ABSTRACT

System identification aims to develop mathematical models for dynamical systems using measured input and output signals. Model structure selection is one of the important steps in a system identification process. Several important criteria for a desirable model structure include its accuracy in model output, model residuals, final prediction error (FPE) and loss function. For this project, linear model was used that is ARX model. Dynamic models are processed before attempting model parameter or structure estimation procedures. This project was carried out using the Graphical User Interface (GUI) in MATLAB application. This project explores the performance of difference parameter estimation method in identification. Parameter estimation is one of the most important things that need to be considered in accurately describing system behavior through mathematical models such as statistical probability distribution functions, parametric dynamic models, and data-based models. The effectiveness of least square and instrumental variable (IV) estimator are investigated. This project brief overview of the system identification process. Vital to pay consideration on all parts of this methodology, from the experiment design to model validation in order to get best results. The usefulness and applicability of both methods are discussed based on the results of model output, model residuals, final prediction error (FPE) and loss function. In conclusion, this investigation found that instrumental variable performs slightly better because the noise is purposely made correlated with input. In opposition, least square is known be better when noise is uncorrelated. This prove the theoretical that instrumental variable method can balance the impacts of a more general class of noise signals while least square method can minimizes the sum of the squared prediction errors between the prediction model and the output data .

ABSTRAK

Pengenalpastian sistem bertujuan untuk membangunkan model matematik untuk sistem dinamik menggunakan isyarat input dan output yang diukur. Pemilihan struktur model adalah salah satu langkah penting dalam proses pengenalan sistem. Beberapa kriteria penting untuk struktur model wajar termasuk ketepatan dalam output model, sisa model, ralat ramalan akhir (FPE) dan fungsi kerugian. Untuk projek ini, model linear telah digunakan iaitu model ARX. Model dinamik diproses sebelum cuba model parameter atau struktur prosedur anggaran. Projek ini telah dijalankan dengan menggunakan muka grafik pengguna (GUI) dalam aplikasi MATLAB. Projek ini meneroka prestasi kaedah anggaran perbezaan parameter dalam pengenalan. Anggaran parameter adalah salah satu perkara yang paling penting yang perlu dipertimbangkan dalam tingkah laku sistem menggambarkan dengan tepat melalui model matematik seperti fungsi taburan kebarangkalian statistik, model dinamik parametrik, dan model berasaskan data.Keberkesanan "kurangnya pembolehubah persegi" dan "instrumental penganggar" disiasat. Projek gambaran ringkas mengenai proses pengenalan sistem. Penting untuk menekankan pertimbangan pada semua bahagian metodologi ini, dari reka bentuk eksperimen untuk model pengesahan untuk mendapatkan hasil yang terbaik. Kegunaan dan kesesuaian kedua-dua kaedah akan dibincangkan berdasarkan keputusan output model, sisa model, ralat ramalan akhir (FPE) dan fungsi kerugian. Kesimpulannya, penyiasatan ini mendapati bahawa pembolehubah instrumental melakukan lebih baik kerana bunyi dibuat berkait rapat dengan input. Manakala, persegi ini dikenali menjadi lebih baik apabila bunyi adalah korelasi. Ini membuktikan teori bahawa kaedah pembolehubah instrumental boleh mengimbangi kesan daripada kelas yang lebih umum isyarat bunyi manakala kaedah persegi kurangnya boleh mengurangkan jumlah kesilapan ramalan kuasa dua antara model ramalan dan data output.

ACKNOWLEDGEMENTS

First and above all, I praise Allah the almighty for providing me this opportunity and granting me the capability to proceed successfully.

This project appears in its current form due to the assistance and guidance of several people. Special appreciation goes to my supervisor, Dr. Md Fahmi bin Abd Samad @ Mahmood, for his relentless support, thoughtful guidance and correction of the project. I attribute the level of my bachelor degree to his encouragement and effort and without him, this project would not have been completed or written. One simply could not wish for a better or friendlier supervisor. I am grateful to Dr. Mohd Basri Bin Ali for his encouragement and practical advice. I am also thankful to him for reading my reports, commenting on my views and helping me understand and enrich my ideas. My thanks also go to Dr Siti Hajar for providing many valuable comments that improved the presentation and contents of this dissertation.

My deepest gratitude goes to my beloved parents, Mr. Hj Mohd Rafie bin Hj Tibek and Mrs. Hjh Noorafizah bt Abdul Aziz for their material and spiritual support in all aspects of my life. Thank you both for giving me strength to reach for the stars and chase my dreams. I also would like to thank to all my siblings for their endless love, prayers and encouragement. I am particularly grateful to my sweetheart for his devoted support throughout my study.

Last but not least, thanks to all my close friends Nurfa bt Fahar and Siti Nur Ariffah Alia for the joyful gatherings and all their supports. I deeply appreciate to all my friends that helped me stay through these difficult years. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

NUR AFIFAH FAIZAH BT HJ MOHD RAFIE, JUNE 2016

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR - AutoRegressive

ARMAX - AutoRegressive Moving Average with eXogenous input

ARX - AutoRegresive with eXogenous input

FPE- Final Prediction Error

GUI - Graphical User Interface

IV- Instrumental Variable

LS - Least squares

MATLAB- Matrix Laboratory

NAR - Nonlinear AutoRegressive

NARMAX - Nonlinear AutoRegressive Moving Average with eXogenous input

NARX - Nonlinear AutoRegresive with eXogenous input

NOE - Nonlinear output order

SI - System Identification

LIST OF SYMBOLS/NOTATIONS

$\hat{\theta}$ = Estimated parameter vector

ϕ = Matrix regressor

ε_M =Model set

σ^2 =Constant variance

ψ = Constraints

θ = Parameter vector

D_M =Open subset

Z =Vector output

CHAPTER ONE

INTRODUCTION

1.1 Background Study

The aim of this project is to perform system identification using linear difference equation model and to identify the performance of different parameter estimation method in identification. System identification is the process of building mathematical models of dynamical systems from observations and prior knowledge (Norton,1988). The basic of a system identification experiment is the set of system inputs and outputs. System defined as a set of element standing in interrelation among themselves and with environment. The input is generated by the environment and influences the system, whereas the output is generated by the system and influences the environment. The system identification is important to estimate the dynamics of the input and output data, and to get Equation 1more insight into the system under investigation. The mathematical model is a basic simulator and theoretical analysis of dynamic systems, as the control system. Mathematical models are the foundation of most scientific and engineering methods. Models can be obtained by either a theoretical approach based on physical laws, or an experimental approach based on obtained measurements from the system. System identification deals with the problem of building mathematical models of dynamic systems based on observed data from the system, and is thus an experimental modelling method. Experimental modelling consists in adjusting the parameters of a chosen model structure in order to fit the model output to measured system output data. The fitting can be done either manually by graphical methods, or by having a computer perform an optimization procedure. In this project, fit the model will be performed using GUI ‗ident' in MATLAB. Parameter is a set of methods to obtain a physical model of a dynamic system from experimental data. In this project will explore the various existing methods of parameter estimation but more focused on instrumental variable and least square method. Both methods are based on a prediction model structure which is linearly parameter.

1.2 Problem Statement

Many mathematical models related to linear and non-linear systems have been suggested in assisting the system identification problem. One of the major problems of system identification is to find an optimal model, which is the simplest model that can adequately represent the dynamic systems or a parsimonious model. Basically, there are two main problems in the system identification (Hong et al, 2008). First is to determine which model structure describes the functional relationship between input and output variables of the system dynamics and second to estimate the parameters of model that determine chosen or elected model structure. There are many different methods that can be used to determine the parameter of a model, and also there are different criteria on which method should be selected. This project will investigate how the instrumental variable and least square method can be applied. This will be followed by a comparison of performance between the two methods as parameter estimation method. In theoretical, least square method the parameter estimates minimizes the entirety of the squared prediction errors between the prediction model and output data. While, instrumental variable method has ability to counteract the effects of a more general class of noise signals. For more detailed information, the difference between least square and instrumental variable will be discussed in Chapter 3.

1.3 Objective

There are an objectives that need to be achieved at the end of this project, which are:

i) To perform system identification using linear differential equation model.

ii) To investigate the performance of difference parameter estimation method in identification.

1.4 Scope of Project

The scope of this project are:

i) Identification will be performed using GUI ‗ident' in MATLAB.

ii) Model to be used is autoregressive with exogeneous input (ARX) model.

iii) Performance comparison will be made from various perspective such as time plot and data spectra.

CHAPTER TWO

LITERATURE REVIEW

2.1 System Identification

One of the first system identification experiments was carried out by Hailey in 1704 when he realized comet sightings in the years 1531, 1607 and 1682 related to an object. In order to further understand the behavior of these objects orbit he calculated the parameters by using the Newton and gravity theoretical (Murdin and Allen,1979). In the case of Hailey model structure provided by an aggregate work the Gallileo, Kepler and Newton. Hailey's combined wealth of background knowledge with a collection of observations and concluded that the appropriate model systems. System identification is the process of developing or improving a quantitative numerical model from a set of input and output data that represents the response of a dynamic system. It is necessary to use model to describe the relationships among the system variables. The developed model has the trademark performance similar like the unknown system. There are four main steps involved in system identification and these are data acquisition, model structure selection, parameter estimation and model validation (Ljung, 1999). The first step is collecting useful data. Second step is choosing a convenient model set. Define the form of the model to match the physics of the system. Use priori knowledge and engineering intuition. Model structure selection stage can be divided into two sequential steps, selection on the type of model to represent the system and construction of the correct or optimum model structure

(Abd Samad,2014). The third step in system identification is the parameter estimation where the estimations of parameters taking into account the selected model structure are assessed. The fourth step is the model validity tests expected to confirm or validate the representation of the last model chose. The flow chart is given as Figure 2.1.

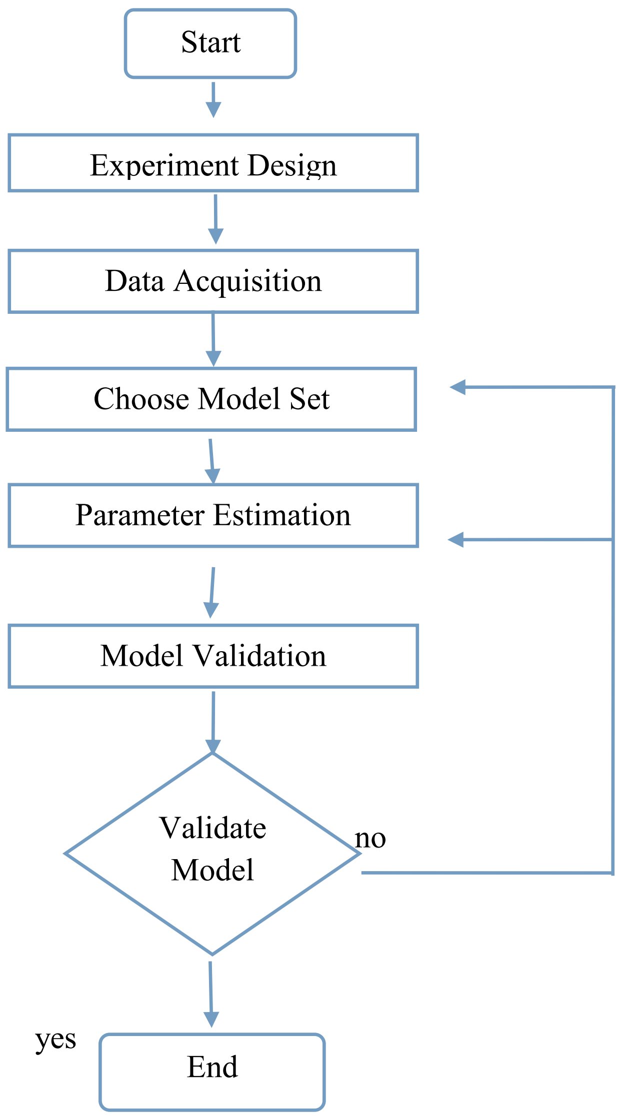

Figure 2.1: The system identification loop

2.2 Mathematical Modeling

Mathematical models of natural and man-made systems play an essential role in today‘s science and technology. Mathematical model is a portrayal of a system utilizing mathematical concepts and language. Mathematical modeling is the process of developing a mathematical model. Choosing a suitable model structure is important procedure before its estimation. The choise of model structure is based upon comprehension of the physical systems.

2.2.1 Types of Mathematical Models

A model is a mathematical representation of a physical, biological or data system. Models permit us to reason about a system and make predictions about how a system will act. There are a variety of model structures available to assist in modeling a system. There are three main types of models are common in system identification: theoretical or white-box model, empirical or black-box model and semi-empirical or gray-box model. System identification methods can deal with an extensive variety of system dynamics without knowledge of the actual system physics,there by lessening the designing exertion required to create models.

2.2.1.1 Theoretical or White-box Identification

White box models are completely derived from first principles in the respective scientific fields. Typically examples are found in mechanical and electrical system where the physical laws, can be used to predict the effects based on data. White box models are models in which priori knowledge dominates the model. The strength of this box identification is that it will permit the user to endeavor invariant prior knowledge.

2.2.1.2 Empirical or Black box Identification

An assortment of parametric model structures are accessible to help in modeling an unknown system. Parametric models despict systems in terms of differential equations and transfer functions. These models give understanding into the system physics and compact model structures. It is often valuable to test a various structures to decide the best one. A parametric model structure is otherwise called a black-box model. The black box is an abstract class that represents a concrete open system that can be viewed purely in terms of stimulus input and output reaction. Both model structure and parameter are determined from experimental modeling. For building black box model, no or very little prior knowledge is exploited. The model parameters have no direct relationship to first principle.

2.2.1.3 Gray box Identification

Gray box model is a combination of white box and black box models. The amount of first principles used is in balance with the amount of experimental data. Besides the knowledge from first principles and the information contained in the measurement data other knowledge sources such as qualitative knowledge formulated in rules may also be utilized in gray box models. Typically, the determination of the model structure relies strongly on prior knowledge while the model parameters are mainly determined by measurement data.

2.2.1.4 Comparison of White Box, Black Box and Gray Box Model.

Table 2.1 gives an overview of the differences between these modeling approaches. The blank fields for gray box models can be almost any combination of white and black box models. The advantages and drawbacks of gray box models lie somewhere between white and black box models, make them a good compromise in practice.

Table 2.1: Differences between white box, gray box and back box models.

	White Box	Gray Box	Black Box
Information sources	First principles insights	Qualitative knowledge Some insight and some data	Experiments data
Features	Good extrapolation, Good understanding high reliability, scalable.	-	Short development time. Little domain expertise requirement can be used in addition even for not understood process

Drawbacks	Time consuming, detailed domain expertise requirement knowledge restricts accuracy only for well understood process.	-	No reliable extrapolation, Not scalable, Data restricts accuracy, Little understanding
Application areas	Planning, construction, design rather simple processes	-	Only for existing processes rather complex processes

2.2.2 Classification of Mathematical Models

Mathematical models may be classified according to their structure and nature. Thus, mathematical models may be linear or non-linear, static or dynamic,deterministic or stochastic and discrete or continuous.

Linear or Non linear, according to the basic equation describing them are linear or non linear. If all the operators in mathematical model shows linearity, the subsequent mathematical model is defined as linear. A linear model uses parameters that are consistent and not vary throughout a simulation. A non linear model presents dependent parameters that are permitted to vary throughout the course of a simulation run, and its utilization gets to be vital where interdependencies between parameter cannot be considered.

Static or Dynamic, according to the time variations in the system are not or are taken into account. A static model or steady-state model computes the system in equilibrium, and along these lines is time-invariant. A static model cannot be changed, and one cannot enter alter mode when static model is open for detail view. A dynamic model covers for time-dependent changes in the state of the system. Dynamic models are normally represented by differential equations.

Deterministic or Stochastic, as indicated by the chance variables are not or are considered. A deterministic model is one in which each arrangement of variable states is uniquely determined by parameters in the model and by sets of past states of these variables. Deterministic models depict conduct on the basic of some physical laws. A stochastic model is one where exact prediction is impractical and irregularity is present and variable states are not described by unique values, but rather by probability distributions.

Discrete or Continuous, according to the variable involved discrete or continuous. A mathematical model is discrete if the state variable does not depends on the space variable. Otherwise the mathematical model is continuous. A discrete model regards objects as discrete, for example the particles in a molecular model. A clock is a case of discrete model because the clock skips to the next event start time as the simulation proceeds. A continuous model represents the objects in a continuous manner, for example the velocity field of fluid in pipe or channels, temperature and electric field.

All dynamic system models may be isolated into two major types based on whether they characterize a continuous-time or discrete-time process (Unbehauen and Rao, 1987). Much of the current literature on identification is currently worried with the identification of continuous-time models (Subrahmanyam, 1993). Furthermore, for many applications, continuous time models are more appealing to engineers than discrete-time models on the grounds that they are closely related to the underlying physical systems, while discrete-time model is considered to be defined at a sequence of time-instants related to measurement. In other hand, the most common classification of models is based on whether the model represents time-varying or time-invariant systems. For time-invariant systems, difference equation models are normally preferred. For time-varying systems, among prominent choices are cascaded block model, neural network, wavelet network and cellular automata (Abd Samad,2014). In difference equation it can be separated into linear and nonlinear models. For linear difference equation model, ARX (Average Model with eXternal input), ARMAX (AutoRegressive Moving Average with eXogenous input) and Output-Error model (OE) are the most used model. While examples of nonlinear difference equation model include NARX (Nonlinear AutoRegressive with eXogenous input), NARMAX (Nonlinear ARMAX) and NOE (Nonlinear Output Error) models.

Since the system identification handles a wide variety of different model structures, it is important that these can be defined in a flexible way. For ARX model, the most used model structure is the simple linear difference equation is (Paktos and Fassois, 2003):

$$y(t)+a_1 y(t-1)+...+a_{na} y(t-na)=b_1 u(t-nk)+...+b_{nb} u(t-nk-nb+1) \tag{2.1}$$