signal $u_t^{[r]}$ is an exogenous deterministic signal for each r independent experiment.
- **A2** The nominal model does not change from one experiment to another, whilst the error-model $G_\Delta(z^{-1}, \eta^{[r]})$ may change for each experiment and all the realizations of $\eta = \{\eta^{[r]}\}_{r=1}^M$ are drawn from the same PDF parameterized by γ.
- **A3** The vector of parameters β_0, the input signal $u_t^{[r]}$ and the noise $\omega_t^{[r]}$ in (4.1) satisfy regularity conditions, guaranteeing that the ML estimate β_{ML} converges (in probability or a.s.) to the *true* solution β_0 as $N \to \infty$.
- **A4** The orders n_a, n_b, n_c, n_d and n_Δ of the polynomials of system (4.1) and the number of components κ of the error-model distribution (4.3) are known.
- **A5** The system (4.1) is asymptotically stable, its models $G(z^{-1}, \theta)$ and $H(z^{-1}, \theta)$ have no poles-zeros on the unit circle and have no pole-zero cancellations. The noise system model $H(z^{-1}, \theta)$ is also a minimum-phase system.

Assumption **A2** is needed to develop an ML methodology based on the uncertainty modeling with SE approach. Assumption **A3** is necessary to develop an estimation algorithm that holds

[2] In this chapter we use capital letters to denote the vector of signals, u_t or ω_t for $t = 1, ..., N$.

the asymptotic properties of the ML estimator. Assumption **A4** can be relaxed by considering a dynamic system model that includes parametric uncertainties with a different error-model structure that we assume in (4.2). We will address this case in Section 4.6. Assumption **A5** is necessary to obtain an asymptotically unbiased ML estimator [103] and a system model that is controllable [104].

In the following section we consider different approaches for modeling the error-model for linear dynamic systems.

4.3 Uncertainty modeling frameworks for linear dynamic systems

4.3.1 Set Membership (SM) approach

A paradigm to describe the uncertainty in a system model on a deterministic hypothesis is to consider that the error-model is unknown-but-bounded (UBB). A typical technique to address this approach is Set Membership (SM) identification [20, 124]. SM method provides efficient algorithms for estimating the set of feasible models, compatible with the available data and the UBB error assumption [20]. The choice of the nominal model is usually performed by minimizing a cost function related to the feasible set. The feasible set itself gives the size of the uncertainty associated with the nominal model.

Let consider a linear dynamic system as follows:

$$y_t = G_T(z^{-1})u_t + \omega_t, \tag{4.13}$$

where $G_T(z^{-1})$ denotes the *true* system model, $y_t \in \mathbb{R}$ denotes the output signal, $u_t \in \mathbb{R}$ is the input deterministic signal, and $\omega_t \in \mathbb{R}$ is a noise sequence. The *true system*, $G_T(z^{-1})$, belongs to a set $\mathcal{K}$ (prior information of the system) and the noise sequence $\omega_{1:N}$ is bounded in some norm $\mathcal{Y}$. i.e., $\|\omega_{1:N}\|_{\mathcal{Y}} \leq \delta$, $\delta > 0$ [20, 125, 126]. Based on these assumptions, it is possible to define the feasible system set (FSS) as follows:

$$\text{FSS} = \{S \in \mathcal{K} : \|y_{1:N} - S(u_{1:N})\|_{\mathcal{Y}} \leq \delta\}, \tag{4.14}$$

which is the set of systems that are compatible with the measured data and the prior assumptions [126]. Notice that $S(u_{1:N})$ is the output of the system model S evaluated with the input $u_{1:N}$. In addition, the definition of the FSS involves a difficulty based on the system structure of $\mathcal{K}$ and the definition of the norm $\|\cdot\|_{\mathcal{Y}}$ (e.g. non-linear systems or linear variant time systems).

Since the FSS contains the information provided by the measurement data, it is natural to evaluate the quality of a nominal model, $G_0(z^{-1}, \theta)$, according to its worst-case error with respect to the elements of FSS. Then, the *identification error* associated with $G_0(z^{-1}, \theta)$, $E(G_0(z^{-1}, \theta))$, is given by [126]:

$$E(G_0(z, \theta)) = \sup_{S \in \text{FSS}} \left\| S - G_0(z^{-1}, \theta) \right\|_{\mathcal{D}}, \tag{4.15}$$

where $\|\cdot\|_{\mathcal{D}}$ is a suitable norm in the system space and θ is the vector of parameters that defines the nominal model.

In order to obtain an estimation of $G_0(z^{-1},\theta)$, a model structure for the nominal system model needs to be chosen. A common criterion is to consider a simple structure for the nominal model, e.g. FIR filters, low dimensional, linearly parameterized [20]. If we consider $\mathcal{M}$ as a set of nominal models parameterized by $\theta \in \mathbb{R}^{n_0}$, the problem of choosing a model in $\mathcal{M}$ according to the criterion in (4.15) can be solved the conditional central estimate as follows [125]:

$$\hat{G}_0(z^{-1},\theta) = \arg \inf_{G_0(z^{-1},\theta)\in\mathcal{M}} \max_{S\in\text{FSS}} \left\| S - G_0(z^{-1},\theta) \right\|_{\mathcal{D}}. \tag{4.16}$$

Since the min-max optimization problem in (4.16) may be computationally infeasible, sub-optimal estimators are considered [125, 126]. For example, in the case that the noise is ℓ_∞-bounded and the FSS is a polytope in $\mathbb{R}^{n_0}$, it can be recursively approximated by simpler regions such as ellipsoids or parallelotopes [127] and the center of this approximating sets may be chosen as an estimate of the nominal model. More sophisticated set approximation strategies, based on inner and outer bounding via polytopes, have been proposed in [128]. They provide a viable computational complexity, which is a key issue in SM identification.

4.3.2 Model error modeling (MEM)

Model error modeling (MEM) [21] is a framework used for modeling the uncertainty in linear dynamic systems. The key idea is to use standard PEM algorithms to estimate a nominal model, $\hat{G}_0(z^{-1},\theta)$, from input-output measurements in time domain [1, 2] in (4.13). The estimated nominal model is then used to obtain the *residuals*, ε_t, as follows [21, 126]:

$$\varepsilon_t = y_t - \hat{G}_0(z^{-1},\theta)u_t. \tag{4.17}$$

Then, a model structure for the *residuals* is chosen as follows:

$$\varepsilon_t = F(z^{-1},\vartheta)u_t + H(z^{-1},\vartheta)v_t, \tag{4.18}$$

where $F(z^{-1},\vartheta)$ and $H(z^{-1},\vartheta)$ are rational transfer functions parameterized by ϑ and v_t is a zero-mean Gaussian white noise with variance σ_v^2. The parameter of the error-model in (4.18) can be estimated by using classical PEM algorithms [1, 2]. This procedure provides an error-model description that captures the reliability of the nominal model. This result can be used, for example, to generate a confidence set for model validation or for other specific applications, such as robust control [23].

Finally, the MEM methodology can be summarized as follows [126]:

- **(i)** Compute the *residuals*, ε_t, from (4.17).
- **(ii)** Compute the error-model in (4.18) using a PEM algorithm with input data u_t and, as output data, the *residuals*, ε_t.
- **(iii)** Compute the uncertainty region by adding frequency by frequency the error-model to the nominal model. This gives a region where the *true* system is supposed to be found.

4.3.3 Stochastic embedding (SE) approach

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 PDF [14, 22]. The original work on SE assumed that the *true* system $G_T(z^{-1})$ in (4.13) could be described as follows [14, 22]:

$$G_T^{[r]}(z^{-1}) = \begin{cases} G_0(z^{-1},\theta) + G_\epsilon(z^{-1},\eta^{[r]}), & \text{(4.19a)} \\ G_0(z^{-1},\theta)(1 + G_\Delta(z^{-1},\eta^{[r]})), & \text{(4.19b)} \end{cases}$$

where $G_0(z^{-1},\theta)$ is the nominal model parameterized by θ, $G_\epsilon(z^{-1},\eta^{[r]})$ is an additive error-model, and $G_\Delta(z^{-1},\eta^{[r]})$ is a multiplicative error-model both parameterized by $\eta^{[r]}$. The error-model is viewed as a realization of a stochastic vector of parameters $\eta^{[r]}$. In addition, the PDF that models $\eta^{[r]}$ is parameterized by γ.

To illustrate the stochastic embedding approach, the magnitude of the frequency response of a linear dynamic system (assumed to be the *true* system) is shown in Figure 4.2, where $G_T^{[r]}(z^{-1})$ is given by

$$G_T^{[r]}(z^{-1}) = G_0(z^{-1}) + G_0(z^{-1})G_\Delta^{[r]}(z^{-1}), \tag{4.20}$$

where $G_0(z^{-1})$ is a fixed nominal model, and $G_\Delta^{[r]}(z^{-1})$ denotes the rth realization of a relative error-model for the *true* system. The red-shaded region represents the area in which possible *true* systems, $G_T^{[r]}(z^{-1})$, i.e, realizations of the *true* system (represented by black dotted lines), can lie. The error-model can introduce model parameter uncertainties at low-frequency range and model structure uncertainties at high-frequency range [119]. This system model behavior can be encountered, for instance, in an aeroservoelastic system model (see e.g., [118]). Aeroservoelastic systems include dynamic coupling due to structural, control, sensor, and actuator dynamics that cover both the low-frequency and high-frequency range. This means that, in an ML framework, the estimates can change considerably from one experiment to another, even if the number of measurements used to obtain an ML estimation is large, due to the structural and parametric uncertainties in the *true* system model.

There are authors which address this approach using ML estimation or a Bayesian perspective under Gaussian assumptions for the error-model distribution [23, 24, 26]. In the following section, we consider to address an estimation of the system (4.1) using the ML method with SE approach.

4.4 Maximum Likelihood estimation using SE approach with GMMs

4.4.1 Maximum Likelihood formulation with SE approach

In this section, we develop an ML estimation algorithm for system (4.1) using the SE approach with GMMs to estimate the parameters that define the nominal model, the PDF of the error-model, and the measurement noise. We consider that system (4.1) can be described as a linear regression in (4.12). We define the vector of parameters to be estimated

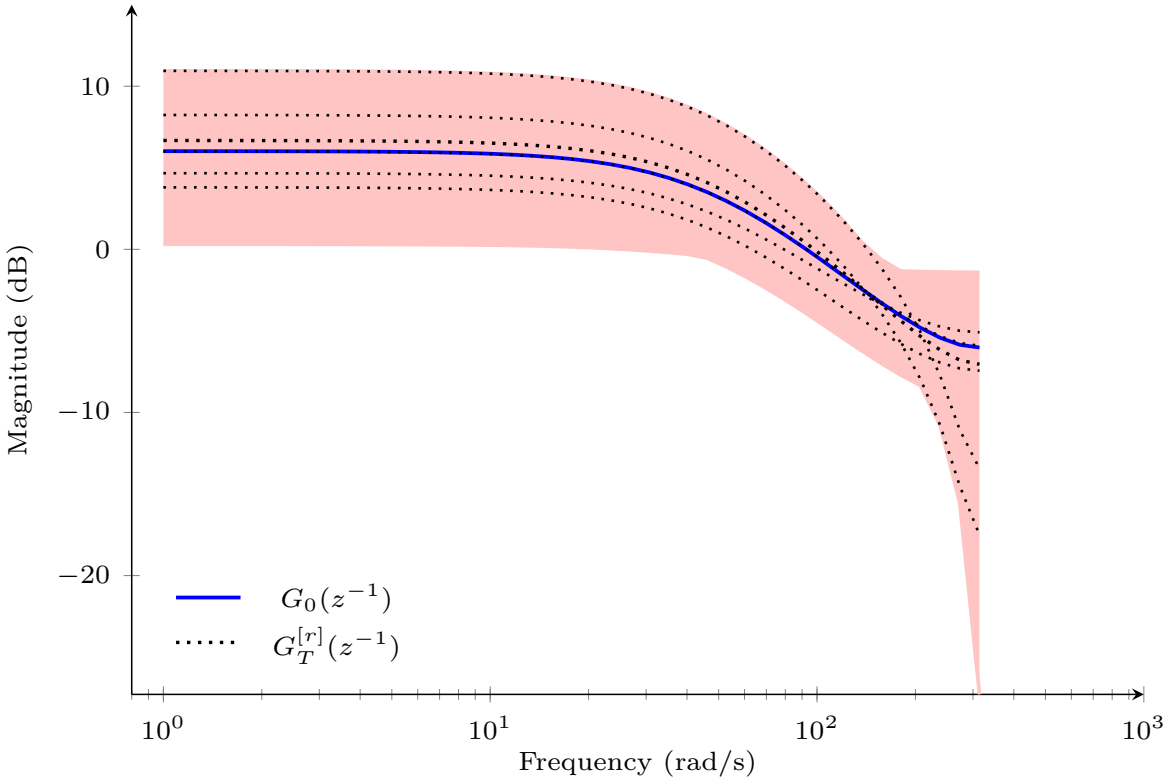


Figure 4.2: Magnitude of the frequency response for the linear dynamic system in (4.20).

as $\beta = [\ \theta^T\ \gamma^T\ \sigma^2\ ]^T$ in order to formulate the ML estimator for the system (4.12):

$$\theta = [b_1\ a_1\ c_1\ d_1\ \cdots\ b_{n_b}\ a_{n_a}\ c_{n_c}\ d_{n_d}]^T, \tag{4.21}$$

where θ is the vector of parameters of the nominal model, γ is the vector of parameters of the error-model distribution as a GMM in (4.3), and σ^2 is the noise variance. For the model in (4.12) using the GMM in (4.3), the likelihood function, $\mathcal{L}(\beta)$, can be obtained by marginalizing the *hidden* variable, $[\eta^{[1]}\ \cdots\ \eta^{[M]}]$, as follows[3]:

$$\begin{aligned}
\mathcal{L}(\beta) &= p(Y^{[1]}, \ldots, Y^{[M]}|\beta), && (4.22)\\
&= \prod_{r=1}^{M} p(Y^{[r]}|\beta), && (4.23)\\
&= \prod_{r=1}^{M} \int_{-\infty}^{\infty} p(Y^{[r]}|\eta^{[r]}, \beta) p(\eta^{[r]}|\beta) d\eta^{[r]}, && (4.24)
\end{aligned}$$

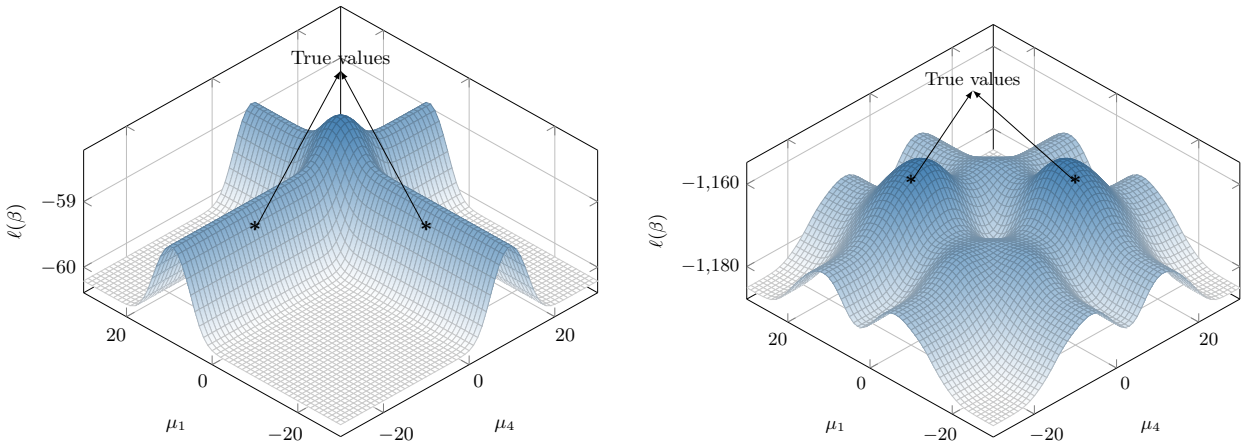
where

$$p(Y^{[r]}|\eta^{[r]}, \beta) = \mathcal{N}(Y^{[r]}; \mathcal{G}(\theta)U^{[r]} + \Psi^{[r]}(\theta)\eta^{[r]}, \sigma^2 I_N), \tag{4.25}$$

$$p(\eta^{[r]}|\beta) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i), \tag{4.26}$$

with $\alpha_i > 0$, $\sum_{i=1}^{\kappa} \alpha_i = 1$.

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



(a) $M = 1$ (b) $M = 20$

Figure 4.3: Log-likelihood function using SE approach with GMMs.

The log-likelihood function, $\ell(\beta)$, is then given by

$$\ell(\beta) = \sum_{r=1}^{M} \log \left\{ \sum_{i=1}^{\kappa} \alpha_i \int_{-\infty}^{\infty} p(Y^{[r]}|\eta^{[r]}, \beta) p(\eta^{[r]}|\beta) d\eta^{[r]} \right\}. \tag{4.27}$$

Finally, the ML estimator is given by

$$\hat{\beta}_{\text{ML}} = \arg\max_{\beta} \ \ell(\beta), \quad \text{s.t.} \quad \sum_{i=1}^{\kappa} \alpha_i = 1, \ 0 \leq \alpha_i \leq 1. \tag{4.28}$$

Remark 12. *The classical ML formulation considers the measurement data of one experiment to obtain an estimate of a system model (see e.g. [1, 2]). In contrast, the ML estimator in (4.28) simultaneously considers the measurements of M independent experiments to obtain an estimation of the system model with SE approach, i.e, the nominal model and error-model distribution as a GMM.* ▽

4.4.2 Likelihood function with GMMs

The ML estimation problem is solved by maximizing the log-likelihood function in (4.27). The optimization problem is typically solved using gradient-based methods. However, as explained in Chapter 3, the solution may be difficult to obtain when the number of components κ in the GMM and/or the number of independent experiments M increases. To illustrate this behavior, we consider the system model (4.2) with a simple FIR structures as follows:

$$G_0(z^{-1}, \theta_0) = g_0 + g_1 z^{-1}, \qquad H(z^{-1}, \theta_0) = 1, \tag{4.29}$$

$$G_\Delta(z^{-1}, \eta^{[r]}) = \eta_0^{[r]} - \eta_0^{[r]} z^{-1}, \tag{4.30}$$

where $\theta_0 = [1 \ \ 0.5]^T$, $\omega_t^{[r]} \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = 1$, $u_t^{[r]} \sim \mathcal{N}(0, \sigma_u^2)$, $\sigma_u^2 = 10$ and $N = 100$. The error-model is parameterized by $\eta^{[r]} = [\eta_0^{[r]} \ \ -\eta_0^{[r]}]^T$ and it is Gaussian mixture distributed with $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.25$, $\Gamma_1 = \Gamma_2 = \Gamma_3 = \Gamma_4 = 10$, $\mu_1 = -5$, $\mu_2 = 5$, $\mu_3 = -15$

and $\mu_4 = 15$. In Figure 4.3(a) we show the log-likelihood function for two mean values of the GMM (μ_1 and μ_4) when the remainder of the parameters are fixed at their corresponding *true* values. We observe local maxima in the log-likelihood function. Moreover, we observe the difficulty to obtain the *true* values of the error-model distribution when the number of realizations (experiments) is low (e.g. $M = 1$), i.e., there is not enough statistical information to describe the corresponding error-model distribution. In contrast, in Figure 4.3(b) we show the log-likelihood function behavior when the number of realizations is increased (e.g. $M = 20$). We observe that the *true* values of the GMM correspond to the maximum values of the log-likelihood function. However, there are several local maxima and to solve the optimization problem in (4.28) using standard gradient-based methods may be cumbersome. An iterative procedure based on the EM algorithm [25] can provide a simpler technique to deal with this difficulty since it typically provides closed form estimators for the GMM parameters [67, 68].

4.4.3 A comparison analysis with SE approach and Bayesian estimation

The estimation of the system in (4.12) can be handled in a variety of approaches [24]. One aspect to consider for the formulation of the estimation problem with SE approach is the selection of the complexity of the system model for both the nominal model and the error-model, i.e., the vector of parameters θ and $\eta^{[r]}$ that we use to address the uncertainty modeling problem. A Bayesian view is to consider that both vector of parameters, $\{\theta, \eta^{[r]}\}$, can be modeled as realizations of random vectors with certain prior distributions [24, 26]. Under this Bayesian framework, it is possible to obtain an unified system model of both the nominal model and error-model if all system models are considered FIR systems. To illustrate the latter idea, we consider the *true* system in (4.2) with (4.29) and (4.30) as follows:

$$G_T^{[r]}(z^{-1}) = \underbrace{(g_0 + g_0\eta_0^{[r]})}_{\bar{\theta}_0^{[r]}} + \underbrace{(g_1 - g_0\eta_0^{[r]} + g_1\eta_0^{[r]})}_{\bar{\theta}_1^{[r]}} z^{-1} - \underbrace{(g_1\eta_0^{[r]})}_{\bar{\theta}_2^{[r]}} z^{-2}. \tag{4.31}$$

Then, the system model in (4.1) can be expressed as follows:

$$y_t^{[r]} = \varphi_t^{[r]}\bar{\theta}^{[r]} + \omega_t^{[r]}, \tag{4.32}$$

where

$$\varphi_t^{[r]} = [u_t^{[r]} \; u_{t-1}^{[r]} \; u_{t-2}^{[r]}]^T, \tag{4.33}$$

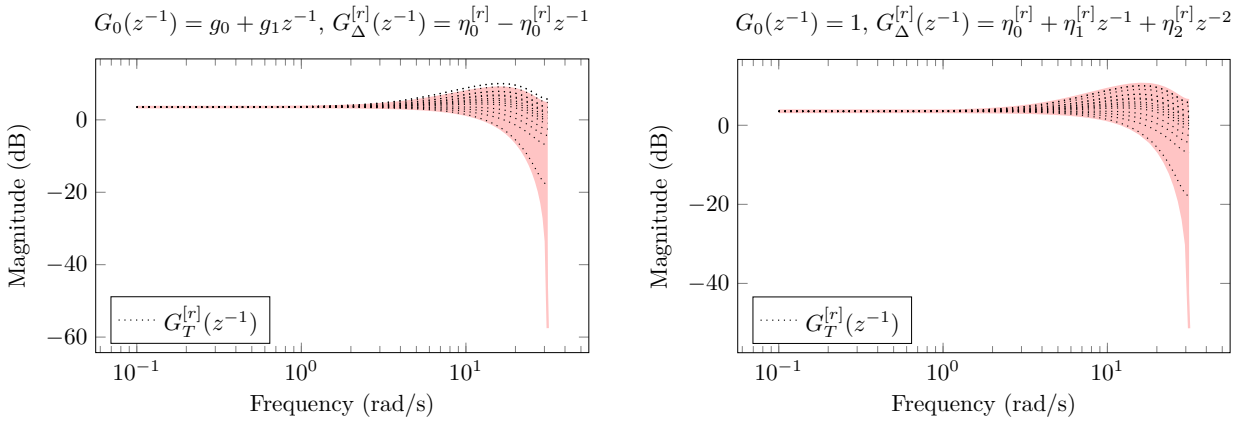
$$\bar{\theta}^{[r]} = [\; \bar{\theta}_0^{[r]} \;\; \bar{\theta}_1^{[r]} \;\; \bar{\theta}_2^{[r]} \;]^T. \tag{4.34}$$

This formulation, then, becomes a standard Bayesian estimation problem that can be formulated with the Maximum a posteriori (MAP) approach as follows [129, 130]:

$$\bar{\theta}_{\text{MAP}}^{[r]} = \arg \max_{\bar{\theta}^{[r]}} \; \bar{\ell}(\bar{\theta}^{[r]}), \tag{4.35}$$

where

$$\bar{\ell}(\bar{\theta}^{[r]}) = -\frac{N}{2}\log[\sigma^2] - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(y_t - \varphi_t^{[r]}\bar{\theta}^{[r]})^2 + \log[p(\bar{\theta}^{[r]})], \tag{4.36}$$



(a) SE approach

(b) Bayesian approach

Figure 4.4: Frequency response of the magnitude of the *true* system models and the uncertainty region.

and $p(\bar{\theta}^{[r]})$ is a given prior distribution of the vector of parameters, $\bar{\theta}^{[r]}$. Notice that this Bayesian framework is closely related with SE approach for uncertainty quantification in linear dynamic systems. In particular, we can obtain an error-model, $G_\Delta^{[r]}(z^{-1})$, in (4.20) as a stochastic system in (4.31), considering a nominal model equals to one ($G_0(z^{-1}) = 1$). In this sense, there are different system model structures that can describe the same uncertainty region of the *true* system with SE approach. In Figure 4.4 we show the magnitude of the frequency response of the *true* system for different nominal model structures. The red-shaded regions (uncertainty regions) represent the area in which realizations of the *true* system (black dotted lines) can lie. In this case, we observe that there is more than one representation of the uncertainty region for the same *true* system, i.e., there is a trade-off between choosing a nominal model and an error-model structure with SE approach. Notice that in the method proposed in this thesis, the complexity and structure of both the nominal model and error-model is defined by the user. Notice that our proposed technique is not limited to a particular model structure, providing more flexibility to obtain more suitable models. This flexibility might result in identifiability issues. However, we select in this thesis the system model structure that can better describe the uncertainty region of the *true* system with the smallest number of parameters for the error-model.

4.5 An EM algorithm using SE approach with GMMs

The log-likelihood function in (4.27) is closely related with the optimization cost given in (2.15). Then, we can formulate an EM-based algorithm to solve the estimation problem with the systematic procedure described in Chapter 2.

4.5.1 EM-based algorithm formulation

From (4.25) and (4.26), we define the following:

$$\mathcal{K}(\beta_i, \eta^{[r]}) = \alpha_i \mathcal{N}(Y^{[r]}; \mathcal{G}(\theta)U^{[r]} + \Psi^{[r]}(\theta)\eta^{[r]}, \sigma^2 I_N)\mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i). \tag{4.37}$$

Then, the log-likelihood function in (4.27) can be expressed as

$$\ell(\beta) = \sum_{r=1}^{M} \log\left[\mathcal{V}^{[r]}(\beta)\right], \tag{4.38}$$

with

$$\mathcal{V}^{[r]}(\beta) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \mathcal{K}(\beta_i, \eta^{[r]}) d\eta^{[r]}. \tag{4.39}$$

The logarithm of the expression in (4.39) can be written as follows:

$$\log\left[\mathcal{V}^{[r]}(\beta)\right] = \mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)}) - \mathcal{H}^{[r]}(\beta, \hat{\beta}^{(m)}), \tag{4.40}$$

where

$$\mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log[\mathcal{K}(\beta_i, \eta^{[r]})] \frac{\mathcal{K}(\hat{\beta}_i^{(m)}, \eta^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} d\eta^{[r]}, \tag{4.41}$$

$$\mathcal{H}^{[r]}(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log\left[\frac{\mathcal{K}(\beta_i, \eta^{[r]})}{\mathcal{V}^{[r]}(\beta)}\right] \frac{\mathcal{K}(\hat{\beta}_i^{(m)}, \eta^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} d\eta^{[r]}, \tag{4.42}$$

where $\hat{\beta}^{(m)}$ is the current estimate. From Lemma 2, the function $\mathcal{H}^{[r]}(\beta, \hat{\beta}^{(m)})$ is a decreasing function for any value of β.

In order to develop the estimation algorithm, we first obtain the following result. This will be used to compute the auxiliary function $\mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)})$ in (4.41).

Lemma 6. *The expression in (4.37) evaluated at the m-th estimate, $\hat{\beta}^{(m)}$, can be written as follows:*

$$\mathcal{K}(\hat{\beta}_i^{(m)}, \eta^{[r]}) = \hat{\alpha}_i^{(m)} \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}^{(m)}, \bar{\Sigma}_{ir}^{(m)}) \mathcal{N}(\eta^{[r]}; \tilde{\mu}_{ir}^{(m)}, \tilde{\Sigma}_{ir}^{(m)}), \tag{4.43}$$

with

$$\bar{\mu}_{ir}^{(m)} = \Psi^{[r]}(\hat{\theta}^{(m)}) \hat{\mu}_i^{(m)} + \mathcal{G}(\hat{\theta}^{(m)}) U^{[r]}, \tag{4.44}$$

$$\bar{\Sigma}_{ir}^{(m)} = [\hat{\sigma}^2]^{(m)} I_N + \Psi^{[r]}(\hat{\theta}^{(m)}) \hat{\Gamma}_i^{(m)} \left[\Psi^{[r]}(\hat{\theta}^{(m)})\right]^T, \tag{4.45}$$

$$\mathcal{P}_{ir}^{(m)} = \hat{\Gamma}_i^{(m)} \left[\Psi^{[r]}(\hat{\theta}^{(m)})\right]^T \left(\bar{\Sigma}_{ir}^{(m)}\right)^{-1}, \tag{4.46}$$

$$\tilde{\mu}_{ir}^{(m)} = \hat{\mu}_i^{(m)} + \mathcal{P}_{ir}^{(m)} \left(Y^{[r]} - \bar{\mu}_{ir}^{(m)}\right), \tag{4.47}$$

$$\tilde{\Sigma}_{ir}^{(m)} = \left(I_{n_\Delta} - \mathcal{P}_{ir}^{(m)} \Psi^{[r]}(\hat{\theta}^{(m)})\right) \hat{\Gamma}_i^{(m)}. \tag{4.48}$$

Proof. The result is directly obtained from (4.37) by using the following identities:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}, \tag{4.49}$$

$$(I + BD^{-1}C)^{-1} = I - B(D + CB)^{-1}C, \tag{4.50}$$

$$\det(A)\det\Big(D - CA^{-1}B\Big) = \det(D)\det\Big(A - BD^{-1}C\Big). \tag{4.51}$$

Let us consider the following identity for a random variable x:

$$\mathbb{E}\left\{(b - Ax)^T(b - Ax)\right\} = (b - A\hat{x})^T(b - A\hat{x}) + \mathbf{tr}\left\{A\Sigma A^T\right\}, \tag{4.52}$$

where $\hat{x} = \mathbb{E}\{x\}$ and $\Sigma = \mathbb{E}\left\{(x - \hat{x})(x - \hat{x})^T\right\}$.

Using (4.49)–(4.52) we can express $\log[\mathcal{K}(\beta_i, \eta^{[r]})]$ in (4.41) as follows:

$$\begin{aligned}\log[\mathcal{K}(\beta_i, \eta^{[r]})] = \frac{N}{2}\log[\sigma^2] - \frac{1}{2}\log[\det(\Gamma_i)] - \frac{1}{2\sigma^2}\Big[(Y^{[r]} - \mathcal{G}(\theta)U^{[r]} - \Psi^{[r]}(\theta)\tilde{\mu}_{ir}^{(m)})^T \\ (Y^{[r]} - \mathcal{G}(\theta)U^{[r]} - \Psi^{[r]}(\theta)\tilde{\mu}_{ir}^{(m)}) + \mathbf{tr}\left(\Psi^{[r]}(\theta)\tilde{\Sigma}_{ir}^{(m)}[\Psi^{[r]}(\theta)]^T\right)\Big] - \\ \frac{1}{2}\Big[\mathbf{tr}(\Gamma_i^{-1}\tilde{\Sigma}_{ir}^{(m)}) + (\tilde{\mu}_{ir}^{(m)} - \mu_i)^T\Gamma_i^{-1}(\tilde{\mu}_{ir}^{(m)} - \mu_i)\Big].\end{aligned} \tag{4.53}$$

□

From Lemma 6, substituting (4.53) in (4.41), the auxiliary function $\mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)})$ can be expressed as:

$$\mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa}\log[\mathcal{K}(\beta_i, \eta^{[r]})]\hat{\zeta}_{ir}^{(m)}, \tag{4.54}$$

with

$$\hat{\zeta}_{ir}^{(m)} = \frac{\hat{\alpha}_i^{(m)}\mathcal{N}\left(Y^{[r]}; \bar{\mu}_{ir}^{(m)}, \bar{\Sigma}_{ir}^{(m)}\right)}{\sum_{l=1}^{\kappa}\hat{\alpha}_l^{(m)}\mathcal{N}\left(Y^{[r]}; \bar{\mu}_{lr}^{(m)}, \bar{\Sigma}_{lr}^{(m)}\right)}. \tag{4.55}$$

Finally, we can formulate the following iterative algorithm:

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{r=1}^{M}\mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)}), \tag{4.56}$$

$$\hat{\beta}^{(m+1)} = \arg\max_{\beta}\mathcal{Q}(\beta, \hat{\beta}^{(m)}), \quad \text{s.t.} \quad \sum_{i=1}^{\kappa}\alpha_i = 1, \quad 0 \leq \alpha_i \leq 1. \tag{4.57}$$

Notice that (4.56) and (4.57) are closely related to the *E-step* and the *M-step*, respectively [25].

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

In order to solve the optimization problem in (4.57), we consider the coordinate descent algorithm [109] using the following steps:

(I) Fix the vector of parameters θ at its value from the current iteration $\hat{\theta}^{(m)}$ to optimize (4.56) with respect to the GMM parameters γ and the noise variance σ^2.

(II) Fix the GMM parameters and the noise variance at their values from the iteration $\hat{\gamma}^{(m+1)}$ and $[\hat{\sigma}^2]^{(m+1)}$ and solve the optimization problem in (4.57) to obtain $\hat{\theta}^{(m+1)}$.

For the maximization problem in (4.57), we can obtain closed form expressions for γ and σ^2. The optimization of the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ can be carried out as described below.

Lemma 7. *Consider the maximization problem stated in (4.57) using (4.56). Under the standing assumptions, the M-step in (4.57) is computed using the following steps:*

(i) Solve (4.57) using (4.56) with $\theta = \hat{\theta}^{(m)}$:

$$\hat{\alpha}_i^{(m+1)} = \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} \right) / M, \tag{4.58}$$

$$\hat{\mu}_i^{(m+1)} = \left(\sum_{r=1}^{M} \tilde{\mu}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \right) / \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} \right), \tag{4.59}$$

$$\hat{\Gamma}_i^{(m+1)} = \left(\sum_{r=1}^{M} \left[(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \tilde{\Sigma}_{ir}^{(m)} \right] \hat{\zeta}_{ir}^{(m)} \right) / \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} \right), \tag{4.60}$$

$$\hat{\sigma}^{2^{(m+1)}} = \left(\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathcal{M}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \right) / NM, \tag{4.61}$$

with

$$\mathcal{M}_{ir}^{(m)} = \left(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)}) U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)}) \tilde{\mu}_{ir}^{(m)} \right)^T \left(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)}) U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)}) \tilde{\mu}_{ir}^{(m)} \right) + \mathbf{tr} \left(\Psi^{[r]}(\hat{\theta}^{(m)})^T \tilde{\Sigma}_{ir}^{(m)} \Psi^{[r]}(\hat{\theta}^{(m)}) \right). \tag{4.62}$$

(ii) Solve (4.57) using (4.56) with $\gamma = \hat{\gamma}^{(m+1)}$:

$$\hat{\theta}^{(m+1)} = \arg\min_{\theta} \sum_{r=1}^{M} \sum_{i=1}^{\kappa} \hat{\zeta}_{ir}^{(m)} \mathcal{B}_{ir}(\theta, \theta^{(m)}), \tag{4.63}$$

where

$$\mathcal{B}_{ir}(\theta, \theta^{(m)}) = \left(Y^{[r]} - \mathcal{G}(\theta) U^{[r]} - \Psi^{[r]}(\theta) \tilde{\mu}_{ir}^{(m+1)} \right)^T \left(Y^{[r]} - \mathcal{G}(\theta) U^{[r]} - \Psi^{[r]}(\theta) \tilde{\mu}_{ir}^{(m+1)} \right) + \mathbf{tr} \left(\Psi^{[r]}(\theta)^T \tilde{\Sigma}_{ir}^{(m+1)} \Psi^{[r]}(\theta) \right). \tag{4.64}$$

Proof. See Appendix 4.A □

From Lemma 7, we obtain closed form expressions for the GMM and noise variance estimators. They are computed using the procedure described in Section 2.5 where an auxiliary function is built without explicitly defining a *hidden* variable. An alternative solution is to formulate a classical EM algorithm with GMMs in which a *hidden* variable, modeled as an indicator, is considered to determine from which GMM component a set of observations comes from [68]. However, when the integral equation in (4.41) has closed form solution, the auxiliary function resulting is equal to (4.56) and the estimators are obtained from (4.58)–(4.63).

Finally, the estimation algorithm is summarized in *Algorithm 4.1*.

Algorithm 4.1 *(EM algorithm using SE with GMMs)*

Inputs $Y^{[1:M]}$, $U^{[1:M]}$, M, κ , $\hat{\theta}^{(0)}$, $\hat{\gamma}^{(0)}$ and $\hat{\sigma}^{2^{(0)}}$.
Outputs $\hat{\theta}$, $\hat{\gamma}$ and $\hat{\sigma}^2$.

$m \leftarrow 0$
E-step:
  Compute $\bar{\mu}_{ir}^{(m)}$ and $\bar{\Sigma}_{ir}^{(m)}$ from (4.44) and (4.45).
  Compute $\tilde{\mu}_{ir}^{(m)}$ and $\tilde{\Sigma}_{ir}^{(m)}$ from (4.47) and (4.48).
  Compute $\hat{\zeta}_{ir}^{(m)}$ from (4.55).
M-step:
Estimate $\hat{\gamma}^{(m+1)}$, $\hat{\sigma}^{2^{(m+1)}}$ from (4.58)–(4.61).
Compute $\mathcal{B}_{ir}(\theta, \theta^{(m)})$ from (4.64) using $\hat{\gamma}^{(m+1)}$ and $\hat{\theta}^{(m)}$.
Estimate $\hat{\theta}^{(m+1)}$ by solving (4.63).
if stopping criterion is not satisfied **then**
  $m \leftarrow m + 1$,
  return to 2
else
  $\hat{\theta} \leftarrow \hat{\theta}^{(m+1)}$, $\hat{\gamma} \leftarrow \hat{\gamma}^{(m+1)}$, $\hat{\sigma}^2 \leftarrow \hat{\sigma}^{2^{(m+1)}}$
end if
End

4.6 Numerical examples

In this section, we present two numerical examples with different approaches to illustrate the benefits of our proposal for modeling both the nominal model and the error-model. In the first example, we consider a variant of the example used in [26] holding all standing assumptions stated in Section 4.2.2. The nominal system model ($G_0(z^{-1}, \theta)$ and $H(z^{-1}, \theta)$) corresponds to a Box-Jenkins model. For simplicity, we also consider that the error-model ($G_\Delta(z^{-1}, \eta^{[r]})$) corresponds to a second-order FIR system in (4.11) with an overlapped Gaussian mixture error-model distribution. We assume that there are not dynamic uncertainties at low frequencies, i.e., the error-model has zero static gain. In this case, we do not establish a comparison analysis with other approaches since the simulated data is generated using our proposed framework with SE and GMMs.

In contrast, we consider a second example to illustrate the flexibility of our proposed algorithm for modeling the error-model for system (4.1). Assumption **A4** is relaxed and the model structure does not correspond to (4.2). Instead, we consider (4.1) with an output-error (OE) model structure where the system model, $G^{[r]}(z^{-1}, \theta)$, is given by the discrete-time model of a simple continuous-time Resistor-Capacitor system model. Here, we consider that the time constant τ is given by $\tau = \tau_0(1 + \delta^{[r]})$, where τ_0 represents the time constant computed with the nominal values of the resistor and capacitor, and $\delta^{[r]}$ denotes the tolerance (uncertainty) with a non-Gaussian-sum distribution. For comparison purposes, we consider a MEM approach described in Section 4.3.2 to obtain an uncertainty quantification of the corresponding system model.

For both examples, the simulation setup is as follows:

(1) The data length is $N = 100$.
(2) The number of experiments is $M = 100$.

(3) The number of Monte Carlo (MC)[4] simulation is 25.
(4) The stopping criterion is satisfied when:

$$\frac{\left\|\hat{\beta}^{(m)} - \hat{\beta}^{(m-1)}\right\|}{\left\|\hat{\beta}^{(m)}\right\|} < 10^{-6},$$

or when 2000 iterations of the EM algorithm have been reached.

The initialization of the estimation algorithm is as follows: For the nominal model initialization, we estimate a system model for each independent experiment using PEM and a sufficiently flexible (high order) structure. Then, we choose the estimated system model and the corresponding estimated noise variance with the smallest number of parameters that better described all the previously estimated models. For the error-model FIR transfer function, we adopt a similar approach for different FIR filter orders (n_Δ). Here, we choose a sufficiently complex FIR system by assuming the nominal model is equal to one ($G_0(z^{-1}, \theta) = 1$). The number of Gaussian mixture model components (κ) is defined by the user, and it is typically a large number. Once the parameters have been estimated, it is possible to discard the Gaussian component with a small estimated weight (e.g., $1/(100\kappa)$). Finally, for a given κ, the initial distribution of the parameters of the error-model is chosen as follows:

(1) From all the estimated FIR models, compute the average value of the coefficient corresponding to each tap.
(2) Evenly space the initial mean values of the n_Δ-dimensional GMM between the estimated maximum and minimum values of each tap.
(3) Set the variances of the n_Δ-dimensional GMM equal to a diagonal covariance matrix with the sample variance of each tap on the diagonal.
(4) Set the mixing weight for each GMM component equal to $1/\kappa$.

4.6.1 Example 1: A general system with a Gaussian mixture error-model distribution

Consider the *true* system $G^{[r]}(z^{-1})$ in (4.1) with a nominal system model as follows:

$$G_0(z^{-1}, \theta) = \frac{b_1^0 z^{-1}}{1 + a_1^0 z^{-1}}, \quad H(z^{-1}, \theta) = \frac{1 + c_1^0 z^{-1}}{1 + d_1^0 z^{-1}}, \tag{4.65}$$

and the error-model (4.11) given by

$$G_\Delta(z^{-1}, \eta^{[r]}) = \eta_0^{[r]} + \eta_1^{[r]} z^{-1}, \tag{4.66}$$

where $b_1^0 = 1$, $a_1^0 = -0.5$, $c_1^0 = 0.1$ and $d_1^0 = 0.8$. The input signal is $u_t^{[r]} \sim \mathcal{N}(0, \sigma_u^2)$, $\sigma_u^2 = 10$, and the noise source $\omega_t^{[r]} \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = 0.1$. In this example, we focus on uncertainties in the system model at high frequencies, i.e., $\eta_1 = -\eta_0$. We also consider that in each experiment, $\eta^{[r]} = [\eta_0 \ \eta_1]^T$ is drawn from a Gaussian mixture distribution (4.3) with $\alpha_1 = \alpha_2 = 0.5$, $\Gamma_1 = \Gamma_2 = 0.2$, $\mu_1 = -1$ and $\mu_2 = 1$. The vector of parameters to

[4]Each MC simulation corresponds to an estimation obtained using the data from M independent experiments.

Table 4.1: Nominal model parameters and noise variance estimated for Example 1

Parameter	True value	Estimated value
b_1	1	$0.9999 \pm 1.10 \times 10^{-3}$
a_1	-0.5	$-0.5001 \pm 0.60 \times 10^{-3}$
c_1	0.1	$0.1010 \pm 1.28 \times 10^{-2}$
d_1	0.8	$0.8021 \pm 7.20 \times 10^{-3}$
σ^2	0.1	$0.1001 \pm 1.40 \times 10^{-3}$

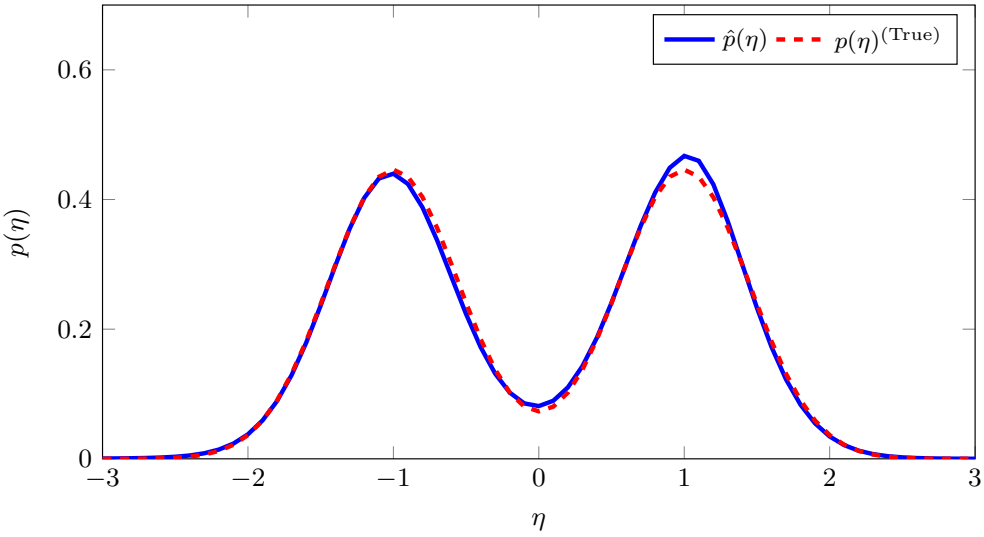


Figure 4.5: Estimated error-model distribution for the Example 1.

estimate is $\beta = \{\theta, \gamma, \sigma^2\}$ with $\theta = \{b_1^0, a_1^0, c_1^0, d_1^0\}$ and $\gamma = \{\alpha_i, \mu_i, \Gamma_i\}_{i=1}^{\kappa}$ with $\kappa = 2$. The initial values are $\hat{\theta}^{(0)} = [1.2510 \ -0.4523 \ 0.0772 \ 0.7546]^T$, $\hat{\sigma}^{2^{(0)}} = 0.0612$, $\hat{\alpha}_1^{(0)} = \hat{\alpha}_2^{(0)} = 0.5$, $\hat{\mu}_1^{(0)} = -1.1962$, $\hat{\mu}_2^{(0)} = 2.9673$, $\hat{\Gamma}_1^{(0)} = \hat{\Gamma}_2^{(0)} = 0.8651$.

Table 4.1 shows the estimation results of the nominal model parameters and the noise variance. Figure 4.5 shows the results of the estimation of the error-model distribution. The blue line corresponds to the average of all the GMMs estimated. It is clear that the estimated PDF is similar to the *true* PDF.

On the other hand, we note that the SE method does not directly provide an uncertainty region. However, we compute M realizations of the model

$$\hat{G}_0(z^{-1}, \hat{\theta}) + \hat{G}_0(z^{-1}, \hat{\theta}) G_\Delta(z^{-1}, \eta^{[r]}), \tag{4.67}$$

with $r = 1, ..., M$, where $\hat{G}_0(z^{-1}, \hat{\theta})$ is the estimated nominal model, and $\eta^{[r]}$ are obtained from M different realizations using the GMM in (4.3) with the estimated parameters $\hat{\gamma}$. In addition, we use PEM [1, 2] to estimate the system model $G^{[r]}(z^{-1})$ with a high order FIR model. We consider the measurements of 25 independent experiments, and we estimate a 10th order FIR system model, $\hat{G}(z^{-1}, \theta)^{(\text{FIR})}$, for each independent experiment in order to validate the estimated uncertainty region.

Figure 4.6 shows the magnitude of the frequency response corresponding to the average

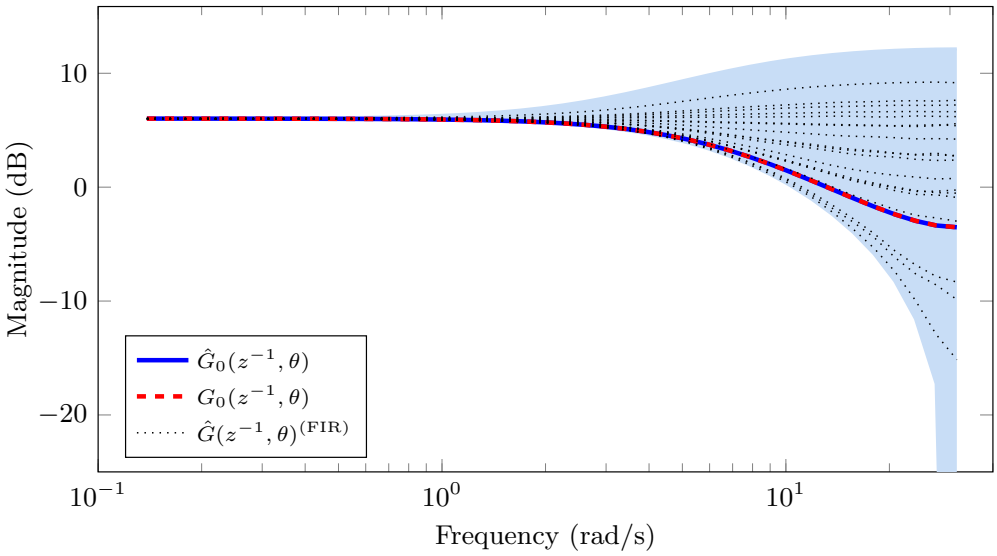


Figure 4.6: Estimated nominal model $G_0(z^{-1},\theta)$ for the Example 1.

of all MC simulations for the estimated nominal model (blue line). The red-dashed line corresponds to the nominal system model in (4.65), i.e. $G_0(z^{-1},\theta)$. The blue-shaded region represents the area of the corresponding estimated uncertainty region of the *true* system. We observe an accurate estimation of the nominal system model. We also observe that the estimated FIR models (black-dotted lines) lie in the uncertainty region obtained with the proposed method.

4.6.2 Example 2: A Resistor-Capacitor system model with non-Gaussian uncertainties

Consider the system (4.1) as follows:

$$G^{[r]}(z^{-1},\tau^{[r]}) = \frac{z^{-1}\left(1-e^{-T_s/\tau^{[r]}}\right)}{1-\left(e^{-T_s/\tau^{[r]}}\right)z^{-1}}, \qquad H(z^{-1}) = 1, \tag{4.68}$$

where $T_s = 0.01$ is the sampled period, and $\tau^{[r]} = \tau_0(1+\delta^{[r]})$ is the time constant with $\tau_0 = 10\times 10^{-3}$ and $\delta^{[r]}$ is uniformly distributed as $\mathcal{U}[-0.4, 0.4]$. We consider that the sampled period T_s is known.

For the SE method, we consider an OE nominal model with a FIR error-model as follows:

$$G_0(z^{-1},\theta) = \frac{z^{-1}\left(1-e^{-T_s/\tau_0}\right)}{1-\left(e^{-T_s/\tau_0}\right)z^{-1}}, \tag{4.69}$$

$$G_\Delta(z^{-1},\eta^{[r]}) = \eta_0 - \eta_0 z^{-1}. \tag{4.70}$$

The distribution of the parameters $\eta^{[r]}$ of the error-model is modeled using a GMM with $\gamma = \{\alpha_i,\mu_i,\Gamma_i\}_{i=1}^{\kappa}$, $\kappa = 2$ components. The vector of parameter to be estimated is $\beta = \{\tau_0,\gamma,\sigma^2\}$. As in the previous example, the uncertainty region is obtained from (4.67) with $r = 1,...,M$, where $\eta^{[r]}$ are obtained from M different realizations using the GMM in (4.3)

with the estimated parameters $\hat{\gamma}$. In addition, the initial values used for our proposed EM algorithm are given by $\hat{\tau}_0^{(0)} = 9 \times 10^{-3}$, $\hat{\sigma}^{2^{(0)}} = 0.1247$, $\hat{\alpha}_1^{(0)} = \hat{\alpha}_2^{(0)} = 0.5$, $\hat{\mu}_1^{(0)} = -0.0163$, $\hat{\mu}_2^{(0)} = 0.8025$, $\hat{\Gamma}_1^{(0)} = \hat{\Gamma}_2^{(0)} = 0.1092$.

For comparison purposes, we estimate the system uncertainty and the nominal model using MEM approach [21]. The estimated nominal model, $\hat{G}_0(z^{-1}, \theta)^{(\text{MEM})}$, corresponds to the average of all PEM estimations obtained using the original structure given in (4.68), i.e $n_a = n_b = 1$ and $n_c = n_d = 0$, from the all independent experiments. In order to obtain the residuals, we compute $\varepsilon_t = y_t - \hat{G}_0(z^{-1}, \theta)^{(\text{MEM})} u_t$ for each independent experiment. The error-model, $G_\varepsilon(z^{-1}, \theta_\varepsilon)$, is obtained from the residual data as follows:

$$\varepsilon_t = G_\varepsilon(z^{-1}, \theta_\varepsilon) u_t + v_t, \tag{4.71}$$

where θ_ε is the vector of parameters of the error-model, and v_t is a zero-mean Gaussian noise sequence which is assumed to be uncorrelated with the input signal u_t. Thus, we consider a 10th order FIR system model for $G_\varepsilon(z^{-1}, \theta_\varepsilon)$ and we use the PEM [2] to estimate the error-model parameters θ_ε for each experiment. The uncertainty region is computed by adding frequency by frequency the average of the all error-model estimated to the nominal model.

Finally, we consider the measurements of 25 independent experiments and we obtain an estimation of the system model (4.68), $\hat{G}(z^{-1}, \theta)^{(\text{PEM})}$, using PEM with $n_a = n_b = 1$ and $n_c = n_d = 0$ for each independent experiment in order to validate the uncertainty region estimated for both SE and MEM approach.

Figure 4.7 shows the magnitude of the frequency response corresponding to the nominal model and the uncertainty region estimated using SE approach. The red-dashed line corresponds to the system model (4.68) considering the time constant without uncertainty, i.e. $G(z^{-1}, \tau_0)$. The blue line corresponds to the average of all MC simulations for the estimated nominal model. We observe an accurate estimation for the nominal model with $\hat{\tau}_0 = 9.7 \times 10^{-3} \pm 0.17 \times 10^{-3}$ and noise variance $\hat{\sigma}^2 = 0.133 \pm 5.9 \times 10^{-3}$. The blue-shaded area corresponds to the uncertainty region estimated for the *true* system. Similarly, Figure 4.8 shows the magnitude of the frequency response corresponding to the nominal model estimated (blue line) and the uncertainty region estimated (blue-shaded area) from the residuals using MEM. We also observe an accurate estimation of the nominal model with $\hat{\tau}_0 = 10.1 \times 10^{-3} \pm 2.30 \times 10^{-3}$ and noise variance $\hat{\sigma}^2 = 0.099 \pm 2.85 \times 10^{-3}$. In addition, we observe in both Figure 4.7 and Figure 4.8 that PEM estimations, $\hat{G}(z^{-1}, \theta)^{(\text{PEM})}$, lie in the uncertainty region obtained using both SE approach and MEM. However, SE method describes better the uncertainty region than MEM method, specifically, the error-model at low frequencies.

4.7 On the uncertainty modeling for continuous-time (CT) systems

Several methods to identify continuous-time (CT) linear systems have been developed from the corresponding discrete-time model (DT) assuming that the CT model structure is known. However, when the model structure is unknown, orthonormal Basis Functions (BFs) can be

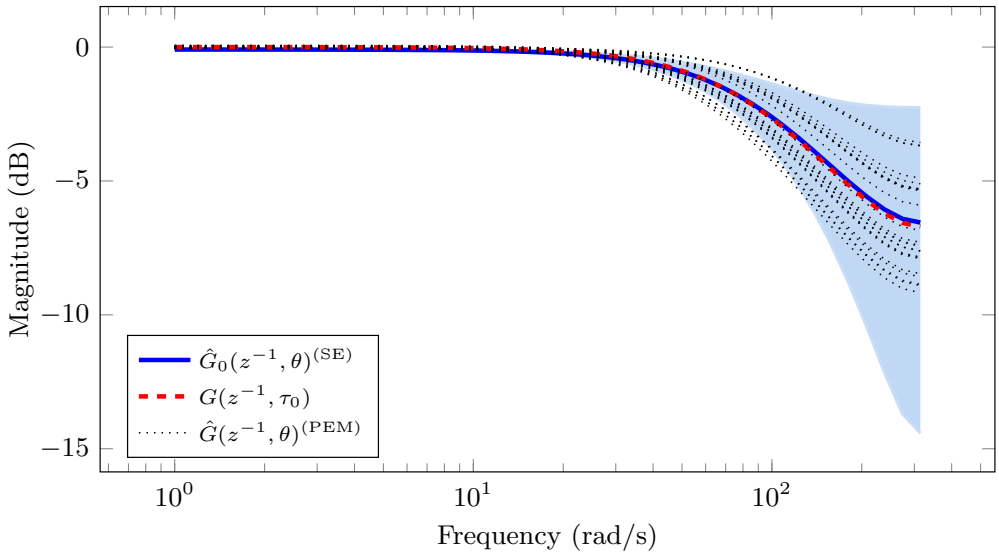


Figure 4.7: Estimated nominal model $G_0(z^{-1},\theta)$ using SE approach for the Example 2.

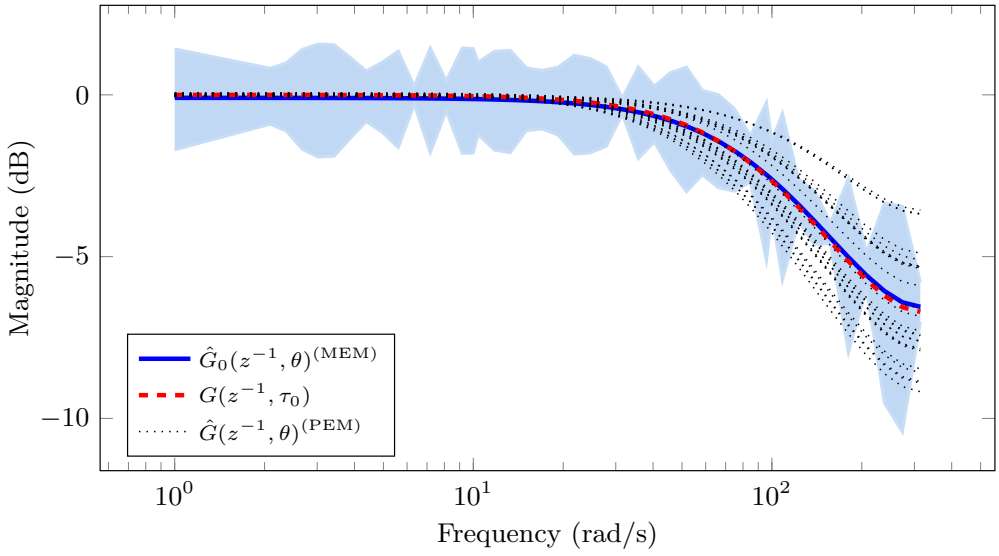


Figure 4.8: Estimated nominal model $G_0(z^{-1},\theta)$ using MEM for the Example 2.

used to obtain accurate models with low complexity. In this section, we use the ideas described in previous sections for developing an error-model identification algorithm for a CT system utilizing the SE approach with sampled data. A linear combination of orthonormal BFs is used to model both the CT nominal model and the stochastic error-model. In particular, the distribution of the stochastic weights associated with the error-model BFs is modeled as a GMM. Then, an ML estimation algorithm is developed to obtain an estimation of both the CT nominal model and the error-model distribution.

4.7.1 CT system of interest with SE approach and sampled data

Here, a general description of a CT system with SE approach using orthonormal BFs is stated. We also list the standing assumptions considered to formulate the ML estimation algorithm. Finally, the exact DT system model is presented as a linear regression using BFs.

General description

Consider the following CT system:

$$y(t)^{[r]} = G^{[r]}(s)u(t)^{[r]}, \tag{4.72}$$

where $r = 1, ..., M$, denotes the r-th realization of the system, with M the number of independent experiments, $y(t)^{[r]}$ is the output signal, $u(t)^{[r]}$ is the input signal, s is the time-derivative operator ($s = \frac{d}{dt}$) or the Laplace transform variable, and $G^{[r]}(s)$ is the rth realization of the *true* system transfer function (TF). We assume that $G^{[r]}(s)$ can be described as follows:

$$G^{[r]}(s) = G_0(s, \theta) + G_\epsilon^{[r]}(s, \eta^{[r]}), \tag{4.73}$$

where $G_0(s, \theta)$ is the nominal model and $G_\epsilon^{[r]}(s, \eta^{[r]})$ is the additive error-model given by the following basis expansions [131]:

$$G_0(s, \theta) = \sum_{j=1}^{n_0} \theta_j \mathcal{B}_j(s), \tag{4.74}$$

$$G_\epsilon^{[r]}(s, \eta^{[r]}) = \sum_{l=1}^{n_\epsilon} \eta_l^{[r]} \mathcal{P}_l(s), \tag{4.75}$$

where n_0 and n_ϵ are the number of orthonormal BFs of $\mathcal{B}_j(s)$ and $\mathcal{P}_l(s)$ used for the nominal model and the error-model, respectively, θ_j is the weight associated with the j-th BF for the nominal model and $\eta_l^{[r]}$ is the stochastic weight associated with the l-th BF for the error-model. A PDF that models $\eta^{[r]} = [\eta_1^{[r]} \ \cdots \ \eta_{n_\epsilon}^{[r]}]^T$ as a GMM is given by:

$$p(\eta^{[r]}|\gamma) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i), \tag{4.76}$$

$$\gamma = [\underbrace{\alpha_1 \ \mu_1 \ \Gamma_1}_{\gamma_1} \ \cdots \ \underbrace{\alpha_\kappa \ \mu_\kappa \ \Gamma_\kappa}_{\gamma_\kappa}]^T, \tag{4.77}$$

where $\alpha_i > 0$, $\sum_{i=1}^{\kappa} \alpha_i = 1$ and $\mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i)$ represents an n_ϵ dimensional Gaussian distribution with mean μ_i and covariance matrix Γ_i.

Remark 13. *Notice that M independent realizations of system (4.73) are used to obtain an estimation of the error-model PDF since one realization does not provide sufficient information to obtain an accurate stochastic description. This approach differs from the SE framework in [22] where a Gaussian stochastic assumption is considered to obtain error-model bounds from measurements of one realization of the DT true system.* ▽

The exact discrete-time (DT) system model

Let us consider that the DT system model from the CT system in (4.72) is given by:

$$y_k^{[r]} = \bar{G}^{[r]}(z^{-1})u_k^{[r]} + \omega_k^{[r]}, \tag{4.78}$$

where $y_k^{[r]} \in \mathbb{R}$ denotes the sampled output signal, $u_k^{[r]} \in \mathbb{R}$ is the sampled input signal[5] and $\omega_k^{[r]} \in \mathbb{R}$ is a zero-mean Gaussian white noise sequence with variance σ^2. We utilized a Zero Order Hold (ZOH) and instantaneous sampling with period T_s in order to obtain the sampled output signal. Notice that $\omega_k^{[r]}$ appears for the first time in (4.78), i.e, it does not come from the discretization of the CT system model in (4.72). The corresponding sampled data TF can be obtained as follows [132][6]:

$$\bar{G}^{[r]}(z^{-1}) = \left(1 - z^{-1}\right) \mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{G^{[r]}(s)}{s}\right]\Bigg|_{t=kT_s}\right\}, \tag{4.79}$$

where $\mathcal{L}^{-1}$ is the inverse Laplace transform and $\mathcal{Z}$ is the Z-transform. Considering the DT model from (4.74) and (4.75), using (4.79), the system (4.78) can be written as follows:

$$y_k^{[r]} = \varphi_k^{[r]}\theta + \Psi_k^{[r]}\eta^{[r]} + \omega_k^{[r]}, \tag{4.80}$$

with

$$\varphi_k^{[r]} = [\bar{\mathcal{B}}_1(z^{-1})u_k^{[r]} \ \cdots \ \bar{\mathcal{B}}_{n_0}(z^{-1})u_k^{[r]}], \tag{4.81}$$

$$\Psi_k^{[r]} = [\bar{\mathcal{P}}_1(z^{-1})u_k^{[r]} \ \cdots \ \bar{\mathcal{P}}_{n_\epsilon}(z^{-1})u_k^{[r]}], \tag{4.82}$$

$$\theta = [\theta_1 \ \cdots \ \theta_{n_0}]^T, \tag{4.83}$$

where $\varphi_k^{[r]}\theta$ represents the output response corresponding to the nominal model, and $\Psi_k^{[r]}\eta^{[r]}$ corresponds to the signal due to the error-model.

Defining the observed data as $Y^{[r]} = [y_1^{[r]} \ \cdots \ y_N^{[r]}]^T$, we then obtain a regression model from (4.80) as follows:

$$Y^{[r]} = \mathbf{\Phi}^{[r]}\theta + \mathbf{\Psi}^{[r]}\eta^{[r]} + W^{[r]}, \tag{4.84}$$

where $Y^{[r]}, W^{[r]} \in \mathbb{R}^{N\times 1}$, $\theta \in \mathbb{R}^{n_0\times 1}$, $\eta^{[r]} \in \mathbb{R}^{n_\epsilon\times 1}$, $\mathbf{\Phi}^{[r]} \in \mathbb{R}^{N\times n_0}$, $\mathbf{\Psi}^{[r]} \in \mathbb{R}^{N\times n_\epsilon}$, and $W^{[r]} \sim \mathcal{N}(0, \sigma^2 I_N)$.

4.7.2 ML estimation for CT error-model using sampled data and GMMs

In this section, we develop an ML estimation algorithm for the CT system (4.72) utilizing the sampled data model in (4.84). We define the vector of parameters to be estimated as $\beta = [\theta^T \ \gamma^T \ \sigma^2]^T$ in order to formulate the ML estimator for the system (4.84), where γ is the vector of parameters in (4.77), θ are the weights associated with the nominal model in (4.84), and σ^2 is the covariance of the noise sequence $\omega_k^{[r]}$. Under the standing assumptions and marginalizing the *hidden* variable, $\eta^{[r]}$, the log-likelihood function, $\ell(\beta)$, for the DT

[5] The input signal $u_k^{[r]}$ is constant between samples for each experiment.
[6] $\bar{X}(z^{-1})$ denotes the exact discrete-time model of $X(s)$ computed using (4.79).

model (4.84) is given by:

$$\ell(\beta) = \sum_{r=1}^{M} \log \left\{ \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \alpha_i p(Y^{[r]}|\eta^{[r]}, \beta) \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i) d\eta^{[r]} \right\}, \tag{4.85}$$

$$p(Y^{[r]}|\eta^{[r]}, \beta) = \mathcal{N}(Y^{[r]}; \mathbf{\Phi}^{[r]}\theta + \mathbf{\Psi}^{[r]}\eta^{[r]}, \sigma^2 I_N). \tag{4.86}$$

Then, the ML estimator is given by:

$$\hat{\beta}_{\text{ML}} = \arg\max_{\beta} \ \ell(\beta), \quad \text{s.t} \ \ 0 \leq \alpha_i \leq 1, \ \sum_{i=1}^{\kappa} \alpha_i = 1. \tag{4.87}$$

Notice that the ML estimation problem in (4.87) is closely related to the ML formulation stated in (4.28). Hence, the formulation of the estimation algorithm yields closed form expressions for both nominal model and GMM estimators. The *E-step* in the proposed algorithm is given by:

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{r=1}^{M} \sum_{i=1}^{\kappa} \log[\mathcal{K}(\beta_i, \eta^{[r]})] \hat{\zeta}_{ir}^{(m)}, \tag{4.88}$$

where

$$\hat{\zeta}_{ir}^{(m)} = \frac{\hat{\alpha}_i^{(m)} \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}^{(m)}, \bar{\Sigma}_{ir}^{(m)})}{\sum_{l=1}^{\kappa} \hat{\alpha}_l^{(m)} \mathcal{N}(Y^{[r]}; \bar{\mu}_{lr}^{(m)}, \bar{\Sigma}_{lr}^{(m)})}, \tag{4.89}$$

$$\bar{\mu}_{ir}^{(m)} = \mathbf{\Psi}^{[r]} \hat{\mu}_i^{(m)} + \mathbf{\Phi}^{[r]} \hat{\theta}^{(m)}, \tag{4.90}$$

$$\bar{\Sigma}_{ir}^{(m)} = \hat{\sigma}^{2^{(m)}} I_N + \mathbf{\Psi}^{[r]} \hat{\Gamma}_i^{(m)} \mathbf{\Psi}^{[r]^T}, \tag{4.91}$$

$$\mathcal{K}(\beta_i, \eta^{[r]}) = \alpha_i \mathcal{N}(Y^{[r]}; \mathbf{\Phi}^{[r]}\theta + \mathbf{\Psi}^{[r]}\eta^{[r]}, \sigma^2 I_N) \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i). \tag{4.92}$$

Then, the *M-step* results in the following:

$$\hat{\alpha}_i^{(m+1)} = \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} \right) / M, \tag{4.93}$$

$$\hat{\mu}_i^{(m+1)} = \left(\sum_{r=1}^{M} \tilde{\mu}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \right) / \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} \right), \tag{4.94}$$

$$\hat{\Gamma}_i^{(m+1)} = \left(\sum_{r=1}^{M} \left[(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \tilde{\Sigma}_{ir}^{(m)} \right] \hat{\zeta}_{ir}^{(m)} \right) / \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} \right), \tag{4.95}$$

$$\hat{\theta}^{(m+1)} = \left(\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathbf{\Phi}^{[r]^T} \mathbf{\Phi}^{[r]} \hat{\zeta}_{ir}^{(m)} \right)^{-1} \left(\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathbf{\Phi}^{[r]^T} (Y^{[r]} - \mathbf{\Psi}^{[r]} \tilde{\mu}_{ir}^{(m)}) \hat{\zeta}_{ir}^{(m)} \right), \tag{4.96}$$

$$\hat{\sigma}^{2^{(m+1)}} = \left(\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathcal{M}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \right) / NM, \tag{4.97}$$

Algorithm 4.2 *(EM-based algorithm with GMMs for CT systems)*

Inputs $Y^{[1:M]}$, $\mathbf{\Phi}^{[1:M]}$, $\mathbf{\Psi}^{[1:M]}$, M, κ, $\hat{\theta}^{(0)}$, $\hat{\gamma}^{(0)}$ and $\hat{\sigma}^{2^{(0)}}$.
Outputs $\hat{\theta}$, $\hat{\gamma}$ and $\hat{\sigma}^2$.

$m \leftarrow 0$
E-step:
  Compute $\hat{\zeta}_{ir}^{(m)}$ from (4.89).
  Compute $\bar{\mu}_{ir}^{(m)}$ and $\bar{\Sigma}_{ir}^{(m)}$ from (4.90) and (4.91).
  Compute $\tilde{\mu}_{ir}^{(m)}$ and $\tilde{\Sigma}_{ir}^{(m)}$ from (4.98)–(4.100).
M-step:
  Estimate $\hat{\gamma}^{(m+1)}$ from (4.93)–(4.95).
  Estimate $\hat{\theta}^{(m+1)}$ and $\hat{\sigma}^{2^{(m+1)}}$ from (4.96) and (4.97).
if stopping criterion is not satisfied **then**
    $m \leftarrow m + 1$,
  return to 2
else
    $\hat{\theta} \leftarrow \hat{\theta}^{(m+1)}$, $\hat{\gamma} \leftarrow \hat{\gamma}^{(m+1)}$, $\hat{\sigma}^2 \leftarrow \hat{\sigma}^{2^{(m+1)}}$
end if
End

with

$$\mathcal{R}_{ir}^{(m)} = \hat{\Gamma}_i^{(m)}[\mathbf{\Psi}^{[r]}]^T(\bar{\Sigma}_{ir}^{(m)})^{-1}, \tag{4.98}$$

$$\tilde{\mu}_{ir}^{(m)} = \hat{\mu}_i^{(m)} + \mathcal{R}_{ir}^{(m)}(Y^{[r]} - \bar{\mu}_{ir}^{(m)}), \tag{4.99}$$

$$\tilde{\Sigma}_{ir}^{(m)} = (I_{n_\epsilon} - \mathcal{R}_{ir}^{(m)}\mathbf{\Psi}^{[r]})\hat{\Gamma}_i^{(m)}, \tag{4.100}$$

$$\mathcal{M}_{ir}^{(m)} = (Y^{[r]} - \mathbf{\Phi}^{[r]}\hat{\theta}^{(m)} - \mathbf{\Psi}^{[r]}\tilde{\mu}_{ir}^{(m)})^T(Y^{[r]} - \mathbf{\Phi}^{[r]}\hat{\theta}^{(m)} - \mathbf{\Psi}^{[r]}\tilde{\mu}_{ir}^{(m)}) + \mathrm{tr}\Big(\mathbf{\Psi}^{[r]}\tilde{\Sigma}_{ir}^{(m)}\mathbf{\Psi}^{[r]^T}\Big). \tag{4.101}$$

Finally, the estimation algorithm for CT system error-model identification is summarized in *Algorithm 4.2* (For more details, see Appendix 4.B).

4.7.3 Numerical simulation examples for CT error-model modeling

In this section we present two numerical examples to illustrate the benefits of our proposal for the identification of the CT system in (4.72), obtaining both the nominal model and the error-model distribution using sampled data. We consider the input signal $u_k \sim \mathcal{N}(0, \sigma_u^2)$, with $\sigma_u^2 = 10$, and the variance of the DT white Gaussian noise ω_k in (4.78) $\sigma^2 = 0.1$. We also consider the CT Laguerre basis functions in (4.74) given by [133, 134]:

$$\mathcal{B}_i(s, a_0) = \frac{\sqrt{2a_0}}{s + a_0}\left(\frac{s - a_0}{s + a_0}\right)^{i-1}, \tag{4.102}$$

$$\mathcal{P}_l(s, a_\epsilon) = \frac{\sqrt{2a_\epsilon}}{s + a_\epsilon} \left(\frac{s - a_\epsilon}{s + a_\epsilon} \right)^{l-1}, \tag{4.103}$$

where $\{a_0, a_\epsilon\}$ are called the time scaling factors for the Laguerre functions [135].

The simulation setup is as follows:

(1) The data length is $N = 50$.
(2) The sampled period is $T_s = 0.1$ ms.
(3) The number of experiments is $M = 150$.
(4) The number of MC[7] simulations is 50.
(5) The stopping criterion is chosen as:

$$\frac{\left\| \hat{\beta}^{(m)} - \hat{\beta}^{(m-1)} \right\|}{\left\| \hat{\beta}^{(m)} \right\|} < 0.5 \times 10^{-6},$$

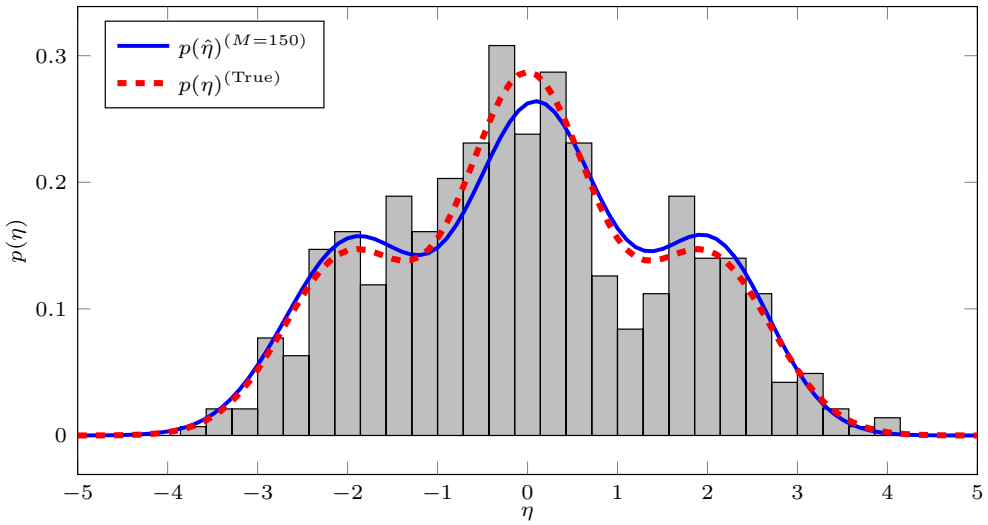
or when 5000 iterations have been reached.

For comparison purposes, we estimate the nominal model and the uncertainty region (UR) using the procedure that combines the SM and MEM approach with BFs proposed in [126]. In order to compare the models obtained by SE and SM methods, we generate M realizations of the model $G_0(s, \hat{\theta}) + G_\epsilon(s, \eta^{[r]})$ with $r = 1, .., M$, where $\eta^{[r]}$ are obtained from M different realizations using the PDF in (4.3) with the estimated parameters $\hat{\gamma}$.

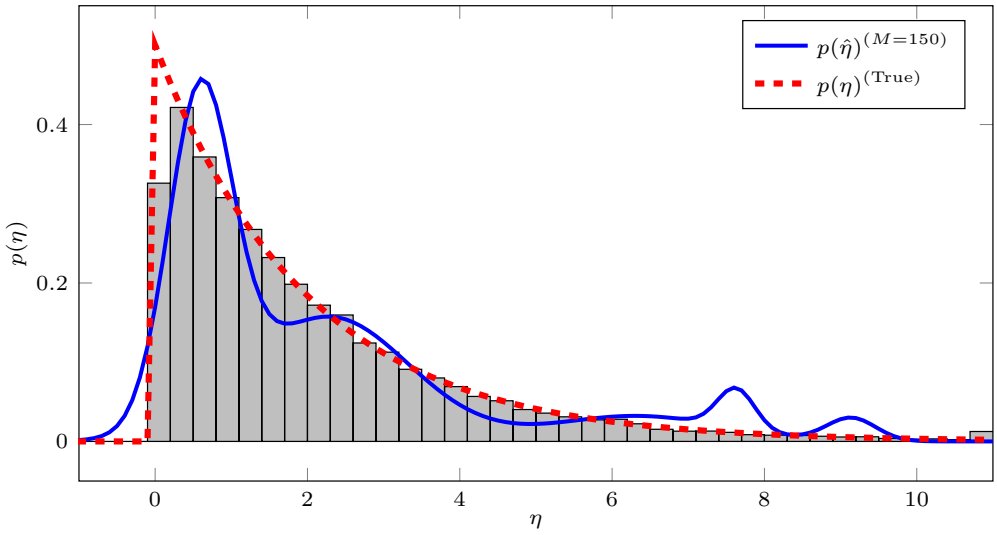
Example 1: A CT Laguerre basis system with an overlapped Gaussian mixture error-model distribution

Consider the CT system (4.73) with the nominal model and the error-model in (4.74) with $n_0 = n_\epsilon = 2$, $a_0 = 2$, and $a_\epsilon = 1$. The nominal model parameters (but unknown) θ are $\theta_0 = [5\ 7]^T$ and $\eta = [\eta_1\ -\eta_1]^T$ is a *hidden* variable. We also consider that in each experiment, η is drawn from a Gaussian mixture distribution (4.3) with $\alpha_1 = \alpha_3 = 0.25$, $\alpha_2 = 0.5$, $\mu_1 = -2$, $\mu_2 = 0$, $\mu_3 = 2$, and $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0.5$.

Figure 4.9(a) shows the results of the estimation of the error-model distribution. The blue-shaded bars represent the histogram of the corresponding error-model observations. The blue line corresponds to the average of all the estimates of the MC simulations. It is clear that the estimated PDF is similar to the *true* PDF when the number of experiments is high ($M = 150$). Figure 4.10(a) shows the magnitude of the frequency response corresponding to the average of all MC simulations for the estimated nominal model using both SE and SM approaches. The blue-shaded region represents the area of the corresponding UR of the *true* system. We observe accurate estimations of the nominal model using both SE and SM. However, the UR is small for high frequencies when our proposed algorithm is used.



(a) Example 1



(b) Example 2

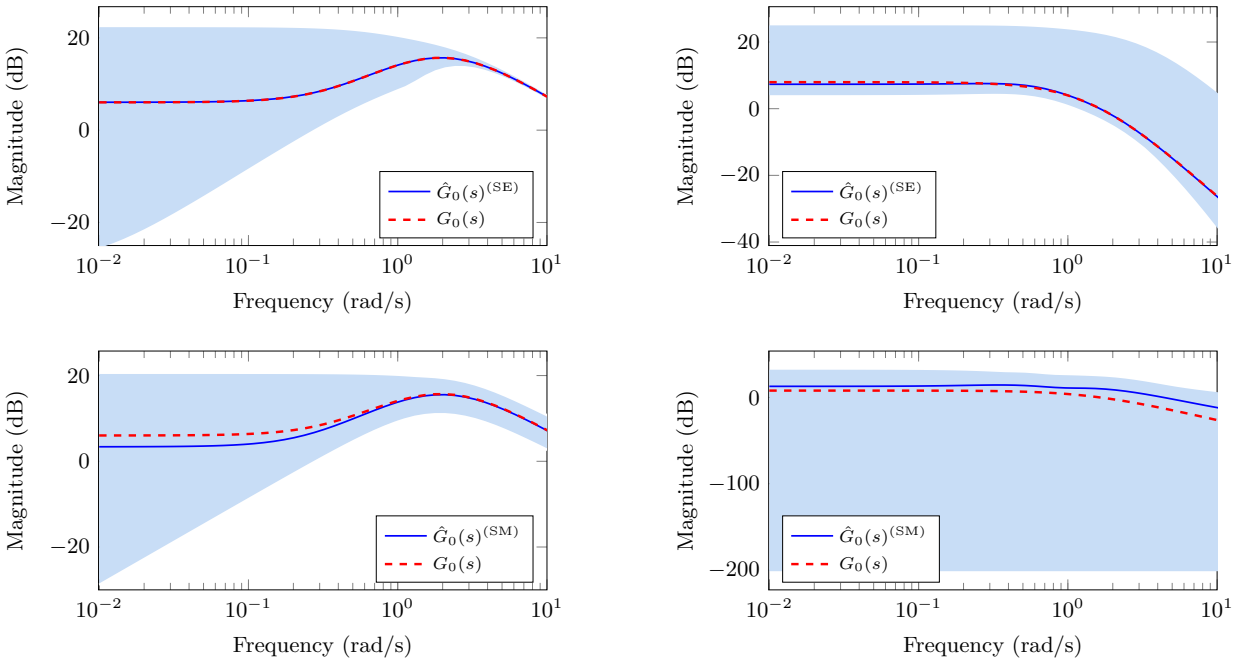
Figure 4.9: The estimated error-model distribution $p(\eta)$ using GMMs.

Example 2: Approximating a CT system with a χ^2 error-model distribution using Laguerre BFs and GMMs

In this example, we show how our proposal can be used to estimate a CT system when the nominal model (4.74) does not correspond to a finite summation of BFs but can be approximated by one. This approach has been tailored in control theory and system identification, specifically, using Laguerre BFs [134]. We also consider that the error-model non-Gaussian distribution does not correspond to a GMM but can be approximated by one. In this example the CT system (4.73) is given by:

$$G_0(s) = \frac{5}{(s+1)(s+2)}, \tag{4.104}$$

[7]Each MC simulation corresponds to an estimation of the nominal model and the error-model distribution obtained from M independent realisations of the error-model TF and the noise sequence.



(a) Example 1 (b) Example 2

Figure 4.10: Bode magnitude plots for the estimated CT nominal system using SE and SM approach.

with $G_\epsilon(s,\eta)$ given by (4.75) with $n_\epsilon = 2$, $a_\epsilon = 3.5$, and $\eta = [\eta_1 \quad -\eta_1]^T$ is a *hidden* variable. The error-model distribution corresponds to a *chi-square* (χ^2) distribution given by

$$p(\eta)^{(\mathrm{True})} = \frac{\eta^{((\nu-2)/2)} e^{-\eta/2}}{2^{(\nu/2)}\tilde{\Gamma}(\nu/2)}, \tag{4.105}$$

where $\nu = 2$ and $\tilde{\Gamma}(\cdot)$ is the Gamma function [136]. To approximate the nominal model (4.104), we consider (4.74) with $n_0 = 5$ and $a_0 = 1$. In addition, the error-model distribution is approximated using a GMM in (4.3) with $\kappa = 5$.

Figure 4.9(b) shows the estimated average error-model PDF for all the MC simulations. The blue-shaded bars represent the histogram of the corresponding error-model observations. We observe that the average of the estimated GMM fits the error-model distribution. Figure 4.10(b) shows the magnitude of the frequency response corresponding to the average of all MC simulations for the estimated nominal model. As in the previous example, the blue-shaded region represents the area in which the corresponding UR of the *true* system lies. We observe an accurate approximation of the nominal model using the Laguerre BFs. In addition, we observe a large UR when the SM approach is used.

4.8 Conclusions

In this chapter, we have addressed the uncertainty modeling problem by combining the SE approach with GMMs. We proposed an identification algorithm using the ML principle to estimate both the nominal model and the distribution of the parameters of the error-model as a GMM. An EM-based algorithm is proposed to solve the ML estimation problem providing

closed form expressions for the estimators of the GMM parameters and the noise variance. In the simulation examples, we considered two scenarios: i) identification of a system model with a structure defined using SE approach with a nominal model and an error-model as a FIR system with Gaussian mixture distribution, and ii) identification of a system model with non-Gaussian uncertainties in the parameters that does not correspond to the SE framework. In both cases, we obtained accurate estimations of the nominal system model. In addition, the proposed method provided a good description of the uncertainty region in the system for the simulation examples. Finally, we extended this approach for uncertainty modeling of CT linear systems using continuous-time BFs and sampled data. We also proposed an iterative algorithm using the exact DT system model to obtain the ML estimation of the nominal model and the error-model distribution as a GMM. In this case, the proposed algorithm resulted in closed form expressions for the estimation of the parameters of the nominal model, noise variance and the GMM.

Appendix

4.A Proof of Lemma 7

Taking the derivative of (4.56) with respect to μ_i and equating to zero yields:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \mu_i} = \sum_{r=1}^{M} \left[\Gamma_i^{-1}(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m+1)})\right] \hat{\zeta}_{ir}^{(m)} = 0. \tag{4.106}$$

Then, we obtain:

$$\hat{\mu}_i^{(m+1)} = \frac{\sum_{r=1}^{M} \tilde{\mu}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)}}{\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}}. \tag{4.107}$$

Taking the derivative of (4.56) with respect to Γ_i^{-1} and equating to zero:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \Gamma_i^{-1}} = \sum_{r=1}^{M} \left(\left[(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \tilde{\Sigma}_{ir}^{(m)}\right] + \hat{\Gamma}_i^{(m+1)}\right) \hat{\zeta}_{ir}^{(m)} = 0. \tag{4.108}$$

From (4.108) we then obtain:

$$\hat{\Gamma}_i^{(m+1)} = \frac{\sum_{r=1}^{M} \left[(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \tilde{\Sigma}_{ir}^{(m)}\right] \hat{\zeta}_{ir}^{(m)}}{\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}}. \tag{4.109}$$

Similarly, taking the derivative of (4.56) with respect to $\rho = \sigma^2$ and equating to zero yields:

$$\begin{aligned}\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \rho} = \sum_{r=1}^{M}\sum_{i=1}^{\kappa} \Big[&(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)})U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)})\hat{\tilde{\mu}}_{ir}^{(m)})^T(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)})U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)})\hat{\tilde{\mu}}_{ir}^{(m)}) \\ &+ \mathbf{tr}\left(\Psi^{[r]}(\hat{\theta}^{(m)})\hat{\tilde{\Sigma}}_{ir}^{(m)}[\Psi^{[r]}(\hat{\theta}^{(m)})]^T\right) + N\hat{\rho}^{(m+1)}\Big] \hat{\zeta}_{ir}^{(m)} = 0.\end{aligned} \tag{4.110}$$

Using $[\hat{\sigma}^2]^{(m+1)} = 1/\hat{\rho}^{(m+1)}$ and $\sum_{i=1}^{\kappa} \hat{\zeta}_{ir}^{(m)} = 1$ with the definition in (4.62), we have:

$$[\hat{\sigma}^2]^{(m+1)} = \frac{\sum_{r=1}^{M}\sum_{i=1}^{\kappa} \mathcal{M}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)}}{NM}. \tag{4.111}$$

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

$$\mathcal{R}(\alpha_i) = \sum_{r=1}^{M}\sum_{i=1}^{\kappa} \log[\alpha_i]\hat{\zeta}_{ir}^{(m)}, \tag{4.112}$$

subject to $\sum_{i=1}^{\kappa} \alpha_i = 1$. Then, using a Lagrange multiplier to deal with the constraint on α_i we define:

$$\mathcal{J}(\alpha_i, \epsilon) = \sum_{r=1}^{M}\sum_{i=1}^{\kappa} \log[\alpha_i]\hat{\zeta}_{ir}^{(m)} + \epsilon\left(\sum_{i=1}^{\kappa} \alpha_i - 1\right). \tag{4.113}$$

Taking the derivative of (4.113) with respect to α_i and ϵ and equating to zero we obtain:

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \alpha_i} = \sum_{r=1}^{M} \frac{\hat{\zeta}_{ir}^{(m)}}{\hat{\alpha}_i^{(m+1)}} - \epsilon = 0, \tag{4.114}$$

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \epsilon} = \sum_{i=1}^{\kappa} \alpha_i - 1 = 0. \tag{4.115}$$

Then, $\hat{\alpha}_i^{(m+1)} = \sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}/\epsilon$. Taking summation over $i = 1, ..., \kappa$ and using (4.115) we have:

$$\sum_{i=1}^{\kappa} \hat{\alpha}_i^{(m+1)} = \sum_{i=1}^{\kappa} \sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}/\epsilon, \tag{4.116}$$

$$\epsilon = \sum_{i=1}^{\kappa} \sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}. \tag{4.117}$$

Notice that $\sum_{i=1}^{\kappa} \hat{\zeta}_{ir}^{(m)} = 1$, then we have $\epsilon = M$. Then, we can obtain:

$$\hat{\alpha}_i{}^{(m+1)} = \frac{\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}}{M}. \tag{4.118}$$

Notice that $0 \leq \hat{\alpha}_i^{(m+1)} \leq 1$ holds even though we did not explicitly consider it in (4.113).

Finally, substituting (4.53) in (4.41) with $\gamma = \hat{\gamma}^{(m+1)} = \{\hat{\alpha}_i^{(m+1)}, \hat{\mu}_i^{(m+1)}, \hat{\Gamma}_i^{(m+1)}\}$ we can directly obtain $\mathcal{B}_{ir}(\theta, \theta^{(m)})$ in (4.64) for the parameter θ.

4.B EM-based algorithm for CT system error-model estimation

To formulate an EM-based estimation algorithm, we introduce the following standing assumptions:

(i) The nominal model $G_0(s, \theta)$ in (4.74) does not change from one experiment to another, whilst the error-model $G_\epsilon^{[r]}(s, \eta^{[r]})$ may change for each experiment and all the realizations of η are drawn from the same PDF parameterized by γ.

(ii) The number of CT orthonormal BFs n_0 and n_ϵ in (4.74) and (4.75), their underlying parameters, and the number of components κ of the error-model distribution (4.3) are known.

(iii) The system in (4.72) is operating in open loop and the input signal, $u(t)^{[r]}$, is an exogenous deterministic signal. Also, M independent experiments with different input signals are available.

Under the standing assumptions, the iterative algorithm is computed as follows:

4.B.1 Computing the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

From (4.85) and (4.86), we define the following:

$$\mathcal{K}(\beta_i, \eta^{[r]}) = \alpha_i \mathcal{N}(Y^{[r]}; \mathbf{\Phi}^{[r]}\theta + \mathbf{\Psi}^{[r]}\eta^{[r]}, \sigma^2 I_N)\mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i). \tag{4.119}$$

Then, the log-likelihood function (4.85) can be expressed as

$$\ell(\beta) = \sum_{r=1}^{M} \log\left[\mathcal{V}^{[r]}(\beta)\right], \tag{4.120}$$

with

$$\mathcal{V}^{[r]}(\beta) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \mathcal{K}(\beta_i, \eta^{[r]}) d\eta^{[r]}. \tag{4.121}$$

The expression in (4.121) can be written as $\log[\mathcal{V}^{[r]}(\beta)] = \mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)}) - \mathcal{H}^{[r]}(\beta, \hat{\beta}^{(m)})$, where:

$$\mathcal{Q}^{[r]}(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log[\mathcal{K}(\beta_i, \eta^{[r]})] \frac{\mathcal{K}(\hat{\beta}_i^{(m)}, \eta^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} d\eta^{[r]}, \tag{4.122}$$

$$\mathcal{H}^{[r]}(\beta, \hat{\beta}^{(m)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log\left[\frac{\mathcal{K}(\beta_i, \eta^{[r]})}{\mathcal{V}^{[r]}(\beta)}\right] \frac{\mathcal{K}(\hat{\beta}_i^{(m)}, \eta^{[r]})}{\mathcal{V}^{[r]}(\hat{\beta}^{(m)})} d\eta^{[r]}. \tag{4.123}$$

From Lemma 2, the function $\mathcal{H}^{[r]}(\beta, \hat{\beta}^{(m)})$ is decreasing function for any value of β. From Lemma 6 using (4.119) we can directly obtain:

$$\begin{aligned} \log[\mathcal{K}(\beta_i, \eta^{[r]})] = \frac{N}{2}\log[\sigma^2] - \frac{1}{2}\log[\det(\Gamma_i)] - \frac{1}{2\sigma^2}\Big[(Y^{[r]} - \mathbf{\Phi}^{[r]}\theta - \mathbf{\Psi}^{[r]}\tilde{\mu}_{ir}^{(m)})^T \\ (Y^{[r]} - \mathbf{\Phi}^{[r]}\theta - \mathbf{\Psi}^{[r]}\tilde{\mu}_{ir}^{(m)}) + \mathbf{tr}\left(\mathbf{\Psi}^{[r]}\tilde{\Sigma}_{ir}^{(m)}[\mathbf{\Psi}^{[r]}]^T\right)\Big] - \\ \frac{1}{2}\left[\mathbf{tr}(\Gamma_i^{-1}\tilde{\Sigma}_{ir}^{(m)}) + (\tilde{\mu}_{ir}^{(m)} - \mu_i)^T\Gamma_i^{-1}(\tilde{\mu}_{ir}^{(m)} - \mu_i)\right], \end{aligned} \tag{4.124}$$

where

$$\bar{\mu}_{ir}^{(m)} = \mathbf{\Psi}^{[r]}\hat{\mu}_i^{(m)} + \mathbf{\Phi}^{[r]}\hat{\theta}^{(m)}, \tag{4.125}$$

$$\bar{\Sigma}_{ir}^{(m)} = \hat{\sigma}^{2^{(m)}} I_N + \mathbf{\Psi}^{[r]}\hat{\Gamma}_i^{(m)}\mathbf{\Psi}^{[r]^T}, \tag{4.126}$$

$$\mathcal{R}_{ir}^{(m)} = \hat{\Gamma}_i^{(m)}[\mathbf{\Psi}^{[r]}]^T(\bar{\Sigma}_{ir}^{(m)})^{-1}, \tag{4.127}$$

$$\tilde{\mu}_{ir}^{(m)} = \hat{\mu}_i^{(m)} + \mathcal{R}_{ir}^{(m)}(Y^{[r]} - \bar{\mu}_{ir}^{(m)}), \tag{4.128}$$

$$\tilde{\Sigma}_{ir}^{(m)} = (I_{n_\epsilon} - \mathcal{R}_{ir}^{(m)}\mathbf{\Psi}^{[r]})\hat{\Gamma}_i^{(m)}. \tag{4.129}$$

Substituting (4.124) in (4.122), the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ is given by:

$$\mathcal{Q}(\beta, \hat{\beta}^{(m)}) = \sum_{r=1}^{M}\sum_{i=1}^{\kappa} \log[\mathcal{K}(\beta_i, \eta^{[r]})]\hat{\zeta}_{ir}^{(m)}, \tag{4.130}$$

with

$$\hat{\zeta}_{ir}^{(m)} = \frac{\hat{\alpha}_i^{(m)} \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}^{(m)}, \bar{\Sigma}_{ir}^{(m)})}{\sum_{l=1}^{\kappa} \hat{\alpha}_l^{(m)} \mathcal{N}(Y^{[r]}; \bar{\mu}_{lr}^{(m)}, \bar{\Sigma}_{lr}^{(m)})}. \tag{4.131}$$

4.B.2 Optimization of the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$

Taking the derivative of (4.130) with respect to μ_i and equating to zero yields:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \mu_i} = \sum_{r=1}^{M} \left[\Gamma_i^{-1}(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m+1)})\right] \hat{\zeta}_{ir}^{(m)} = 0. \tag{4.132}$$

From (4.132) we have:

$$\hat{\mu}_i^{(m+1)} = \frac{\sum_{r=1}^{M} \tilde{\mu}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)}}{\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}}. \tag{4.133}$$

Then, taking the derivative of (4.130) with respect to Γ_i^{-1} and equating to zero:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \Gamma_i^{-1}} = \sum_{r=1}^{M} \left(\left[(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \tilde{\Sigma}_{ir}^{(m)} \right] + \hat{\Gamma}_i^{(m+1)} \right) \hat{\zeta}_{ir}^{(m)} = 0. \tag{4.134}$$

From (4.134) we obtain:

$$\hat{\Gamma}_i^{(m+1)} = \frac{\sum_{r=1}^{M} \left[(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})(\tilde{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \tilde{\Sigma}_{ir}^{(m)} \right] \hat{\zeta}_{ir}^{(m)}}{\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}}. \tag{4.135}$$

Similarly, taking the derivative of (4.130) with respect to θ and equating to zero yields:

$$\sum_{r=1}^{M} \sum_{i=1}^{\kappa} [\mathbf{\Phi}^{[r]}]^T (Y^{[r]} - \mathbf{\Psi}^{[r]} \tilde{\mu}_{ir}^{(m)}) \hat{\zeta}_{ir}^{(m)} = \sum_{r=1}^{M} \sum_{i=1}^{\kappa} [\mathbf{\Phi}^{[r]}]^T \mathbf{\Phi}^{[r]} \hat{\theta}^{(m+1)} \hat{\zeta}_{ir}^{(m)}. \tag{4.136}$$

From (4.136) we obtain:

$$\hat{\theta}^{(m+1)} = \left(\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathbf{\Phi}^{[r]^T} \mathbf{\Phi}^{[r]} \hat{\zeta}_{ir}^{(m)} \right)^{-1} \left(\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathbf{\Phi}^{[r]^T} (Y^{[r]} - \mathbf{\Psi}^{[r]} \tilde{\mu}_{ir}^{(m)}) \hat{\zeta}_{ir}^{(m)} \right). \tag{4.137}$$

Using (4.101) and taking the derivative of (4.130) with respect to $\rho = 1/\sigma^2$ and equating to zero yields:

$$\sum_{r=1}^{M} \sum_{i=1}^{\kappa} N \hat{\zeta}_{ir}^{(m)} = \sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathcal{M}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \hat{\rho}^{(m+1)}. \tag{4.138}$$

Notice that $\sum_{i=1}^{\kappa} \hat{\zeta}_{ir}^{(m)} = 1$, then we can directly obtain:

$$[\hat{\sigma}^2]^{(m+1)} = \frac{\sum_{r=1}^{M} \sum_{i=1}^{\kappa} \mathcal{M}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)}}{NM}. \tag{4.139}$$

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

$$\mathcal{R}(\alpha_i) = \sum_{r=1}^{M}\sum_{i=1}^{\kappa} \log[\alpha_i]\hat{\zeta}_{ir}^{(m)}, \tag{4.140}$$

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

$$\mathcal{J}(\alpha_i, \epsilon) = \sum_{r=1}^{M}\sum_{i=1}^{\kappa} \log[\alpha_i]\hat{\zeta}_{ir}^{(m)} - \epsilon\left(\sum_{i=1}^{\kappa}\alpha_i - 1\right). \tag{4.141}$$

Taking the derivative of (4.141) with respect to α_i and ϵ and equating to zero we obtain:

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \alpha_i} = \left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)} / \hat{\alpha}_i^{(m+1)}\right) - \epsilon = 0, \tag{4.142}$$

$$\frac{\partial \mathcal{J}(\alpha_i, \epsilon)}{\partial \epsilon} = \sum_{i=1}^{\kappa}\alpha_i - 1 = 0. \tag{4.143}$$

Taking a summation over $i = 1 \ldots \kappa$ in (4.142) and using (4.143) we can obtain:

$$\hat{\alpha}_i^{(m+1)} = \frac{\left(\sum_{r=1}^{M} \hat{\zeta}_{ir}^{(m)}\right)}{M}. \tag{4.144}$$

Notice that $0 \leq \hat{\alpha}_i^{(m+1)} \leq 1$ holds, even though we did not explicitly consider it in (4.141).

Chapter 5
Conclusions and Future work

5.1 Conclusions

In this thesis the problem of identifying linear dynamic systems i) with structural and parametric non-Gaussian uncertainties, and ii) subject to non-Gaussian noise, has been addressed. For both scenarios, the identification algorithms were developed using the ML principle considering that the non-Gaussian distributions are modeled as GMMs. The main advantage of this approach is that a GMM can be used to obtain an accurate approximation of a non-Gaussian distribution. In general, the simulations exhibited great accuracy in the estimation of the dynamic system parameters.

In the first part of the thesis we addressed the problem of ML estimation with data augmentation approach when the likelihood function is expressed as an infinite mixture distribution. We considered that the *hidden* variables were modeled using GMMs. An EM-based algorithm with GMMs was developed in order to solve the associated ML estimation problem. We presented a systematic procedure to construct an auxiliary function to solve the expectation step of the proposed iterative algorithm. Solving the maximization step of the EM-based algorithm resulted in closed form expressions for the GMM estimators. In the numerical examples, we observed accurate estimations, even when a GMM is used to approximate a non-Gaussian-sum distribution. In addition, we showed how this estimation methodology can be used for estimating the distribution of stellar rotational velocities with a finite mixture distribution. In this problem, the measurements have been treated as random variables drawn from the projected rotational velocity PDF and the rotational velocity PDF was approximated as a finite mixture of Maxwellian distributions. The performance of the proposed algorithm was analyzed utilizing real observed data from three sets of measurements. We observed that the estimated MSA yield a better statistical description of the stellar rotational velocities than the Tikhonov Regularization Method used in the literature.

In the second part of the thesis, we addressed the identification of linear dynamic systems subject to non-Gaussian noise. In particular, we focused on the development of an ML estimator for a general class of linear dynamic systems with non-minimum-phase noise transfer function driven by an exogenous input signal when the noise PDF is a GMM. The ML estimation algorithm was formulated utilizing the prediction error that was computed using causal and anti-causal filtering techniques. In addition, the corresponding initial conditions of the filters of the prediction error were incorporated as deterministic parameters to be estimated. We proposed an EM algorithm with GMMs to solve the ML estimation problem, and the optimization problem was solved using the coordinate descent algorithm. Based on this approach, we obtained closed form expressions for the GMM estimators. We also showed that the proposed algorithm provides more accurate estimations of the system model model parameters than High-Order Moments method. Moreover, the proposed EM algorithm exhibited good accuracy even when the non-Gaussian noise distribution does not correspond to a GMM, but can be approximated by one. In addition, we showed that this proposal can be used to estimate dynamic systems with minimum-phase noise transfer function with

Gaussian mixture noise distributions, obtaining system model estimates more accurate than traditional prediction errors algorithms.

Finally, for the problem of structural and parametric uncertainty modeling in linear dynamic systems, the SE approach was used. We considered that the system model is a realization drawn from an underlying probability space and the dynamic system behavior is modeled with the combination of a nominal model and a Gaussian mixture distributed error-model. In this framework, the error-model parameters were considered as *hidden* variables in order to formulate an ML estimation algorithm to obtain both the nominal model parameters and the error-model Gaussian mixture distribution parameters. We jointly utilized the measurements of a finite number of experiments. The likelihood function was obtained as an infinite mixture distribution and marginalizing with respect to the *hidden* variable. We showed that the likelihood function presents several local maxima and that the estimation can be improved when the number of experiments is increased. To overcome this difficulty, an EM-based algorithm with GMMs was developed to solve the associated ML estimation problem. We obtained closed form expressions for the estimators of the GMM parameters and the noise variance. The simulation results showed accurate estimations of the nominal model including the case where a system model with non-Gaussian-sum uncertainties in the parameters does not correspond to the SE system model assumption.

In addition, we extended the SE with GMMs approach for uncertainty modeling in CT linear dynamic systems utilizing sampled data. In particular, a linear combination of CT orthonormal BFs were used for modeling both the nominal model and the error-model. This approach yield accurate models with a limited number of parameters. An ML estimation algorithm was formulated from the exact DT system model of the CT dynamic system. To solve the associated ML estimation problem an EM-based algorithm with GMMs was developed. The proposed algorithm resulted in closed form expressions to estimate the parameters of the nominal model, noise variance, and the error-model distribution as a GMM. Based on the simulation results, the proposed algorithm for modeling the error-model involved a large computational load with respect to others uncertainty modeling approaches since high number of experiments is needed to obtain accurate estimations of the system model. However, the proposed estimation algorithm exhibited accurate estimations when the number of experiments is high.

5.2 Future work

In this thesis, we showed a paradigm for modeling linear dynamic systems with a nominal model and an error-model using *Stochastic Embedding* approach, which results in an inherent trade-off between the complexity of the nominal model and the complexity of the error model to obtain suitable system models. When the nominal model is less complex than the error-model, then systematic errors arise since the former cannot reproduce properly the dynamics of the *true* system. On the other hand, choosing a more complex nominal model can lead to large variance estimation errors. We could use regularization techniques to deal with the complexity issue by adding a penalty term to the cost function in order to choose the system model than can describe the *true* system with a limited number of parameters for the error-model term. This framework may be addressed utilizing rank-constrained optimization.

In addition, we also showed a systematic procedure for constructing a surrogate function to develop an EM-based algorithm with GMMs when the cost function corresponds to an integral equation. This scenario can be used to address the channel estimation problem in wireless communication. It is well known in the Communications community that the wireless channel corresponds to the superposition of different copies of the transmitted signal that have been reflected, refracted, and scattered. It has been shown empirically that a good model for the multi-path channel typically corresponds to either Rayleigh or Rice distributions. However, there are scenarios in which those distributions do not provide a good model for the channel, for example, the so called *urban scenarios* where the channel exhibits different behaviors in different places. For this case, the wireless channel can be viewed as a problem of uncertainty modeling where there is a nominal model with an error-model that can be estimated with a SE approach. This can be done using a mixture of the most common channel distributions such as Rayleigh, Rice, log-normal, and Nakagami.

On the other hand, there are several control strategies that have been developed from systems in state-space representation. In particular, there are robust control strategies that consider different uncertainty modeling frameworks for developing control techniques with their corresponding robust stability tests. This problem can be addressed with linear fractional transformation for state-space models. The uncertainty modeling framework proposed in this thesis can be extended for linear dynamic systems in state-space representation. Then, this approach can be used for robust control design and stability analysis. Similarly, stochastic model predictive control (SMPC) has been developed for state-space system models when the disturbances and measurement noise are described as random variables. Typically, they are assumed unknown at the current and future instants but have known probability distributions. This control strategy has been tailored for complex dynamics systems with input and state/output constraints. In particular, SMPC exploits the probabilistic uncertainty descriptions to define *chance constraints*, which require the state/output constraints to be satisfied with at least a *priori* specified probability level. These constraints can also be expressed assuming multiplicative uncertainty models, i.e., state-space system models with time-varying uncertainty elements with known PDFs. However, when the initial assumption for the disturbances and measurement noise PDF does not match the actual PDF, the control performance can be degraded. An extension of the uncertainty modeling with GMMs for state-space system models could be used to develop SMPC techniques in a flexible scenario in which unknown PDFs can be modeled as GMMs, incorporating stability, computational load, and feasibility analysis.

Finally, in the Astronomy and Astrophysics field, we can extend the methodology to deconvolve stellar rotational velocities for the problem of mass ratio distribution estimation of binary stars. The binary mass-ratio distribution is important to better understand the evolution of stars in binary systems. However, in most cases, the mass ratio cannot be measured directly, but can only be derived as the convolution of a function that depends on the mass ratio and on the unknown inclination angle of the orbit on the plane of the sky. In this sense, this problem can also be described as an integral equation (similar to the stellar rotational estimation problem addressed in this thesis). We can consider the *true* mass ratio as a *hidden* variable, from which we can formulate an ML estimation algorithm to estimate the mass ratio distribution as a finite mixture of known distributions.

Bibliography

[1] T. Söderström and P. Stoica, *System Identification*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

[2] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[3] G. C. Goodwin and R. Payne, *Dynamic system identification : experiment design and data analysis*. Academic Press New York, 1977.

[4] J. Guo and Y. Zhao, "Recursive projection algorithm on FIR system identification with binary-valued observations," *Automatica*, vol. 49, no. 11, pp. 3396–3401, 2013.

[5] A. Moschitta, J. Schoukens, and P. Carbone, "Parametric System Identification Using Quantized Data," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2312–2322, 2015.

[6] A. 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, 21th IFAC World Congress.

[7] 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.

[8] J. C. Agüero, G. C. Goodwin, and J. Yuz, "System identification using quantized data," in *46th IEEE Conference on Decision and Control*, 2007, pp. 4263–4268.

[9] B. Godoy, G. C. Goodwin, J. C. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no. 9, pp. 1905–1915, 2011.

[10] B. Godoy, J. C. Agüero, R. Carvajal, G. C. Goodwin, and J. Yuz, "Identification of sparse FIR systems using a general quantisation scheme," *Int. J. Control*, vol. 87, no. 4, pp. 874–886, 2014.

[11] J. C. Agüero, K. González, and R. Carvajal, "EM-based identification of ARX systems having quantized output data," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8367 – 8372, 2017, 20th IFAC World Congress.

[12] R. Carvajal, B. Godoy, J. C. Agüero, and G. C. Goodwin, "EM-based sparse channel estimation in OFDM systems," in *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications*, 2012, pp. 530–534.

[13] R. Carvajal, J. C. Agüero, B. Godoy, and G. C. Goodwin, "EM-Based Maximum-Likelihood Channel Estimation in Multicarrier Systems With Phase Distortion," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 152–160, 2013.

[14] G. C. Goodwin and M. Salgado, "A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models," *International Journal of Adaptive Control and Signal Processing*, vol. 3, no. 4, pp. 333–356, 1989.

[15] P. Van den Hof, B. Wahlberg, P. Heuberger, B. Ninness, J. Bokor, and T. Oliveira e Silva, "Modelling and Identification with Rational Orthogonal Basis Functions," *IFAC Proceedings Volumes*, vol. 33, no. 15, pp. 445–455, 2000, iFAC Symposium on System Identification.

[16] P. Peralta, R. Ruiz, and A. Taflanidis, "Bayesian identification of electromechanical properties in piezoelectric energy harvesters," *Mechanical Systems and Signal Processing*, vol. 141, p. 106506, 2020.

[17] G. C. Goodwin, S. Graebe, and M. Salgado, *Control Systems Design*. Upper Saddle River, New Jersey 07458: Prentice Hall, 2001.

[18] S. Douma and P. Van den Hof, "Relations between uncertainty structures in identification for robust control," *Automatica*, vol. 41, no. 3, pp. 439–457, 2005.

[19] G. Calafiore and M. Campi, "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, 2006.

[20] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica*, vol. 27, no. 6, pp. 997–1009, 1991.

[21] L. Ljung, "Model Error Modeling and Control Design," *IFAC Proceedings Volumes*, vol. 33, no. 15, pp. 31–36, 2000, 12th IFAC Symposium on System Identification.

[22] G. C. Goodwin, M. Gevers, and B. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection," *IEEE Transactions on Automatic Control*, vol. 37, no. 7, pp. 913–928, 1992.

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

[24] L. Ljung, G. C. Goodwin, and J. C. Agüero, "Stochastic Embedding revisited: A modern interpretation," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 3340–3345.

[25] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[26] L. Ljung, G. C. Goodwin, J. C. Agüero, and T. Chen, "Model Error Modeling and Stochastic Embedding," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 75 – 79, 2015, 17th IFAC Symposium on System Identification SYSID.

[27] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81–93, 2010.

[28] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *International Journal of Control*, vol. 93, no. 2, pp. 173–180, 2020.

[29] S. Bittanti, M. Lovera, and L. Moiraghi, "Application of non-normal process capability indices to semiconductor quality control," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 2, pp. 296–303, 1998.

[30] M. Campi and E. Weyer, "Non-Asymptotic Confidence Sets for the Parameters of Linear Transfer Functions," *IEEE Transactions on Automatic Control*, vol. 55, no. 12, pp. 2708–2720, 2010.

[31] A. Carè, B. Csáji, M. Campi, and E. Weyer, "Finite-Sample System Identification: An Overview and a New Correlation Method," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 61–66, 2018.

[32] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a non-minimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385–399, 1980.

[33] M. Zidane, S. Said, M. Sabri, and A. Boumezzough, "Impulse Response Identification of Minimum and Non-Minimum Phase Channels," *International Journal of Advanced Science and Technology*, vol. 64, pp. 59–72, 2014.

[34] B. Chen and A. Petropulu, "Frequency domain blind MIMO system identification based on second- and higher order statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1677–1688, 2001.

[35] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 360–377, 1989.

[36] K. Lii and M. Rosenblatt, "An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes," *Journal of Multivariate Analysis*, vol. 43, no. 2, pp. 272 – 299, 1992.

[37] K. Rasmussen, "Maximum likelihood estimation of the parameters of non-minimum phase and non-causal ARMA models," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 209–211, 1994.

[38] O. Shalvi and E. Weinstein, "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 328–339, 1994.

[39] K. Lii and M. Rosenblatt, "Maximum Likelihood Estimation for non-Gaussian Non-Minimum Phase ARMA Sequences," *Statistica Sinica*, vol. 6, no. 1, pp. 1–22, 1996.

[40] M. Rosenblatt, "Non-Gaussian Time Series Models," in *Time Series Analysis and Applications to Geophysical Systems*. Springer New York, 2004, pp. 227–237.

[41] N. Hsu and F. Breidt, "Exact Maximum Likelihood estimation for non-Gaussian moving averages," *Statistica Sinica*, vol. 19, no. 2, pp. 545–560, 2009.

[42] H. Chen, A. Lee, and C. Lee, "Alternative errors-in-variables models and their applications in finance research," *The Quarterly Review of Economics and Finance*, vol. 58, pp. 213 – 227, 2015.

[43] M. Khorasani and E. Weyer, "Non-Asymptotic Confidence Regions for Errors-In-Variables Systems," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 1020 – 1025, 2018, 18th IFAC Symposium on System Identification SYSID 2018.

[44] J. Tugnait, "Identification of non-minimum phase linear stochastic systems," *Automatica*, vol. 22, no. 4, pp. 457 – 464, 1986.

[45] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *International Journal of Adaptive Control and Signal Processing*, vol. 3, pp. 191–229, 1989.

[46] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high-order moments," *IEEE Transactions on Automatic Control*, vol. 35, no. 1, pp. 27–35, 1990.

[47] A. Pandav, D. Mallick, and B. Mohanty, "Effect of limited statistics on higher order cumulants measurement in heavy-ion collision experiments," *Nuclear Physics A*, vol. 991, p. 121608, 2019.

[48] I. Arasaratnam, S. Haykin, and R. Elliott, "Discrete-Time Nonlinear Filtering Algorithms Using Gauss-Hermite Quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[49] M. Balenzuela, J. Dahlin, N. Bartlett, A. Wills, C. Renton, and B. Ninness, "Accurate Gaussian mixture model smoothing using a two-filter approach," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 694–699.

[50] I. Bilik and J. Tabrikian, "Target tracking in glint noise environment using nonlinear non-Gaussian Kalman filter," in *2006 IEEE Conference on Radar*, 2006, p. 6 pp.

[51] W. Li, Y. Jia, J. Du, and J. Zhang, "PHD filter for multi-target tracking with glint noise," *Signal Processing*, vol. 94, pp. 48 – 56, 2014.

[52] S. Bernstein, M. Burrows, J. Evans, A. Griffiths, D. McNeill, C. Niessen, I. Richer, D. White, and D. Willim, "Long-range communications at extremely low frequencies," *Proceedings of the IEEE*, vol. 62, no. 3, pp. 292–312, 1974.

[53] S. Haykin, K. Huber, and Z. Chen, "Bayesian sequential state estimation for MIMO wireless communications," *Proceedings of the IEEE*, vol. 92, pp. 439–454, 2004.

[54] H. Sorenson and D. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.

[55] J. Dahlin, A. Wills, and B. Ninness, "Sparse Bayesian ARX models with flexible noise distributions," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 25 – 30, 2018, 18th IFAC Symposium on System Identification.

[56] G. Bittner, "Identificación de sistemas dinámicos lineales mediante Máxima Verosimilitud con mezcla finita de distribuciones normales," Master's thesis, Universidad Técnica Federico Santa Maria, Valparaíso, Chile, 2020, Magister en Ciencias de la Ingeniería Electrónica.

[57] S. Chandrasekhar and G. Münch, "On the integral equation governing the distribution of the true and the apparent rotational velolicities of stars," *Apj*, vol. 111, p. 142, 1950.

[58] J. Lo, "Finite-dimensional sensor orbits and optimal nonlinear filtering," *IEEE Transactions on Information Theory*, vol. 18, no. 5, pp. 583–588, 1972.

[59] M. Tanner, *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*. Springer, 1991.

[60] B. Carlin and T. Louis, *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*. Taylor & Francis, 2010.

[61] R. Risuleo, G. Bottegal, and H. Hjalmarsson, "A nonparametric kernel-based approach to Hammerstein system identification," *Automatica*, vol. 85, pp. 234 – 247, 2017.

[62] G. Bottegal, H. Hjalmarsson, and G. Pillonetto, "A new kernel-based approach to system identification with quantized output data," *Automatica*, vol. 85, pp. 145 – 152, 2017.

[63] M. Curé, D. Rial, A. Christen, and J. Cassetti, "A method to deconvolve stellar rotational velocities," *Astronomy & Astrophysics*, vol. 85, pp. 1–13, 2014.

[64] A. Christen, P. Escárate, M. Curé, D. Rial, and J. Cassetti, "A method to deconvolve stellar rotational velocities II–The PDF via Tikhonov regularization," *Astronomy & Astrophysics*, vol. 595, no. A50, pp. 1–8, 2016.

[65] H. Suzuki, "A statistical model for urban radio propogation," *IEEE Transactions on Communications*, vol. 25, no. 7, pp. 673–680, 1977.

[66] C. Kraus, B. Bornschein, L. Bornschein, J. Bonn, B. Flatt, A. Kovalik, B. Ostrick, E. Otten, J. Schall, T. Thümmler, and C. Weinheimer, "Final results from phase II of the Mainz neutrino mass searchin tritium β decay," *The European Physical Journal C - Particles and Fields*, vol. 40, no. 4, pp. 447–468, 2005.

[67] G. McLachlan and D. Peel, *Finite Mixture Models*. NJ, USA: Wiley, 2000.

[68] S. Frühwirth-Schnatter, G. Celeux, and C. Robert, *Handbook of Mixture Analysis*. Chapman and Hall/CRC, 2018.

[69] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh-Rice Mixture Parameter Estimation via EM Algorithm for Change Detection in Multispectral Images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5004–5016, 2015.

[70] K. Mengersen, C. Robert, and M. Titterington, *Mixtures: Estimation and Applications*. Wiley, 2011.

[71] D. Alspach, "A Gaussian sum approach to the multi-target identification-tracking problem," *Automatica*, vol. 11, no. 3, pp. 285 – 296, 1975.

[72] Y. Huang, K. Englehart, B. Hudgins, and A. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1801–1811, 2005.

[73] J. Janouek, P. Gajdo, M. Radecký, and V. Snáel, "Gaussian Mixture Model Cluster Forest," in *IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, pp. 1019–1023.

[74] N. Achieser, *Theory of Approximation*. Dover Publications, 1992.

[75] N. Wiener, "Tauberian theorems," *Annals of Mathematics*, vol. 33, no. 1, pp. 1–100, 1932.

[76] H. Cramér, *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press, 1999.

[77] C. Gourieroux and A. Monfort, *Statistics and Econometric Models*, ser. Themes in Modern Econometrics. Cambridge University Press, 1995.

[78] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[79] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, 2nd ed. Hoboken, NJ: Wiley, 2008.

[80] C. Jin, Y. Zhang, S. Balakrishnan, M. Wainwright, and M. Jordan, "Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences." Curran Associates Inc., 2016, pp. 4123–4131.

[81] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667 – 1682, 2005.

[82] R. 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.

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

[84] R. Durrett, *Probability: Theory and Examples*, 4th ed. USA: Cambridge University Press, 2010.

[85] 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, pp. 1–24, 2018.

[86] L. Shampine, "Vectorized adaptive quadrature in MATLAB," *Journal of Computational and Applied Mathematics*, vol. 211, no. 2, pp. 131 – 140, 2008.

[87] M. Jordan and R. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm," *Neural Comput.*, vol. 6, no. 2, pp. 181–214, 1994.

[88] G. Celeux and G. Soromenho, "An entropy criterion for assessing the number of clusters in a mixture model," *Journal of Classification*, vol. 13, no. 2, pp. 195–212, 1996.

[89] R. Neal, "Slice sampling," *Ann. Statist.*, vol. 31, no. 3, pp. 705–767, 2003.

[90] L. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, p. 745, 1974.

[91] P. Hansen, *Discrete Inverse Problems: Insight and Algorithms*. USA: Society for Industrial and Applied Mathematics, 2010.

[92] A. Deutsch, "Maxwellian Distributions for Stellar Rotations," in *Stellar Rotation, Proceedings of the IAU colloquium*, 1970, pp. 207–218.

[93] Y. Gaigé, "Stellar rotational velocities from the V sin i observations - Inversion procedures and applications to open clusters," *Astronomy & Astrophysics*, vol. 269, no. 1-2, pp. 267–281, 1993.

[94] J. Carvalho, J. do Nascimento, R. Silva, and J. De Medeiros, "Non-Gaussian Statistics and Stellar Rotational Velocities of Main-Sequence Field Stars," *The Astrophysical Journal*, vol. 696, no. 1, pp. L48–L51, 2009.

[95] J. Mermilliod, M. Mayor, and S. Udry, "Catalogues of radial and rotational velocities of 1253 F-K dwarfs in 13 nearby open clusters," *Astronomy & Astrophysics*, vol. 498, pp. 949–960, 2009.

[96] O. Ramírez-Agudelo, S. Simón-Díaz, H. Sana, A. de Koter, C. Sabín-Sanjulían, S. de Mink, P. Dufton, G. Gräfener, C. Evans, A. Herrero, N. Langer, D. Lennon, J. Maíz Apellániz, N. Markova, F. Najarro, J. Puls, W. Taylor, and J. Vink, "The VLT-FLAMES Tarantula Survey - XII. Rotational velocities of the single O-type stars," *Astronomy & Astrophysics*, vol. 560, p. A29, 2013.

[97] B. Nordström, M. Mayor, J. Andersen, J. Holmberg, F. Pont, B. Jorgensen, E. Olsen, S. Udry, and N. Mowlavi, "The Geneva-Copenhagen survey of the Solar neighbourhood - Ages, metallicities, and kinematic properties of ˜14000 F and G dwarfs," *Astronomy & Astrophysics*, vol. 418, no. 3, pp. 989–1019, 2004.

[98] J. Holmberg, B. Nordström, and J. Andersen, "The Geneva-Copenhagen survey of the Solar neighbourhood II- New uvby calibrations and rediscussion of stellar ages, the G dwarf problem, age-metallicity diagram, and heating mechanisms of the disk," *Astronomy & Astrophysics*, vol. 475, no. 2, pp. 519–537, 2007.

[99] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986.

[100] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, "Numerical Recipes, The Art of Scientific Computing ," *The American Mathematical Monthly*, vol. 94, no. 9, pp. 889–893, 1987.

[101] A. Tikhonov and V. Arsenin, "Solutions of Ill-Posed Problems," *SIAM Review*, vol. 21, no. 2, pp. 266–267, 1979.

[102] V. Ivanov, V. Vasin, and V. Tanana, *Theory of Linear Ill-Posed Problems and its Applications*. De Gruyter, 2013.

[103] C. Yu and M. Verhaegen, "Data-driven fault estimation of non-minimum phase LTI systems," *Automatica*, vol. 92, pp. 181–187, 2018.

[104] M. 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.

[105] R. A. Fisher, "Moments and product moments of sampling distributions," *Proceedings of the London Mathematical Society*, vol. s2-30, no. 1, pp. 199–238, 1930.

[106] D. Brillinger, "An Introduction to Polyspectra," *The Annals of Mathematical Statistics*, vol. 36, no. 5, pp. 1351–1374, 1965.

[107] D. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of kth-order spectra," *Proceedings of the National Academy of Sciences*, vol. 57, no. 2, pp. 206–210, 1967.

[108] A. Oppenheim and R. Schafer, *Discrete-time signal processing*. Prentice Hall, 1989.

[109] S. Wright, "Coordinate descent algorithms," *Mathematical Programming*, vol. 151, no. 1, pp. 3–34, 2015.

[110] 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.

[111] J. Ma, L. Xu, and M. Jordan, "Asymptotic Convergence Rate of the EM Algorithm for Gaussian Mixtures," *Neural Computation*, vol. 12, no. 12, pp. 2881–2907, 2000.

[112] I. Naim and D. Gildea, "Convergence of the EM Algorithm for Gaussian Mixtures with Unbalanced Mixing Coefficients," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, 2012, pp. 1427–1431.

[113] 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.

[114] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[115] K. Burnham and D. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer New York, 2002.

[116] D. Spiegelhalter, N. Best, B. Carlin, and A. van der Linde, "The deviance information criterion: 12 years on," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 3, pp. 485–493, 2014.

[117] C. Jeff Wu, "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95 – 103, 1983.

[118] R. Jategaonkar, *Flight Vehicle System Identification: A Time-Domain Methodology*. Progress in Astronautics and Aeronautics, 2nd ed., 2015.

[119] P. Hamel and R. Jategaonkar, "Evolution of flight vehicle system identification," *Journal of Aircraft*, vol. 33, no. 1, pp. 9–28, 1996.

[120] R. Jategaonkar and F. Thielecke, "Aircraft parameter estimation–A tool for development of aerodynamic databases," *Sadhana*, vol. 25, no. 2, pp. 119–135, 2006.

[121] B. Rosi and J. Diekmann, "Methods for the Uncertainty Quantification of Aircraft Simulation Models," *Journal of Aircraft*, vol. 52, no. 4, pp. 1247–1255, 2015.

[122] R. Patton, P. Frank, and R. Clark, *Issues of Fault Diagnosis for Dynamic Systems*, 1st ed. Springer Publishing Company, Inc., 2010.

[123] P. Heuberger, P. Van den Hof, and O. Bosgra, "A Generalized Orthonormal Basis for Linear Dynamical Systems," *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 451–465, 1995.

[124] E. Walter, J. Norton, H. Piet-Lahanier, and M. Milanese, *Bounding Approaches to System Identification*. Perseus Publishing, 1996.

[125] A. Garulli, A. Vicino, and G. Zappa, "Conditional central algorithms for worst case set-membership identification and filtering," *IEEE Transactions on Automatic Control*, vol. 45, no. 1, pp. 14–23, 2000.

[126] W. Reinelt, A. Garulli, and L. Ljung, "Comparing different approaches to model error modeling in robust identification," *Automatica*, vol. 38, no. 5, pp. 787–803, 2002.

[127] E. Fogel and Y. Huang, "On the value of information in system identification Bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, 1982.

[128] M. Milanese and M. Taragna, "Suboptimality evaluation of approximated models in H_∞ identification," in *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 2, 1999, pp. 1494–1499.

[129] A. Spanos, *Probability Theory and Statistical Inference: Econometric modeling with observational data*. London: Cambridge University Press,, 1999.

[130] M. Volkhardt, S. Kalesse, S. Müller, and H. Gross, "Maximum a Posteriori Estimation of Dynamically Changing Distributions." Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 484–491.

[131] P. Heuberger, P. Van den Hof, and B. Wahlberg, *Modelling and Identification With Rational Orthogonal Basis Functions*. Springer, 2005.

[132] G. C. Goodwin, J. C. Agüero, M. 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.

[133] B. Wahlberg, "Laguerre and Kautz Models," *IFAC Proceedings Volumes*, vol. 27, no. 8, pp. 965–976, 1994, IFAC Symposium on System Identification.

[134] B. Wahlberg and P. Mäkilä, "On Approximation of Stable Linear Dynamical Systems using Laguerre and Kautz Functions," *Automatica*, vol. 32, no. 5, pp. 693–708, 1996.

[135] L. Wang, *Model Predictive Control System Design and Implementation Using MATLAB*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[136] P. Davis, "Leonhard Euler's Integral: A Historical Profile of the Gamma Function," *The American Mathematical Monthly*, vol. 66, no. 10, pp. 849–869, 1959.