**ONLINE LTI SYSTEM IDENTIFICATION AND TIME DELAY ESTIMATION**


by

**Seyed Hossein Rahnamaee**

Bachelor of Science Degree in Electrical Engineering

Azad University of Najafabad, 2006


A Thesis

presented to Ryerson University

in partial fulfillment of the requirements for the degree of

Master of Applied Science in the

Program of Electrical and Computer Engineering

Toronto, Ontario, Canada

AUTHOR'S DECLARATION FOR SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

# Abstract

## Online LTI System Identification and Time Delay Estimation

© Seyed Hossein Rahnamaee, 2016  Master of Applied Science

Electrical and Computer Engineering

Ryerson University

Model order selection for linear time-invariant (LTI) systems is an important system modeling concern and has been widely investigated through past decades. Different approaches of order selection such as Akaike information criterion (AIC), Bayesian information criterion (BIC), minimum description length (MDL) and reconstruction error LTI system identification (RE-LTI) propose different criteria to select the optimum order of a system. In many real life applications of model order selection the size of an observed data set is increasing. Thus, order selection methods need to adopt the best fit of a model as the data set size is increasing. This is our motivation to extend RE-LTI order selection for online application of order selection with lower computational cost and complexity. It has been shown previously that AIC, BIC, two-stage MDL and many existing order selection criteria are special cases of RE-LTI method. Our online order selection approach reduces the computational complexity of the offline approach from $O(N^3)$ to $O(N^2)$. It should be noted that RE-LTI and MNDL order selection methods have same fundamentals and consequently extending RE-LTI to online RE-LTI also extends MNDL to online MNDL.

Another crucial issue in system identification and modeling is estimating the time delay of a system's impulse response (or determining the start of its non-zero part). This problem is addressed in various areas including radar, sonar, acoustic source tracking, multipath channel identification, as well as many automatic control applications. Utilizing fundamentals of RE-LTI approach, here we introduce a new time-delay estimator. Simulation results show advantages of the proposed method and its superiority to existing approaches in accuracy and robustness in terms of the FIT index.

# Acknowledgment

I would like to express my deepest gratitude to my supervisor, Dr. Soosan Beheshti, for her immense knowledge, great patience and support. It was my privilege and honor to work under her supervision.

My sincere gratitude goes to my thesis committee members, Dr. Lian Zhao, Dr. Xiao-Ping (Steven) Zhang and Dr. Vadim Geurkov for their insightful comments and feedback.

I would also like to give my special thanks to my beloved father, mother and wife for their support and love.

# Contents

# List of Tables

# List of Figures

# List of Acronyms

| | |
|---|---|
| AGC | Automatic Generation Control |
| AIC | Akaike Information Criterion |
| BIC | Bayesian Information Criterion |
| CRLB | Cramer-Rao Lower Bound |
| ED | Economic Dispatch |
| EEF | Exponentially Embedded Family |
| ESPRIT | Estimation of Signal Parameters via Rotational Invariance Technique |
| GIC | Generalized Information Criterion |
| GPR | Ground Penetrating Radar |
| KIC | Kullback Information Criterion |
| LSE | Least Square Estimator |
| LTI | Linear Time-Invariant |
| MDL | Minimum Description Length |
| MNDL | Minimum Noiseless Description Length |
| MLE | Maximum Likelihood Estimator |
| PDF | Probability Density Function |
| RE-LTI | Reconstruction Error-Linear Time-Invariant |
| SNR | Signal-to-Noise Ratio |
| SVM | Support Vector Machines |
| TDE | Time-Delay Estimation |
| TWR | Through Wall Radar |

# Notations

$A_{S_M}$ :      An $N \times M$ Toeplitz matrix generated by the input

$A_{S_m}$ :      First $m$ columns of $A_{S_M}$

$A_{S_{d_{m^\star}}}$ :      An $N \times m^\star$ Toeplitz matrix generated by the input

$d^\star$ :      Optimum estimated time delay

$d$ :      True time delay

$\overline{y}[n]$ :      Noiseless data

$y[n]$ :      Noisy data

$\hat{y}[n]$ :      Data estimate

$u[n]$ :      Input data

$w[n]$ :      i.i.d. white Gaussian noise

$\sigma_w$ :      Noise standard deviation

$m^\star$ :      Optimum subspace order

$m$ :      Subspace order

$a_i^\star$ :      Real-valued coefficients of the system's impulse response

$\|.\|_2$ :      $L^2$ norm

$\widehat{\theta}_{S_M}$ :      Parameter estimate of order $M$

$\theta^\star$ :      True parameter

$z_{S_m}$ :      Reconstruction error

$x_{S_m}$ :      Data error

$Q(\alpha)$ :      Validation probability

$Q(\beta)$ :      Confidence probability

$\Delta_{S_m}$ :      Vector of length $N - m$, corresponding to unmodeled coefficients

$Nz_{S_m}$ :      Normalized reconstruction error

$\epsilon$ :      Stop-ID threshold

# Chapter 1

# Introduction

System modeling and identification is an important subject with various applications [3], [4]. One crucial subject in this type of modeling that has received much attention in the literature is model order selection. In model order selection the number of parameters involved in modeling is not known. Over the past years, many different criteria and prior assumptions have been proposed for this important question. [5]. It has been shown that reconstruction error is an order selection method that is consistent and can cover many other order selection approaches such as AIC and two-stage MDL [6].The first focus of this thesis is on online order selection. Online order selection aims to find the system order and identify the system as more and more data becomes available. With the fast developments in technology and consequently the fast growth of the amount of data to be processed, a consistent online order selection approach is highly desirable as it would be the answer to many applications needing real-time processes.

Model order selection has a rich history. Back in 1970s, Akaike, a pioneer in model selection, introduced a minimum information criterion estimate (MAICE) based on the previously proposed minimum information criterion (AIC) to provide an objective judgment among the competing models [7]. Although AIC has received much attention in literature, it tends to overestimate as data set size increases or when noise variance is very small. Consequently, AIC is not consistent [8]. How-

ever, some other works have tried to solve the AIC inconsistency issue by modifying its penalty term to achieve a more generalized information criterion (GIC) [9], [10], [11]. Nevertheless, the true model being a subset of the estimated model set is an unrealizable assumption of AIC and the other methods expanding AIC to more consistent criteria. A more parsimonious criterion in comparison with AIC is Bayesian information criterion (BIC) [12]. It has the same structure of penalized likelihood as AIC but penalizes the model order more than AIC. Therefore, for underlying models with finite number of parameters BIC outperforms AIC, but the problem is we do not have such information of true model in real application . This method like AIC leads to overestimation and as a result many modifications such as e-BIC, lp-BIC, and h-BIC in literature have been proposed as the better understanding of BIC framework with more consistency in high signal to noise ratios (SNRs) [13], [14]. Another well-known and more recent criterion of model selection is minimum description length (MDL). It measures the probability of models' goodness considering a code length that must be minimized [15]. Among all code length formulations, two-stage MDL as the best criterion has the same form of penalized likelihood as BIC. Although MDL is consistent for large data set sizes, in high SNRs it tends to overestimate the model and like AIC is inconsistent [8]. In [5] an informative review of previously mentioned criteria is presented.

Exponentially embedded family (EEF) is another method of model selection for linear models [16]. Even though the high SNR consistency of EEF is shown in the literature [8], the asymptotic behavior of EEF is like BIC that means its consistency is met under the assumption of large enough data sizes compared to the number of parameters. This assumption is not always realizable in real applications [17].

Kullback information criterion (KIC) and its different variants as model selection criteria, based on Kullback divergence, have been investigated in [18]. They are used for autoregressive models. It has been shown that KIC outperforms AIC for large samples in that KIC overestimates less. However, KIC and its variants like the AIC have the assumption of the true model being a member of a family of searched estimated models, that cannot be fulfilled.

2

Reconstruction error model order selection introduced in [6] has solved the shortcomings of all previously mentioned methods. This approach has utilized a different type of error along with feasible assumptions. This method has proved its consistency and unwillingness to overestimation in different scenarios for data set size and SNR. Despite the rich history of order selection, to our knowledge none of the most used order selection approaches mentioned above has been modified for fast and efficient online procedures.

The second focus of this thesis is on time delay estimation (TDE). Several TDE methods are investigated in literature for different applications. A review of some of these methods is provided here.

Ground penetrating radars (GPR) or through wall radars (TWR) localize hidden targets behind the wall or in subsurfaces of the ground by estimating the time delay between backscattered echoes from different layers. In [19] a model-based method of TDE, Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT), is used for the purpose of subsurface investigations by GPR. This method has two preprocessing steps prior to time delay estimation procedure . First, a deconvolution of the transmitted pulse in frequency domain is done to eliminate the effect of ESPRIT's assumption of time-shifted Dirac echoes and then to create the correlation matrix a subband smoothing method (ISS) is applied to prevent the failure of ESPRIT in the case of strongly correlated GPR echoes. The support vector machines (SVMs) in [20] have been chosen among supervised machine learning algorithms to estimate the time delays of correlated GPR signals with low computational time. Estimation results of this algorithm have shown good accuracy when signals are correlated. In [21] a time delay estimation method based on compressive sensing (CS) along with a clustering principle to enhance the echoes received in GPRs has been investigated. In this paper, weak noisy correlated echoes are enhanced with the Karhunen-Loeve transform (KLT) and then the estimation problem is done by an optimization.

To identify channel characteristics in wireless communications, the overlapped signals must be recovered. In [22] to estimate the time delay of overlapped signals, the least squares estimator with

two preprocessing steps is proposed.The preprocessing steps, under-sampling and interpolation, are used to compensate errors caused by low sampling rate in ultra-wideband systems. It has been shown that the proposed time domain algorithm has a good performance in terms of lower mean square error for different SNRs. In biomedical applications, artificial time delays resulted from artifacts (e.g. coughing, sneezing, choking) must be estimated. This is an important part of intensive care to decrease the discomfort of patients. This kind of time delay during ICU anesthesia has been estimated in [23] by using a cross-correlation analysis over a sliding window. The sliding window provides an online way of considering the changes in time delays.

Considering a signal corrupted by different types of additive independently and identically distributed (i.i.d) noise (like uniform, Gaussian, mixed Gaussian), [24] has estimated its time delay by using maximum likelihood function. This method has shown that the variance of its estimator satisfies the Cramer-Rao lower bound (CRLB) to prove better performance of maximum log-likelihood estimate over the other methods. In [25] a time delay estimation methods in the frequency domain has been introduced. The method estimate the time delay between two received band-pass signals by averaging their phase differences or their different frequency components in the pass-band.

In [26] a detailed review of methods for estimating the time delay has been presented. These methods are classified into four groups as time delay approximation methods, explicit time delay estimation, area and moments and finally higher order statistics methods. Different goals of time delay estimation have been determined by the author as well. It has been specified that in some applications the estimation of true time delay is needed while in some other applications the most suitable time delay for a given model is estimated. Another objective of time delay estimation has been defined in [2]. This objective is estimating a time delay that provides the best fit of a model in terms of the system identification process.

In this thesis, we are focused on the type of time delay estimation that gives the best fit of a model in the system identification process.

## 1.1 Thesis Contributions

Considering two crucial issues, online order selection and time delay estimation, in Chapter 3 we develop an online reconstruction error linear time-invariant (RE-LTI ) model order selection procedure. We show the one order decrease in the complexity order of our online procedure compared with RE-LTI offline procedure. Stop-ID threshold is defined in this chapter to validate our online model order selection method. In Chapter 4 using RE-LTI model order selection, we propose a method of time delay estimation with respect to the definition of system identification and we show its superiority compared with existing methods by using the FIT index.

The remainder of this thesis is arranged as follows. In Chapter 2 a review of popular model order selection criteria is presented. Also, there is a brief study of RE-LTI order selection as the state of the art and basis of our online procedure. This chapter ends with a short review of the time delay estimation concept with respect to te the definition of system identification. In Chapter 3 our online RE-LTI order selection has been introduced. The simulation results are provided to show the consistency and fast computational time of our online procedure. This chapter ends with a case study and its simulation results. We start chapter 4 by proposing a time delay estimation method based on RE-LTI reconstruction error. The simulation results are presented at the end of this chapter. Finally, in Chapter 5 we come to our conclusions and suggest some future works.

# Chapter 2

# Background

## 2.1 System Identification and Online Order selection

### 2.1.1 System Identification

In system identification, a mathematical model is formulated based on prior knowledge of a system. In many cases, the available input and output of the system play the critical role in this procedure. Applications of system identification in various fields of science and engineering are in numerous examples [4]. For instance, in signal processing, applications of this field are such as fault detection, pattern recognition, and spectral analysis. The applications are also extended in other areas such as biology and econometrics that require predictions or decisions based on the system identification models [3].

As the exact underlying system is not available in this scenario, the aim of system identification is finding the best fit of a true system which is analytically well-defined. Generally, a system identification approach requires the following:

1. Recording input-output data of the system to be modeled.

2. Considering a set of model structures that can define the relationship between input and output such as linear difference equation.

3. Some assessment criteria like the least squares or maximum likelihoods to find the best model in the set.

Going through the steps of the above procedure we reach to a model that needs to pass a validation test to be accepted as a good enough and true description of the system, otherwise the procedure must be repeated until the validation test is satisfied [4].

Due to wide applications of system identification, linear or non-linear, in different fields such as automatic control, biology, economics, geology and many others on one hand, and on the other hand the need to deal with more complicated systems as a result of rapidly developing technology, this topic has received much attention in the literature and has been developed through past decades. In [5] a review of system identification methods is provided. Methods including least squares and maximum likelihood are the well-studied traditional methods for linear system identification while neural networks, fuzzy logic, genetic algorithm, swarm intelligence optimization algorithms are modern system identification methods of non-linear systems [5,27,28,29,30,31,32].

Linear time-invariant (LTI) systems as the idealizations of real-life processes are the most important model structures considered in the field of system identification; while time-varying and nonlinear models also have their uses in describing systems. Representing LTI systems is through their impulse responses or their transfer functions and determining these definitions could be done with selecting a finite set of parameters as the best fit for a system or directly without a set of possible models, known as parametric or nonparametric identification methods respectively [4]. LTI system identification field has an important role yet still has its own challenges. The focus of this thesis is on LTI system identification. However, proposing an online approach for this type of system identification is also beneficial in slowly varying scenarios. In the following one general

formulation of LTI system identification, which is used in this thesis, is derived.

The input-output relation of a causal and finite impulse response duration (FIR) LTI system can be defined by its impulse response in the discrete time domain as:

$$y[n] = \sum_{i=0}^{M-1} a_i^\star u[n-i] + w[n] \tag{2.1}$$

where $u[n]$ , $y[n]$, $a_i^\star$s, $w[n]$ and $M$ are input, observed output, impulse response coefficients, additive white Gaussian noise with zero mean and variance of $\sigma_w^2$, and number of impulse response's coefficients or system's order respectively.

In this situation, if $M$ in (2.1) is assumed to be known the optimum parameter model order selection is typical of parametric approaches of system identification while it is not the case in nonparametric methods because they do not use finite-dimensional vectors to find the best model for a system [5]. Since maximum likelihood (ML) estimates of a system parameters has been proved to be asymptotically efficient [33], it is usually used in parameter estimation and consequently, order selection methods are mostly tied to it. In ML estimate, a likelihood function of observed data $y$ depending on a parameter vector $\theta = [a_0, a_1, ..., a_{m-1}]^T$ is maximized to obtain the estimate of $\theta$, as [6]:

$$\hat{\theta} = arg \max_{\theta} \ f(y; \theta) \tag{2.2}$$

when $f(y; \theta)$ is the probability density function (PDF) of observed data $y$ depending on the parameter $\theta$. If the assumptions are such that observed data of length $N$, $y^N$, is normally distributed or in other words its PDF is:

$$f(y; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{\|y^N - A_{S_M}\theta\|_2^2}{2\sigma_w^2}} \tag{2.3}$$

where $A_{S_M}$ is an $N \times M$ Toeplitz matrix generated by the available input data of finite length $N$, $u^N = [u[1], ..., u[N]]^T$, and the observed data, $y^N$, is the noisy data as:

$$y[n] = A_{S_M}\theta^\star + w[n] \tag{2.4}$$

consequently, the ML estimate in (2.2) is reduced to the form of least square (LS) estimate as:

$$\widehat{\theta}_{S_M} = arg \min_{\theta} \ \|y^N - A_{S_M}\theta\|_2^2 \tag{2.5}$$

$$= (A_{S_M}^T A_{S_M})^{-1} A_{S_M}^T y^N \tag{2.6}$$

This system identification problem becomes much more challenging if no information on the value of $M$ is available. In this scenario, if absolutely no information on this value is available, $M$ can grow up to data length $N$. In this case, the parameter estimate for the different values of m is calculated as follows:

$$\widehat{\theta}_{S_m}(y^N) = (A_{S_m}^T A_{S_m})^{-1} A_{S_m}^T y^N \tag{2.7}$$

where $A_{S_m}$ is the matrix with the first $m$ columns of Toeplitz matrix $A_{S_M}$.

Using the available data error defined as:

$$x_{S_m} = \frac{1}{N} \|y^N - \widehat{y}_{S_m}^N\|_2^2 \tag{2.8}$$

where $\widehat{y}_{S_m}^N$ is:

$$\widehat{y}_{S_m}^N = A_{S_m} \widehat{\theta}_{S_m} \tag{2.9}$$

combining the order selection with maximum likelihood estimator provides the following known solution:

$$N = arg \min_m \min_{\widehat{\theta}} \|y^N - A_{S_m} \widehat{\theta}\|_2^2 \tag{2.10}$$

This is a well-known problem with using only ML estimator and data error. Illustration of this is given in Figure (2.1) which shows a typical behavior of this error as a function of $m$ that is $1 \leq m \leq M$. As the figure shows in this case that $N = 100$ even though the true order $M = 50$, ML estimation and using the data error as a criterion picks $m = 100$ as the order of the system since the data error is zero at that point.

So while data error is an important available value to be used in order selection, by minimizing itself it will not provide the best fit of a system. Consequently, the following section concentrates on the methods answering the order selection problem.

9

Figure 2.1: Typical behavior of data error as the function of m

### 2.1.2 Order Selection

Choosing the best model among the estimated models of different orders based on a set of observed data or in other words model order selection has been investigated for years. Model order selection has many practical applications. Well-known methods of order selection in literature are Akaike information criterion (AIC) [7], Bayesian information criterion (BIC) [12], the Minimum Description Length (MDL) [15] and generalized information criterion (GIC) [9]. The common part of all these methods is the ML estimator as an unbiased efficient estimate of a defined likelihood function depending on an unknown vector of parameters, namely true parameters, for which occurrence of the observed data has the highest probability. The next step in these methods which is the part that makes the difference among them is the way of finding the optimum order among the subspaces of different orders estimated for the set of observed data of a system.

Maximum likelihood estimator as the basis of the above methods looks for the best estimate of a model parameters by minimizing the estimation error. Estimation error is a decreasing function

of model order,therefore, minimizing this error does not help us with choosing the optimum order. Thus a penalty term ,which is an increasing function of the model order, must be added to create the bias-variance tradeoff [34] that leads to the optimum choice.

A general form, penalized log likelihood (PLL), of AIC, BIC, GIC and two-stage MDL is:

$$-\frac{1}{N}\ln f_m(y,\hat{\theta}_m) + P(m,N) \tag{2.11}$$

where $f_m(y,\hat{\theta}_m)$ is the likelihood function of observed data and $\hat{\theta}_m$ is its maximizer in different subspaces of order $m$. The second term of this general form, $P(m,N)$, is the penalty term as a function of subspace order $m$ and data length $N$. The penalty term of each rule is derived differently as [9]:

$$\text{AIC} \quad : \quad P(m,N) = \frac{m}{N} \tag{2.12}$$

$$\text{BIC} \quad : \quad P(m,N) = \frac{\ln N}{N} \tag{2.13}$$

$$\text{GIC} \quad : \quad P(m,N) = \frac{\rho+1}{N} \tag{2.14}$$

$$\text{MDL(two} - \text{stage)} \quad : \quad P(m,N) = \frac{m\log N}{2N} \tag{2.15}$$

All these rules are based on two important assumptions. First, the data length, $N$, is large enough in comparison with subspace order, $m$, and then they assume that true model, $\theta^\star$, is a subset of estimated models:

$$m << N \tag{2.16}$$

$$\theta^\star \in S_m \tag{2.17}$$

In many real applications, the set of available observed data has a finite length which means that the first assumption is not always met. On the other hand because true model is unknown the second assumption is not fulfilled. Obviously, out of the scope of these assumptions mentioned methods are not consistent [35]. The inconsistency problem in high SNR conditions when data set size is finite and noise variance tends towards zero has been addressed by many recent works.

Seminal article [8] has explored inconsistency or overestimation of MDL and AIC in case of finite data sample sizes or when the noise variance is very small $\sigma_w^2 \to 0$, in other words in high

SNR conditions. Authors defined the probability of choosing model order $H_j$ when the true model order is $H_p$ as the probability of wrong model order, $P_e$. They introduced a lower bound of the probability of choosing the wrong model order in case of MDL and AIC. They proved this lower bound is fixed by the size of observed data and is not dependent on noise variance in case of MDL. In terms of AIC, this lower bound is constant for all different data sizes which shows the inconsistency of AIC for both finite and infinite data set sizes.

Inconsistency of BIC when the length of available data, $N$, is finite and noise variance is very small was discussed in [14] and then considering two different conditions of large $N$ and small noise variance, $\sigma_w^2 << 1$, authors introduced a proper form of BIC for only high SNR cases as $\text{BIC}_{\text{SNR}}(n)$. Since this form of BIC is only valid for high SNRs and fails in case of large $N$ to estimate the model order, $n$, authors tried to combine the $\text{BIC}_{SNR}(n)$ with its basic form $\text{BIC}_N(n)$ to obtain a compact form $\text{BIC}_{N,SNR}(n)$. In the combined form, the penalty term is the largest one between $\text{BIC}_N$ and $\text{BIC}_{SNR}$ penalty terms. Although, in different combinations of observed data size, $N$, and noise variance, $\hat{\sigma}_w^2$, they have shown acceptable choice of penalty term in their method, in a small $N$ and low SNR case their method is equal to $\text{BIC}_N(n)$ which is not a good choice. With the basis of BIC, in [13] another supplement to proper forms of BIC studied in [14] has been presented by introducing three model order selection methods, namely empirical BIC (e-BIC), hyper BIC (h-BIC) and laplace BIC (lp-BIC). Estimating the hyperparameter g of the g-prior when the SNR is unknown yields e-BIC, considering g as a random variable with a defined pdf out of a marginal likelihood results in h-BIC, and Laplace approximation for $\ln g$ shows the lp-BIC, all has been illustrated as improvements to $\text{BIC}_{N,SNR}$ which perform well under conditions of high or low SNRs.

In [36] the consistency of penalized likelihood order selection methods and their limitations in the absence of a prior upper bound of model order with respect to penalty terms of different orders have been explored and showed that penalty terms of order $\log \log n$ are minimal while penalties of order $n$ are maximal. Necessary and sufficient conditions to provide penalty functions for PLL based order estimation approaches such as AIC,BIC and MDL, that result in high SNR consistency

are looked for in [11]. Achieving this goal, authors considered two events of overestimation (O) and underestimation (U) and investigated conditions that make the probability of occurring these two events zero when the variance of noise is very small (high SNR conditions). Satisfying the following assumptions is the necessary and sufficient conditions proposed in [11]:

$$h(k, \sigma^2) - h(k_0, \sigma^2) \xrightarrow{P} \infty \text{ as } \sigma^2 \to 0; \text{ for } k > k_0 \tag{2.18}$$

$$\sigma^2 \left[ h(k_0, \sigma^2) - h(k, \sigma^2) \right] \xrightarrow{P} 0 \text{ as } \sigma^2 \to 0; \text{ for } k < k_0 \tag{2.19}$$

where $h$ is the penalty function, $k_0$ is true order and $\sigma^2$ is noise variance which is known. In [10] a review of information theoretic criteria (ITC) such as AIC, BIC, MDL or more generally generalized information criterion (GIC) has been presented and analyzed the performance of those criteria with respect to the probability of correct detection of model order for different SNRs:

$$P_c = P_c(q, SNR, \nu) = 1 - P_{over} - P - under \tag{2.20}$$

$$P_{over} \simeq P \left( \bigcup_{i=1}^{q_{max}-q} \{ITC(q+1) < ITC(q)\} \right) \tag{2.21}$$

$$P_{under} \simeq P \left( \bigcup_{i=1}^{q} \{ITC(q-1) < ITC(q)\} \right) \tag{2.22}$$

where q is the model order, $\nu$ is the parameter of GIC which must be selected properly, and $P_{over}$ and $P_{under}$ are probabilities of overestimation and underestimation respectively. As they have proved for high SNR underestimation probability is almost zero and consequently due to non-zero overestimation probability these criteria always select the model order incorrectly. Based on the assumption that true model is a member of the estimated model set, to overcome this shortcoming in case of finite sample sizes authors proposed setting the parameter $\nu$ in GIC as:

$$\tilde{\nu} = \arg \max_{\nu} \max_{q,SNR} \left\{ P_{over}(q, SNR, \nu) \mid P_{over} \leq P_{over}^{MAX} \right\} \tag{2.23}$$

Another approach to model order selection proposed in [16] is done by embedding exponentially two or more probability density functions, $p_i(X)$, (PDFs) into a family of pdfs:

$$p_i(x; \eta) = exp \left[ \eta \ln \frac{p_i(X; \theta_i)}{p_0(X)} - \ln \int p_1^\eta(X) p_0^{1-\eta}(X) dx + \ln p_0(X) \right] \tag{2.24}$$

13

where $\theta_i$ is a vector of unknown model parameters, $0 \leq \eta < 1$, and $p_0(X)$ is given. Thus the model order selection method is choosing the maximum $p_i(X; \hat{\eta})$ among all models:

$$\text{EEF}(i) = \max_{\eta} \left[ \eta \ln \left( \frac{1}{p_{T_i}(t(X); 0)} \right) - K_0(\eta) \right] \tag{2.25}$$

A detailed study on the performance of EEF as an order selection criterion with a generalized likelihood ratio (GLR) base has been presented in [17] where authors illustrated that estimating model order, $n$, is through maximizing $\hat{f}(n)$ which is defined as:

$$\max_{n} \hat{f}(n) = 2 \ln \left( \frac{\hat{p}_{n+1}}{\hat{p}_1} \right) - n \left[ 1 + \ln \left( \frac{2 \ln \left( \frac{\hat{p}_{n+1}}{\hat{p}_1} \right)}{n} \right) \right] \tag{2.26}$$

In this work, it has been shown that the asymptotic behavior of EEF is like BIC and similarly this method holds as long as the observed data size is large enough compared to the model order or in the other words $N >> n$.

From what have been discussed so far, we can understand that underestimation or overestimation issue of model order selection's forefathers, AIC, BIC and MDL, is because of their penalty terms. Therefore, many works have been done during the past years in literature to address these problems by modifying the penalty terms used in those criteria. These modifications have led to many variants of previously mentioned methods. However, a novel approach to model order selection has been introduced in [6], namely reconstruction error model order selection (RE). This method unlike the previous criteria, instead of using data error and then trying to find a penalty term to create the tradeoff between the estimated fit goodness and its complexity, has used another error. The error used in [6] has the desired tradeoff in itself, therefore, it does not need an additional penalty term. The error used in this method is the difference between the observed output data and the noiseless output. Since the noiseless output as the representative of the true model is not available, this error has been bounded probabilistically. Furthermore, unlike the above order selection criteria, this method does not consider the true model to be a subset of estimated model sets, which is a realizable assumption. Last but not least, this approach has proved that AIC, BIC, and two-stage MDL are its special cases. Considering these specifications, RE-LTI model order

14

selection is the state of the art to the best of our knowledge and we develop our online procedure based on this method.

## 2.1.3   Reconstruction Error (RE)

Here we briefly review RE as the one powerful method of model order selection. Note that this method is closely related to the MNDL approach [37] :

$$\text{MNDL}(y^N, \sigma_w^2) = \min_{S_m} \overline{\text{DL}_{\sigma_w^2}(\overline{y}^N; \hat{y}_{S_m}^N)} \tag{2.27}$$

where description length (DL) is:

$$\text{DL}(\overline{y}^N; \hat{y}_{S_m}^N) = \log_2 \sqrt{2\pi\sigma_w^2} + \frac{\log_2 e}{2\sigma_w^2} z_{S_m} \tag{2.28}$$

and reconstruction error, $z_{S_m}$, is:

$$z_{S_m} = \frac{1}{N} \|\overline{y}^N - \hat{y}_{S_m}^N\|_2^2 \tag{2.29}$$

In the absence of noise if we consider a class of estimated model sets with different orders for an LTI system, obviously the highest order among them would introduce the least estimation error while in the presence of noise estimating more parameters of impulse responses, in other words models with higher orders, always do not result in smaller estimation errors since they are noisier than the lower orders [6]. As a result, to compare model sets with different orders a metric which shows parameter estimation error as a decreasing function of model sets orders by using available noisy finite-length input-output data of the system is needed. Considering the input-output relation for a causal LTI system with finite length impulse response as follows:

$$\overline{y}(n) = \sum_{i=0}^{M-1} a_i^* u[n-i] \tag{2.30}$$

where $u[n]$ is the input and independent of additive noise with length $N$ ,$a_i^*$ represents the real-valued coefficients of the impulse response which has the maximum length of $M$ and $\overline{y}$ is the

noiseless output of length $N$ which is not available. Available noisy version of the output is:

$$y[n] = \overline{y}[n] + w[n] \tag{2.31}$$

$w[n]$ is additive white Gaussian noise with zero mean and variance of $\sigma_w^2$.

To write (2.30) as a linear transformation using Toeplitz matrix, we consider the vector of impulse response coefficients as:

$$\theta^* = [a_0^*, a_1^*, ..., a_{M-1}^*]^T, \theta^* \in S_M \tag{2.32}$$

where $\theta^*$, true parameter, is a member of a compact set $S_M$ which is a subset of $R^M$. It is assumed that $M$ is an upper bound for the order of the true parameter. Then:

$$\overline{y}^N = A_{S_M}\theta^* \tag{2.33}$$

Now considering $\hat{\theta}_{S_M}$ as the parameter estimate of order $M$, a Maximum Likelihood estimator of $\theta^*$ as previously defined in (2.6) is:

$$\widehat{\theta}_{S_M}(y^N) = (A_{S_M}^T A_{S_M})^{-1} A_{S_M}^T y^N = \tag{2.34}$$
$$(A_{S_M}^T A_{S_M})^{-1} A_{S_M}^T (\overline{y} + w^N)$$

which is the projection of $y^N$, the sum of noiseless output and additive Gaussian noise into the parameters space. As we can see in (2.34) when $w^N = 0$ ,in the absence of noise, the parameter estimate is true parameter. On the other hand, when $w^N \neq 0$, some amounts of additive noise are added to the elements of true parameter which must be especially noted when the amount of each element is zero or much smaller than the amount of corresponding fitting element of noise, in this case in order to reduce the effect of noise and decrease the estimation error, it would be better to look for an estimate of a subset of true parameter which means setting those elements to zero or in other words having them unmodeled. When estimating in subspaces of $S_M$ , we can rewrite true parameter vector as :

$$\theta^* = \begin{bmatrix} \theta_{S_m}^* \\ \Delta_{S_m} \end{bmatrix} \tag{2.35}$$

where $\theta^*_{S_m}$ is the vector of true parameter in the subspace $S_m$ with length $m$, $1 \leq m \leq M$, and $\Delta_{S_m}$ is a vector of length $M - m$ for the rest of parameters not modeled in the subspace $S_m$. Therefore, elements of each subspace $S_m$ would be:

$$\theta = [a_0, a_1, ..., a_{m-1}, 0, ..., 0]^T \in S_m, \ S_m \subset S_M \tag{2.36}$$

Now, looking at the first $m$ columns of $A_{S_M}$ matrix as $A_{S_m}$ and the rest of $M - m$ columns as $B_{S_m}$, we can rewrite (2.33) as :

$$\overline{y}^N = \begin{bmatrix} A_{S_m} & B_{S_m} \end{bmatrix} \begin{bmatrix} \theta^*_{S_m} \\ \Delta_{S_m} \end{bmatrix} = \begin{bmatrix} A_{S_m}\theta^*_{S_m} + B_{S_m}\Delta_{S_m} \end{bmatrix} \tag{2.37}$$

and for parameter estimate of order $m$ in subspace $S_m$ we just replace $A_{S_m}$ instead of $A_{S_M}$ in (2.34):

$$\widehat{\theta}_{S_m}(y^N) = \begin{bmatrix} \theta^*_{S_m} + (A^T_{S_m}A_{S_m})^{-1}A^T_{S_m}(B_{S_m}\Delta_{S_m} + w^N) \\ 0_{(N-m)\times 1} \end{bmatrix} \tag{2.38}$$

and using estimation of true parameter in subspace $S_m$ we can define estimate of observed output data as:

$$\widehat{y}^N_{S_m} = A_{S_m}\widehat{\theta}_{S_m} = \begin{bmatrix} A_{S_m} & B_{S_m} \end{bmatrix} \widehat{\theta}_{S_m} = \tag{2.39}$$

$$\begin{bmatrix} A_{S_m}(\theta^*_{S_m} + (A^T_{S_m}A_{S_m})^{-1}A^T_{S_m}(B_{S_m}\Delta_{S_m} + w^N)) \\ 0_{(N-m)\times 1} \end{bmatrix}$$

To compare the estimates of true parameter in subspaces with different lengths of $m$ and choose the optimum number of elements to be estimated, we can use the parameter estimation error. Since unmodeled coefficients are not available, parameter estimation error cannot be calculated with available data. In each subspace, observed output and its estimate are available so a possible error to calculate is data error, $x_{S_m}$, (2.8). As it has been shown in Figure (2.1) data error is a decreasing function of $m$ or in other words it cannot be used as a comparing error among the subspaces to find the optimum length of true parameter estimate. According to this error the lowest amount of error is obtained for the highest amount of $m$ which is the length of true parameter itself. Obviously,

an increasing function of $m$ as a penalty term must be added to this error then it could be used as a comparing measurement. Having data error available and knowing its distribution we can find lower and upper bounds for parameter estimation error that is our comparing criterion to find the optimum $m$. In order to get to that point another error which is the transformation of parameter error into the output space by using Toeplitz matrix $A_{S_M}$ is defined and called reconstruction error:

$$z_{S_m} = \frac{1}{N}\|\bar{y}^N - \widehat{y}_{S_m}^N\|_2^2 = \frac{1}{N}\|A_{S_M}(\theta^\star - \widehat{\theta}_{S_m})\|_2^2 = \tag{2.40}$$

$$\frac{1}{N}\|G_{S_m}B_{S_m}\Delta_{S_m} + C_{S_m}w^N\|_2^2$$

where

$$G_{S_m} = I - A_{S_m}(A_{S_m}^T A_{S_m})^{-1}A_{S_m}^T \tag{2.41}$$

$$C_{S_m} = A_{S_m}(A_{S_m}^T A_{S_m})^{-1}A_{S_m}^T \tag{2.42}$$

are projection matrices with ranks $N - m$ and m respectively.

Bounding $Z_{S_m}$ in order to compare the competing model sets, we have to consider the worst case value of this error which is the upperbound of $z_{S_m}$.

$$\underline{z_{S_m}(Q(\beta), x^N, y^N, Q(\alpha))} \le z_{S_m} \le \overline{z_{S_m}(Q(\beta), x^N, y^N, Q(\alpha))} \tag{2.43}$$

Using available observed data and data error the probabilistic bounds of $z_{S_m}$ are possible to be calculated and its upperbound as the worst case is:

$$\overline{z_{S_m}(Q(\beta), x^N, y^N, Q(\alpha))} = U_{S_m}(Q(\alpha), x^N, y^N) + \frac{m}{N}\sigma_w^2 + \beta\frac{\sqrt{2m}\sigma^2}{N} \tag{2.44}$$

where $U_{S_m}$ an upperbound of unmodeled coefficients effect is defined as:

$$U_{S_m}(x^N, y^N, Q(\alpha)) = x_{S_m} - m_w + \frac{2\alpha^2\sigma_w^2}{N} + K_{S_m}(\alpha) \tag{2.45}$$

and $x_{S_m}$ is the available data error and $K_{S_m}$ and $m_w$ are:

$$K_{S_m}(\alpha) = 2\alpha\frac{\sigma_w}{\sqrt{N}}\sqrt{\frac{\alpha^2\sigma^2}{N} + x_{S_m} - \frac{1}{2}m_w} \tag{2.46}$$

$$m_w = \left(1 - \frac{m}{N}\right)\sigma_w^2 \tag{2.47}$$

18

$Q(\alpha)$ and $Q(\beta)$ are chosen confidence and validation probabilities such that:

$$Q(\beta) = Pr\left\{\mid Z_{S_m} - E(Z_{S_m}) \mid \leq \beta \frac{\sqrt{2m}}{N} \sigma_w^2\right\} \tag{2.48}$$

$$Q(\alpha) = Pr\left\{\mid X_{S_m} - E(X_{S_m}) \mid \leq \alpha \sqrt{var X_{S_m}}\right\} \tag{2.49}$$

Confidence and validation probabilities are desired to be as high as possible but under certain conditions to provide tight bounds on $z_{S_m}$ [6], therefore conditions are:

$$\lim_{N\to\infty} \alpha_N = \infty \tag{2.50}$$

$$\lim_{N\to\infty} \beta_N = \infty \tag{2.51}$$

$$\lim_{N\to\infty} \frac{\alpha_N}{\sqrt{N}} = 0 \tag{2.52}$$

$$\lim_{N\to\infty} \frac{\beta_N}{\sqrt{N}} = 0 \tag{2.53}$$

Setting the confidence and validation probabilities to zero causes the upper and lower bounds converge together therefore an estimate of the reconstruction error is:

$$z_{S_m} \approx x_{S_m} + (\frac{2m}{N} - 1)\sigma_w^2 \tag{2.54}$$

which in this format is like AIC defined by data error and a penalty term as a function of model order and data length. On the other hand setting the confidence probability to zero and setting $\beta$ as $\sqrt{m} \log(N)$ for the validation probability will make the upper bound of reconstruction error like the criterion used in BIC or two-stage MDL.

The bias-variance tradeoff in $z_{S_m}$ can be illustrated with its expectation and variance:

$$E(z_{S_m}) = \frac{m}{N}\sigma_w^2 + \frac{1}{N}\|G_{S_m} B_{S_m} \Delta_{S_m}\|_2^2 \tag{2.55}$$

$$Var(z_{S_m}) = \frac{2m}{N^2}(\sigma_w^2)^2 \tag{2.56}$$

$\Delta_{S_m}$ as the effect of unmodelled coefficients would be zero for orders greater than true parameter's order where only the first term of $E(Z_{S_m})$ is not zero which is like a penalty increasing by $m$. But in the limit this term would become zero as well which means in the limit for subspaces of orders greater than true parameter's order $E(Z_{S_m})$ is zero.

Finally, summarizing the proposed method in [6] to choose the optimum subspace order, $m$, would be as follows:

$$S_{m^*} = arg \min_{S_m} \overline{z_{S_m} \left( Q(\beta), x^N, y^N, Q(\alpha) \right)} \tag{2.57}$$

where $m^\star$ is the optimum order and $\widehat{\theta}_{S_{m^\star}}$ is its corresponding least noisy estimated impulse response.

## 2.2  Time Delay Estimation (TDE)

As discussed in Chapter 1, many methods of time delay estimation (TDE) are available in literature. Time delay can be estimated in different domains such as time domain or frequency domain. If the time delay is not an explicit parameter of a model, the model parameters must be estimated first and then the time delay is estimated. Otherwise, the time delay and model parameters can be estimated at the same time [26]. In the time domain, a delayed impulse response has a zero part at its beginning. Therefore, some methods such as Cumulative Sum (CUSUM) [2] try to detect the beginning of non-zero part. These methods need to first estimate the impulse response parameters. Then, they use some techniques to reduce the uncertainty of the estimated parameters and finally, some thresholding methods are applied to detect the delay at the beginning of the impulse response. In the frequency domain, the real part of the Fourier transform of an impulse response consists of a sinusoid with the frequency of time delay and a phase shift. The spectrum of this real part has a peak at the frequency of time delay [26].

Having described the above time delay estimation methods, we are going to estimate the time delay in the time domain when it is not an explicit parameter of a model. Therefore, in Chapter 4 we use RE-LTI order selection to estimate the delayed impulse response. Then, we propose a method to estimate the number of zero taps at the beginning of the estimated impulse response. The estimated time delay is going to provide the best fit of the model. In order to show that estimated

delay leads to the best fit of the model, we use fit index [2] as:

$$fit = 100 \left( 1 - \frac{\| \hat{y} - y \|}{\| y - \overline{y} \|} \right) (\%) \qquad (2.58)$$

# Chapter 3

# Online RE- LTI Order Selection

How should the least noisy estimated parameters or order of a linear time-invariant system be modified when the available data set enlarges?

The above question would be addressed through online model order selection.

Among the information theoretic criteria discussed in Chapter 2, we have chosen the RE-LTI model order selection method [6] as it outperforms the other criteria in finding the best fit of a true model from the available data set. Utilizing (2.57) for an online model order selection would be time-consuming and cannot be implemented in an online procedure. Therefore, in order to be able to keep up with the incoming information flow we should introduce a recursive function of previously defined $S_{m^*}(N)$ and newly arrived data as:

$$S_{m^*}(N+1) = f\left(S_{m^*}(N), u[N+1], y[N+1]\right) \tag{3.1}$$

In this way, $f$ should be defined as a recursive function with small and simple steps as possible that responds fast to the changes of input data set size, especially when the data set size becomes extremely large. A metric providing an upper bound on the growth rate of a function is big O notation. This notation is used in the following sections of this chapter to show the proposed

online algorithm improvement.

According to (2.57), the optimum $m$ is the one that minimizes the upper bound of reconstruction error, $z_{S_m}$, for all nested subspaces $m$, $1 \leq m \leq M$. As the data length grows, using this equation to find the least noisy estimate is time-consuming in particular when the data length becomes extremely large.

Considering the upper bound of $z_{S_m}$ with respect to the computational complexity order, we find that computing the data error, $x_{S_m}$, is the most complicated part due to finding the estimate of true parameter for every $m$ as data length grows. Thus to achieve a fast online model order selection algorithm, we should estimate $\widehat{\theta}_{S_M}(y^{N+1})$ recursively. But prior to that, we go briefly through the definition of $\widehat{\theta}_{S_M}(y^N)$.

Parameter estimate of order M, $\widehat{\theta}_{S_M}(y^N)$ is defined in (2.6) as:

$$\widehat{\theta}_{S_M}(y^N) = \left(A_{S_M}^T(N)A_{S_M}(N)\right)^{-1} A_{S_M}^T(N)y^N \tag{3.2}$$

where $A_{S_M}$ is the $N \times M$ Toeplitz matrix generated by the available input $u^N$ as:

$$\begin{bmatrix}
u_1 & 0 & 0 & 0 & \dots & 0 \\
u_2 & u_1 & 0 & 0 & \dots & 0 \\
u_3 & u_2 & u_1 & 0 & \dots & 0 \\
. & . & . & . & \dots & . \\
. & . & . & . & \dots & . \\
. & . & . & . & \dots & . \\
u_N & u_{N-1} & u_{N-2} & u_{N-3} & \dots & u_{N-M+1}
\end{bmatrix} \tag{3.3}$$

and $y^N$ is the observed output.

When the length of available data is increased by one, $u_N$ becomes $u_{N+1}$ and $y_N$ becomes $y_{N+1}$, the new $\widehat{\theta}_{S_m}(y^{N+1})$ can be calculated by using $\widehat{\theta}_{S_m}(y^N)$ in a recursive manner. In this way, the number of steps to calculate new $\widehat{\theta}_{S_m}(y^{N+1})$ will be decreased substantially and consequently, as shown in the next section, we will reach a lower complexity order in comparison with simply defining all matrices from the beginning and using (2.57).

*Important Notation:* In the following sections, $\widehat{\theta}_{S_m}(y^{N+1})$ is denoted by $\widehat{\theta}_{S_m}(N+1)$ and also $\widehat{\theta}_{S_m}(N)$ represents $\widehat{\theta}_{S_m}(y^N)$ to simplify notations.

## 3.1 Online Parameter Estimation $\widehat{\theta}_{S_m}(N+1)$

Increasing data length by one adds a new row and column to $A_{S_M}(N)$ and $A_{S_M}^T(N)$ respectively. Instead of multiplying the new matrices of $A_{S_M}(N+1)$ and $A_{S_M}^T(N+1)$, we can add to the previously multiplied $A_{S_M}$ and $A_{S_M}^T$ a new matrix as the added new row and column effect, and find the same result. Therefore, the complexity order of matrix multiplication of the size of $A_{S_M}$ and its transpose in each increasing data length step, $N$ to $N+1$, reduces from $O(N^3)$ to $O(N)$ as shown in the following:

$$F_N = A_{S_M}^T(N)A_{S_M}(N) \tag{3.4}$$

Now adding a new row and column to $A_{S_M}(N)$ and $A_{S_M}^T(N)$ respectively, and then multiplying them to find the $F_{N+1}$ can be shown as:

$$F_{N+1} = F_N + \begin{bmatrix} u_{N+1} \\ u_N \\ u_{N-1} \\ u_{N-2} \\ . \\ . \\ . \\ u_{N-M+2} \end{bmatrix} \begin{bmatrix} u_{N+1} & u_N & u_{N-1} & u_{N-2} & \ldots & u_{N-M+2} \end{bmatrix} \tag{3.5}$$

and if we define an $M \times 1$ column vector U as:

$$U^T = \begin{bmatrix} u_{N+1} & u_N & u_{N-1} & u_{N-2} & \ldots & u_{N-M+2} \end{bmatrix} \tag{3.6}$$

then $F_{N+1}$ is:

$$F_{N+1} = F_N + UU^T = \tag{3.7}$$

$$F_N + \begin{bmatrix} u_{N+1}^2 & u_{N+1}u_N & u_{N+1}u_{N-1} & \ldots & u_{N+1}u_{N-M+2} \\ u_N u_{N+1} & u_N^2 & u_N u_{N-1} & \ldots & u_N u_{N-M+2} \\ u_{N-1}u_{N+1} & u_{N-1}u_N & u_{N-1}^2 & \ldots & u_{N-1}u_{N-M+2} \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ u_{N-M+2}u_{N+1} & u_{N-M+2}u_N & u_{N-M+2}u_{N-1} & \ldots & u_{N-M+2}^2 \end{bmatrix}$$

where $F_N$ is an $M \times M$ matrix defined by $u_1$ to $u_N$ and an $M \times M$ matrix with rank one. As (3.5) indicates, updating $F_N$ to $F_{N+1}$ needs $N$ multiplications at most.

Although $F_{N+1}$ is done in $N$ steps as the data length goes to infinity, its inverse still has a complexity order of $O(N^3)$. So the lemma is to calculate $F_{N+1}^{-1}$ based on $F_N^{-1}$.

As proved in [38], because the rank of matrix $UU^T$ is one:

$$F_{N+1}^{-1} = (F_N + UU^T)^{-1} = F_N^{-1} - \frac{1}{1+g}F_N^{-1}UU^T F_N^{-1} \tag{3.8}$$

where

$$g = \text{trace}(UU^T F_N^{-1}) \tag{3.9}$$

$$g \neq -1$$

Updating the inverse of $F_N$ by using (3.8) and finding $F_{N+1}^{-1}$ decreases the complexity order of matrix inversion from $O(N^3)$ to $O(N^2)$ [38], which decreases the processing time of our online order estimation procedure significantly.

Finally, calling the previously calculated $A_{S_M}^T(N)y^N$ as matrix $C_N$:

$$C_N = A_{S_M}^T(N)y^N \tag{3.10}$$

we can define $C_{N+1}$ as:

$$C_{N+1} = C_N + Uy_{N+1} = C_N + \begin{bmatrix} u_{N+1}y_{N+1} \\ u_N y_{N+1} \\ u_{N-1}y_{N+1} \\ . \\ . \\ . \\ u_{N-M+2}y_{N+1} \end{bmatrix} \tag{3.11}$$

$Uy_{N+1}$ is the effect of the new column and row added to $A^T_{S_M}(N)$ and $y^N$ respectively in their multiplication. It decreases the big O notation of $C_{N+1}$ from $O(N^2)$ to $O(N)$.

Now substituting $F^{-1}_{N+1}$ and $C_{N+1}$ in $\widehat{\theta}_{S_M}(N)$ we have:

$$\widehat{\theta}_{S_M}(N+1) = F^{-1}_{N+1}C_{N+1} = \left( F^{-1}_N - \frac{1}{1+g}F^{-1}_N UU^T F^{-1}_N \right)(C_N + Uy_{N+1}) \tag{3.12}$$

$$\widehat{\theta}_{S_M}(N+1) = \left( I - \frac{1}{1+g}F^{-1}_N UU^T \right)\widehat{\theta}_{S_M}(N) + \tag{3.13}$$
$$\left( I - \frac{1}{1+g}F^{-1}_N UU^T \right) F^{-1}_N Uy_{N+1}$$

and consequently, parameter estimate in subspaces of order $m$, $1 \le m \le N$, is as follows:

$$\widehat{\theta}_{S_m}(N+1) = \tag{3.14}$$

$$\begin{bmatrix} \left( I - \frac{1}{1+g}(A^T_{S_m}A_{S_m})^{-1}U_m U^T_m \right)\widehat{\theta}_{S_m}(N) + \left( I - \frac{1}{1+g}(A^T_{S_m}A_{S_m})^{-1}U_m U^T_m \right)(A^T_{S_m}A_{S_m})^{-1}U_m y_{N+1} \\ 0_{(N+1-m)} \end{bmatrix}$$

Getting back to Chapter 2, the reconstruction error upper bound, as the worst case is used to compare subspaces of different orders. To calculate $z_{S_m}$ recursively we should also define data error, $x_{S_m}$, $\widehat{y}_{S_m}$ and $\widehat{\theta}_{S_m}(N)$ recursively. Replacing (3.14) and $A_{S_M}(N+1)$ in (2.37) we have:

$$\widehat{y}_{S_m}(N+1) = A_{S_M}(N+1)\widehat{\theta}_{S_m}(N+1) \tag{3.15}$$

26

and then updated $x_{S_m}$ would be:

$$x_{S_m}(N+1) = \frac{1}{N}\|y^{N+1} - \hat{y}_{S_m}^{N+1}\|_2^2 \tag{3.16}$$

substituting (3.16) in (2.44):

$$\overline{z_{S_m}(Q(\beta), u^{N+1}, y^{N+1}, Q(\alpha))} = U_{S_m}(Q(\alpha), u^{N+1}, y^{N+1}) + \frac{m}{N+1}\sigma_w^2 + \beta\frac{\sqrt{2m}\sigma^2}{N+1} \tag{3.17}$$

where $U_{S_m}(N+1)$ is:

$$U_{S_m}(u^{N+1}, y^{N+1}, Q(\alpha)) = x_{S_m}(N+1) - m_w(N+1) + \frac{2\alpha^2\sigma_w{}^2}{N+1} + K_{S_m}(\alpha, N+1) \tag{3.18}$$

## 3.2   Online RE-LTI Order Selection Procedure

Through the previous sections we derived the reconstruction error upper bounds based on its previous amounts for subspaces of different orders when the data set size increases by one . Now, putting all those equations together we can modify the proposed offline model selection process in [6] to get the online RE-LTI order selection procedure as follows:

- Step 1: Optimum subspace order, $m^\star$, is chosen for any data length of $N$ through RE-LTI system identification approach (2.57).

- Step 2: For any new data point arrived, any subspace's impulse response is updated using (3.14). Then $\hat{y}_{S_m}(N+1)$, $x_{S_m}(N+1)$ and $\overline{z_{S_m}}(N+1)$ are calculated using (3.15), (3.16), and (3.17) respectively.

- Step 3: Optimum subspace order, $m^\star$, is updated for any new data point arrived, by comparing the probabilistic worst cases for all subspaces of $S_1, S_2, ..., S_M$:

$$S_{m^*}(N+1) = arg\min_{S_m} \overline{z_{S_m}(Q(\beta), u^{N+!}, y^{N+1}, Q(\alpha))} \tag{3.19}$$

where $m^\star$ is the updated optimum order and $\widehat{\theta}_{S_{m^\star}}$ is its corresponding least noisy estimated impulse response.

- Step 4: Go to step 2.

As discussed in Chapter 2, MNDL and RE-LTI model order selections are closely related in that MNDL and RE-LTI use the same reconstruction error, $z_{S_m}$. Therefore, following the above steps also make the MNDL online.

*Important Notation:* In terms of $z_{S_m}$, because the worst case is its upper bound which is the basis of all decisions in RE-LTI, in the remainder of this thesis $\overline{z_{S_m}}$ is denoted by $z_{S_m}$.

## 3.3 Complexity Order of Online RE-LTI Order Selection Procedure

In Section 3.1, we showed that updating $\widehat{\theta}_{S_m}(N)$ to $\widehat{\theta}_{S_m}(N+1)$ is done through a recursive function 3.14 with the complexity order of $O(N^2)$. As a result, updating $x_{S_m}(N)$ to $x_{S_m}(N+1)$ through 3.16 also has the complexity order of $O(N^2)$ because $\hat{y}_{S_m}(N+1)$ can be updated recursively as in 3.15. The upper bound of reconstruction error, $\overline{z_{S_m}}$, is calculated through 3.17 and 3.18. In these equations, the highest complexity order is for updating data error which is $O(N^2)$. Finally, since the optimum order, $m^\star$, in online procedure is chosen based on the upper bound of reconstruction error 3.19, the complexity order of our online procedure is $O(N^2)$ which shows one order decrease compared with the offline RE-LTI order selection algorithm.

## 3.4 Online RE-LTI System Identification and Slowly Varying Systems

As it was described, RE-LTI system identification approach chooses the optimum $m^\star$ for any data of length $N$. In this chapter, we have shown how to update the value of $m^\star$ as the data length

grows. The procedure for each $N$ provides an $m^\star$ (3.19) and consequently a value of $z_{S_{m^\star}}$ (3.17). In online system identification our confidence in the calculation of $m^\star$ will improve as the data length increases. Where should be the stopping point of identification depends on the application. Defining the normalized reconstruction error as follows:

$$Nz_{S_m} = \frac{z_{S_m}}{\|y\|_2^2} \tag{3.20}$$

for each given data set of length $N$, the optimum normalized reconstruction error is calculated as follows:

$$Nz_{S_{m^\star}} = arg \min_{m^\star} Nz_{S_m} \tag{3.21}$$

We call the tolerance for this error Stop-ID threshold. Stop-ID threshold is provided by the user to be an $\epsilon$ value. Once the normalized $z_{S_{m^\star}}$ becomes less than this value the online system identification procedure will halt. Consequently, another potential application of this method will be for slowly varying LTI systems. The online system identification procedure keeps searching for the optimum system until the $\epsilon$ normalized reconstruction error is reached. Then the procedure uses that estimated system until the value of normalized reconstruction error becomes larger than $\epsilon$. This indicates that the system has slowly been changed . Note that up to this stage the $z_{S_m}$ will be updated using the already estimated parameters. At this point, the system identification procedure gets activated till the normalized reconstruction error again becomes less than $\epsilon$. The following steps summarize this procedure:

- Step 1: Value of Stop-ID threshold, $\epsilon$, for the optimum normalized reconstruction error, $z_{S_{m^\star}}$, is selected by the user.

- Step 2: Optimum subspace order, $m^\star$, is chosen for any data length of $N$ through RE-LTI system identification approach (2.57).

- Step 3: Optimum subspace order, $m^\star$, is updated through online RE-LTI order selection procedure (3.2) for any new data point.

- Step 4: At the end of each iteration of online RE-LTI order selection procedure, the normalized reconstruction error is calculated (3.20) and then the optimum normalized reconstruction error is obtained by (3.22)

- Step 5: If $Nz_{S_{m^\star}} < \epsilon$ then the online procedure will halt at this point to update the parameters and optimum order. From now it just updates the $z_{S_m}$ for new arriving data points using the already estimated parameters up to the stopping point and as soon as $Nz_{S_{m^\star}}$ becomes larger than $\epsilon$ the online procedure gets back to its full operation. Otherwise, the online procedure keeps searching for the optimum system.

## 3.5   Simulation Results

Considering an independent identically distributed (i.i.d.) Bernoulli sequence of $\pm 1$ with increasing length from $N = 50$ to $N = 400$ data samples and three different SNRs of 10 dB, 15 dB, and 40 dB as the input data of an LTI system with the following structure:

$$y[n] = \sum_{i=0}^{M} a_i^\star u[n-1] + w[n] \tag{3.22}$$

where $a_i^\star$s are the coefficients of an FIR filter, $h[n]$, of length 50 as:

$$a_i^\star = 0.3(0.5)^{i-1} + 3(i-1)(0.8)^{i-1} \,,\; 0 \le i \le 50 \tag{3.23}$$

Figure 3.1 shows the impulse response as defined in (3.22). Reconstruction error simulation results of such a system are provided in Figures 3.2, 3.3, and 3.4 when $\alpha$ and $\beta$ used in validation and confidence probabilities are chosen as:

$$\alpha = 15\, ln(ln(N)) \tag{3.24}$$

$$\beta = 25\, ln(N) \tag{3.25}$$

Different data lengths and SNRs are used for these figures to illustrate the asymptotic behavior and the robustness of our online order selection method. As $m$ increases, more noise is fitted and the

30

number of unmodeled coefficients is decreased. Therefore, we see the $z_{S_m}$ bias-variance tradeoff behavior which helps us to find the optimum order. For example, in Figure 3.3, when data length has arrived to $N = 400$ estimating more than 20 taps will increase the $z_{S_m}$ as a result of fitting more noise. Comparing these three figures, reveals that when data length is increased because of noise effect reduction (2.55) reconstruction error is decreased. Figure 3.5 shows optimum order as a function of data length for three different SNRs. This figure illustrates that for higher SNRs the chance of choosing the correct order is higher. Furthermore, this figure shows the convergence of the optimum order to the true model order as data length and SNR are increased.

Figure 3.6 shows the minimums of normalized reconstruction error as a function of data length for three different SNRs (3.20). $Nz_{S_{m^*}}$ decreases as data length and SNR are increased because of noise effect reduction. On this figure, the $\epsilon$ is $10^{-2}$ as an example to show the Stop-ID definition. When SNR is 40dB, the $Nz_{S_{m^*}}$ is less than our defined threshold by arriving only 50 samples so the online procedure should be stopped after arriving 50 samples. For 10 dB SNR, the $Nz_{S_{m^*}}$ passes the threshold after arriving 250 samples.
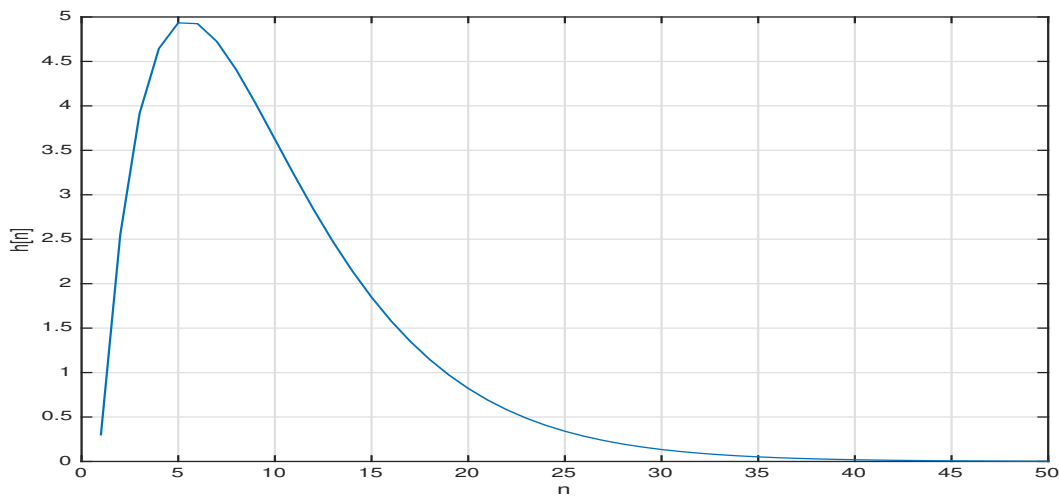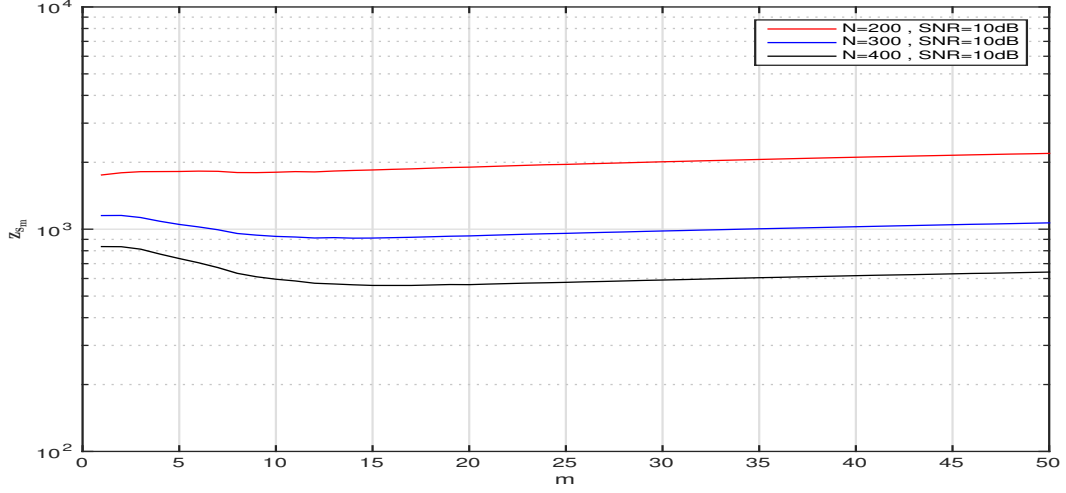


Figure 3.1: True Impulse Response

Figure 3.2: Reconstruction error as a function of $m$ , for data lengths $N = 200$, $N = 300$, and $N = 400$ when SNR=10 dB.
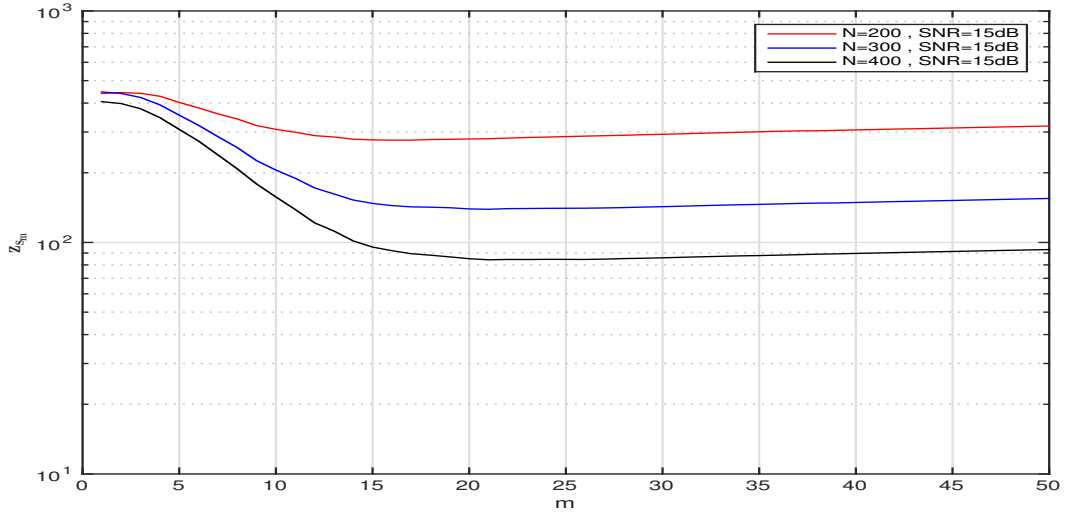


Figure 3.3: Reconstruction error as a function of $m$ , for data lengths $N = 200$, $N = 300$, and $N = 400$ when SNR=15 dB.
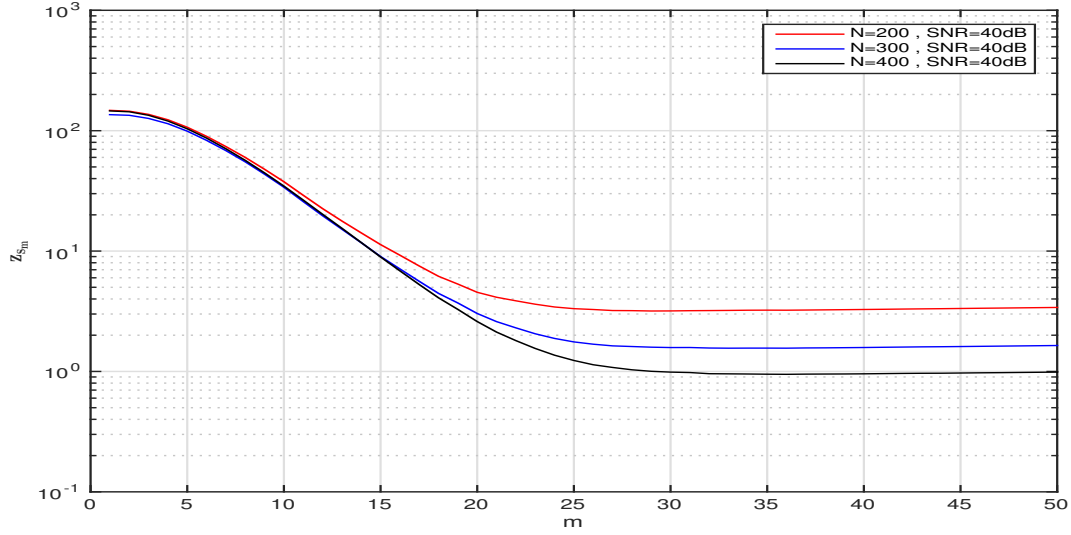
Figure 3.4: Reconstruction error as a function of $m$ , for data lengths $N = 200$, $N = 300$, and $N = 400$ when SNR=40 dB.
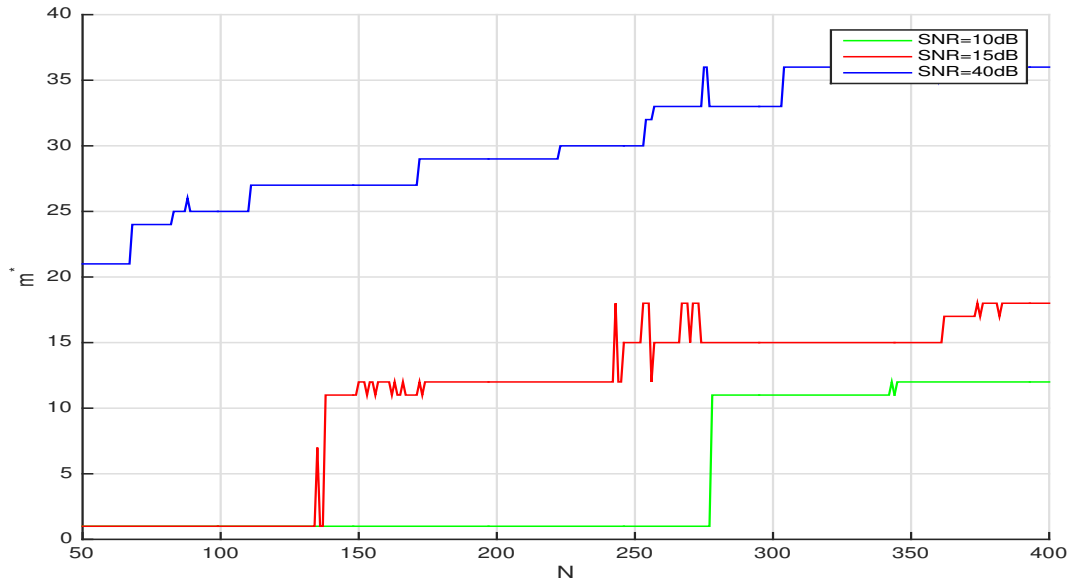


Figure 3.5: Optimum subspace order $m^\star$ as a function of data length $N$ for three different SNRs of 10 dB, 15 dB, and 40 dB.
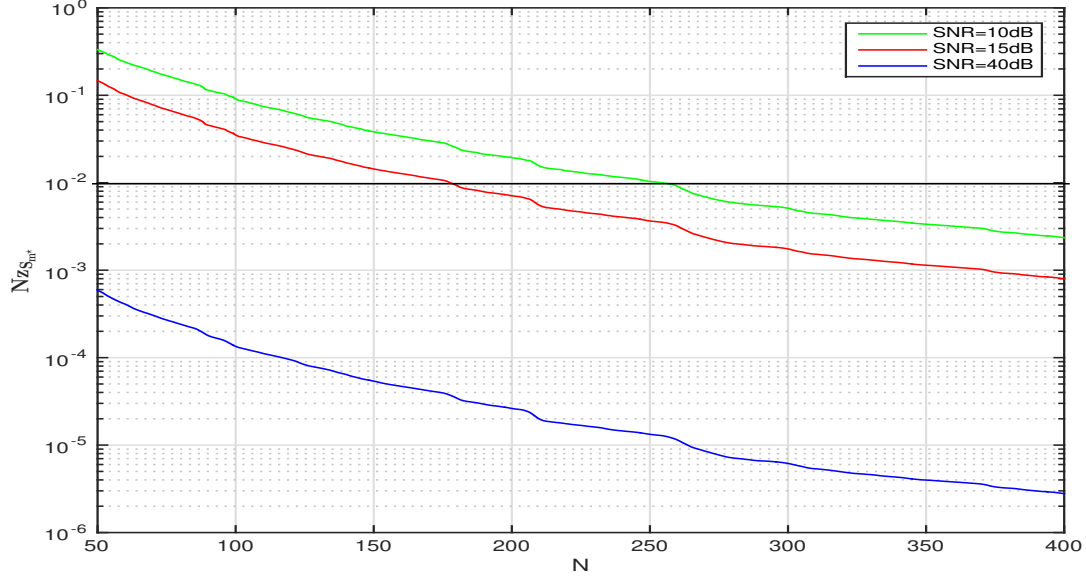
Figure 3.6: Normalized reconstruction error $z_{S_{m^\star}}$ as the function of data length for three different SNRs of 10 dB, 15 dB, and 40 dB.

In Table (3.1), the performance of the proposed online procedure is illustrated in terms of elapsed time as data samples arrive. In this table, $\epsilon$ is $10^{-2}$ and the number of data points needed to pass the $\epsilon$ and their corresponding optimum orders are shown. Four last columns of this table are the elapsed times at $\epsilon$ and at $N = 200$. To show the low complexity and fast computational time of our online procedure, we used the offline RE-LTI order selection in an online manner. Although for larger data sizes the time gap between online and offline algorithm would be greater, still we can see our online procedure is almost 1.7 times faster than the offline algorithm.

| SNR (dB) | Stop-ID Threshold $\epsilon$ | Stop-ID Threshold at N | $m^\star$ at Stop-ID Threshold | Time Passed at $\epsilon$ (Online) (Sec.) | Time Passed at $\epsilon$ (Offline) (Sec.) | Time Passed at N = 200 (Online) (Sec.) | Time Passed at N = 200 (Offline) (Sec.) |
|---|---|---|---|---|---|---|---|
| **10** | $10^{-2}$ | 254 | 16 | 0.9205 | 1.7054 | 2.0029 | 3.4209 |
| **15** | $10^{-2}$ | 178 | 19 | 0.4922 | 0.9804 | 2.015 | 3.4032 |
| **40** | $10^{-2}$ | 50 | 24 | 0.0122 | 0.0122 | 2.0062 | 3.4173 |

Table 3.1: Online RE-LTI vs. Offline RE-LTI Procedure

## 3.6   Practical Application in Power System

In a power system, electrical loads vary considerably throughout the day. These variations are shown in load curves such as in Figure 3.7. Therefore, the load of a power system is not constant that leads to a mismatch between real power generated at the power system and the amount of power consumed at the load side. In power systems, load variation is detected by monitoring the variations in frequency. For instance, when load increases the power system frequency would decrease. Thus additional generation is needed to supply required power to consumers [39].

To overcome the mismatch problem, an adjusting system for the output power of interconnected generators is designed. This unit is known as automatic generation control (AGC). AGC has three kinds of control loops, namely primary, secondary and tertiary ( [39]). Each control loop has different time duration as shown in Table 3.2. There is a primary control loop or governor for every generator. The primary loop controls the variations of frequency around a specified frequency which is defined by the secondary control loop. The secondary control loop or AGC has a proportionalintegral (PI) controller which resets frequency droops to maintain a balance between generation and load. A group of generators which are controlled by a regulation unit and have the same frequency variations is named an area.
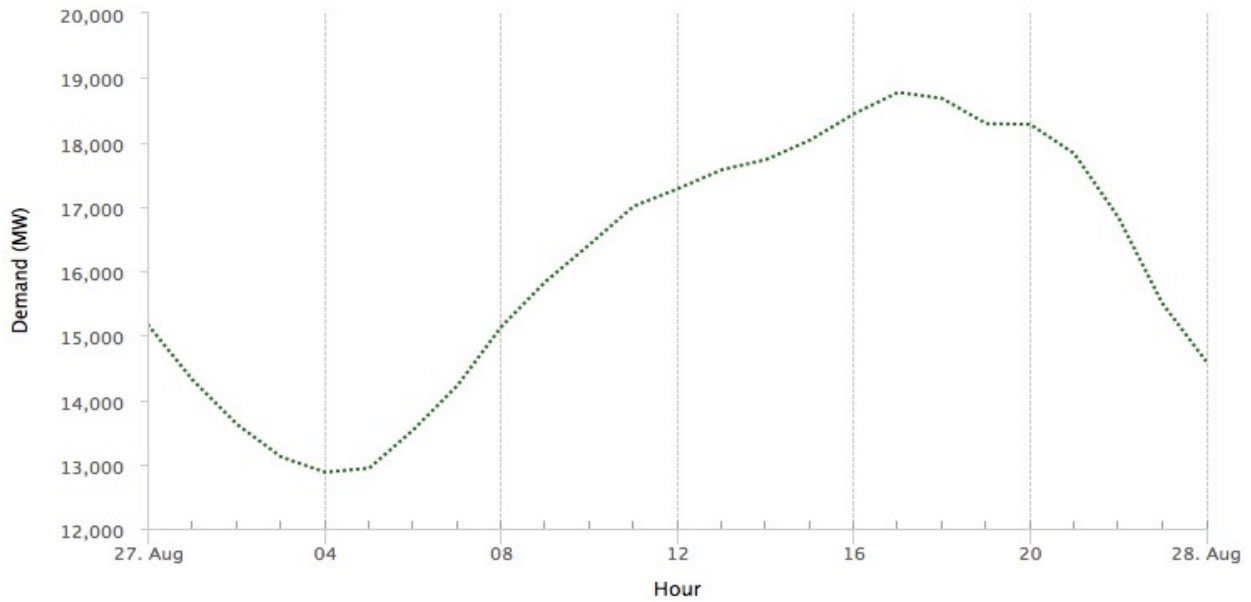
Figure 3.7: Ontario Demand At 12:00pm EST July 26, 2016 [1]

| Control Loop | Time Duration |
|---|---|
| Primary (or Governor) | A few seconds |
| Secondary (or AGC) | Less than one minute |
| Tertiary (or ED) | More than 12 minutes and less than an hour |

Table 3.2: AGC Control Loops' Time Duration

An area is controlled by AGC that is located on the regulation unit. AGC should update PI controller parameters because of load changes in an area. These parameters depend on load-generation scenario which is determined by the third control loop. The tertiary control loop or economic dispatch (ED) determine the generation scenario for every generator based on the amount of load [40]. Therefore, PI controller parameters should be updated at least every 12 minutes.

36

Due to the nonlinearity of power systems, the system is tried to be linearized around its operating point [40]. Then the linear model is used in designing the AGC controller parameters. Although, there are methods of designing robust automatic generation controllers against the changes, another way is designing controller parameters through online algorithms. Considering the second solution, there is a need for an online procedure capable of linearly modeling the changes in system frequency caused by load variation in a fast and consistent enough way.

### 3.6.1   Power System Simulation Results

In this section we are going to prove the ability of our online procedure in a practical application which is a system of two interconnected power plants, hydro and nuclear. This system has a transfer function model as shown in Figure 3.8. We simulated this system in MATLAB/Simulink with a $10\%$ load variation in hydro plant and we collected the frequency variations as the output data from both plants. Having this input and output data available, we can apply our online RE-LTI order selection procedure to estimate a linear model which is used by AGC to redesign its controller parameters over the period of arriving new data. As an example, we consider the frequency variation at hydro plant and we show the simulation results.

Figure 3.9 illustrates $z_{S_m}$ as a function of $m$ when length of data is 3000 and for three different SNRs. For example, when SNR is 40 dB, $z_{S_m}$ starts increasing after 35 estimated parameters of the system impulse response. This tradeoff in behavior of the reconstruction error for lower SNRs can be seen for a smaller number of estimated impulse response taps, which mean some coefficients are set to zero in order to reduce the effect of the noise in the estimated model while achieving the best fit.

Figure 3.10 provides us with a better understanding of the SNR and data length effects on the estimated optimum order. This figure shows the convergence of the optimum order to the true model order as more data becomes available. As SNR grows, this figure shows that the optimum

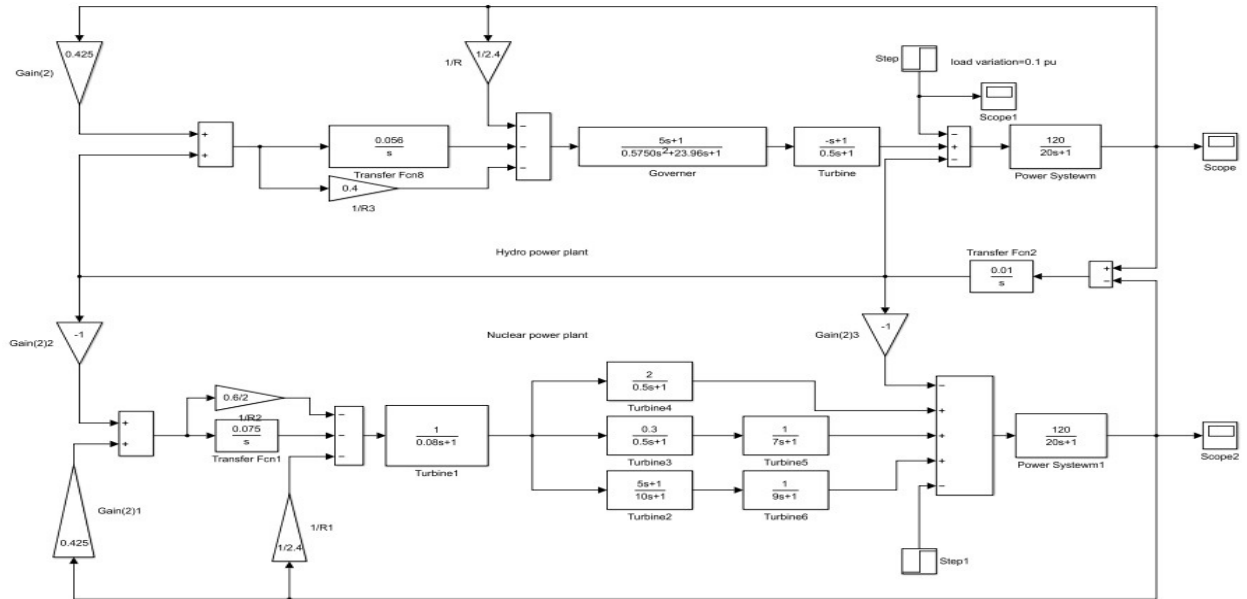order goes faster towards the true model order.



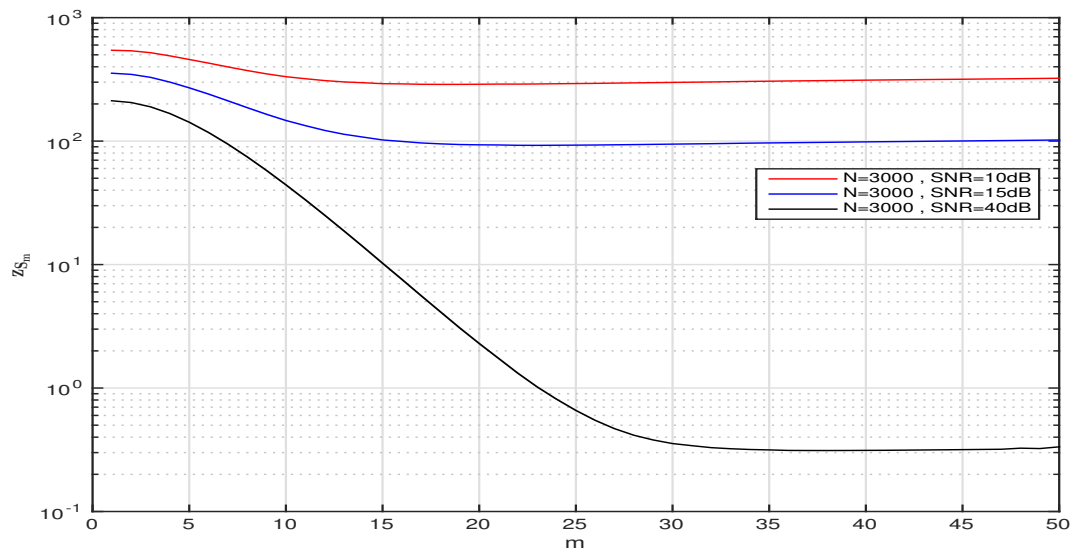Figure 3.8: Transfer function model of an interconnected two-area Hydro-Nuclear system [41]



Figure 3.9: Reconstruction error as a function of $m$ , for data length $N = 3000$ and three different
SNRs of 10 dB, 15dB and 40dB.

Figure 3.10: Optimum subspace order $m^\star$ as a function of data length $N$ for three different SNRs of 10 dB, 15 dB, and 40 dB.

Normalized reconstruction error as a function of $N$ is shown in Figure 3.11 for three different SNRs. An $\epsilon$ ,as the Stop-ID threshold defined by the user, on this figure can validate the estimated model and stop the online procedure after arriving a number of samples.

According to Table 3.2, in tertiary loop we should be able to update the PI controller parameters in less than an hour. While Table 3.3 shows that our online procedure needs almost 16 minutes to model 10000 data samples. On the other hand Figure 3.11 shows after arriving 3000 data samples $Nz_{S_{m^\star}}$ is very small meaning that the PI controller parameters have been updated fast enough to control the frequency deviation.
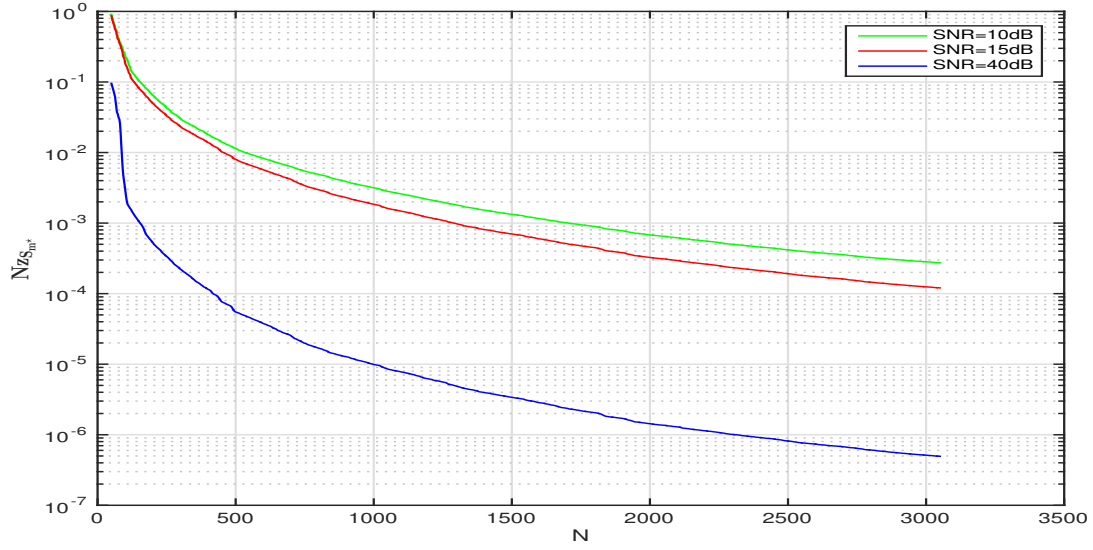
Figure 3.11: Normalized reconstruction error $z_{S_{m^\star}}$ as the function of data length for three different SNRs of 10 dB, 15 dB, and 40 dB.

| Number of Samples Arrived (N) | Elapsed Time at N |
|:---:|:---:|
| 3000 | 85.97 Sec. |
| 5000 | 235.48 Sec. |
| 10000 | 983.21 Sec. |

Table 3.3: Elapsed Time

# Chapter 4

# Optimal Time Delay Estimation

We reviewed different classifications and concepts of time delay estimation in Chapter 1 . Some of the methods looked for the true time delay estimation such as time delay estimated in radars target localization [42], channel equalization in communications [43], echo cancellation [44] and ultrasonic ranging [45]. A few of them looked for the most suitable time delay of a given model. The others estimated the time delay that provides the best fit of a model [2]. As we mentioned in Chapter 2, in this chapter, we propose a method of estimating time delay which results in the best fit of a model.

Considering the model structure in (2.1), if the impulse response of this model has a number of zero coefficients at its beginning, then the input-output relation of that model would be as follows:

$$y[n] = \sum_{i=0}^{M-1} h[n-d]u[n-i] + w[n] = \sum_{i=0}^{M-1} h[n]u[n-d-i] + w[n] \tag{4.1}$$

where $h[n-d]$ is the impulse response with delay as:

$$h[n-d] = [\underbrace{0, \ldots, 0}_{d}, a_0^\star, a_1^\star, \ldots, a_{M-1-d}^\star] \tag{4.2}$$

$u^N$ and $y^N$ are available data, $d$ is the true delay, and $w[n]$ is additive white Gaussian noise with zero mean and variance of $\sigma_w^2$. Here the definition of time delay is the starting point of impulse re-

sponse or the beginning of non-zero part. Employing RE-LTI model order selection, we introduce a process of estimating optimum time delay as:

$$d^\star = arg \min_d \overline{z_{S_{d_{m^\star}}}(Q(\beta), u^N, y^N, Q(\alpha))} \tag{4.3}$$

where $d^\star$ is the optimum estimate of true delay $d$, $S_{d_m}$s for $1 \leq d_m \leq m^\star$ are the competing subspaces with different delays of $0 \leq d \leq m^\star - 1$, $Q(\alpha)$ and $Q(\beta)$ are chosen such that conditions in (2.51), (2.52), (2.53), (2.54) are satisfied, $u^N$ and $y^N$ are available data, and $m^\star$ is the optimum order of the model structure with delay as in (4.1) chosen by RE model order selection in (2.57).

## 4.1   RE Time-Delay Estimation

If $\widehat{\theta}_{S_{m^\star}}$ is the optimum parameter estimate of order $m^\star$ for model structure in (4.1) obtained by minimizing the reconstruction error over all subspaces, we are going to seek in subspaces of $m^\star$ for an optimum estimate of delayed parameter, $\widehat{\theta}_{S_{d_{m^\star}}}$ using (4.3). In other words, the proposed process is twofold:

- Step 1: Estimating an optimum order and parameters of the delayed model based on (2.57).

- Step 2: Estimating the optimum number of zero taps at the beginning of the estimated model in Step 1.

Step 1 has been studied in detail in [6]. In this section, we go through details of step 2 as our contribution to the time delay estimation problem.

## 4.1.1 Estimation of Model Parameters in Subspaces of $m^\star$

To estimate the optimum delay, $d^\star$, we consider subspaces of optimum estimated order $m^\star$ as $1 \leq d_m \leq m^\star$, with the following structure of $\widehat{\theta}_{S_{m^\star}}$:

$$\widehat{\theta}_{S_{m^\star}} = \begin{bmatrix} d_{S_{d_m}} \\ \theta_{S_{d_m}} \end{bmatrix} \tag{4.4}$$

where $d_{S_{d_m}}$ is a zero vector of length $d$ and $\theta_{S_{d_m}}$ is a vector of non-zero parameters of the optimum estimated delayed model with the length of $m^\star - d$. In order to estimate the number of zero coefficients, $d$, and consequently the non-zero parameters of $\theta_{S_{d_m}}$ we use the least square estimator as in (2.38). Thus, we define the Toeplitz matrix generated by the input, $u^N$, with the following format:

$$A_{S_{d_{m^\star}}} = \begin{bmatrix} u_1 & 0 & 0 & \ldots & 0 \\ u_2 & u_1 & 0 & \ldots & 0 \\ u_3 & u_2 & u_1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ u_N & u_{N-1} & u_{N-2} & \ldots & u_{N-m^\star+1} \end{bmatrix} \tag{4.5}$$

when $A_{S_{d_{m^\star}}}$ is an $N \times m^\star$ matrix, it can be represented as:

$$A_{S_{d_{m^\star}}} = \begin{bmatrix} A_{S_{d_m}} & B_{S_{d_m}} \end{bmatrix} \tag{4.6}$$

where $A_{S_{d_m}}$ is the first $d$ columns of Toeplitz matrix $A_{S_{d_m^\star}}$ and $B_{S_{d_m}}$ is the rest of $m^\star - d$ columns of that Toeplitz matrix. Now the estimated parameters of delayed model would be:

$$\widehat{\theta}_{S_{d_m}} = \begin{bmatrix} 0_{(d \times 1)} \\ (B_{S_{d_m}}^T B_{S_{d_m}})^{-1} B_{S_{d_m}}^T y^N \end{bmatrix} \tag{4.7}$$

$B_{S_{d_m}}$ matrices for different $d_m$s are like Toeplitz matrices generated by delayed version of $u^N$s for different lengths of delay, $d$.Therefore, the results of least square error correspond to $d_m$ coefficients of non-zero part of the delayed true impulse response. The rest of $d$ coefficients at the beginning of each vector of estimated impulse response corresponds to the number of zero taps which is the estimated delay.

## 4.2   Simulation Results

To generate Figures 4.1, 4.2, 4.3, and 4.4 Our true model is the delayed version of FIR filter in (3.23) for 23 taps. We considered an i.i.d. Bernoulli sequence of $\pm 1$ with length $N = 300$ and three different SNRs. The true impulse response and its estimates for three SNRs of 10 dB, 15 dB and 40 dB are shown in Figure 4.1, 4.2 and 4.3. As it has been shown, the optimum estimated delay for SNR=10 dB and SNR=15 dB is 25 and for SNR=40 dB is 24. Obviously, for SNR=40 dB our optimum estimated delay is closer to the true time delay than the two other SNRs. This is the result of noise effect reduction. However, it does not mean that our method is not accurate for low SNRs. It must be noted that this method looks for the time delay which results in the best fit of the true model in the presence of the noise. In the next section, we use an index to illustrate our method accuracy or superiority better.
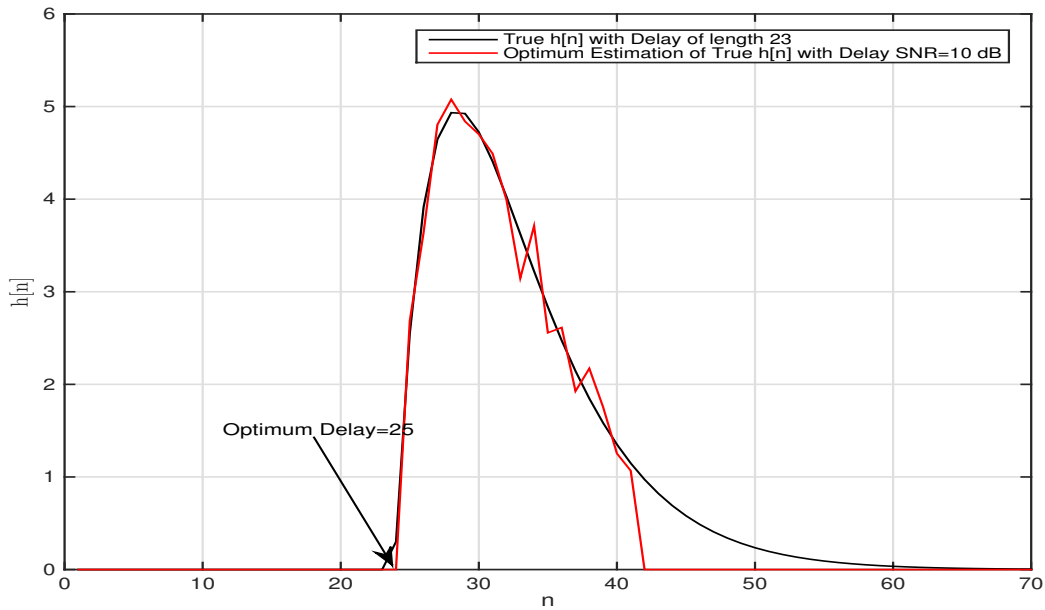


Figure 4.1: True Impulse Response and Optimum Time Delay Estimated Impulse Response with $N = 300, Td = 23$, and $d^\star = 25$
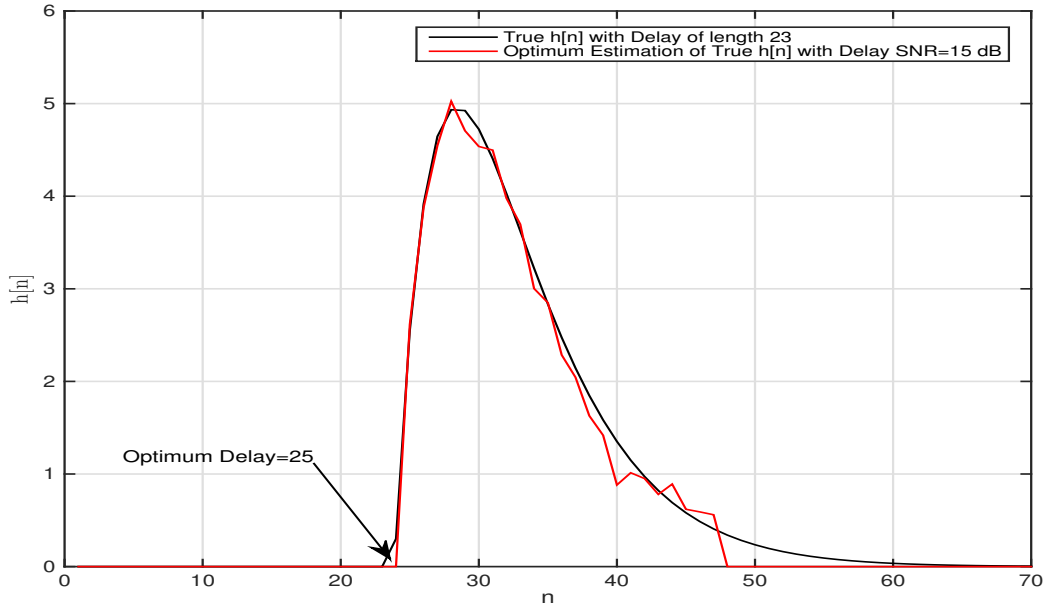
Figure 4.2: True Impulse Response and Optimum Time Delay Estimated Impulse Response with $N = 300$, $Td = 23$, and $d^\star = 25$

In Figure 4.4, reconstruction errors as functions of $m$ for both steps of time delay estimation are illustrated. There are two graphs for each SNR in this figure, one graph corresponds to the $z_{S_m}$ used in model order selection step and the other corresponds to the $z_{S_m}$ used in time delay estimation step. For example, the black graph with SNR=15 dB shows that more than 40 taps estimation increases the reconstruction error as the effect of fitting more noises. The red graph with the same SNR as the black one shows the calculated reconstruction error in step 2 of time delay estimation. The red graph has its minimum at $m = 23$ which is the number of zero taps at the beginning of the estimated impulse response.
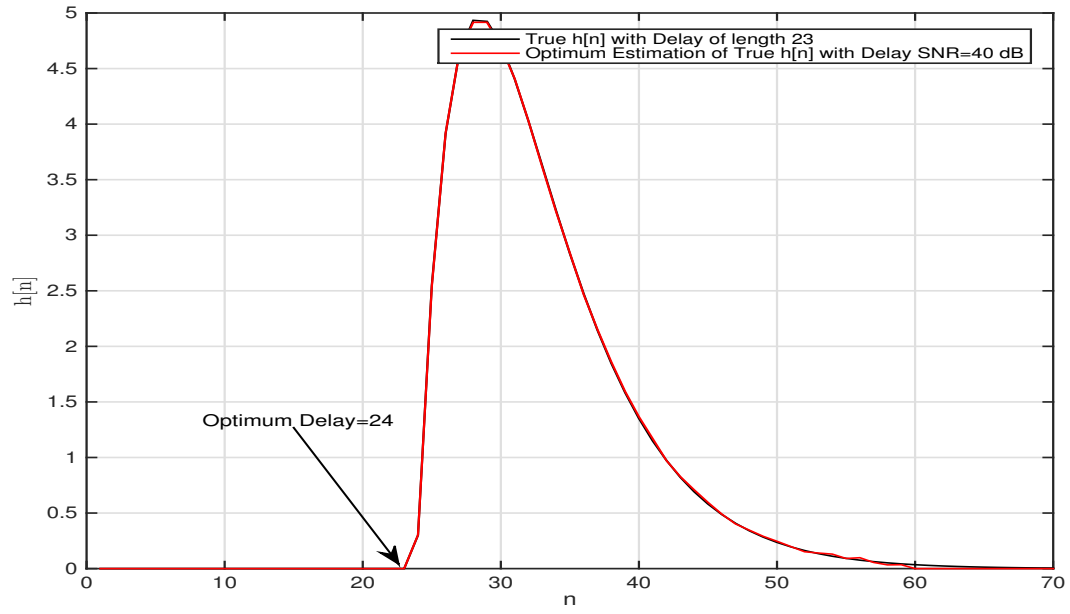
Figure 4.3: True Impulse Response and Optimum Time Delay Estimated Impulse Response with $N = 300$, $Td = 23$, and $d^{\star} = 24$



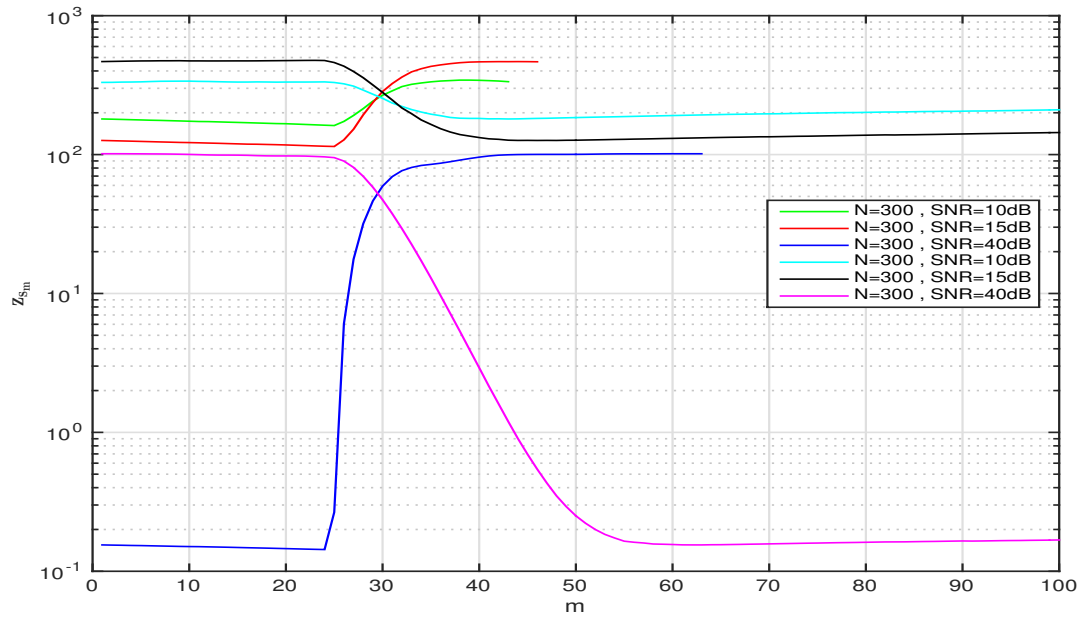Figure 4.4: Reconstruction error $z_{S_m}$ as the function of $m$ for three different SNRs of 10 dB, 15 dB, and 40 dB. and Td=23

## 4.2.1 Comparison between Proposed and Existing Methods

An index to measure the quality of an estimated time delay as the best fit in a system identification scenario is presented in [2] as follows:

$$fit = 100 \left( 1 - \frac{\|\hat{y} - u\|}{\|y - \overline{y}\|} \right) (\%) \tag{4.8}$$

where $y$ is the observed output, $\hat{y}$ is the estimated output and $\overline{y}$ is the noiseless output. Some time-delay estimation methods investigated thoroughly in [26], such as the Cumulative Sum (CUSUM), delayest algorithm and separating time delay from dynamics in the frequency domain, are compared in terms of the fit index when the process model is:

$$G(n) = 0.256(0.909)^n u(n) - 0.333(0.815)^n u(n) \tag{4.9}$$

when the true time delay at the beginning of (4.9) is 11. To excite this model, 1000 samples of a pseudo-random binary sequence (PRBS) are used. Three scenarios are considered. First, the order of (4.9) is identified correctly by all methods. Second, A wrong order for our considered model is identified by all methods. Finally, an output error model with correct order is applied. In these three scenarios considered methods are: CUSUM-CrossCorr, CUSUM-FIR and CUSUM-ARX-High, frequency domain technique, delayest and OPT which are studied completely in [26] and [2]. CUSUM-CrossCorr use crosscorrelation to estimate the impulse response, CUSUM-FIR uses an FIR model to model the impulse response and CSUM-ARX-High uses a high order autoregressive to model the impulse response. OPT is a two-step optimization problem which first tries to predict the best model and then it estimates the time delay. All techniques are applied when the true time delay, $n_k$ is 11. Data in Table 4.1 are adopted from [2]. The last row of this table shows the estimation of the time delay for the same settings as the other methods by our proposed RE time-delay estimation method. Since our method is not order dependent and it selects the best order

using reconstruction error, therefore the second and third scenario does not apply to our method and just SNR variations make changes in our results.

In the absence of the noise, like OPT and delayest, our method provides the true delay as the best fit. In case of SNR=3 dB and SNR=30dB, RE-LTI order selection method will set some impulse response parameters to zero in order not to fit too much noise to parameters which are too small. Therefore, the estimated result will not fit $100\%$ the true model and as we mentioned earlier, it would be the best fit of the model. As the results show, RE time-delay estimation method would provide better fit for low SNRs than OPT and delayest.

| | Example 1 - ARX $[n_a n_b]^T = [22]$ | | | | | | Example 2 - ARX $[n_a n_b]^T = [11]$ | | Example 3 - OE $[n_b n_f]^T = [22]$ | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | SNR $\infty$ | | SNR 30 | | SNR 3 | | SNR30 | | SNR 30 | | SNR 3 | |
| | $n_k$ | FIT | $n_k$ | FIT | $n_k$ | FIT | $n_k$ | FIT | $n_k$ | FIT | $n_k$ | FIT |
| $n_{k_0}$ | 11 | 100 | 11 | -14.5 | 11 | -11.1 | 18 | 7.76 | 11 | 98.1 | 11 | 94.0 |
| $n_k^\star$ | 11 | 100 | 12 | 34.3 | 13 | 25.0 | 12 | 14.9 | 11 | 98.1 | 11 | 94.0 |
| CUSUM-CrossCorr | 14 | 23.5 | 14 | 20.3 | 15 | 2.33 | 14 | 5.27 | 14 | 61.3 | 15 | 60.6 |
| CUSUM-FIR | 14 | 23.5 | 14 | 20.3 | 14 | 4.96 | 14 | 5.27 | 14 | 61.3 | 14 | 61.0 |
| CUSUM-ARX-High | 14 | 23.5 | 14 | 20.3 | 15 | 2.33 | 14 | 5.27 | 14 | 61.3 | 15 | 60.6 |
| Freq-Domain | 12 | $\infty$ | 12 | 34.3 | 12 | 6.41 | 12 | 14.9 | 12 | 65.3 | 12 | 65.2 |
| delayest | 11 | 100 | 11 | -14.5 | 11 | -11.1 | 11 | -26.6 | 11 | 98.1 | 11 | 94.0 |
| OPT | 11 | 100 | 13 | 19.1 | 20 | 11.1 | 21 | 7.60 | 11 | 98.1 | 11 | 94.0 |
| RE | 11 | 100 | 12 | 96.29 | 15 | 67.23 | 12 | 96.29 | 12 | 96.29 | 15 | 67.23 |

Table 4.1: Comparison of Time Delay Estimation Techniques - Noise-Free FIT(%) and $n_k$ Values [2]

# Chapter 5

# Conclusions and Future Works

In this thesis, we studied one powerful method of model order selection: RE-LTI order selection. We started Chapter 1 with an overview of information theoretic criteria such as AIC, BIC, two-stage MDL and we discussed some of their variants that have been proposed to improve their performances. We continued this chapter with the review of time delay estimation methods. In Chapter 2 we presented a short review of system identification. We studied existing model order selection criteria more in detail and we discussed their shortcomings. In this chapter, we provided a brief review of RE-LTI model order selection as the basis of our contributions. At the end of this chapter, we stated the type of time delay we wanted to estimate.

In Chapter 3 we proposed our Online RE-LTI Order Selection method. In our online procedure, we updated the optimum order and its corresponding estimated parameters recursively. We showed one order decrease, from $O(N^3)$ to $O(N^2)$, in complexity order of our online procedure compared with the offline RE-LTI order selection method. Stop-ID threshold as a criterion to validate our online order selection procedure was introduced in this chapter. Stop-ID threshold defined by the user can halt the procedure and restart it. We discussed the advantages of our online procedure in case of slowly varying systems. Our simulation results showed the improvement in consistency and computational time of our online procedure compared with offline RE-LTI method. Estimating

the impulse response of an AGC in a power system of two interconnected hydro and nuclear plants was presented at the end of this chapter. We suggested updating the controller parameters of AGC by using our online RE-LTI order selection procedure to improve the power system dynamic performance.

In Chapter 4 we used the RE-LTI order selection in two steps to estimate the time delay that provides the best fit of a true model. First, we estimated the optimum order of the true model and then we estimated the number of zero taps at the beginning of the estimated model. We compared the simulation results of our method with the existing methods in the literature, in terms of FIT index, and showed the superiority of our proposed method.

Applying RE-LTI online order selection procedure to updating orders of real-time applications would lead to great results in future.

# Bibliography

[1] The Independent Electricity System Operator, "Ontario demand at 12:00pm est august 26." `http://www.ieso.ca/pages/power-data/default.aspx`, 2016.

[2] V. A. O. Alves, R. J. C. de Godoy, and C. Garcia, "Optimal time delay estimation for system identification," in *2013 American Control Conference*, pp. 95–100, June 2013.

[3] T. Soderstrom and P. G. Stoica, *System Identification*. Prentice Hall, 1989.

[4] L.Ljung, *System Identification, Theory for the User*. Prentice Hall, 1999.

[5] L. Fu and P. Li, "The research survey of system identification method," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on*, vol. 2, pp. 397–401, Aug 2013.

[6] S. Beheshti and M. A. Dahleh, "Noisy data and impulse response estimation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 510–521, Feb 2010.

[7] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, Dec 1974.

[8] Q. Ding and S. Kay, "Inconsistency of the mdl: On the performance of model order selection criteria with increasing signal-to-noise ratio," *IEEE Transactions on Signal Processing*, vol. 59, pp. 1959–1969, May 2011.

[9] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, pp. 36–47, July 2004.

[10] A. Mariani, A. Giorgetti, and M. Chiani, "Model order selection based on information theoretic criteria: Design of the penalty," *IEEE Transactions on Signal Processing*, vol. 63, pp. 2779–2789, June 2015.

[11] S. Kallummil and S. Kalyani, "High snr consistent linear model order selection and subset selection," *IEEE Transactions on Signal Processing*, vol. 64, pp. 4307–4322, Aug 2016.

[12] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[13] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "Bayesian model comparison and the bic for regression models," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6362–6366, May 2013.

[14] P. Stoica and P. Babu, "On the proper forms of bic for model order selection," *IEEE Transactions on Signal Processing*, vol. 60, pp. 4956–4961, Sept 2012.

[15] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Transactions on Information Theory*, vol. 44, pp. 2743–2760, Oct 1998.

[16] S. Kay, "Exponentially embedded families - new approaches to model order estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, pp. 333–345, Jan 2005.

[17] P. Stoica and P. Babu, "On the exponentially embedded family (eef) rule for model order selection," *IEEE Signal Processing Letters*, vol. 19, pp. 551–554, Sept 2012.

[18] A. K. Seghouane, "Vector autoregressive model-order selection from finite samples using kullback's symmetric divergence," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, pp. 2327–2335, Oct 2006.

[19] L. Qu, Q. Sun, T. Yang, L. Zhang, and Y. Sun, "Time-delay estimation for ground penetrating radar using esprit with improved spatial smoothing technique," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 1315–1319, Aug 2014.

[20] C. L. Bastard, Y. Wang, V. Baltazart, and X. Drobert, "Time delay and permittivity estimation by ground-penetrating radar with support vector regression," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 873–877, April 2014.

[21] J. Li, C. L. Bastard, Y. Wang, G. Wei, B. Ma, and M. Sun, "Enhanced gpr signal for layered media time-delay estimation in low-snr scenario," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 299–303, March 2016.

[22] F. Fereidoony, M. A. Sebt, S. Chamaani, and S. A. Mirtaheri, "Model-based super-resolution time-delay estimation with sample rate consideration," *IET Signal Processing*, vol. 10, no. 4, pp. 376–384, 2016.

[23] C. M. Ionescu, R. Hodrea, and R. D. Keyser, "Variable time-delay estimation for anesthesia control during intensive care," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 363–369, Feb 2011.

[24] G. K. Rohde, F. Bucholtz, and J. M. Nichols, "Maximum empirical likelihood estimation of time delay in independently and identically distributed noise," *IET Signal Processing*, vol. 8, pp. 720–728, September 2014.

[25] Y. Huang, J. Lin, and M. Giess, "High accuracy time delay measurements for band-pass signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, pp. 2998–3005, Nov 2013.

[26] S. Bjorkund, *A Survey and Comparison of Time-Delay Estimation Methods in Linear Systems*. PhD thesis, Linkopings universitet, Department of Electrical Engineering, Jun 2003.

[27] M. Davanipour, M. Zekri, and F. Sheikholeslam, "Fuzzy wavelet neural networks with hybrid algorithm in nonlinear system identification," in *Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on*, vol. 1, pp. 153–156, June 2011.

[28] P. M. Makila, "On optimal lti approximation of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1178–1182, July 2004.

[29] M. Salimifard and A. A. Safavi, "Nonlinear system identification based on a novel adaptive fuzzy wavelet neural network," in *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, pp. 1–5, May 2013.

[30] L. Ljung, "Approaches to identification of nonlinear systems," in *Proceedings of the 29th Chinese Control Conference*, pp. 1–5, July 2010.

[31] A. A. Adeniran and S. E. Ferik, "Modeling and identification of nonlinear systems: A review of the multimodel approach–part 1," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP, no. 99, pp. 1–11, 2016.

[32] G. Pillonetto, F. Dinuzzo, T. Chen, G. D. Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657 – 682, 2014.

[33] M. Barkat, *Signal Detection and Estimation*. Artech House, 2005.

[34] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform cr bound," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2026–2041, Aug 1996.

[35] S. Beheshti and M. A. Dahleh, "Lti systems, additive noise, and order estimation," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 6, pp. 6491–6496 Vol.6, Dec 2003.

[36] E. Gassiat and R. van Handel, "Consistent order estimation and minimal penalties," *IEEE Transactions on Information Theory*, vol. 59, pp. 1115–1128, Feb 2013.

[37] S. Beheshti and M. A. Dahleh, "A new information-theoretic approach to signal denoising and best basis selection," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3613–3624, Oct 2005.

[38] K. S. Miller, "On the inverse of the sum of matrices," *Mathematics Magazine*, vol. 54, pp. 67–72, Mar. 1981.

[39] N. Hasan, S. Khatoon, Ibraheem, Nizamuddin, and Y. Singh, "Automatic generation control problem in interconnected power systems," *National Conference on Emerging Trends in Electrical, Instrumentation and Communication Engineering*, vol. 3, no. 2, pp. 46–52, 2013.

[40] M. Azzam, "Robust automatic generation control," *Energy Conversion and Management*, vol. 40, pp. 1413–1421, Dec 1998.

[41] N. Hasan, Ibraheem, and S. Farooq, "Real time simulation of automatic generation control for interconnected power system," *International Journal on Electrical Engineering and Informatics*, vol. 4, pp. 40–51, Mar 2012.

[42] A. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 527–533, Jun 1981.

[43] G. L. Turin, "Introduction to spread-spectrum antimultipath techniques and their application to urban digital radio," *Proceedings of the IEEE*, vol. 68, pp. 328–353, March 1980.

[44] Y. Lu, R. Fowler, W. Tian, and L. Thompson, "Enhancing echo cancellation via estimation of delay," *IEEE Transactions on Signal Processing*, vol. 53, pp. 4159–4168, Nov 2005.

[45] A. K. Nandi, "On the subsample time delay estimation of narrowband ultrasonic echoes," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, pp. 993–1001, Nov 1995.

[46] J. Ianniello, "Time delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, pp. 998–1003, Dec 1982.