

Article

# Robust Parametric Identification for ARMAX Models with Non-Gaussian and Coloured Noise: A Survey

Jesica Escobar <sup>1,\*</sup>  and Alexander Poznyak <sup>2,†</sup>

<sup>1</sup> Instituto Politecnico Nacional ESIME Zacatenco, Unidad Profesional Adolfo Lopez Mateos, Av. IPN S/N, Mexico City 07738, Mexico

<sup>2</sup> Department of Automatic Control, CINVESTAV-IPN A.P. 14-740, Mexico City 07000, Mexico; apoznyak@ctrl.cinvestav.mx

\* Correspondence: jmescoar@esimez.mx

† These authors contributed equally to this work.

**Abstract:** In this paper the Cramer-Rao information bound for ARMAX (Auto-Regression-Moving-Average-Models-with-Exogenous-inputs) under non-Gaussian noise is derived. It is shown that the direct application of the Least Squares Method (LSM) leads to incorrect (shifted) parameter estimates. This inconsistency can be corrected by the implementation of the parallel usage of the MLMW (Maximum Likelihood Method with Whitening) procedure, applied to all measurable variables of the model, and a nonlinear residual transformation using the information on the distribution density of a non-Gaussian noise, participating in Moving Average structure. The design of the corresponding parameter-estimator, realizing the suggested MLMW-procedure is discussed in details. It is shown that this method is asymptotically optimal, that is, reaches this information bound. If the noise distribution belongs to some given class, then the Huber approach (min-max version of MLM) may be effectively applied. A numerical example illustrates the suggested approach.



**Citation:** Escobar, J.; Poznyak, A. Robust Parametric Identification for ARMAX Models with Non-Gaussian and Coloured Noise: A Survey. *Mathematics* **2022**, *10*, 1291. <https://doi.org/10.3390/math10081291>

Academic Editors: Alfonso Baños, Francisco German Badía and María D. Berrade

Received: 7 February 2022

Accepted: 29 March 2022

Published: 13 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** parameter estimation; least squares method; whitening filter; Fisher information; maximum likelihood method; nonlinear residual transformation

**MSC:** 93E03; 93E10; 93E11; 93E24

## 1. Introduction

### 1.1. Road Map of This Survey

The topic of parameter identification in a large class of linear models with external noise acting on-line and perturbing the dynamics of an investigated system is addressed in this overview. The considered models are classified as ARMAX (auto-regression-moving average with exogenous inputs) and are commonly expressed in discrete-time format by recurrent linear stochastic difference equations. The class of distribution functions is supposed to be known a priori but not its exact analytical expression: such models contain “uncertainties” in their descriptions, which are associated with unknown parameters and probabilistic characteristics of external noise (perturbations): only a class of distribution functions is supposed to be known a priori but not its exact analytical expression. As a result, any identification technique that may be used in such a circumstance should be robust (resilient) with respect to existing uncertainty. The focus of this work is on a critical examination of robust parametric identification techniques, highlighting a gap in the current literature in which the vast majority of publications adopt the traditional assumption that external stochastic perturbations are independent and Gaussian (have a “normal distribution”). Only a few papers deal with a different non-standard assumption about the available stochastic characteristics of external noisy disturbances. Here, we expose readers to the underlying difficulties (mathematical and computational) and discuss a few distinct ways that have shown to be successful in the absence of available probabilistic data.

The structure of the paper is as follows:

- Review of publications:
  - It contains the descriptions of the important survey published in the 1970's–1990's (Åström, Becky, Ljung and Gunnarson, Billings among others).
  - Nongaussian noises have been studied by Huber, Tsytkin and Polyak.
- Problem formulation and model description:
  - The ARMAX model with correlated non-Gaussian noise, generated by a stable and non-minimal phase filter, is introduced.
- Some classes  $\mathcal{P}_{3be}$  of noise p.d.f.:
  - In a rigorous mathematical manner several classes of random stationary sequences with different p.d.f. as an input of a forming filter are considered (all symmetric distributions non-singular in origin, all symmetric distributions with a bounded variance, all symmetric “approximately normal” distributions and “approximately uniform” distributions).
- Main assumptions:
  - These concern the martingale difference property with conditional bounded second moment for stochastic sequences in the input of the forming filter, stability and minimal-phase property for this filter, independent of this sequence with other measurable inputs).
- Regression representation format:
  - The extended regression form of the considered model is introduced.
- Main contribution of the paper:
  - The exact presentation of the main contribution of the paper.
- Why LSM does not work for the identification of ARMAX models with correlated noise:
  - A simple example exhibiting the lack of workability of this technique in the case of dynamic (autoregression) models is described in detail for a reader who is not actively involved in the least-squares method.
- Some other identification techniques:
  - Identification of non-stationary parameters and the Bayesian method, matrix forgetting factor and its adaptive version are reviewed.
- Regular observations and information inequality for observations with coloured noise:
  - the Cramér–Rao bound (CRB) and the Fisher information, characterising the maximal possible rate of estimation under the given information resource, are presented.
- Robust version of the maximum likelihood method with whitening (MLMW procedure):
  - This approach is demonstrated to reach the CRB bound, indicating that it is asymptotically the best among all identification procedures.
- Recurrent identification procedures with nonlinear residual transformations: static (regression) and dynamic (autoregression) models:
  - Within a specified noise p.d.f. class, it is proven that such a strategy with particular selection of nonlinear residual transformation is resilient (robust) optimum in achieving min–max error variance in CRB inequality.
- Instrumental variables method (IVM):
  - IV or total least-squares estimators is the method which also recommends to estimate parameters in the presence of coloured noises with a finite correlation.
- Joint parametric identification of ARMAX model and the forming filter:
  - The “generalised residual sequence” is introduced, which is shown to be asymptotically closed to the independent sequence acting in the input of the forming filter, which helps to resolve the identification problem in an extended parametric space.

- Numerical example.
- Discussion and conclusions.
- Appendix A and abbreviations:
  - This part offers proofs of some of the article’s claims that appear to be significant from the authors’ perspective, as well as a list of acronyms used throughout the work.

### 1.2. Review of the System Identification Literature

A mathematical model is a simplified mathematical structure connected to a component of reality and produced for a specific purpose of system analysis [1]. Differential equations, state space models and transfer functions are all examples of mathematical models of dynamic systems that are useful in a variety of fields [2]. System identification, which can be applied to nearly any system and give models that explain the system behaviour, is an alternative to modelling.

### 1.3. Classical Surveys on Identification

The least-squares method (LSM), as well as some of its variants, have been extensively researched in the past, according to the survey given by Åström [3]. The least squares, maximum likelihood, instrumental variables and tally principle are examples of these variety. Several approaches for the identification of dynamic systems using computer techniques, such as spectral analysis, certain gradient methods, quasi-linearization and stochastic approximations, were provided by Becky [4]. In the case of a time-varying situation, Ljung and Gunnarson investigated several methods for developing identification algorithms that could take into account the time-varying dynamics of systems as well as signals [5]. Some mean square expressions were examined in this study. In [6] by Billings, several methods for the nonlinear case were described; these algorithms were based on the functional expansion of Wiener and Volterra, block-oriented and bi-linear systems, structure detection and some catastrophe theory. System identification is a vast field of study, with a variety of methods based on the models to be estimated: linear, nonlinear, continuous, discrete, time-varying and so on. Ljung’s survey [7] demonstrates that, despite the wide range of techniques, the field can be defined by a few key principles, such as data information, validation and model complexity and offers some basic principles and results, as well as a method for solving real-world problems. As it is mentioned by Ljung in [8], system identification is a well established research area, whose paradigms are most of the time based on classical statistical methods. Some recent techniques are based on kernel regularisation methods. The paper presented by Ljung presents some of the main ideas and results of kernel-based regularisation methods for system identification.

### 1.4. Identification under Correlated Noise Perturbations

In the measurements and modelling of dynamic systems [9], the unpredictability inherent in physical processes always creates inaccuracy. Stochastic processes make it possible to model these random events and create more realistic models [10]. It is commonly assumed that only white noise is presented in stochastic systems, however, there are also cases where the noise is correlated or “coloured”. Coloured noise is prevalent in linear and pseudo-linear regression identification models, where one of the challenges is the presence of unknown inner variables and immeasurable noise components [11]. The stochastic gradient algorithm proved to be a useful technique for those cases. In situations where there are a lot of noisy sources, noise suppression is crucial. In many practical circumstances, coloured noise may be converted to white noise [12] by passing it through an invertible time-invariant linear (“whitening”) filter. The existence of coloured noise typically leads to robust identification theory, which was first proposed by P. Huber [13] and Ya. Z. Tsympkin-B. Polyak [14] over fifty years ago. In [15], the basic principles of robust identification were presented, as well as the identification methods for auto regression with exogenous input (ARX) models. The suggested approach used a whitening procedure and a variant of the maximum likelihood method in parallel to conduct asymptotically the estimation of

unknown parameters. The need of system identification has grown in some areas, such as robotics, due to the increasing interest in showing movement accuracy in the industry. The application of system identification in DC servomotors is widely used in robotics; in [16], a study on identifying two model structures, ARX and ARMAX, of the system to test and compare their performance on validation criterion is presented.

### 1.5. Identification of ARMAX and NARMAX Models

For many years, the parameter estimation for autoregressive moving average exogenous input (ARMAX) models has been investigated [17,18]. The importance of parameter-bounding methods in the identification process is highlighted in [19] and stands for such models. These methods offer a radical alternative to compute parameter point estimates and covariances. They require constraints in an effectively deterministic model formulation instead of p.d.f. or mean and covariance for the noise and previous parameter estimations.

In the context of parameter-bounding identification, this study offers a preliminary assessment of certain typical tasks, such as experiment design, testing for outliers, tolerance prediction and worst-case control design. In [20], the parametrisation of ARMAX models was also discussed. The results reported in this work were aimed at developing a technique for modelling and fitting multivariable time-series data based on spatial approach and parametrisation, with tolerance for missing or incomplete data.

ARMAX models are widely used in industrial modelling nowadays [21–23]. For example, functional time series are the realisation of a stochastic process where each observation is a continuous function defined on a finite interval. These processes are commonly used in electricity markets and are gaining more importance as more market data become available and markets head toward continuous time-marginal pricing approaches. In [24], the authors propose a new functional forecasting method that attempts to generalise the standard seasonal ARMAX time-series model to the  $L_2$  Hilbert space; the proposed approach is tested by forecasting the daily price profile of the Spanish and German electricity markets, and it is compared with other functional reference models. A physic-based ARMAX model of room temperature in office buildings was presented in [25], where thermodynamic equations are used to determine the structure and order of the model. In this study, extensive measurements over 109 days are used to develop and validate the model. This model can be used to predict the variations of the room temperature accurately in short-term, and long-term periods and has shown to be suitable for real-time fault detection and control applications.

Traditional stochastic information gradient methods for ARMAX identification have a lower computational cost, but its convergence speed is still low, in [26], a two-step algorithm based on gradient acceleration strategies is proposed to deal with this problem. When in the ARMAX process, the noises presented are additive; it is possible to introduce additional information to the estimation problem using nuisances variables to model the output noises (see [27]). Then, a regularised estimator suppresses the adverse effects of the noises and provides minimum variance estimates.

In [18], a technique for concurrently picking the order and identifying the parameters of an ARMAX model was explored, and it was also assessed by computational experiments. The technique presented in that work was based on reformulating the issue for a standard state space, then implementing a bank of Kalman filters, identifying the true model and utilizing multi-model partitioning theory to solve it. In the study presented by Correa and Poznyak in 2001 (see [28]), the problem of simultaneous robust state and parameter estimation for some class of MIMO non-linear systems under mixed uncertainties (unmodeled dynamics as well as observation noises) is presented. A switching gain robust “observer-identifier” is introduced to obtain the estimation. This is achieved by applying an observer to the so-called nominal extended system, obtained from the original system without any uncertainties and considering the parameters as additional constant states. As it was shown in general the extended systems, these can lose the global observability property, supposed to be valid for the original non-extended system, and a special procedure is needed to

provide a good estimation process in this situation [29]. The suggested adaptive observer has the Luenberger-type observer structure with switching matrix gain that guarantees a good enough upper bound for the identification error performance index [30]. The Van der Monde generalised transformation is introduced to derive this bound which turns out to be “tight” (it is equal to zero in the absence of both noises and unmodeled dynamics). One approach for dealing with coloured noises is to utilise parameter estimate algorithms based on Kalman filters. The Kalman filter is frequently used for control and estimate (see [31]), and this technique may be thought of as Hammerstein–Wiener ARMAX models. An extended Kalman filter, or the unscented Kalman filter, can be implemented to extend this approach to the nonlinear situation. For the nonlinear autoregressive moving average with exogenous inputs (NARMAX) models, Kalman filters are a commonly used identification method. The off-line observer/Kalman filter presented in [32] was implemented as an identification method, since it has shown a good initial guess of the NARMAX model to reduce the on-line system identification process time, this method showed to be effective in the case of system faults and input failures. In the case of Hammerstein nonlinear systems with coloured noises, a maximum likelihood-based stochastic gradient algorithm was implemented in [33], where the unknown noises were replaced in the information vector by their estimates and through these, one can obtain the parameters. For multivariable Hammerstein controlled autoregressive moving average systems, an interactive maximum likelihood estimation method was implemented in [34]. In that paper, the logarithmic likelihood function over multiple parameter vectors is maximised; the proposed method overcomes the limit on an autoregressive model form with one parameter vector.

In this survey, we present a compendium of some of the existing literature regarding identification in ARMAX models and some of the techniques used in these type of models. We also obtain the Cramer–Rao information bound for ARMAX models with non-Gaussian noises and show that the maximum likelihood method with whitening procedure (MLMW) reaches this low bound, or in other words, is asymptotically optimal.

## 2. Problem Formulation

### 2.1. Robust Parametric Identification Model Description

Consider the following ARMAX (autoregression moving average exogenous input) model given by

$$y_n = \sum_{l=1}^L a_l y_{n-l} + \sum_{k=0}^K b_k w_{n-k} + \eta_n, \quad n \geq 0, \tag{1}$$

where

- $\{y_n\} \in R^1$  is scalar sequence of available on-line state variables.
- $\{w_n\}$  is a measurable input sequence (deterministic or, in general, stochastic).
- $\{\eta_n\} \in R^1$  is a noise sequence (not available during the process) generated by the exogenous system

$$\eta_n + \sum_{s=1}^{K_2} d_{2,s} \eta_{n-s} = \zeta_n + \sum_{s=1}^{K_1} d_{1,s} \zeta_{n-s}, \tag{2}$$

which can be symbolically represented as the forming filter

$$\eta_n = H(q^{-1})\zeta_n \tag{3}$$

with the transition function

$$\left. \begin{aligned} H(q^{-1}) &= H_1(q^{-1})/H_2(q^{-1}), \\ H_1(q^{-1}) &= 1 + \sum_{s=1}^{K_1} d_{1,s} q^{-s}, \quad H_2(q^{-1}) = 1 + \sum_{s=1}^{K_2} d_{2,s} q^{-s} \end{aligned} \right\} \tag{4}$$

$q^{-1}$  is the one-step delay operator acting as  $y_{k-1} = q^{-1}y_k$ ,

- $\{\xi_n\}$  as an independent zero mean stationary sequence with the probability density function (p.d.f.)  $p_\xi(x)$  which may be unknown but belonging to some given class  $\mathcal{P}_\xi$  of p.d.f., that is,

$$p_\xi(x) \in \mathcal{P}_\xi.$$

2.2. Some Classes  $\mathcal{P}_\xi$  of p.d.f.

All possible classes  $\mathcal{P}_\xi$  of p.d.f., considered in practical applications, are related with a priori information on stationary generating sequence  $\{\xi_n\}$ . Here we present some of them which look natural from practical point of views.

- Class  $\mathcal{P}_\xi^1$  (of all symmetric distributions non singular in the point  $x = 0$ ):

$$\mathcal{P}_\xi^1 = \left\{ p_\xi : p_\xi(0) \geq \frac{1}{2a} > 0 \right\}. \tag{5}$$

We deal with this class if there is not any a priori information on a noise distribution  $p_\xi$ .

- Class  $\mathcal{P}_\xi^2$  (of all symmetric distributions with a bounded variance):

$$\mathcal{P}_\xi^2 = \left\{ p_\xi : \int_{\mathbb{R}} x^2 p_\xi(x) ds \leq \sigma^2 < \infty \right\}. \tag{6}$$

- Class  $\mathcal{P}_\xi^3$  (of all symmetric “approximately normal” distributions):

$$\mathcal{P}_\xi^3 = \left\{ p_\xi : p_\xi(x) = (1 - \alpha)p_{\mathcal{N}(0,\sigma^2)}(x) + \alpha q(x) \right\}, \tag{7}$$

where  $p_{\mathcal{N}(0,\sigma^2)}(x)$  is the centred Gaussian distribution density with the variance defined by  $\sigma^2$  and  $q(x)$  is another distribution density. The parameter  $\alpha \in [0, 1]$  characterises the level of the effect of a “dirty” distribution  $q(x)$  to the basic Gaussian distribution  $p_{\mathcal{N}(0,\sigma^2)}(x)$ .

- Class  $\mathcal{P}_\xi^4$  (of all symmetric “approximately uniform” distributions):

$$\mathcal{P}_\xi^4 = \left\{ p_\xi : p_\xi(x) = (1 - \alpha)p_{U(0,a)}(x) + \alpha q(x) \right\} \tag{8}$$

where

$$p_{U(0,a)}(x) := \frac{1}{2a} \chi(|x| \leq a)$$

$$\chi(|x| \leq a) = \begin{cases} 1 & \text{if } |x| \leq a \\ 0 & \text{if } |x| > a \end{cases}$$

is the centred uniform distribution and  $q(x)$  is one process with a different distribution density. The parameter  $\alpha \in [0, 1]$  characterises the level of the effect of a “dirty” distribution  $q(x)$  to the basic one  $p_{U(0,a)}(x)$ .

2.3. Main Assumptions

1. All random variables  $\{w_n, \xi_n\}$  are defined on the probability space  $(\Omega, \mathcal{F}, P)$  with the  $\sigma$ -algebras flow  $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$

$$\mathcal{F}_{n-1} = \sigma(y_{-l}, \dots, y_{-1}, \dots, y_{n-1}; w_0, \dots, w_n; \eta_{-K_2}, \dots, \eta_{n-1}; \xi_{-K_1}, \dots, \xi_{n-1}). \tag{9}$$

2. For all  $n$

$$E\{\xi_n | \mathcal{F}_{n-1}\} \stackrel{a.s.}{=} 0, \quad E\{\xi_n^2 | \mathcal{F}_{n-1}\} \stackrel{a.s.}{=} \sigma_\xi^2 < \infty. \tag{10}$$

3. The measurable input sequence  $\{w_n\}_{n \geq 0}$  is of bounded power:

$$E\{w_n^2 | \mathcal{F}_{n-1}\} \stackrel{a.s.}{=} \sigma_{w,n}^2 < \infty, \tag{11}$$

and is independent of  $\{\xi_n\}$ , i.e.,

$$E\{w_n \xi_n \mid \mathcal{F}_{n-1}\} \stackrel{a.s.}{=} w_n E\{\xi_n \mid \mathcal{F}_{n-1}\} \stackrel{a.s.}{=} 0. \tag{12}$$

4. It is assumed that the forming filter is stable and “minimal-phase”, that is, both polynomials  $H_1(q^{-1})$  and  $H_2(q^{-1})$  are Hurwitz, i.e., have all roots inside of the unite circle in the complex plain.
5. The ARMAX plant (1) is stable: the polynomial

$$A(q^{-1}) = 1 - \sum_{l=1}^L a_l q^{-l} \tag{13}$$

is Hurwitz.

**Remark 1.** *As it follows from the assumptions above, the noise sequence admits to be non-Gaussian and correlated (coloured).*

#### 2.4. Regression Format Representation

The system (1) can be represented in the, regression format as

$$y_n = z_n^T c + \eta_n, \tag{14}$$

where the extended vector

$$c = (a_1, \dots, a_L; b_0, \dots, b_K)^T \in R^{L+K+1}, \tag{15}$$

represents the collection of unknown parameters to be estimated, and the vector

$$z_n = (y_{n-1}, \dots, y_{n-L}; w_n, \dots, w_{n-K})^T \in R^{L+K+1}, \tag{16}$$

is referred to as the generalised regression measurable (available on-line) input.

#### 2.5. Robust Parametric Identification Problem Formulation

**Problem 1.** *We need to estimate the vector  $c$  of unknown parameters based on available data  $\{z_n\}$  and a priory knowledge of the p.d.f. class  $\mathcal{P}_\xi$  of the stationary noise sequence  $\{\xi_n\}$  in the input of the forming filter. Two possible cases may be considered:*

- *the parameters  $(d_{1,s}, d_{2,s})$  of the forming filter  $H(q^{-1})$  are known.*
- *The parameters  $(d_{1,s}, d_{2,s})$  of the forming filter  $H(q^{-1})$  are also unknown.*

#### 2.6. Main Contribution of the Paper

- The Cramer–Rao information bound for ARMAX (autoregression moving average models with exogenous inputs) under non-Gaussian noises is derived.
- It is shown that the direct implementation of the least-squares method (LSM) leads to an incorrect (shifted) parameter estimation.
- This inconsistency can be corrected by the implementation of the parallel use of the MLMW (maximum likelihood method with whitening) procedure, applied to all measurable variables of the model, and a nonlinear residual transformation using the information on the distribution density of a non-Gaussian noise, participating in moving average structure.
- The design of the corresponding parameter estimator, realising the suggested MLMW procedure, containing a parallel on-line “whitening” process as well as a nonlinear residual transformation, is presented in detail.
- It is shown that the MLMW procedure attains the obtained information bound, and hence, is asymptotically optimal.

### 3. Why LSM Does Not Work for the Identification of ARMAX Models with Correlated Noise

The problem of LSM estimation and identification in ARMAX models has been widely studied in the past. The estimation of the noise-induced bias was presented, for example, in [35], where a unique structure of the ARMAX model was proposed, utilising extra outputs delay. Let us show in this section that for dynamic models (in particular for ARMA-models) the least-squares method (LSM) does not work properly, this means, it leads to biased estimates!

Consider the simplest stable ARMA model with the 1-step correlated noise given by

$$\left. \begin{aligned} y_{n+1} &= ay_n + \zeta_n + d\zeta_{n-1}, \quad y_0 \in \mathbb{R} \text{ is given,} \\ |a| < 1, \quad d \in \mathbb{R}, \quad E\{\zeta_n\} &= 0, \quad E\{\zeta_n^2\} = \sigma^2 > 0 \end{aligned} \right\} \tag{17}$$

where  $\{\zeta_n\}$  is a sequence of independent random variables. Then, the LSM estimate, realising

$$a_n = \arg \min_{a \in \mathbb{R}} \sum_{t=1}^n (y_{t+1} - ay_t)^2,$$

is

$$a_n = \left[ \sum_{t=1}^n y_t y_{t+1} \right] \left[ \sum_{t=1}^n y_t^2 \right]^{-1} \tag{18}$$

and under by the strong version of large number law (LNL) [36] it becomes

$$\begin{aligned} a_n &\stackrel{a.s.}{=} \frac{\frac{1}{n} \sum_{t=1}^n E\{y_t y_{t+1}\}}{\frac{1}{n} \sum_{t=1}^n E\{y_t^2\}} + o_\omega(1), \\ o_\omega(1) &\xrightarrow{n \rightarrow \infty} 0 \text{ (with Prob.1)} \end{aligned}$$

or, equivalently,

$$\begin{aligned} a_n &\stackrel{a.s.}{=} \frac{\frac{1}{n} \sum_{t=1}^n E\{y_t (ay_t + \zeta_t + d\zeta_{t-1})\}}{\frac{1}{n} \sum_{t=1}^n E\{y_t^2\}} + o_\omega(1) = \\ &a \frac{\frac{1}{n} \sum_{t=1}^n E\{y_t (\zeta_t + d\zeta_{t-1})\}}{\frac{1}{n} \sum_{t=1}^n E\{y_t^2\}} + o_\omega(1) = a + d \frac{\frac{1}{n} \sum_{t=1}^n E\{y_t \zeta_{t-1}\}}{\frac{1}{n} \sum_{t=1}^n E\{y_t^2\}} + o_\omega(1). \end{aligned}$$

So, the corresponding identification error comes to be as

$$a_n - a \stackrel{a.s.}{=} d \frac{\frac{1}{n} \sum_{t=1}^n E\{y_t \zeta_{t-1}\}}{\frac{1}{n} \sum_{t=1}^n E\{y_t^2\}} + o_\omega(1)$$

For stable models with  $|a| < 1$  there exist limits

$$\lim_{n \rightarrow \infty} E\{y_n \zeta_{n-1}\} \text{ and } \lim_{n \rightarrow \infty} E\{y_n^2\}$$

and hence, by the Kronecker lemma

$$a_n - a \stackrel{a.s.}{=} d \frac{\lim_{n \rightarrow \infty} E\{y_n \zeta_{n-1}\}}{\lim_{n \rightarrow \infty} E\{y_n^2\}} + o_\omega(1) \tag{19}$$

Let us calculate these limits. From (17), it follows

$$E\{y_{n+1}\xi_n\} = aE\{y_n\xi_n\} + E\{\xi_n^2\} + dE\{\xi_{n-1}\xi_n\} = \sigma^2 \tag{20}$$

$$\begin{aligned} E\{y_{n+1}^2\} &= a^2E\{y_n^2\} + E\{\xi_n^2\} + d^2E\{\xi_{n-1}^2\} + \\ &2aE\{y_n\xi_n\} + 2adE\{y_n\xi_{n-1}\} + 2dE\{\xi_{n-1}\xi_n\} = \\ &a^2E\{y_n^2\} + (1 + d^2)\sigma^2 + 2adE\{y_n\xi_{n-1}\} = \\ &a^2E\{y_n^2\} + (1 + d^2)\sigma^2 + 2ad\sigma^2 \end{aligned} \tag{21}$$

Since, for the stable linear recursion

$$z_{n+1} = \bar{a}z_n + c, |\bar{a}| < 1$$

we have

$$\begin{aligned} z_{n+1} &= \bar{a}z_n + c = \bar{a}(\bar{a}z_{n-1} + c) + c = \\ \bar{a}^2z_{n-1} + c + \bar{a}c &= \dots = \bar{a}^nz_1 + c + \bar{a}c + \bar{a}^2c + \dots + \bar{a}^nc = \\ \bar{a}^nz_1 + c\left(\frac{1 - \bar{a}^{n+1}}{1 - \bar{a}}\right) &\xrightarrow{n \rightarrow \infty} \frac{c}{1 - \bar{a}}. \end{aligned}$$

Then, for (21), we get

$$E\{y_n^2\} \rightarrow \frac{(1 + d^2) + 2ad}{1 - a^2} \sigma^2 = \frac{(1 - a^2) + (a^2 + 2ad + d^2)}{1 - a^2} \sigma^2 = \left[1 + \frac{(a + d)^2}{1 - a^2}\right] \sigma^2 \tag{22}$$

Substitute the obtained limits (20) and (22) into (19) leads to

$$\begin{aligned} a_n - a &\stackrel{a.s.}{\approx} d \frac{\sigma^2}{\left[1 + \frac{(a + d)^2}{1 - a^2}\right] \sigma^2} + o_\omega(1) = \\ d \frac{1}{1 + \frac{(a + d)^2}{1 - a^2}} + o_\omega(1) &\xrightarrow[n \rightarrow \infty]{a.s.} d \frac{1}{1 + \frac{(a + d)^2}{1 - a^2}}. \end{aligned}$$

The derivative calculation of the limit value with respect to  $d$  then gives

$$\left( d \frac{1}{1 + \frac{(a + d)^2}{1 - a^2}} \right)' = (a^2 - 1) \frac{d^2 - 1}{(d^2 + 2ad + 1)^2}$$

So, the extremal points are  $d = \pm 1$ , and hence,

$$\left( d \frac{1}{1 + \frac{(a + d)^2}{1 - a^2}} \right)_{d=1} = \frac{1}{2} - \frac{1}{2}a, \left( d \frac{1}{1 + \frac{(a + d)^2}{1 - a^2}} \right)_{d=-1} = -\frac{1}{2}a - \frac{1}{2}$$

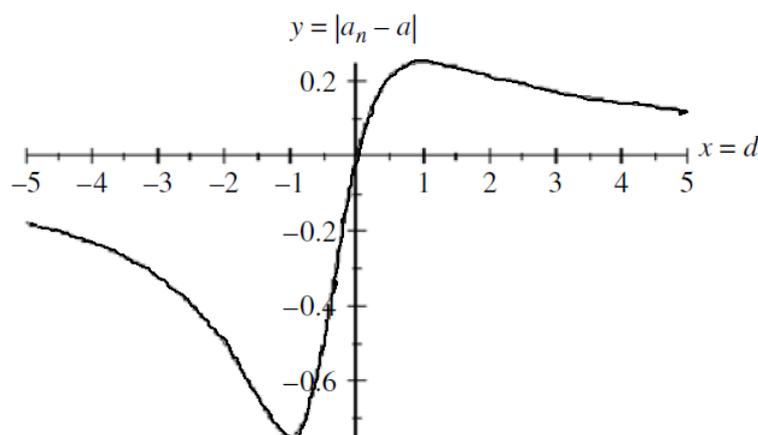
These relations imply the following conclusion: the maximum bias of the LSM estimate is

$$\max_d \lim_{n \rightarrow \infty} |a_n - a| = \frac{1}{2} \max\{|1 - a|; |1 + a|\}$$

The illustrative graphic ( $x := d, y := |a_n - a|$  for  $a = 0.5$ ) is shown in Figure 1.

Conclusion: Be careful! The LS method does not work for identification of parameters of dynamic models with correlated noises!

As a result, certain unique approaches, distinct from LSM, must be developed.



**Figure 1.** The bias dependence on the correlation coefficient  $d$ .

#### 4. Some Other Identification Techniques

##### *Identification of Non-Stationary Parameters and Bayesian Method*

Some other parameter estimation/identification methods have been proposed for more complex situations. In [37], a combination of recursive version of the IV method with the matrix forgetting factor was presented for identification of time-varying parameters for ARMAX models, showing that the identification error in average has a bound dependent on the rate of the parameter variation, as well as on the variance of the noise. The version of IV method with adaptive matrix forgetting factor was studied in [38]. In some cases equation-error and output-error approaches have been used to deal with the problem where all the observed variables are corrupted by noise. The parameter bounding by the bounded equation error and the bounded errors in variables based on these approaches was studied in [39]. Bayesian parameter estimation and prediction of lineal-in-parameters models under the presence of coloured noise is addressed in [40], and it is based on a model called ARMAX. This model is a finite mixture of ARMAX elements with a common ARX part. This ARX part described a fixed deterministic input–output relationship: this model is estimated using a recursive quasi-Bayes algorithm that relies on a classical Bayesian solution without restriction on the MA component. The proposed model provides flexibility with respect to varying characteristics of the model noise. The measurement errors that affect data entries make the estimation problem more complicated. A solution to this problem was proposed in [41] by enhancing the ARMAX models by including some additive error terms on the output, and then developing a moving horizon estimator for the extended ARMAX model. The proposed method then models the measurement errors as nuisance variables and these are estimated simultaneously with the states, and the identifiability was achieved by regularising the LS cost with the  $\ell_2$  norm of the nuisance variables, leading to an optimisation problem with an analytical solution. The nuisance variables have been recently used to model the output noise, as well as the potentially existing outliers (see [27]). These outliers are regularised with the  $\ell_2$  norm for the estimation, and the regularised estimator suppresses the influence of the output noise and provides a minimum-variance estimate.

For the continuous-time case, the LSM with forgetting factor has been implemented for estimating constant and time-varying parameters ([42–44]). The proposed algorithms in [45] showed a good performance, but the bias, as in the discrete-time case, affects the estimation. The estimation algorithm was implemented for additive and multiplicative noises (see [46]), and in both scenarios, LSM is affected by the noise level, showing that is not the best method for stochastic systems, either in discrete or continuous time. To deal with the bias problem, a method combining the equivalent control with LSM was proposed, these two algorithms working in parallel reduce the bias in the estimation even in the presence of coloured noises (see [47]).

### 5. Regular Observations and Information Inequality for Observations with Coloured Noise

In estimation theory and statistics [48,49], the Cramér–Rao bound (CRB) expresses a lower bound on the variance of unbiased estimators  $c_n$  of a deterministic (fixed, though unknown) parameter  $c$ , stating that the variance of any such estimator is at least as high as the inverse of the Fisher information (FIM)  $\mathbb{I}_F^{-1}(c, n)$ . Namely, for every unbiased estimator  $c_n$  ( $n$  is the number of available regular observations), an inequality of the type

$$\text{Var}_c(c_n) \geq \mathbb{I}_F^{-1}(c, n) \tag{23}$$

for every  $c$  in the parameter space  $C$ , it is called an information inequality, which plays a very important role in parameter estimation. The early works of Cramér (1946) [50] and Rao (1945) [51] introduced the Cramer–Rao inequality for regular density functions. Later, Vincze (1979) [52] and Khatri (1980) (see [53]) introduced information inequalities by imposing the regularity assumptions on a priori distribution rather than on the model. An unbiased estimator which achieves this lower bound is usually said to be (fully) efficient. This is a solution that achieves the lowest possible mean squared error among all unbiased methods, and therefore is the minimum variance unbiased (MVU) estimator. However, in some cases, there are no unbiased techniques that achieve this bound. This may occur either if for any unbiased estimator there exists another estimator with a strictly smaller variance, or if an MVU estimator exists but its variance is strictly greater than the inverse of the Fisher information. The Cramér–Rao bound (37) can also be used to bound the variance of biased estimators of given bias. In some cases, a biased approach can result in both a variance and a mean squared error that are below the unbiased Cramér–Rao lower bound.

Recall some important definitions.

#### 5.1. Main Definitions and the Cramer–Rao Information Inequality

In a general case, the observable output sequence  $\mathbf{y}_n := \{y_1, y_2, \dots, y_n\}$  may be of a vector type ( $y_i \in \mathbb{R}^L$ ) containing the information on the parameter  $c \in \mathbb{R}^N$ . The function  $p(\mathbf{y}_n | c)$ ,  $c \in C \subseteq \mathbb{R}^N$  is called the joint density of the distribution of the vector  $\mathbf{y}_n$ . Any Borel function  $c_n = c_n(\mathbf{y}_n) \in \mathbb{R}^N$  can be considered as an estimate of the parameter  $c$ .

**Definition 1.** *The vector-function*

$$m_n(c) = \mathbb{E}\{c_n\} = \int_{Y_n} c_n(\mathbf{y}_n) p(\mathbf{y}_n | c) d\mathbf{y}_n \in \mathbb{R}^N, \\ Y_n = \{\mathbf{y}_n | p(\mathbf{y}_n | c) > 0, c \in C\},$$

is referred to as the averaged value of the estimate  $c_n$ , based on available observations  $\mathbf{y}_n$ ;

- If  $m_n(c) = c$ , then the estimate  $c_n$  is called unbiased and asymptotically unbiased if  $\lim_{n \rightarrow \infty} m_n(c) = c$ .
- The observations  $\mathbf{y}_n$  are referred to be as regular on the class  $C$  of parameters if

$$\sup_{c \in C} \mathbb{E}\{\|\ln p(\mathbf{y}_n | c)\|^2\} = \sup_{c \in C} \int_{Y_n} \|\ln p(\mathbf{y}_n | c)\|^2 p(\mathbf{y}_n | c) d\mathbf{y}_n < \infty, \tag{24}$$

and for any  $c \in C$

$$\mathbb{I}_F(c, n) = \mathbb{E}\{\nabla_c \ln p(\mathbf{y}_n | c) \nabla_c^T \ln p(\mathbf{y}_n | c)\} = \int_{Y_n} [\nabla_c \ln p(\mathbf{y}_n | c) \nabla_c^T \ln p(\mathbf{y}_n | c)] p(\mathbf{y}_n | c) d\mathbf{y}_n > 0. \tag{25}$$

- The matrix  $\mathbb{I}_F(c, n)$  is called the Fisher information matrix for the set of available observations  $\mathbf{y}_n$ .

As it was mentioned in [54], when the Fisher information takes into account the stochastic complexity and the associated universal processes are derived for a class of para-

metric processes. The main condition required is that the maximum likelihood estimates satisfy the central limit theorems.

In some cases, the Fisher information matrix (FIM) is required to be non-singular (25) to guarantee the observability of the system (see [55]). The algebraic properties of FIM for stationary processes have been widely studied, for example there is a survey paper written by André Klein where this study is presented [56]. The FIM is necessary for the Cramer–Rao inequality; it is a basic tool for estimation theory in mathematical statistics, and in stationary processes is related to the solution of Stein equations. A procedure to compute the theoretical periodic autocovariance function in terms of the parameters of the periodic model for periodic autoregressive moving average models was presented in [57], where the necessary and sufficient condition for non-singular FIM of a periodic ARMA model was calculated. So, the Fisher information matrix for the Gaussian case in ARMAX processes has been previously studied. In [58], an algorithm composed by Chandrasekhar recursion equations at a vector-matrix level was proposed, where the recursions consist of derivatives based on appropriate differential rules that are applied to a state space model for a vector process. The recursions obtained were given in terms of expectations of derivatives of innovations.

**Theorem 1.** *Cramer–Rao information inequality. For any set  $Y_n$  of regular observations, and for any estimate  $c_n$  with differentiable averaged value function  $m_n(c)$  the following inequality holds*

$$E\{(c_n - c)(c_n - c)^T\} \geq [m_n(c) - c][m_n(c) - c]^T + \nabla m_n(c) \mathbb{I}_F^{-1}(c, n) \nabla^T m_n(c). \tag{26}$$

**Corollary 1.** *For unbiased estimates satisfying  $m_n(c) = c$ ,  $\nabla m_n(c) = I_{n \times n}$ , the Cramer–Rao inequality becomes*

$$E\{(c_n - c)(c_n - c)^T\} \geq \mathbb{I}_F^{-1}(c, n). \tag{27}$$

This inequality is widely used in discrete-time systems for various purposes. The posterior Cramer–Rao bound on the mean square error in tracking the bearing, bearing rate and power level of a narrowband source is developed in [59]. Their formulation used a lineal process model with additive noise and a general nonlinear measurement model, where the measurements are the sensor array data. This bound can be applied to multi-dimensional nonlinear and possibly non-Gaussian systems. In [60], the case of a singular conditional distribution of the one-step-ahead state vector given the present state was considered. The bound was evaluated for recursive estimation of slowly varying parameters of AR processes, tracking a slowly varying single cisoid in noise and tracking the parameters of a sinusoidal frequency with a sinusoidal phase modulation. A variation of the Cramer–Rao inequality is the Cramer–Rao–Frechet inequality, which has been applied for discrete-time nonlinear filtering. In [61], this inequality was reviewed and extended to track fitting, where it is shown that the inequality does not cause the limitations of the resolution of the track fits with a certain number of observations, and that the inequality remains valid even in irregular models supporting the similar improvement of resolution for realistic models.

### 5.2. Fisher Information Matrix Calculation

Using the Bayes formula

$$p(\mathbf{y}_n | c) = p(y_n | \mathbf{y}_{n-1}; c) p(\mathbf{y}_{n-1} | c) = \dots = \left[ \prod_{k=1}^n p(y_k | \mathbf{y}_{k-1}; c) \right] p(y_0 | c),$$

for the likelihood function  $L_n(\mathbf{y}_n | c) = -\ln p(\mathbf{y}_n | c)$  we have the following representation:

$$L_n(\mathbf{y}_n | c) = -\ln p(\mathbf{y}_n | c) = -\sum_{k=1}^n \ln p(y_k | \mathbf{y}_{k-1}; c) - \ln p(y_0 | c). \tag{28}$$

Define also

$$u_t(c) = \nabla_c L_t(\mathbf{y}_t | c) - \nabla_c L_{t-1}(\mathbf{y}_{t-1} | c) = \frac{\nabla_c p(y_t | \mathbf{y}_{t-1}; c)}{p(y_t | \mathbf{y}_{t-1}; c)} = -\nabla_c \ln p(y_t | \mathbf{y}_{t-1}; c),$$

which is a martingale difference, since  $E\{u_t(c) | \mathcal{F}_{t-1}\} \stackrel{a.s.}{=} 0$ , and satisfies the property

$$\nabla_c u_t(c) = -\frac{\nabla_c^2 p(y_t | \mathbf{y}_{t-1}; c)}{p(y_t | \mathbf{y}_{t-1}; c)} + \frac{\nabla_c p(y_t | \mathbf{y}_{t-1}; c) \nabla_c^T p(y_t | \mathbf{y}_{t-1}; c)}{p^2(y_t | \mathbf{y}_{t-1}; c)}. \tag{29}$$

For regular unbiased observations, the Fisher information matrix  $\mathbb{I}_F(c, n)$  can be calculated as

$$\mathbb{I}_F(c, n) = \sum_{k=1}^n E\{u_k u_k^T\} = \sum_{k=1}^n E\{\nabla_c u_k(c)\} = \left. \begin{aligned} E\left\{ \sum_{k=1}^n \frac{\nabla_c p(y_k | \mathbf{y}_{k-1}; c) \nabla_c^T p(y_k | \mathbf{y}_{k-1}; c)}{p^2(y_k | \mathbf{y}_{k-1}; c)} \right\} = \\ E\left\{ \sum_{k=1}^n \nabla_c \ln p(y_k | \mathbf{y}_{k-1}; c) \nabla_c^T \ln p(y_k | \mathbf{y}_{k-1}; c) \right\}. \end{aligned} \right\} \tag{30}$$

### 5.3. Asymptotic Cramer–Rao Inequality

Multiplying both sides of (27) by  $n$  we get

$$nE\{(c_n - c)(c_n - c)^T\} \geq \left(\frac{1}{n} \mathbb{I}_F(c, n)\right)^{-1}.$$

Taking  $n \rightarrow \infty$  we get

$$\liminf_{n \rightarrow \infty} nE\{(c_n - c)(c_n - c)^T\} \geq \mathbb{I}_F^{-1}(c), \tag{31}$$

where

$$\mathbb{I}_F(c) := \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{I}_F(c, n) = \limsup_{n \rightarrow \infty} E\left\{ \frac{1}{n} \sum_{k=1}^n u_k(c) u_k^T(c) \right\} > 0.$$

**Remark 2.** In view of (29) it follows

$$\mathbb{I}_F(c) = \limsup_{n \rightarrow \infty} \frac{1}{n} E\{\nabla_c^2 L_n(\mathbf{y}_t | c)\}. \tag{32}$$

### 5.4. Whitening Process for Stable and Minimal-Phase Forming Filters

Although additive Gaussian white noise is widely used, in many research areas the present noises are non-Gaussian. In some cases, detectors are used to whiten the data and then the estimation/identification is performed (see for example [62]). In the presence of non-white noises, one of the most common methods to deal with this perturbation is a whitening filter. A transfer function of an estimated noise can be used to filter the input–output data of the system and presents a filtering-based recursive analogue of the LSM algorithm for the ARMAX model. In [63], it is shown that through data filtering one can obtain two identification models, the first one including the parameter of the system model and the second including the parameter of the noise model; this can lead to a more accurate parameter estimation. A whitening filter can be applied in coloured Gaussian noises when there is a residual white noise component present. The existence of a realisable whitening filter is demonstrated in [64].

The model (14) can be symbolically represented as

$$y_n = z_n^\top c + \eta_n = z_n^\top c + \frac{H_1(q)}{H_2(q)} \zeta_n. \tag{33}$$

In view of the Assumption 4, the polynomials  $H_1(q)$  and  $H_2(q)$  are stable and, hence, we are able to apply the inverse operator  $\frac{H_2(q)}{H_1(q)}$  to both sides of the model (33), obtaining

$$\begin{aligned} \tilde{y}_n &= \frac{H_2(q)}{H_1(q)} y_n, \quad \tilde{y}_{-s} := 0, \quad s = 0, 1, \dots, K_1, \\ \tilde{z}_n &= \frac{H_2(q)}{H_1(q)} z_n, \quad \tilde{z}_{-s} := 0, \quad s = 0, 1, \dots, K_1, \\ \tilde{\zeta}_n &:= \frac{H_2(q)}{H_1(q)} \frac{H_1(q)}{H_2(q)} \zeta_n \stackrel{a.s.}{=} \zeta_n + O_\omega(\lambda^n), \quad |\lambda| < 1, \end{aligned} \tag{34}$$

where  $\lambda$  is one of the eigenvalues of the polynomials  $H_1(q)$  and  $H_2(q)$  which is most close to the unitary circle. The function  $O_\omega(\lambda^n)$  is a random process, defined on  $(\Omega, \mathcal{F}, P)$  and such that

$$0 < \liminf_n^{a.s.} O_\omega(\lambda^n)/\lambda^n \leq \limsup_n O_\omega(\lambda^n)/\lambda^n \leq \text{const}(\omega) \stackrel{a.s.}{<} \infty.$$

So, finally, after the “whitening process” (inverse operator) application we get

$$\tilde{y}_n = \tilde{z}_n^\top c + \tilde{\zeta}_n. \tag{35}$$

**Remark 3.** This means that on-line application of the “whitening process” to the initial model (33) permits considering the corresponding transformed model (34) which deals with “quasi” white noise  $\tilde{\zeta}_n$  exponentially quickly, tending to the exact white noise  $\zeta_n$ , fulfilling

$$\|\tilde{\zeta}_n - \zeta_n\| = O_\omega(\lambda^n) \stackrel{a.s.}{\rightarrow} 0$$

when  $n \rightarrow \infty$ . This permits to represent (35) as

$$\tilde{y}_n = \tilde{z}_n^\top c + \zeta_n + O_\omega(\lambda^n). \tag{36}$$

5.5. Cramer–Rao Inequality for ARMAX Models with a Generating Noise from the Class  $\mathcal{P}_\zeta$  of p.d.f.

**Theorem 2.** The Cramer–Rao inequality (see [15]) in the form (31) is

$$\left. \begin{aligned} \liminf_{n \rightarrow \infty} n E\{(c_n - c)(c_n - c)^\top\} &\geq \sup_{p_\zeta \in \mathcal{P}_\zeta} \mathbb{I}_F^{-1}(c) = \\ &\sup_{p_\zeta \in \mathcal{P}_\zeta} (I_{F,\zeta}(p_\zeta) \mathcal{R}(p_\zeta))^{-1} \end{aligned} \right\} \tag{37}$$

where

$$\left. \begin{aligned} \mathcal{R}(p_\zeta) &:= \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E\{\tilde{z}_k \tilde{z}_k^\top\}, \\ I_{F,\zeta}(p_\zeta) &= E\left\{ \left( \frac{\partial}{\partial \zeta} \ln p_\zeta(\zeta(\omega)) \right)^2 \right\} = \int_{x \in \mathbb{R}^1} \frac{\left( \frac{\partial}{\partial x} p_\zeta(x) \right)^2}{p_\zeta(x)} dx. \end{aligned} \right\} \tag{38}$$

**Remark 4.** In the regression case ( $a_l = 0, l = 1, \dots, L$ ) the matrix  $\mathcal{R}$  (38) does not depend on  $p_\zeta$ .

Conclusion. According to the information inequality (37), the “best” (asymptotically optimum or efficient) estimate  $c_n^*$  of the parameter  $c \in C$  is the one that achieves equality in the (37). The inequality given by (37) implies that after  $n$  regular observations  $y_n$  the covariance matrix of the estimation error  $(c_n - c)$ , which defines the quality of the estimation process, can not be less than the corresponding Fisher information matrix (25). In other

words, the Fisher information matrix (25) will define the maximal possible quality of the identification process, which can not be improved by any other identification algorithm.

## 6. Robust Version of Maximum Likelihood Method with Whitening: MLMW Procedure

For parameter estimation and system modelling, the maximum likelihood technique is critical. The maximum of the likelihood function in Gaussian case is equivalent to minimising the least-squares cost function [65]. In this paper, a recursive maximum likelihood least-squares identification algorithm for systems with autoregressive moving average noises was derived. The maximum likelihood has been widely implemented under Gaussian perturbations, for example in [66], the Gaussian likelihood function was studied when data are generated by a high-order continuous-time ARMAX model, and these data are observed as stocks and flows at different frequencies. The maximum likelihood method can be modified using the stochastic gradient; this modification was presented in [67], where this modification was proposed for ARMAX models. In this case, the modified algorithm can estimate unknown parameters and the unknown noise simultaneously, with less computational cost and better accuracy. Non-asymptotic deviations bounds for least-squared estimation in Gaussian AR processes have been recently studied (see [68]). The study relies on martingale concentration inequalities and tail bound for  $\chi^2$ -distributed variables; in the end, they provided a concentration bound for the sample covariance matrix of the process output.

### 6.1. Whitening Method and Its Application

The whitening method is commonly used to prevent the bias problem [69]. A modified version of direct whitening method, which is called MDWM, was proposed as an ARMA model parameter estimation technique in [70]. The proposed direct whitening method (DWM) provides the parameter estimates which make the prediction errors uncorrelated, in some cases this algorithms might fall at local minima and give parameter estimates. To deal with this problem, an MDWM which chooses the consistent estimates among a large number of DWM estimates can be implemented. Pre-whitening can be performed with first order differentiation of signals and/or the implementation of an inverse filter based on linear prediction, as it is shown in [71], where the whitening was the previous step in a cross-correlation method for identifying aircraft noise, showing that whitening can be successfully developed for real-time operation in the detection of correlation peaks. An iterative procedure for minimising and whitening the residual of the ARMAX model was presented in [72], since usually when the system is identified from input–output data in the time domain, it is assumed that the data is enough and the ARX model order is large enough. The results show that in the residual whitening method we can use an ARMAX model that includes the noise dynamics, instead of an ARX model, and the properties of the residual sequence, such as the orthogonal conditions, can convert to the optimal properties of the Kalman filter. The influence function is an analysis tool in robust statistics we used to formulate a recursive solution for ARMAX processes filtering in [73], in particular for a  $t$ -distribution noise. The filter was formulated as a maximum likelihood problem, where an influence function approximation was used to obtain a recursive solution to reduce computational load and facilitate the implementation.

Whitening techniques have also been implemented for noise cancellation, in [74] is the base for an approach to adaptive white noise cancellation based on adaptive control principles. In this case, the goal was to create a physical noise-reduced environment at the vicinity of noisy machinery for a stochastic machine noise. Another method implemented for filtering is signal smoothing when the data are generated (or represented) by an autoregressive moving average with exogenous inputs (ARMAX) model. In the case presented in [75], the original ARMAX recurrence relation is used and combined with a constrained LS optimisation to filter the system as well as the measurement noise components and estimate the desired signal in the form of a block-wise matrix formulation.

Whitening processes are a very useful pre-processing technique to deal with the presence of non-white noises, and the improve the estimation results regardless of the estimation algorithm used [76]. In [77], a residual whitening method enforces the properties of the Kalman filter for a finite set of data. This technique has been implemented in ARMAX models for the identification of inductor motor systems. The importance of the study and development of estimation/identification algorithms for ARMAX models with coloured noises relies on the importance of this model in various areas of study and its many applications. The identification of the ARMAX models allows the implementation of control techniques, such as the predictive control presented in [78], which is applied for the control of a pneumatic actuator based on an ARX model built by a neural network. There, the control showed a quick response and an accurate tracking. The estimation has been implemented for electromechanical modes and mode shapes for multiple synchrophasors (see [79]). Their approach was based on identifying the transfer function of the state space model of a linearised power system through the estimation of a multichannel ARMAX mode, and it was simulated using data from a reduced-order model of the Western Electricity Coordinating Council (WECC) system. The ARMAX model has been used to model an outlet temperature of the first-stage cyclone pre-heater in the cement firing system (see [80]). In that case, a Butterworth low-pass filter and normalized processing are used to process a cement firing system data, and the input variables modelled are selected by the Pearson correlation analysis. The parameters of the model were identified using a recursive maximum likelihood algorithm, and the results validated with a residual analysis method.

Econometrics is an area where the estimation/identification of ARMAX models (the integral version of the ARMAX model) has great importance (see [66,81,82]). The integration of macroeconomic indicators in the accuracy of throughput time-series forecasting model can be addressed using ARMAX models, as it is shown in [83]. There, the dynamic factors are extracted from external macroeconomic indicators influencing the observed throughput, and then a family of ARMAX models is generated based on derived factors. This model can be used to produce future forecasts. Some variations of the ARMAX model, such as the autoregressive moving average explanatory input model of the Koyck kind (KARMAX) are also used in econometrics. Another interesting application in econometrics is presented in [84], where it is shown how the recent deregulation of the electricity industry and reliance on competitive wholesale markets has generated significant volatility in wholesale electricity prices. Due to the importance of short-term price forecasts, an estimation and evaluation of the forecasting performance of four ARMAX–GARCH models for five MISO pricing hubs (Cinergy, First Energy, Illinois, Michigan and Minnesota) using hourly data from 1 June 2006 to 6 October 2007 is given. In this study, the importance of the patterns of the electricity price volatility is shown, as well as the volatility dynamics regulated by the states.

In [85], an identification algorithm is presented, where the debt management in indebted poor countries is studied, using data from the World Bank database from 1970 to 2018 based on the maximum likelihood method, and then comparing the results with prediction error and the instrumental variable methods.

6.2. Recurrent Robust Identification Procedures with Whitening and a Nonlinear Residual Transformation

Consider the following class of recurrent identification procedures [15] which may be applied to the transformed model (36):

$$\left. \begin{aligned} c_n &= c_{n-1} + \Gamma_n \tilde{z}_n \varphi(\tilde{y}_n - \tilde{z}_n^T c_{n-1}) \\ c_0 &\text{—any given value} \end{aligned} \right\} \tag{39}$$

$$\Gamma_n = \left( \sum_{t=0}^n \tilde{z}_t \tilde{z}_t^T \right)^{-1}, \quad n \geq n_0 := \left\{ \min_k \sum_{t=0}^k \tilde{z}_t \tilde{z}_t^T \right\} > 0.$$

**Remark 5.** Notice that  $\Gamma_n$  in (39) can be calculated recursively (as in the least-square method)

$$\Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1} \tilde{z}_n \tilde{z}_n^T \Gamma_{n-1}}{1 + \tilde{z}_n^T \Gamma_{n-1} \tilde{z}_n}, \quad n \geq n_0, \tag{40}$$

and possesses (in the accepted assumptions) the following property

$$\Gamma_n \stackrel{a.s.}{\simeq} \frac{1}{n} \mathcal{R}^{-1} \tag{41}$$

$$\mathcal{R} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \{ \tilde{z}_k^T \tilde{z}_k \} > 0.$$

**Theorem 3 ([15]).** If

1.  $\xi_n$  is i.i.d. sequence with

$$\mathbb{E} \{ \xi_n \} = 0, \quad \mathbb{E} \{ \xi_n^2 \} = \sigma^2 > 0, \quad \mathbb{E} \{ \xi_n^4 \} = \mathbb{E} \{ \xi_1^4 \} < \infty.$$

2. The nonlinear transformation  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  satisfies the conditions

$$x\psi(x) \geq \delta x^2, \quad \delta > 0, \quad \psi(0) = 0, \quad S(x) \leq k_0 + k_1 x^2,$$

with

$$\psi(x) = \mathbb{E} \{ \varphi(x + \xi_n) \}, \quad S(x) := \mathbb{E} \{ \varphi^2(x + \xi_n) \},$$

then

$$\Delta_n = c_n - c \stackrel{a.s.}{\xrightarrow{n \rightarrow \infty}} 0. \tag{42}$$

Following to Lemma 13.7 in [36] and defining a new process  $\{ \tilde{\Delta}_n \}_{n \geq 0}$  as

$$\tilde{\Delta}_n = \left[ 1 - \frac{\psi'(0)}{n} \right] \tilde{\Delta}_{n-1} + \frac{1}{n} \mathcal{R}^{-1} \tilde{z}_n (o_\omega(1) + \zeta_n), \quad \tilde{\Delta}_0 = \Delta_0, \tag{43}$$

we may formulate the following auxiliary result.

**Theorem 4** (on  $\sqrt{n}$ -equivalency). Under the assumptions of Theorem 3, the process (42) is  $\sqrt{n}$ -equivalent to the process (43), that is,

$$\sqrt{n}(\Delta_n - \tilde{\Delta}_n) \stackrel{a.s.}{\xrightarrow{n \rightarrow \infty}} 0. \tag{44}$$

The property of the asymptotic normality of the process  $\{ \sqrt{n} \Delta_n \}_{n \geq 0}$  helps us to estimate the exact rate of convergence (not only the order of convergence, but also its constant) of the identification procedure (39).

**Theorem 5** (on asymptotic normality). Suppose that the conditions of Theorem 3 are fulfilled and, additionally,

$$\psi(0) = 0, \quad S(0) > 0, \quad \psi'(0) > 1/2. \tag{45}$$

Then, the process  $\{ \sqrt{n} \Delta_n \}_{n \geq 0}$  is asymptotically normal

$$\sqrt{n} \Delta_n \stackrel{d}{\xrightarrow{n \rightarrow \infty}} \mathcal{N}(0, V), \tag{46}$$

with the covariation matrix  $V$ , equal to

$$V = \frac{S(0)}{2\psi'(0) - 1} \mathcal{R}^{-1}. \tag{47}$$

It results directly from Theorem 13.6 in [36].

**Remark 6.** The matrix  $V$  defines the rate of the convergence of the procedure (39), that is

$$\Delta_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, n^{-1}V).$$

As it follows from (47),  $V$  depends on a real noise density distribution  $p_{\xi}$  (since  $S(0)$ ,  $\psi'(0)$  and  $\mathcal{R}$  may be dependent on  $p_{\xi}$ ) and on a nonlinear function  $\varphi$  (through  $S(0)$  and  $\psi'(0)$ ). That’s why, to emphasise this dependence, we use the notation

$$V = V(p_{\xi}, \varphi).$$

Following [13,14], let us introduce the main definition of this section.

**Definition 2.** The pair of functions given by  $(p_{\xi}^*, \varphi^{**})$  define the estimating procedure (54) with the nonlinear residual transformation  $\varphi^*$ , which is robust with respect to a distribution  $p_{\xi}$ , belonging to a class  $\mathcal{P}_{\xi}$ , if for any admissible  $\varphi$ , satisfying the assumptions of Theorem 5, and any generating noise distribution  $p_{\xi} \in \mathcal{P}_{\xi}$  the following “saddle-point” inequalities hold:

$$V(p_{\xi}, \varphi^{**}) \leq V(p_{\xi}^*, \varphi^{**}) \leq V(p_{\xi}, \varphi^*). \tag{48}$$

Here, both inequalities should be treated in a “matrix sense”, that is,

$$A = A^T \leq B = B^T \text{ if } B - A \geq 0.$$

In other words:

- The distribution  $p_{\xi}^*$  is the “worst” within the class  $\mathcal{P}_{\xi}$ .
- The nonlinear transformation  $\varphi^{**}$  is “the best one” oriented on the “worst” noise with the distribution  $p_{\xi}^*$ .

This can be expressed mathematically as follows:

$$\left. \begin{aligned} \varphi^{**} &= \arg \inf_{\varphi} \sup_{p_{\xi} \in \mathcal{P}_{\xi}} V(p_{\xi}, \varphi), \\ p_{\xi}^* &= \arg \sup_{p_{\xi} \in \mathcal{P}_{\xi}} \inf_{\varphi} V(p_{\xi}, \varphi), \end{aligned} \right\} \tag{49}$$

so that

$$\inf_{\varphi} \sup_{p_{\xi} \in \mathcal{P}_{\xi}} V(p_{\xi}, \varphi) = \sup_{p_{\xi} \in \mathcal{P}_{\xi}} \inf_{\varphi} V(p_{\xi}, \varphi) := V^*. \tag{50}$$

According to (37), for any fixed  $p_{\xi} \in \mathcal{P}_{\xi}$

$$\inf_{\varphi} V(p_{\xi}, \varphi) = \inf_{\varphi} \left( \frac{S(0)}{2\psi'(0) - 1} \mathcal{R}^{-1} \right) \geq \sup_{p_{\xi} \in \mathcal{P}_{\xi}} [I_{F,\xi}(p_{\xi}) \mathcal{R}(p_{\xi})]^{-1} \tag{51}$$

**Lemma 1 ([15]).** The low bound in (51) coincides with the Cramer–Rao bound (37) and is achieved when the nonlinear function in (39) is

$$\varphi^{**}(v) = -I_{F,\xi}^{-1}(p_{\xi}^*) \frac{d}{dv} \ln p_{\xi}^*(v) \tag{52}$$

with

$$p_{\xi}^* = \arg \sup_{p_{\xi} \in \mathcal{P}_{\xi}} [I_{F,\xi}(p_{\xi}) \mathcal{R}(p_{\xi})]^{-1} \tag{53}$$

In other words, this lemma states that the nonlinear residual transformation  $\varphi^{**}$  (52) is robust with respect to distributions  $p_{\zeta} \in \mathcal{P}_{\zeta}$ .

So, the asymptotically optimal recurrent robust identification procedure (39) for coloured noise perturbations in (33) is

$$\left. \begin{aligned} c_n &= c_{n-1} - \Gamma_n \tilde{z}_n I_{F,\zeta}^{-1} \left( p_{\zeta}^* \right) \frac{d}{dv} \ln p_{\zeta}^*(v) \Big|_{v=\tilde{y}_n - \tilde{z}_n^T c_{n-1}} \\ \Gamma_n &= \Gamma_{n-1} - \frac{\Gamma_{n-1} \tilde{z}_n \tilde{z}_n^T \Gamma_{n-1}}{1 + \tilde{z}_n^T \Gamma_{n-1} \tilde{z}_n}, \quad n \geq n_0, \end{aligned} \right\} \tag{54}$$

which in fact is the maximum likelihood recurrent procedure with the worst p.d.f.  $p_{\zeta}^*(v)$  on the given class  $\mathcal{P}_{\zeta}$ .

**Remark 7.** Notice that for the class of ARMAX models with coloured noises and regular observations, there does not exist any other algorithm providing, asymptotically, a rate of convergence better than the suggested procedure (54).

As we can see, whitening is a pre-processing step in the estimation process which can be applied simultaneously with the identification procedure (54). In [86], this step was implemented in the blind source separation process, where a robust whitening is based on the eigenvalue decomposition of a positive definite linear combination of correlation matrices.

This problem can be addressed analysing the noise power spectra density. This has been implemented by identifying the noise power spectral density of interferometric detectors using parametric techniques (see [87]). This is an adaptive technique used to identify and to whiten data provided by the interferometric detectors. The least-squares lattice filter proved to be the best among the analysed filters. One of the applications for this technique was presented in [88] where it was implemented for gravitational data wave analysis. There, it is shown how it is possible to estimate the noise power spectral density of gravitational wave detectors using parametric techniques, and it also shows how is it possible to whiten the noise data before they pass the detection algorithms.

### 6.3. Particular Cases for Static (Regression) Models

Recall that for regression models ( $a_l = 0, l = 1, \dots, L$ ) the matrix  $\mathcal{R}$  does not depend on p.d.f.  $p_{\zeta}$ , and the relation (53) becomes

$$p_{\zeta}^*(v) = \arg \sup_{p_{\zeta} \in \mathcal{P}_{\zeta}} [I_{F,\zeta}(p_{\zeta})]^{-1} = \arg \inf_{p_{\zeta} \in \mathcal{P}_{\zeta}} I_{F,\zeta}(p_{\zeta}).$$

**Lemma 2.** In the class  $\mathcal{P}_{\zeta}^1 := \left\{ p_{\zeta} : p_{\zeta}(0) \geq \frac{1}{2a} > 0 \right\}$  (5) the worst distribution density  $p_{\zeta}^*(x)$  is the Laplace p.d.f. given by

$$p_{\zeta}^*(x) = \arg \inf_{p_{\zeta} \in \mathcal{P}_{\zeta}^1} I_{F,\zeta}(p_{\zeta}) = \frac{1}{2a} \exp \left\{ -\frac{|x|}{a} \right\}. \tag{55}$$

See Figure 2.

**Corollary 2.** The robust on  $\mathcal{P}_{\zeta}^1$  version of the procedure (54) contains

$$\varphi^{**}(x) = -I_{F,\zeta}^{-1} \left( p_{\zeta}^* \right) \frac{d}{dv} \ln p_{\zeta}^*(v) = a \operatorname{sign}(x) \tag{56}$$

**Lemma 3.** In the class  $\mathcal{P}_{\zeta}^2 := \left\{ p_{\zeta} : \int_{\mathbb{R}} x^2 p_{\zeta}(x) ds \leq \sigma^2 < \infty \right\}$  (6), the worst distribution density  $p_{\zeta}^*(x)$  is the Laplace p.d.f. given by

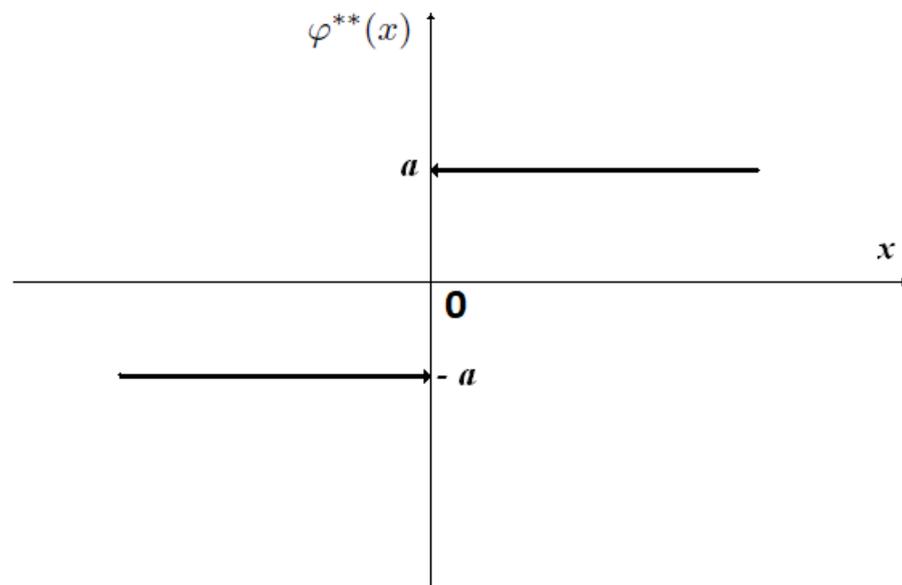
$$p_{\xi}^*(x) = \arg \inf_{p_{\xi} \in \mathcal{P}_2} I_F(p_{\xi}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \tag{57}$$

that is, the worth on  $\mathcal{P}_{\xi}^2$  distribution density is the Gaussian p.d.f. (57).

**Corollary 3.** The robust on  $\mathcal{P}_{\xi}^2$  version of the procedure (54) contains

$$\varphi^*(x) = -I_F^{-1}\left(p_{\xi}^*\right) \frac{d}{dv} \ln p_{\xi}^*(v) = x \tag{58}$$

which means that the standard LSM algorithm with linear residual transformation is robust within the class  $\mathcal{P}_2$ .



**Figure 2.** The nonlinear transformation  $\varphi^{**}$  for the class  $\mathcal{P}_{\xi}^1$ .

**Lemma 4.** In the class  $\mathcal{P}_{\xi}^3 := \{p_{\xi} : p_{\xi}(x) = (1 - \alpha)p_{\mathcal{N}(0,\sigma^2)}(x) + \alpha q(x)\}$  (7) (of all symmetric “approximately normal” p.d.f.), the worth distribution density  $p_{\xi}^*(x)$  is Gaussian p.d.f. within some zone  $\Delta$  and the Laplace p.d.f. out of this zone:

$$p_{\xi}^*(x) = \arg \inf_{p_{\xi} \in \mathcal{P}_{\xi}^3} I_{F,\xi}(p_{\xi}) = \begin{cases} \frac{1-\alpha}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} & \text{for } |x| \leq \Delta \\ \frac{1-\alpha}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\Delta|x|}{\sigma^2} + \frac{\Delta^2}{2\sigma^2}\right\} & \text{for } |x| > \Delta \end{cases} \tag{59}$$

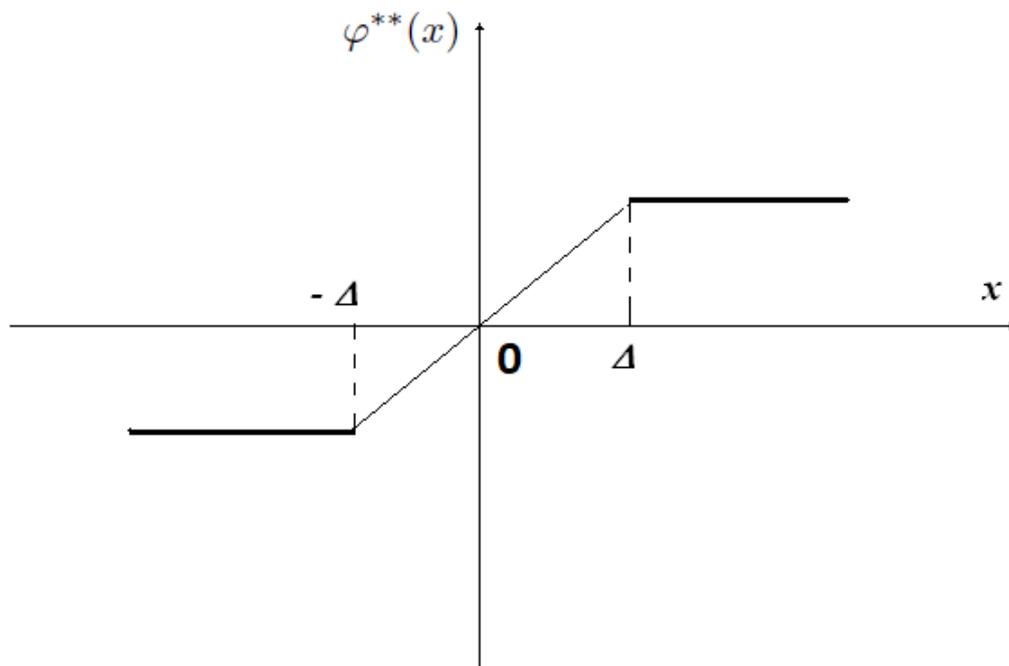
The parameter  $\alpha \in [0, 1]$  characterises the level of the effect of a “dirty” distribution  $q(x)$  to the basic one  $p_{\mathcal{N}(0,\sigma)}(x)$ , and  $\Delta$  is a solution of the transcendent equation

$$\frac{1}{1-\alpha} = \int_{-\Delta}^{\Delta} p_{\mathcal{N}(0,\sigma)}(x) dx + 2p_{\mathcal{N}(0,\sigma)}(\Delta) \frac{\sigma^2}{\Delta} \tag{60}$$

that is, the worth on  $\mathcal{P}_{\xi}^3$  distribution density is the Gaussian one for  $|x| \leq \Delta$  and the Laplace type for  $|x| > \Delta$ , (see Figure 3).

**Corollary 4.** The robust on  $\mathcal{P}_\xi^3$  version of the procedure (54) contains

$$\varphi^{**}(x) = -I_F^{-1}\left(p_\xi^*\right) \frac{d}{dv} \ln p_\xi^*(v) = \begin{cases} x & \text{for } |x| \leq \Delta \\ \Delta \operatorname{sign}(x) & \text{for } |x| > \Delta \end{cases} \quad (61)$$



**Figure 3.** The nonlinear transformation  $\varphi^{**}$  for the class  $\mathcal{P}_\xi^3$ .

**Lemma 5.** In the class

$$\mathcal{P}_\xi^4 = \left\{ p_\xi : p_\xi(x) = (1 - \alpha)p_{U(0,a)}(x) + \alpha q(x) \right\},$$

$$p_{U(0,a)}(x) := \frac{1}{2a} \chi(|x| \leq a)$$

(7) (of all symmetric “approximately uniform” distributions) the worth distribution density  $p_\xi^*(x)$  is

$$p_\xi^*(x) = \arg \inf_{p_\xi \in \mathcal{P}_\xi^4} I_{F,\xi}(p_\xi) = \begin{cases} \frac{1-\alpha}{2a} & \text{for } |x| \leq a \\ \frac{1-\alpha}{2a} \exp\left\{-\left(1-\alpha\right)\frac{|x|-a}{aa}\right\} & \text{for } |x| > a > 0 \end{cases} \quad (62)$$

that is, the worth on  $\mathcal{P}_\xi^4$  distribution density is the uniform p.d.f. for  $|x| \leq a$  and the Laplace type for  $|x| > a$ .

**Corollary 5.** The robust on  $\mathcal{P}_\xi^4$  version of the procedure (54) contains

$$\varphi^{**}(x) = -I_F^{-1}\left(p_\xi^*\right) \frac{d}{dv} \ln p_\xi^*(v) = \begin{cases} 0 & \text{for } |x| \leq a \\ \frac{1-\alpha}{\alpha a} \operatorname{sign}(x) & \text{for } |x| > a \end{cases} \quad (63)$$

See Figure 4.

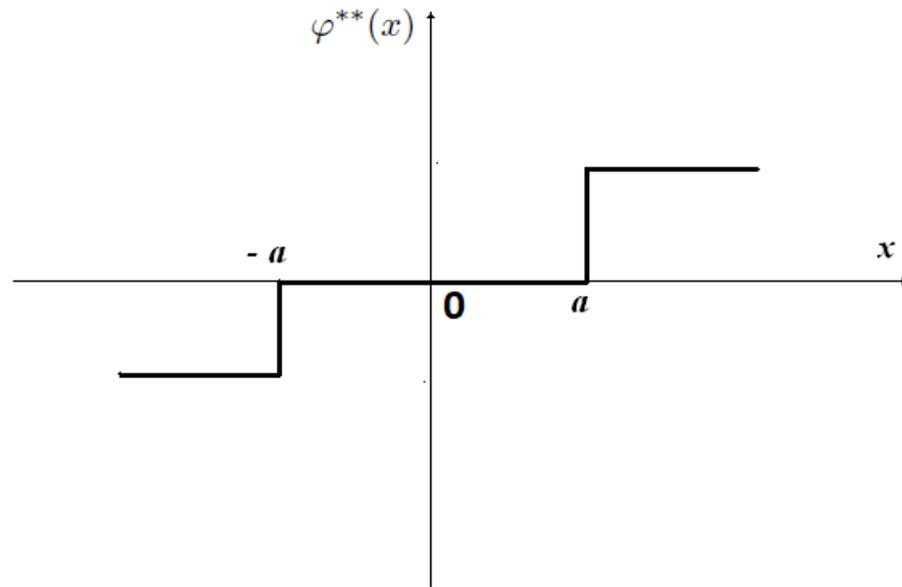


Figure 4. The nonlinear dead-zone transformation  $\varphi^{**}$  for the class  $\mathcal{P}_{\xi}^4$ .

6.4. Robust Identification of Dynamic ARX Models

In the case of dynamic autoregression models (ARX model) where the generalised inputs are dependent on the state of the system, the matrix  $\mathcal{R}$  depends on  $p_{\xi}$ , too, and therefore, we deal with the complete problem, namely, we need to calculate

$$\sup_{p_{\xi} \in \mathcal{P}} [I_F(p_{\xi})\mathcal{R}(p_{\xi})]^{-1} \tag{64}$$

For the ARX model (65) (for simplicity we put here  $b_k = 0, (k = 0, \dots, K)$ , so  $c_l = a_l (l = 1, \dots, L)$ ) the relation (36) becomes

$$\left. \begin{aligned} \tilde{y}_n &= a^T \tilde{v}_n + \xi_n + O_{\omega}(\lambda^n) \\ a^T &= (a_1, \dots, a_L), \tilde{v}_n^T = (y_{n-1}, \dots, y_{n-L}) \end{aligned} \right\} \tag{65}$$

Here we have

$$\frac{1}{n} \sum_{t=0}^n \mathbb{E}\{\tilde{v}_t \tilde{v}_t^T\} \rightarrow \mathcal{R}(p_{\xi})$$

where  $\mathcal{R}(p_{\xi})$  satisfies

$$\mathcal{R}(p_{\xi}) = A\mathcal{R}(p_{\xi})A^T + \sigma^2\Xi_0 \tag{66}$$

with

$$A = \left\| \begin{array}{cccccc} a_0 & a_1 & \dots & \dots & a_{L_a} & \\ 1 & 0 & \dots & \dots & 0 & \\ 0 & 1 & 0 & \dots & 0 & \\ 0 & \dots & \ddots & 0 & 0 & \\ 0 & \dots & 0 & 1 & 0 & \end{array} \right\|, \Xi_0 := \left\| \begin{array}{cccccc} 1 & 0 & \dots & \dots & 0 & \\ 0 & 0 & \dots & \dots & 0 & \\ 0 & 0 & 0 & \dots & 0 & \\ 0 & \dots & \ddots & 0 & 0 & \\ 0 & \dots & 0 & 0 & 0 & \end{array} \right\|$$

Obviously,  $\mathcal{R}(p_{\xi})$  can be represented as  $\mathcal{R}(p_{\xi}) = \sigma^2\mathcal{R}_0(p_{\xi})$ , where  $\mathcal{R}_0$  is the solution of

$$\mathcal{R}_0(p_{\xi}) = A\mathcal{R}_0(p_{\xi})A^T + \Xi_0. \tag{67}$$

In this case, the problem (64) is reduced to

$$\sup_{p_{\xi} \in \mathcal{P}_{\xi}} [\sigma^2(p_{\xi})I_F(p_{\xi})]^{-1}$$

or equivalently, to

$$\inf_{p_{\xi} \in \mathcal{P}_{\xi}} [\sigma^2(p_{\xi}) I_F(p_{\xi})] \tag{68}$$

Consider now some classes  $\mathcal{P}_{\xi}$  of a priori informative generating noise distributions and solutions of the problem (68) within these classes.

- (1) Class  $\mathcal{P}_{\xi}^{ARX-1}$  (containing among others the Gaussian distribution  $p_{\mathcal{N}(0, \sigma_0^2)}(x)$ ).

**Lemma 6.** For the class  $\mathcal{P}_{\xi}^{ARX-1}$

$$p_{\xi}^*(x) = \arg \inf_{p_{\xi} \in \mathcal{P}_{\xi}^{ARX-1}} [\sigma^2(p_{\xi}) I_F(p_{\xi})] = p_{\mathcal{N}(0, \sigma_0^2)}(x) \tag{69}$$

that is, the worth on  $\mathcal{P}_{\xi}^{ARX-1}$  p.d.f. is exactly that the Gaussian distribution  $p_{\mathcal{N}(0, \sigma_0^2)}(x)$ .

**Proof.** Taking in (A1)

$$f(x) = x, \quad \varphi(x) = p'_{\xi}(x)/p_{\xi}(x)$$

we get

$$\sigma^2 I_F(p_{\xi}) \geq \left( \int_{\mathbb{R}} x p'_{\xi}(x) dx \right)^2 = \left( \int_{\mathbb{R}} p_{\xi}(x) dx \right)^2 = 1$$

such that the equality is attained when  $p'_{\xi}(x)/p_{\xi}(x) = \lambda x$ , which leads to

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi/\lambda}} \exp\left\{-\frac{\lambda x^2}{2}\right\}$$

But since  $I_F(p_{\mathcal{N}(0, \sigma_0^2)}) = \sigma_0^{-2}$  from the inequality above we get

$$\sigma^2(p_{\xi}) I_F(p_{\xi}) \geq 1 = \sigma_0^2 I_F(p_{\mathcal{N}(0, \sigma_0^2)})$$

which means that  $p_{\xi}^*(x) = p_{\mathcal{N}(0, \sigma_0^2)}(x)$ .  $\square$

**Corollary 6.** The robust on  $\mathcal{P}_{\xi}^{ARX-1}$  version of the procedure (54) contains

$$\varphi^{**}(x) = -I_F^{-1}(p_{\xi}^*) \frac{d}{dv} \ln p_{\xi}^*(v) = x.$$

- (2) Class  $\mathcal{P}_{\xi}^{ARX-2}$  (containing all centred distributions with a variance not less than a given value):

$$\mathcal{P}_{\xi}^{ARX-2} = \left\{ p_{\xi} : \int_{\mathbb{R}} x^2 p_{\xi}(x) dx \geq \sigma_0^2 \right\} \tag{70}$$

**Lemma 7.** For the class  $\mathcal{P}_{\xi}^{ARX-2}$

$$p_{\xi}^*(x) = \arg \inf_{p_{\xi} \in \mathcal{P}_{\xi}^{ARX-2}} I_F(p_{\xi}) \tag{71}$$

that is, the worth on  $\mathcal{P}_{\xi}^{\text{ARX-2}}$  distribution density  $p_{\xi}^*(x)$  coincides with the worth p.d.f. on the classes  $\mathcal{P}_{\xi}^i$  ( $i = 1, \dots, 4$ ) characterising distribution uncertainties (if additional information is available) for static regression models provided that

$$\sigma^2(p_{\xi}^*(x)) = \sigma_0^2 \quad (72)$$

**Proof.** It follows directly from the inequality  $\sigma^2(p_{\xi})I_F(p_{\xi}) \geq \sigma_0^2 I_F(p_{\xi})$ .  $\square$

**Remark 8.** Notice that all of the preceding analysis is based on the assumption that the transfer function (4) of the forming filter (3) is known a priori, allowing the parallel whitening process (34) to be applied and the information Cramer–Rao bound (37) to be reached, resulting in the asymptotically effective (the “best” ones) procedure, which is robust on given p.d.f. classes  $\mathcal{P}_{\xi}$  of generating noises.

Below, we look at a considerably more challenging scenario where the forming filter (3) is unknown a priori. In this situation, nobody can definitely achieve the information Cramer–Rao bound (37) and build an asymptotically successful parametric estimate technique in this circumstance. However, the problem can be handled utilising alternative techniques of identification.

## 7. Instrumental Variables Method for ARMAX Model with Finite Noise Correlation

### 7.1. About IVM

Instrumental variables (IV) or total least-squares estimators is the method which also recommends to estimate parameters in the presence of white or coloured noises [89–91]. Even if the accuracy of the estimator for errors-in-variables models cannot be handled with a conventional analysis, the results produced by any of these estimators in practice demonstrate that their response can be well theoretically anticipated. The instrumental variables algorithms have been implemented for multivariable model forms, such as ARMAX models, dynamic adjustments with autoregressive errors and multivariable transfer functions (see [92]), where the IV algorithm provides asymptotically efficient estimation results and a low variance. The IV method can be adapted to work with the maximum likelihood method [93]. An analysis of the refined instrumental variable-approximate maximum likelihood (IVAML) method was presented. The proposed technique proved to be asymptotically efficient and to approach minimum variance estimation of the model parameters, even with a low sample size and low signal noise ratios. An unified refined instrumental variable (RIV) approach was proposed in [94] for the estimation of discrete and continuous-time transfer functions. The estimator was based on the formulation of a pseudo-linear regression involving an optimal prefiltering process derived from a Box–Jenkins transfer function model. This method showed a reliable solution to the maximum likelihood optimisation equations, and the estimates are optimal in the maximum likelihood sense. The optimal refined instrumental variables for Box–Jenkins models has been studied on various occasions, for example in [95]. There, in contrast to the most common forms of the algorithm used in ARMAX models, a modification that facilitates the representation of the more general noise component of the Box–Jenkins model was proposed, and that could also be used as an adaptive filter and as a state variable feedback control. For the nonlinear case, the instrumental variable method has been used in particular for nonlinear Hammerstein models. The nonlinear recursive instrumental variables method has been used to deal with these models due to its simplicity in practical applications (see [96]). The recursive IV method also proves to be superior to the recursive LSM in terms of accuracy and convergence under the presence of coloured noises, and this is valid either for discrete or continuous-time [97].

Consider now the system (1) in the regression format

$$\left. \begin{aligned} y_n &= z_n^T c + \eta_n, \\ c &= (a_1, \dots, a_L; b_0, \dots, b_K)^T \in R^{L+K+1}, \\ z_n &= (y_{n-1}, \dots, y_{n-L}; w_n, \dots, w_{n-K})^T \in R^{L+K+1} \end{aligned} \right\} \quad (73)$$

where the exogenous noise input  $\eta_n$  has a finite correlation, that is, the transfer matrix of the forming filter has only the nominator:

$$H(q^{-1}) = H_1(q^{-1}) = 1 + \sum_{s=1}^{K_1} d_{2,s} q^{-s}, \quad H_2(q^{-1}) = 1 \quad (74)$$

It is acknowledged that the parameters  $d_{2,s}$  ( $s = 1, \dots, K_1$ ) may be unknown a priori.

### 7.2. Instrumental Variables and the System of Normal Equations

Let  $v_n \in R^{L+K+1}$  be an auxiliary vector variable (an instrumental variable) depending on information available up to time  $n$ . Considering the moments  $t = 1, \dots, n$  and multiplying both sides of (73) by  $v_t$  we get the so-called system of "normal equations":

$$\left. \begin{aligned} v_1 y_1 &= v_1 z_1^T c + v_1 \eta_1, \\ v_2 y_2 &= v_2 z_2^T c + v_2 \eta_2, \\ &\vdots \\ v_n y_n &= v_n z_n^T c + v_n \eta_n \end{aligned} \right\} \quad (75)$$

Summing these relations, after multiplying by  $n^{-1}$ , we obtain

$$n^{-1} \sum_{t=1}^n v_t y_t = \left( n^{-1} \sum_{t=1}^n v_t z_t^T \right) c + n^{-1} \sum_{t=1}^n v_t \eta_t \quad (76)$$

Define the instrumental variable estimate  $c_n^{IV}$  of the vector  $c$  as a vector which in each time  $n$  satisfies the relation

$$\sum_{t=1}^n v_t y_t = \left( \sum_{t=1}^n v_t z_t^T \right) c_n^{IV}. \quad (77)$$

If the matrix  $\left( \sum_{t=1}^n v_t z_t^T \right)$  is invertible, that is, the matrix  $\Gamma_n^{IV} := \left( \sum_{t=1}^n v_t z_t^T \right)^{-1}$  exists (for all  $n \geq n_0$ ), then  $c_n^{IV}$  can be expressed as

$$c_n^{IV} = \Gamma_n^{IV} \sum_{t=1}^n v_t y_t, \quad (78)$$

or in the recurrent form

$$\left. \begin{aligned} c_n^{IV} &= c_{n-1}^{IV} + \Gamma_n^{IV} v_n (y_n - z_n^T c_{n-1}^{IV}), \\ \Gamma_n^{IV} &= \Gamma_{n-1}^{IV} - \frac{\Gamma_{n-1}^{IV} v_n z_n^T \Gamma_{n-1}^{IV}}{1 + z_n^T \Gamma_{n-1}^{IV} v_n}, \quad z_n^T \Gamma_{n-1}^{IV} v_n \neq -1 \end{aligned} \right\} \quad (79)$$

**Remark 9.** Notice that if  $v_n = z_n$  the estimates  $c_n^{IV}$  (77)–(79) coincide with LSM estimates.

As it follows from (76) and (77), the estimation error  $\delta_n = c_n^{IV} - c$  satisfies

$$-n^{-1} \sum_{t=1}^n v_t \eta_t = \left( n^{-1} \sum_{t=1}^n v_t z_t^T \right) \delta_n \quad (80)$$

Under the main assumptions accepted above, in view of the strong large number law (see Theorem 8.10 in [36]), we have

$$\begin{aligned} n^{-1} \sum_{t=1}^n v_t y_t &\stackrel{a.s.}{\rightarrow} n^{-1} \sum_{t=1}^n E\{v_t y_t\} + o_\omega(1), \\ n^{-1} \sum_{t=1}^n v_t z_t &\stackrel{a.s.}{\rightarrow} n^{-1} \sum_{t=1}^n E\{v_t z_t\} + o_\omega(1), \\ n^{-1} \sum_{t=1}^n v_t \eta_t &\stackrel{a.s.}{\rightarrow} n^{-1} \sum_{t=1}^n E\{v_t \eta_t\} + o_\omega(1) \end{aligned}$$

*a.s.* means almost sure or with probability 1.

and the relation (80) becomes

$$-n^{-1} \sum_{t=1}^n E\{v_t \eta_t\} \stackrel{a.s.}{\rightarrow} \left( n^{-1} \sum_{t=1}^n E\{v_t z_t\} \right) \delta_n + o_\omega(1)$$

from which one may conclude that if

(1)

$$n^{-1} \sum_{t=1}^n E\{v_t z_t\} \rightarrow \mathcal{R}_{IV}, \det \mathcal{R}_{IV} \neq 0; \tag{81}$$

(2)

$$n^{-1} \sum_{t=1}^n E\{v_t \eta_t\} \rightarrow 0, \tag{82}$$

then the estimate  $c_n^{IV}$  is asymptotically consistent with probability 1, namely,  $\delta_n \xrightarrow{a.s.} 0$ .

**Corollary 7.** Evidently, the condition (82) holds if the instrumental variable  $v_t$  and the external noise  $\eta_t$  are not correlated:

$$E\{v_t \eta_t\} = 0 \text{ for all } t = 1, \dots$$

So, in the example (17), instead of the LSM estimate (18) we need to use (see [98]) the IV estimate  $c_n^{IV}$  (78) with  $v_t = y_{t-k}$  ( $k \geq 1$ ):

$$a_n = \left[ \sum_{t=1}^n y_{t-k} y_{t+1} \right] \left[ \sum_{t=1}^n y_t y_{t-k} \right]^{-1} \xrightarrow{a.s.} a$$

In general cases for the model (73) and (74) with a finite correlation ( $E\{\eta_t \eta_{t-k}\} = 0, k > K_1$ ) we may use the following IV estimate  $c_n^{IV}$  with  $v_t = z_{t-k}$  ( $k \geq K_1$ ):

$$\left. \begin{aligned} c_n^{IV} &= c_{n-1}^{IV} + \Gamma_n^{IV} z_{t-k} (y_n - z_n^\top c_{n-1}^{IV}), \\ \Gamma_n^{IV} &= \Gamma_{n-1}^{IV} - \frac{\Gamma_{n-1}^{IV} z_{n-k} z_n^\top \Gamma_{n-1}^{IV}}{1 + z_n^\top \Gamma_{n-1}^{IV} z_{n-k}}, \quad z_n^\top \Gamma_{n-1}^{IV} z_{n-k} \neq -1. \end{aligned} \right\} \tag{83}$$

### 8. Joint Parametric Identification of ARMAX Model and the Forming Filter

Unfortunately, IVM identification algorithms cannot be applied in the situation when the correlation function of a coloured noise is not finite. Below, we treat exactly this case considering that the transfer function of a finite-dimensional forming filter is completely unknown, including both numerator and denominator parameters in (4). So, here our problem under the consideration is as follows: based on the available data (16) we need to construct an identification procedure, generating some parameter estimates  $\hat{a}_{n,i}$  ( $i = 1, \dots, L$ ),  $\hat{b}_{n,i}$  ( $i = 0, \dots, K$ ),  $\hat{h}_{n,1i}$  ( $i = 0, \dots, K_1$ ) and  $\hat{h}_{n,2i}$  ( $i = 1, \dots, K_2$ ) which asymptotically convergence with probability 1 (or almost sure) to the real values, namely,

$$\hat{a}_{n,i} \xrightarrow{a.s.} a_i, \hat{b}_{n,i} \xrightarrow{a.s.} b_i, \hat{h}_{n,1i} \xrightarrow{a.s.} h_{1,i}, \hat{h}_{n,2i} \xrightarrow{a.s.} h_{2,i} \text{ when } n \rightarrow \infty. \tag{84}$$

8.1. An Equivalent ARMAX Representation

Multiplying (1) by  $H_2(q) = 1 + \sum_{i=1}^{K_2} h_{2,i}q^i$  we obtain the corresponding ARMAX (autoregression with moving average noise term model)

$$\left(1 + \sum_{i=1}^{K_2} h_{2,i}q^i\right)\left(1 + \sum_{i=1}^L a_iq^i\right)y_n = \left(1 + \sum_{i=1}^{K_2} h_{2,i}q^i\right)\left(\sum_{i=0}^K b_iq^i\right)u_n + \left(h_{1,0} + \sum_{i=1}^{K_1} h_{1,i}q^i\right)\xi_n$$

or, equivalently, in the “open format” (with  $m_A := \max\{K_2, L\}$  and  $m_B := \max\{K_2, K\}$ )

$$\left. \begin{aligned} &\left(1 + \sum_{i=1}^{m_A} q^i [a_i\chi(i \leq n_A) + h_{2,i}\chi(i \leq n_{D_2})] + \left(h_{2,i}\chi(i \leq n_{D_2}) \sum_{j=1}^{m_A} a_j\chi(j \leq n_A)q^j\right)\right)y_n = \\ &\left(b_0 + \sum_{i=1}^{m_B} q^i \left[b_i\chi(i \leq M_B) + h_{2,i}\chi(i \leq n_{D_2}) \sum_{j=0}^{m_B} b_j\chi(j \leq n_B)q^j\right]\right)u_n \\ &\quad + \left(1 + \sum_{i=1}^{n_{D_1}} h_{1,i}q^i\right)\xi_n, \end{aligned} \right\} \tag{85}$$

where

$$\chi(\mathcal{A}) = \begin{cases} 1 & \text{if the event } \mathcal{A} \text{ is valid} \\ 0 & \text{if not} \end{cases}$$

**Remark 10.** Notice that since the polynomial  $H_2(q)$  is stable, the reactions  $\{y_\tau\}_{\tau=1, n}$  of both difference Equations (4) and (85) on the same inputs  $\{u_\tau\}_{\tau=-2m_B, n}$  and  $\{\xi_\tau\}_{\tau=-K_1, n}$  are asymptotically closed, namely, the difference between these reactions tends to zero exponentially quickly with probability one. That is why to obtain the desired property (84), designing the identification procedure using the data of the model (4) can be realised based on data but generated by the ARMAX model (85) (see [99]).

The ARMAX model (85) can be represented in the standard regression format (different from (14)) as

$$y_n = x_n^T c + h_{1,0}\xi_n \tag{86}$$

with

$$x_n = \left(-y_{n-1}, \dots, -y_{n-2m_A}; u_n, \dots, u_{n-2m_B}; \xi_{n-1}, \dots, \xi_{n-n_{K_1}}\right)^T \in R^N \tag{87}$$

$$N := 2m_A + 2m_B + 1 + K_1$$

and

$$c = (\tilde{a}_1, \dots, \tilde{a}_{2m_A}, \tilde{b}_0, \dots, \tilde{b}_{2m_B}, h_{1,1}, \dots, h_{1,K_1})^T \in R^N, \tag{88}$$

containing the components

$$\left. \begin{aligned} &\tilde{a}_1 = a_1 + h_{2,1}, \\ &\tilde{a}_i = a_i\chi(i \leq n_L) + h_{2,i}\chi(i \leq n_{K_2}) + \sum_{k=1}^{m_A} h_{2,k}\chi(k \leq n_{K_2})a_{i-k}\chi(i - k \leq L), \quad i = 2, \dots, 2m_A, \\ &\tilde{b}_0 = b_0, \quad \tilde{b}_i = b_i\chi(i \leq m_B) + \sum_{k=1}^{m_B} h_{2,k}\chi(k \leq n_{K_2})b_{i-k}\chi(i - k \leq K), \quad i = 1, \dots, 2m_B. \end{aligned} \right\} \tag{89}$$

**Remark 11.** Notice that the extended input vector  $x_n$  is not completely available since it contains immeasurable components  $\xi_{n-1}, \dots, \xi_{n-n_{K_1}}$ . This property is the main difference with the standard ARMAX model identification problem where the vector  $x_n$  does not contain these immeasurable term.

### 8.2. Auxiliary Residual Sequence

Now, let us define the “generalised residual sequence” given by the recursion relation

$$\varepsilon_n = y_n - \hat{x}_n^\top c_{n-1} \tag{90}$$

where the “extended vector”  $\hat{x}_n \in R^{2m_A+2m_B+1+K_1}$  is defined as

$$\hat{x}_n = \left( -y_{n-1}, \dots, -y_{n-2m_A}; u_{n-1}, \dots, u_{n-2m_B}; \varepsilon_{n-1}, \dots, \varepsilon_{n-n_{K_1}} \right)^\top \tag{91}$$

with  $\varepsilon_{-1} = \dots = \varepsilon_{-n_{D_1}} = 0$ . Notice that the “extended vector”  $\hat{x}_n$  is measurable on-line.

**Lemma 8 ([99]).** For  $n \rightarrow \infty$

$$\Delta_n = \varepsilon_n - \zeta_n = O\left(|\lambda_{H_1}|^n\right) \xrightarrow{a.s.} 0, \tag{92}$$

where  $\lambda_{H_1}$  is the eigenvalue of the polynomial  $H_1$  with minimal module  $|\lambda_{H_1}| < 1$ .

From (A5) we get

$$y_n = \hat{x}_n^\top c - \sum_{i=1}^{n_{D_1}} h_{1,i} \Delta_{n-i} + h_{1,0} \zeta_n = \hat{x}_n^\top c + h_{1,0} \zeta_n + O\left(|\lambda_{H_1}|^n\right). \tag{93}$$

### 8.3. Identification Procedure

To estimate the extended vector  $c$  from the relation (93) let us apply the least-squares method (LSM), defining the current estimate  $\hat{c}_n$  as

$$\hat{c}_n = \left( \sum_{t=0}^n \hat{x}_t \hat{x}_t^\top \right)^{-1} \sum_{t=0}^n \hat{x}_t y_t, \quad n \geq n_0 = \left\{ \inf n: \sum_{t=t}^n \hat{x}_t \hat{x}_t^\top > 0 \right\} \tag{94}$$

In the recurrent form, this estimate can be represented as in (39):

$$\left. \begin{aligned} \hat{c}_n &= \hat{c}_{n-1} + \Gamma_n \hat{x}_n \varphi(y_n - \hat{x}_n^\top \hat{c}_{n-1}), \\ \Gamma_n &= \Gamma_{n-1} - \frac{\Gamma_{n-1} \hat{x}_n \hat{x}_n^\top \Gamma_{n-1}}{1 + \hat{x}_n^\top \Gamma_{n-1} \hat{x}_n}, \quad n \geq n_0 + 1, \\ \Gamma_{n_0}^{-1} &:= \sum_{t=n_0}^{n_0} \hat{x}_t \hat{x}_t^\top \end{aligned} \right\} \tag{95}$$

Notice that taking  $\Gamma_{n_0}^{-1}$  as

$$\Gamma_{n_0}^{-1} = \rho I_{N \times N} + \sum_{t=t}^{n_0} \hat{x}_t \hat{x}_t^\top, \quad 0 < \rho \ll 1, \tag{96}$$

we can select  $n_0 = 0$ , and the procedure (95) can be applied from the beginning of the process.

**Theorem 6 ([100]).** If

(1) the following “persistent excitation condition” (PEC) holds:

$$\liminf_n \left( \frac{1}{n} \sum_{t=0}^n x_t x_t^\top \right) \stackrel{a.s.}{\geq} \nu I_{M \times M} \stackrel{a.s.}{>} 0, \tag{97}$$

$$M := 2m_A + 2m_B + 1 + D_1.$$

(2)  $\{\zeta_n\}$  is a martingale difference sequence satisfying (10),

then, the LSM procedure (95) and (96) generates the sequence of the estimates  $\{\hat{c}_n\}_{n \geq 0}$ , which is asymptotically consistent with probability 1, that is,  $\hat{c}_n \xrightarrow[n \rightarrow \infty]{a.s.} c$ .

8.4. Recuperation of the Model Parameters from the Obtained Current Estimates

8.4.1. Special Case When the Recuperation Process Can Be Realised Directly

When  $K = 0$  and the gain parameter  $b_0 \neq 0$  are a priori known, the system of algebraic Equations (89) becomes linear with respect to the unknown parameters  $a_i (i = 0, \dots, L)$  and  $h_{2,i} (i = 0, \dots, n_{K_2})$ , and may be resolved analytically without application of any numerical procedure.

8.4.2. General Case Requiring the Application of Gradient Descent Method (GDM)

In view of (89), we can recuperate the parameters  $a_i (i = \overline{0, L})$ ,  $b_i (i = \overline{0, K})$  and  $h_{2,i} (i = 0, \dots, n_{K_2})$  for this purpose using the command Fsolve in Matlab or some numerical method such as GDM.

For example, if we consider the case when  $K = L = n_{K_2} = m_A = m_B = 2$ , the component relations from (89) become

$$\begin{aligned} \tilde{a}_1 &= a_1 + h_{2,1}, \quad \tilde{a}_2 = a_2 + h_{2,2} + h_{2,1}a_1, \\ \tilde{a}_3 &= h_{2,1}a_2 + h_{2,2}a_1, \quad \tilde{a}_4 = h_{2,2}a_2, \\ \tilde{b}_0 &= b_0, \quad \tilde{b}_1 = b_1 + h_{2,1}b_0, \\ \tilde{b}_2 &= b_2 + h_{2,1}b_1 + h_{2,2}b_0, \\ \tilde{b}_3 &= h_{2,1}b_2 + h_{2,2}b_1, \quad \tilde{b}_4 = h_{2,2}b_2. \end{aligned}$$

Since this system is formed by nonlinear equations, and in some particular cases it is actually possible to solve the equations analytically, the gradient descent method (GDM) is implemented to estimate the values from the original system, taking the best average value from the estimated parameters. For this purpose, we define the following objective function:

$$\begin{aligned} F(a_1, a_2, b_1, h_{21}) &= (a_1 + h_{21} - c_1)^2 + \\ & (a_1 h_{21} + a_2 - c_2)^2 + (a_2 h_{21} - c_3)^2 + \\ & (b_1 + h_{21} b_0 - c_5)^2 + (b_1 h_{21} - c_6)^2 \rightarrow \min \end{aligned}$$

The original parameter can be recovered using some of the existing optimisation commands in Matlab, such as Fsolve or optimvar, or some algorithms such as GDM mentioned previously, although some other optimisation techniques could be implemented (see [99]). The performance of Fsolve is good in second or third order systems; in these cases, the command can recover all the original parameters from the nonlinear system. In higher order systems, this method presents problems at recovering the original values, while gradient descent has a good performance with low- and high-order systems. In some cases, such as the example presented before, it is possible to recover the original values by a mathematical simplification. The main condition for a good estimation is that in the objective function one should have at least as many terms as variables to estimate, otherwise it is not possible to recover all the original values.

9. Numerical Example

The algorithms presented in the previous sections are illustrated with a numerical example.

Raised Cosine Distribution

Consider the following system

$$\begin{aligned} y(k) &= 0.85y(k-1) + 2u(k) + \eta(k), \\ \eta(k) &= -0.3\eta(k-1) + \zeta(k) + 0.8\zeta(k-1), \end{aligned}$$

with  $\zeta$  having the raised cosine distribution

$$p_\zeta(v) = \frac{1}{2s} \left[ 1 + \cos\left(\frac{v-\mu}{s} \pi\right) \right], \quad \mu > 0, \quad s > 0,$$

which is a continuously differentiable function supported on the interval  $[\mu - s, \mu + s]$ . The system can be rewritten as follows

$$y(k) = z(k)^\top c + \eta(k),$$

with

$$z(k) = \begin{pmatrix} y(k-1) \\ u(k) \end{pmatrix}, c := \begin{pmatrix} 0.85 \\ 2 \end{pmatrix}.$$

The whitening process is then given by

$$\tilde{y}(k) = H(q^{-1})y(k), \tilde{z}(k) = H(q^{-1})z(k),$$

or in the extended form,

$$\begin{aligned} \tilde{y}(k) + 0.3\tilde{y}(k-1) &= y(k) + 0.8y(k-1), \tilde{y}(0) = y(0), \\ \tilde{z}(k) + 0.3\tilde{z}(k-1) &= z(k) + 0.8z(k-1), \tilde{z}(0) = z(0), \end{aligned}$$

where the “inverse filter” has the transfer function

$$H(q^{-1}) = \frac{1 + 0.8q^{-1}}{1 + 0.3q^{-1}}.$$

The recursive WLSM algorithm with the residual nonlinear transformation is given by

$$\begin{aligned} c_n &= c_{n-1} - I_{F,\xi}^{-1} \Gamma_n \tilde{z}_n \frac{p'_\xi(v)}{p_\xi(v)} \Big|_{v=\tilde{y}_n - \tilde{z}_n^\top c_{n-1}} = \\ c_{n-1} + 2\pi I_{F,\xi}^{-1} \Gamma_n \tilde{z}_n &\frac{\sin\left(\frac{v-\mu}{s}\pi\right)}{1 + \cos\left(\frac{v-\mu}{s}\pi\right)} \Big|_{v=\tilde{y}_n - \tilde{z}_n^\top c_{n-1}} = \\ c_{n-1} + \frac{s}{\pi} \Gamma_n \tilde{z}_n &\frac{\sin\left(\frac{\pi}{s} [\tilde{y}_n - \tilde{z}_n^\top c_{n-1} - \mu]\right)}{1 + \cos\left(\frac{\pi}{s} [\tilde{y}_n - \tilde{z}_n^\top c_{n-1} - \mu]\right)}. \end{aligned} \tag{97}$$

Here, we have used that for the raised cosine distribution

$$I_{F,\xi} = 2 \frac{\pi^2}{s}. \tag{98}$$

The initial conditions are  $c(0) = 2, y(0) = 3, \Gamma(0) = 10^5$ . The Figures 5 and 6 show the estimated parameters  $a$  and  $b$  using LSM and MLLM+ whitening with a nonlinear residual transformation.

In the Figures 5 and 6, one can see that in the LSM case the noise has a strong influence in the estimation results, while in the MLLM+ whitening, the noise influence is minimised in the estimated parameter, reducing the bias, which is the most common problem in parameter estimation using LSM under the presence of the correlated noises. The performance index of the estimated algorithm is illustrated in Figure 7; here, one can see that the MLLM+ whitening is a better option for parameter estimation in systems with coloured noises.

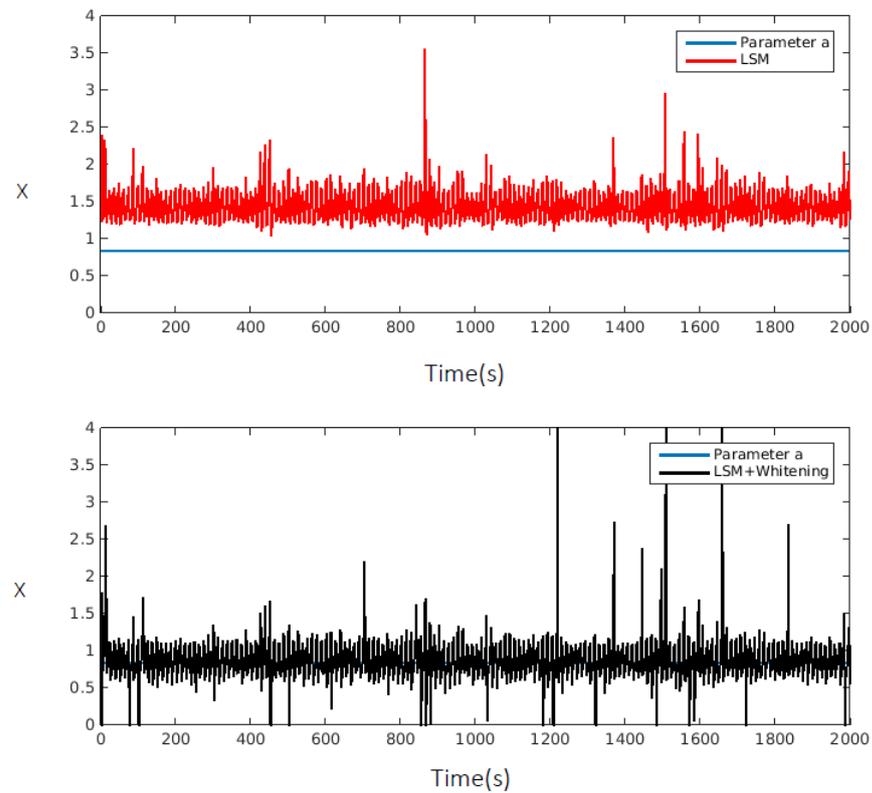


Figure 5. Parameter  $a$  and its estimated using LSM and LSM+ whitening (raised cosine distribution case).

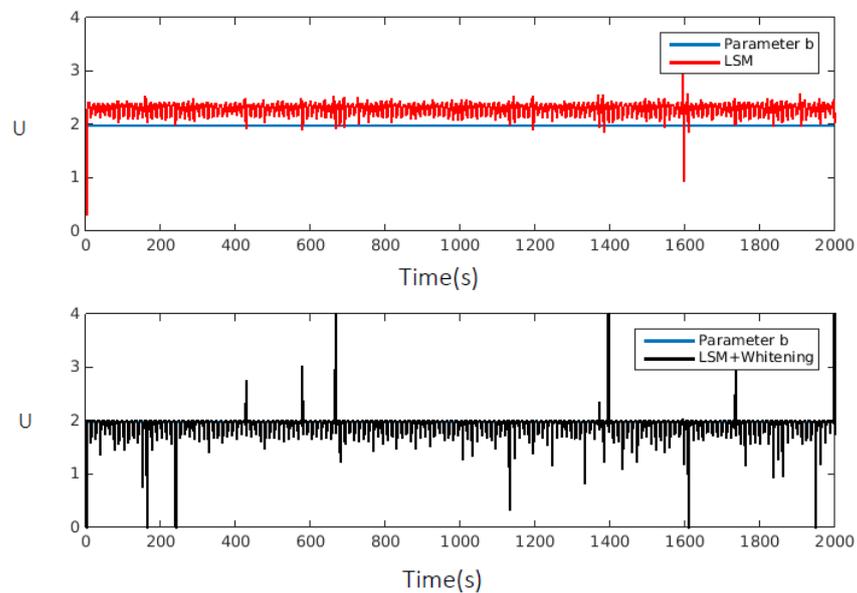
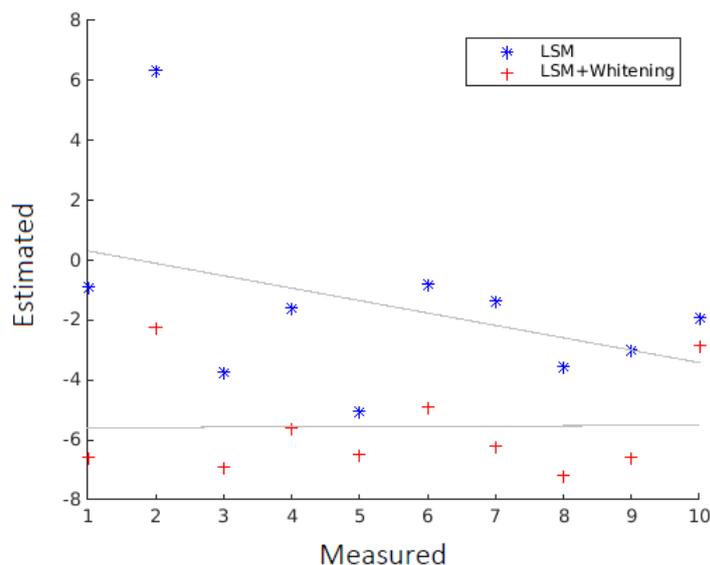


Figure 6. Parameter  $b$  and its estimated using LSM and LSM+ whitening (raised cosine distribution case).

In this case, the filter structure is known; a numerical example where the filter structure is unknown is presented in [99].



**Figure 7.** Performance indexes of the estimation algorithms implemented in a system with a raised cosine distribution.

## 10. Discussion

In this paper, we demonstrated that the traditional LSM algorithm failed to accurately estimate the parameters of ARX (dynamic) models when subjected to a coloured perturbation, and because of this, it is necessary to implement a different estimation strategy. For the identification issue under non-Gaussian and coloured noises, the Cramer–Rao inequality and the related Fisher information limits were explored, when the forming filter (the noise spectral function) is known a priori.

It was shown that a recurrent process, which employs both the whitening technique and the nonlinear residual transformation (operating in parallel), is the asymptotically effective (the “best”) identification algorithm.

The main limitation of the proposed method is that in the case of having a partially unknown filter, the method cannot be implemented. In this case, there are two different identification methods that might be used:

- Instrumental variables method (IVM) for ARMAX models with a finite noise-correlation.
- The nonlinear residual transformation method for simultaneous parametric identification of the ARMAX model and the forming filter.

Both techniques are not asymptotically effective, as they do not achieve the Cramer–Rao information limits.

In a future work, we plan to analyse the case in which the filter is partially known, or even unknown, and if it is possible to achieve the information limits that were previously mentioned.

## 11. Conclusions

In the present work, the limits for the Cramer–Rao inequality and the related Fisher information were explored under coloured noise perturbations, and we demonstrated that the whitening technique and the nonlinear residual transformation working in parallel generate an estimation sequence with the asymptotic convergence rate that proves to be the best identification algorithm for the case studied in this manuscript, reaching the Fisher information bound, which cannot be improved by any other estimation algorithm. The effectiveness of the suggested approach is illustrated by a numerical example with a non-Gaussian noise, having a raised cosine distribution.

**Author Contributions:** A.P. and J.E. conceived of the present idea. A.P. developed the theory and J.E. performed the computations and verified the analytical methods. All the authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Abbreviations**

ARX	Autoregressive model with exogenous variables
ARMAX	Autoregression moving average exogenous input
NARMAX	Nonlinear autoregression moving average exogenous input
LSM	Least-squares method
IVM	Instrumental variables method
LNL	Large number law
NARMAX	Nonlinear autoregressive moving average with
CRB	Cramer–Rao bound
FIM	Fisher information
MVU	Minimum variance unbiased
MLMW	Maximum likelihood method with whitening
DWM	Direct whitening method
WECC	Western Electricity Coordinating Council
KARMAX	Autoregressive moving average explanatory input model of the Koyck kind
IVAML	Instrumental variable approximate maximum likelihood
RIV	Refinen instrumental variables
GDM	Gradient descent method

**Appendix A**

*Appendix A.1. Proof of Lemma 2*

- By the Cauchy–Schwarz inequality

$$\left( \int_{\mathbb{R}^1} f \varphi p_{\xi} dx \right)^2 \leq \left( \int_{\mathbb{R}^1} f^2 p_{\xi} dx \right) \left( \int_{\mathbb{R}^1} \varphi^2 p_{\xi} dx \right) \tag{A1}$$

valid for any p.d.f.  $f, \varphi$ , and any noise density distribution  $p_{\xi}$  (for which the integrals have a sense), for  $f := p'_{\xi}(x)/p_{\xi}(x)$ , after integrating by parts it follows

$$I_{F,\xi}(p_{\xi}) \geq \left( \int_{\mathbb{R}^1} p_{\xi}(x) d\varphi(x) \right)^2 / \int_{\mathbb{R}^1} \varphi^2(x) p_{\xi}(x) dx, \tag{A2}$$

where the equality is attained when  $p'_{\xi}(x)/p_{\xi}(x) = \lambda \varphi(x)$ ,  $\lambda$  is any constant. Taking  $\varphi(x) := \text{sign}(x)$  in (A2) and using the identity  $[\text{sign}(x)]' = 2\delta(x)p_{\xi}(0)$  leads to

$$I_{F,\xi}(p_{\xi}) \geq 4p_{\xi}^2(0) \geq \frac{1}{a^2} \text{ for any } p_{\xi} \in \mathcal{P}_1, \tag{A3}$$

where the equality is attained when  $p'_{\xi}(x)/p_{\xi}(x) = \lambda \text{sign}(x)$ , or equivalently, for  $p_{\xi}(x) = \frac{\lambda}{2} \exp\{-|x|/\lambda\}$ . With  $\lambda = a$  we have

$$p_{\xi}(x) = \frac{a}{2} \exp\{-|x|/a\} = p_{\xi}^*(x). \tag{A4}$$

So,  $I_{F,\xi}(p_{\xi}^*) = \frac{1}{a^2}$  and the worst noise distribution within  $\mathcal{P}_{\xi}^1$  is  $p_{\xi}^*(x)$  (55).

Appendix A.2. Proof of Lemma 3

From (4), we have

$$\left. \begin{aligned} y_n &= x_n^T c + h_{1,0} \xi_n = \hat{x}_n^T c + c^T (x_n - \hat{x}_n) + h_{1,0} \xi_n \\ &= \hat{x}_n^T c + \sum_{i=1}^{n_{K_1}} h_{1,i} (\xi_{n-i} - \varepsilon_{n-i}) + h_{1,0} \xi_n, \end{aligned} \right\} \tag{A5}$$

which implies the following recurrence

$$h_{2,0} \Delta_n + \sum_{i=1}^{n_{K_1}} h_{1,i} \Delta_{n-i} = H_1(q) \Delta_n = 0. \tag{A6}$$

Taking into account that the polynomial  $H_1(q)$  is stable, we get (92).

Appendix A.3. Proof of Lemma 4

- Taking in (A2)  $\varphi(x) = x$  for all  $p_{\xi} \in \mathcal{P}_{\xi}^2$ , we get

$$I_{F,\xi}(p_{\xi}) \geq 1 / \int_{\mathbb{R}} x^2 p_{\xi}(x) dx \geq 1/\sigma^2, \tag{A7}$$

where the equality is attained when

$$p'_{\xi}(x)/p_{\xi}(x) = \lambda x, \lambda \text{ is any constant} \tag{A8}$$

or, equivalently, for

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi/\lambda}} \exp\left\{-\frac{\lambda x^2}{2}\right\} \tag{A9}$$

For  $\lambda = \sigma^{-2}$  we have

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} = p_{\xi}^*(x) \tag{A10}$$

implying

$$I_{F,\xi}(p_{\xi}) \geq 1 / \int_{\mathbb{R}^1} x^2 p_{\xi}(x) dx \geq 1/\sigma^2 = I_F(p_{\xi}^*) \tag{A11}$$

So, the worst noise distribution within  $\mathcal{P}_{\xi}^2$  is  $p_{\xi}^*(x)$ .

Appendix A.4. Proof of Lemma 5

- (without details). From (7) it follows

$$p_{\xi}(x) \geq (1 - \alpha) p_{\mathcal{N}(0,\sigma^2)}(x) \tag{A12}$$

So, we need to solve the following variational problem:

$$\inf_{p_{\xi}: p_{\xi} \geq (1-\alpha) p_{\mathcal{N}(0,\sigma^2)}} I_{F,\xi}(p_{\xi}) \tag{A13}$$

As it is shown in [14], its solution is (59).

## References

1. Bender, E. *An Introduction to Mathematical Modeling*; Dover Publications, Inc.: Mineola, NY, USA, 2012.
2. Hugues, G.; Liuping, W. *Identification of Continuous-Time Models from Sampled Data*; Springer: Berlin/Heidelberg, Germany, 2008.
3. Åström, K.J.; Eykhoff, P. System identification—A survey. *Automatica* **1971**, *7*, 123–162. [[CrossRef](#)]
4. Bekey, G.A. System Identification—An Introduction and a Survey. *Simulation* **1970**, *15*, 151–166. [[CrossRef](#)]
5. Ljung, L.; Gunnarsson, S. Adaptation and tracking in system identification—A survey. *Automatica* **1990**, *26*, 7–21. [[CrossRef](#)]
6. Billings, S.A. Identification of Nonlinear Systems—A Survey. In *Proceedings of the IEE Proceedings D-Control Theory and Applications*; IET: London, UK, 1980; Volume 127, pp. 272–285.
7. Ljung, L. Perspectives on system identification. *Annu. Rev. Control* **2010**, *34*, 1–12. [[CrossRef](#)]
8. Ljung, L.; Chen, T.; Mu, B. A shift in paradigm for system identification. *Int. J. Control* **2020**, *93*, 173–180. [[CrossRef](#)]
9. Tudor, C. *Procesos Estocásticos*; Sociedad Mexicana de Matemáticas: Ciudad de Mexico, Mexico, 1994.
10. Sobczyk, K. *Stochastic Differential Equations: With Applications to Physics and Engineering*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2013.
11. Feng, D.; Liu, P.X.; Liu, G. Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises. *Signal Process.* **2009**, *89*, 1883–1890.
12. Vo, B.-N.; Antonio Cantoni, K.L.T. *Filter Design with Time Domain Mask Constraints: Theory and Applications*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2013; Volume 56.
13. Huber, P. *Robustness and Designs: In “A Survey of Statistical Design and Linear Models”*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1975.
14. Tsytkin, Y.; Polyak, B. Robust likelihood method. *Dyn. Syst. Math. Methods Oscil. Theory Gor’kii State Univ.* **1977**, *12*, 22–46. (In Russian)
15. Poznyak, A.S. Robust identification under correlated and non-Gaussian noises: WMLLM procedure. *Autom. Remote Control* **2019**, *80*, 1628–1644. [[CrossRef](#)]
16. Mokhlis, S.E.; Sadki, S.; Bensassi, B. System identification of a dc servo motor using arx and armax models. In Proceedings of the 2019 International Conference on Systems of Collaboration Big Data, Internet of Things & Security (SysCoBloTS), Granada, Spain, 22–25 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.
17. AldemİR, A.; Hapođlu, H. Comparison of ARMAX Model Identification Results Based on Least Squares Method. *Int. J. Mod. Trends Eng. Res.* **2015**, *2*, 27–35.
18. Likothanassis, S.; Demiris, E. Armax model identification with unknown process order and time-varying parameters. In *Signal Analysis and Prediction*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 175–184.
19. Norton, J. Identification of parameter bounds for ARMAX models from records with bounded noise. *Int. J. Control* **1987**, *45*, 375–390. [[CrossRef](#)]
20. Stoffer, D.S. Estimation and identification of space-time ARMAX models in the presence of missing data. *J. Am. Stat. Assoc.* **1986**, *81*, 762–772. [[CrossRef](#)]
21. Mei, L.; Li, H.; Zhou, Y.; Wang, W.; Xing, F. Substructural damage detection in shear structures via ARMAX model and optimal subpattern assignment distance. *Eng. Struct.* **2019**, *191*, 625–639. [[CrossRef](#)]
22. Ferkl, L.; Široký, J. Ceiling radiant cooling: Comparison of ARMAX and subspace identification modelling methods. *Build. Environ.* **2010**, *45*, 205–212. [[CrossRef](#)]
23. Rahmat, M.; Salim, S.N.S.; Sunar, N.; Faudzi, A.M.; Ismail, Z.H.; Huda, K. Identification and non-linear control strategy for industrial pneumatic actuator. *Int. J. Phys. Sci.* **2012**, *7*, 2565–2579. [[CrossRef](#)]
24. González, J.P.; San Roque, A.M.S.M.; Perez, E.A. Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. *IEEE Trans. Power Syst.* **2017**, *33*, 545–556. [[CrossRef](#)]
25. Wu, S.; Sun, J.Q. A physics-based linear parametric model of room temperature in office buildings. *Build. Environ.* **2012**, *50*, 1–9. [[CrossRef](#)]
26. Jing, S. Identification of an ARMAX model based on a momentum-accelerated multi-error stochastic information gradient algorithm. In Proceedings of the 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS), Suzhou, China, 14–16 May 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1274–1278.
27. Le, Y.; Hui, G. Optimal Estimation for ARMAX Processes with Noisy Output. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 5048–5051.
28. Correa Martinez, J.; Poznyak, A. Switching Structure Robust State and Parameter Estimator for MIMO Nonlinear Systems. *Int. J. Control* **2001**, *74*, 175–189. [[CrossRef](#)]
29. Shieh, L.; Bao, Y.; Chang, F. State-space self-tuning controllers for general multivariable stochastic systems. In Proceedings of the 1987 American Control Conference, Minneapolis, MN, USA, 10–12 June 1987; IEEE: Piscataway, NJ, USA, 1987; pp. 1280–1285.
30. Correa-Martínez, J.; Poznyak, A.S. Three electromechanical examples of robust switching structure state and parameter estimation. In Proceedings of the 38th IEEE Conference on Decision and Control, Phoenix, AZ, USA, 7–10 December 1999; pp. 3962–3963. [[CrossRef](#)]
31. Mazaheri, A.; Mansouri, M.; Shooredeli, M. Parameter estimation of Hammerstein-Wiener ARMAX systems using unscented Kalman filter. In Proceedings of the 2014 Second RSI/ISM International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, 15–17 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 298–303.

32. Tsai, J.S.-H.; Hsu, W.; Lin, L.; Guo, S.; Tann, J.W. A modified NARMAX model-based self-tuner with fault tolerance for unknown nonlinear stochastic hybrid systems with an input—Output direct feed-through term. *ISA Trans.* **2014**, *53*, 56–75. [[CrossRef](#)]
33. Pu, Y.; Chen, J. A novel maximum likelihood-based stochastic gradient algorithm for Hammerstein nonlinear systems with coloured noise. *Int. J. Model. Identif. Control* **2019**, *32*, 23–29. [[CrossRef](#)]
34. Wang, D.; Fan, Q.; Ma, Y. An interactive maximum likelihood estimation method for multivariable Hammerstein systems. *J. Frankl. Inst.* **2020**, *357*, 12986–13005. [[CrossRef](#)]
35. Zheng, W.X. On least-squares identification of ARMAX models. *IFAC Proc. Vol.* **2002**, *35*, 391–396. [[CrossRef](#)]
36. Poznyak, A.S. *Advanced Mathematical Tools for Automatic Control Engineers Volume 2: Stochastic Techniques*; Elsevier: Amsterdam, The Netherlands, 2009.
37. Medel-Juárez, J.; Poznyak, A. S. Identification of Non Stationary ARMA Models Based on Matrix Forgetting. 1999. Available online: <http://repositoriodigital.ipn.mx/handle/123456789/15474> (accessed on 6 February 2022).
38. Poznyak, A.; Medel, J. Matrix Forgetting with Adaptation. *Int. J. Syst. Sci.* **1999**, *30*, 865–878. [[CrossRef](#)]
39. Cerone, V. Parameter bounds for armax models from records with bounded errors in variables. *Int. J. Control* **1993**, *57*, 225–235. [[CrossRef](#)]
40. He, L.; Kárný, M. Estimation and prediction with ARMMAX model: A mixture of ARMAX models with common ARX part. *Int. J. Adapt. Control Signal Process.* **2003**, *17*, 265–283. [[CrossRef](#)]
41. Yin, L.; Gao, H. Moving horizon estimation for ARMAX processes with additive output noise. *J. Frankl. Inst.* **2019**, *356*, 2090–2110. [[CrossRef](#)]
42. Moustakides, G.V. Study of the transient phase of the forgetting factor RLS. *IEEE Trans. Signal Process.* **1997**, *45*, 2468–2476. [[CrossRef](#)]
43. Paleologu, C.; Benesty, J.; Ciochina, S. A robust variable forgetting factor recursive least-squares algorithm for system identification. *IEEE Signal Process. Lett.* **2008**, *15*, 597–600. [[CrossRef](#)]
44. Zhang, H.; Zhang, S.; Yin, Y. Online sequential ELM algorithm with forgetting factor for real applications. *Neurocomputing* **2017**, *261*, 144–152. [[CrossRef](#)]
45. Escobar, J.; Poznyak, A.S. Time-varying matrix estimation in stochastic continuous-time models under coloured noise using LSM with forgetting factor. *Int. J. Syst. Sci.* **2011**, *42*, 2009–2020. [[CrossRef](#)]
46. Escobar, J. Time-varying parameter estimation under stochastic perturbations using LSM. *IMA J. Math. Control Inf.* **2012**, *29*, 35–258. [[CrossRef](#)]
47. Escobar, J.; Poznyak, A. Benefits of variable structure techniques for parameter estimation in stochastic systems using least squares method and instrumental variables. *Int. J. Adapt. Control Signal Process.* **2015**, *29*, 1038–1054. [[CrossRef](#)]
48. Taylor, J. The Cramer-Rao estimation error lower bound computation for deterministic nonlinear systems. *IEEE Trans. Autom. Control* **1979**, *24*, 343–344. [[CrossRef](#)]
49. Hodges, J.; Lehmann, E. Some applications of the Cramer-Rao inequality. In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 31 July–12 August 1950; University of California Press: Berkeley, CA, USA, 1951; pp. 13–22.
50. Cramér, H. A contribution to the theory of statistical estimation. *Scand. Actuar. J.* **1946**, *1946*, 85–94. [[CrossRef](#)]
51. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Reson. J. Sci. Educ.* **1945**, *20*, 78–90.
52. Vincze, I. On the Cramér-Fréchet-Rao inequality in the nonregular case. In *Contributions to Statistics, the J. Hajek Memorial*; Reidel: Dordrecht, The Netherlands; Boston, MA, USA, 1979; pp. 253–262.
53. Khatri, C. Unified treatment of Cramér-Rao bound for the nonregular density functions. *J. Stat. Plan. Inference* **1980**, *4*, 75–79. [[CrossRef](#)]
54. Rissanen, J. Fisher information and stochastic complexity. *IEEE Trans. Inf. Theory* **1996**, *42*, 40–47. [[CrossRef](#)]
55. Jauffret, C. Observability and Fisher information matrix in nonlinear regression. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 756–759. [[CrossRef](#)]
56. Klein, A. Matrix algebraic properties of the Fisher information matrix of stationary processes. *Entropy* **2014**, *16*, 2023–2055. [[CrossRef](#)]
57. Bentarzi, M.; Aknouche, A. Calculation of the Fisher information matrix for periodic ARMA models. *Commun. Stat. Methods* **2005**, *34*, 891–903. [[CrossRef](#)]
58. Klein, A.; Mélard, G. An algorithm for the exact Fisher information matrix of vector ARMAX time series. *Linear Algebra Its Appl.* **2014**, *446*, 1–24. [[CrossRef](#)]
59. Bell, K.L.; Van Trees, H.L. Posterior Cramer-Rao bound for tracking target bearing. In Proceedings of the 13th Annual Workshop on Adaptive Sensor Array Process, Puerto Vallarta, Mexico, 13–15 December 2005; Citeseer: Princeton, NJ, USA, 2005.
60. Tichavsky, P.; Muravchik, C.H.; Nehorai, A. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.* **1998**, *46*, 1386–1396. [[CrossRef](#)]
61. Landi, G.; Landi, G.E. The Cramer—Rao Inequality to Improve the Resolution of the Least-Squares Method in Track Fitting. *Instruments* **2020**, *4*, 2. [[CrossRef](#)]
62. Efron, A.; Jeen, H. Detection in impulsive noise based on robust whitening. *IEEE Trans. Signal Process.* **1994**, *42*, 1572–1576. [[CrossRef](#)]

63. Liao, Y.; Wang, D.; Ding, F. Data filtering based recursive least squares parameter estimation for ARMAX models. In Proceedings of the 2009 WRI International Conference on Communications and Mobile Computing, Washington, DC, USA, 6–8 January 2009; IEEE: Piscataway, NJ, USA, 2009; Volume 1, pp. 331–335.
64. Collins, L. Realizable whitening filters and state-variable realizations. *Proc. IEEE* **1968**, *56*, 100–101. [\[CrossRef\]](#)
65. Wang, W.; Ding, F.; Dai, J. Maximum likelihood least squares identification for systems with autoregressive moving average noise. *Appl. Math. Model.* **2012**, *36*, 1842–1853. [\[CrossRef\]](#)
66. Zdrozny, P. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econom. Theory* **1988**, *4*, 108–124. [\[CrossRef\]](#)
67. Li, L.; Pu, Y.; Chen, J. Maximum Likelihood Parameter Estimation for ARMAX Models Based on Stochastic Gradient Algorithm. In Proceedings of the 2018 10th International Conference on Modelling, Identification and Control (ICMIC), Guiyang, China, 2–4 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
68. González, R.A.; Rojas, C.R. A Finite-Sample Deviation Bound for Stable Autoregressive Processes. In Proceedings of the 2nd Conference on Learning for Dynamics and Control, Berkeley, CA, USA, 10–11 June 2020; Bayen, A.M., Jadbabaie, A., Pappas, G., Parrilo, P.A., Recht, B., Tomlin, C., Zeilinger, M., Eds.; PMLR: Birmingham, UK, 2020; Volume 120, pp. 191–200.
69. Anderson, B.D.; Moore, J.B. State estimation via the whitening filter. In Proceedings of the Joint Automatic Control Conference, Ann Arbor, MI, USA, 26–28 June 1968; pp. 123–129.
70. Seong, S.M. A modified direct whitening method for ARMA model parameter estimation. In Proceedings of the 2007 International Conference on Control, Automation and Systems, Seoul, Korea, 17–20 October 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 2639–2642.
71. Yamada, I.; Hayashi, N. Improvement of the performance of cross correlation method for identifying aircraft noise with pre-whitening of signals. *J. Acoust. Soc. Jpn. (E)* **1992**, *13*, 241–252. [\[CrossRef\]](#)
72. Kuo, C.H. An Iterative Procedure for Minimizing and Whitening the Residual of the ARMAX Model. *Mech. Tech. J.* **2010**, *3*, 1–6.
73. Ho, W.K.; Ling, K.V.; Vu, H.D.; Wang, X. Filtering of the ARMAX process with generalized t-distribution noise: The influence function approach. *Ind. Eng. Chem. Res.* **2014**, *53*, 7019–7028. [\[CrossRef\]](#)
74. Graupe, D.; Efron, A.J. An output-whitening approach to adaptive active noise cancellation. *IEEE Trans. Circuits Syst.* **1991**, *38*, 1306–1313. [\[CrossRef\]](#)
75. Roonizi, A.K. A new approach to ARMAX signals smoothing: Application to variable-Q ARMA filter design. *IEEE Trans. Signal Process.* **2019**, *67*, 4535–4544. [\[CrossRef\]](#)
76. Zheng, H.; Mita, A. Two-stage damage diagnosis based on the distance between ARMA models and pre-whitening filters. *Smart Mater. Struct.* **2007**, *16*, 1829. [\[CrossRef\]](#)
77. Kuo, C.H.; Yang, D.M. Residual Whitening Method for Identification of Induction Motor System. In *Proceedings of the 3rd International Conference on Intelligent Technologies and Engineering Systems (ICITES 2014)*; Springer: Berlin/Heidelberg, Germany, 2016, pp. 51–58.
78. Song, Q.; Liu, F. The direct approach to unified GPC based on ARMAX/CARIMA/CARMA model and application for pneumatic actuator control. In Proceedings of the First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06), Beijing, China, 30 August–1 September 2006; IEEE: Piscataway, NJ, USA, 2006; Volume 1; pp. 336–339.
79. Dosiek, L.; Pierre, J.W. Estimating electromechanical modes and mode shapes using the multichannel ARMAX model. *IEEE Trans. Power Syst.* **2013**, *28*, 1950–1959. [\[CrossRef\]](#)
80. Chen, W.; Han, G.; Qiu, W.; Zheng, D. Modeling of outlet temperature of the first-stage cyclone preheater in cement firing system using data-driven ARMAX models. In Proceedings of the 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 11–13 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 472–477.
81. Akal, M. Forecasting Turkey's tourism revenues by ARMAX model. *Tour. Manag.* **2004**, *25*, 565–580. [\[CrossRef\]](#)
82. Pan, B.; Wu, D.C.; Song, H. Forecasting hotel room demand using search engine data. *J. Hosp. Tour. Technol.* **2012**, *3*, 196–210. [\[CrossRef\]](#)
83. Intihar, M.; Kramberger, T.; Dragan, D. Container throughput forecasting using dynamic factor analysis and ARIMAX model. *Promet-Traffic Transp.* **2017**, *29*, 529–542. [\[CrossRef\]](#)
84. Hickey, E.; Loomis, D.G.; Mohammadi, H. Forecasting hourly electricity prices using ARMAX–GARCH models: An application to MISO hubs. *Energy Econ.* **2012**, *34*, 307–315. [\[CrossRef\]](#)
85. Ekhsosuehi, V.U.; Omoregie, D.E. Inspecting debt servicing mechanism in Nigeria using ARMAX model of the Koyck-kind. *Oper. Res. Decis.* **2021**, *1*, 5–20.
86. Adel, B.; Cichocki, A. Robust whitening procedure in blind source separation context. *Electron. Lett.* **2000**, *36*, 2050–2051.
87. Cuoco, E.; Calamai, G.; Fabbroni, L.; Losurdo, G.; Mazzoni, M.; Stanga, R.; Vetrano, F. On-line power spectra identification and whitening for the noise in interferometric gravitational wave detectors. *Class. Quantum Gravity* **2001**, *18*, 1727–1751. [\[CrossRef\]](#)
88. Cuoco, E.; Losurdo, G.; Calamai, G.; Fabbroni, L.; Mazzoni, M.; Stanga, R.; Guidi, G.; Vetrano, F. Noise parametric identification and whitening for LIGO 40-m interferometer data. *Phys. Rev.* **2001**, *64*, 122022. [\[CrossRef\]](#)
89. Söderström, T.; Mahata, K. On instrumental variable and total least squares approaches for identification of noisy systems. *Int. J. Control* **2002**, *75*, 381–389. [\[CrossRef\]](#)
90. Bowden, R.J.; Turkington, D.A. *Instrumental Variables*; Cambridge University Press: Cambridge, UK, 1990.

91. Martens, E.P.; Pestman, W.R.; de Boer, A.; Belitser, S.V.; Klungel, O.H. Instrumental variables: Application and limitations. *Epidemiology* **2006**, *17*, 260–267. [[CrossRef](#)] [[PubMed](#)]
92. Jakeman, A.; Young, P. Refined instrumental variable methods of recursive time-series analysis Part II. Multivariable systems. *Int. J. Control* **1979**, *29*, 621–644. [[CrossRef](#)]
93. Young, P.; Jakeman, A. Refined instrumental variable methods of recursive time-series analysis Part III. Extensions. *Int. J. Control* **1980**, *31*, 741–764. [[CrossRef](#)]
94. Young, P.C. Refined instrumental variable estimation: Maximum likelihood optimization of a unified Box–Jenkins model. *Automatica* **2015**, *52*, 35–46. [[CrossRef](#)]
95. Wilson, E.D.; Clairon, Q.; Taylor, C.J. Non-minimal state-space polynomial form of the Kalman filter for a general noise model. *Electron. Lett.* **2018**, *54*, 204–206. [[CrossRef](#)]
96. Ma, L.; Liu, X. A nonlinear recursive instrumental variables identification method of Hammerstein ARMAX system. *Nonlinear Dyn.* **2015**, *79*, 1601–1613. [[CrossRef](#)]
97. Escobar, J.; Enqvist, M. Instrumental variables and LSM in continuous-time parameter estimation. *Esaim. Control Optim. Calc. Var.* **2017**, *23*, 427–442. [[CrossRef](#)]
98. Kazmin, S.; Poznyak, A. Recurrent estimates of ARX models with noises described by arma processes. *Autom. Remote Control* **1992**, *53*, 1549–1556.
99. Escobar, J.; Poznyak, A. Parametric identification of ARMAX models with unknown forming filters. *IMA J. Math. Control Inf.* **2021**, *39*, 171–184. [[CrossRef](#)]
100. Poznyak, A.S.; Tikhonov, S. Strong consistency of the extended least squares method with nonlinear error transformation. *Autom. Remote. Control* **1990**, *8*, 119–128.