Nonlinear System Identification and Its Applications in Fault Detection and Diagnosis

Sun, Zhen

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Nonlinear System Identification and Its Applications in Fault Detection and Diagnosis

Zhen Sun

Department of Energy Technology

Aalborg University

A thesis submitted for the degree of

Doctor of Philosophy

07, October, 2013

I would like to dedicate this thesis to my loving parents and my wife. A special feeling of gratitude to my loving parents, Shusheng Sun and Yan Gao, for giving birth to me at the first place and supporting me spiritually throughout my life. Their words of encouragement and push for tenacity ring in my ears is very helpful during my study. My wife, Yuan Tian, have always stand on my side.

I also dedicate this dissertation to my many friends and the whole family who have supported me throughout the process. I will always appreciate all they have done, especially for helping me develop my technology skills.

I dedicate this work and give special thanks to my best friends in China.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Zhenyu Yang for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. I also would like to thank Zhenyu for his help in living in Denmark.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Rafal Wisniewski, Prof. Youmin Zhang and Tommy Mølbak , for their encouragement, insightful comments, and hard questions.

Abstract

Interest in nonlinear system identification has grown significantly in recent years. It is much more difficult to develop general results than the concern for linear models since the nonlinear model structures are often much more complicated. As a consequence, the thesis only considers two different kinds of models, one is a type of state space model which is described by It$\hat{o}$ Stochastic Differential Equations (ISDE), the other one is a nonlinear First Order Plus Dead Time (FOPDT) model. This thesis aims to investigate these two different kinds of nonlinear models and to propose the corresponding methods to deal with their system identifications.

Firstly, the system described by an ISDE model is considered. Extended from conventional stochastic systems, where the random part of the system is often described as a type of normal distribution signal added to the deterministic differential equation, the ISDE model generally consists of not only a structured deterministic part called drift term, but also a structured random part called diffusion term. The model can describe the system in which the random features are correlated with system states (inputs, outputs) and this relationship can be explicitly described by the model itself. The considered nonlinearity of this model can be expressed by the nonlinearity of the system functions. The parameter identification based on a state estimation such as an Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF), is investigated for this type of model in the thesis. Moreover, a new method by combining Maximum Likelihood (ML) technique plus UKF is proposed and its convergence property with regard to the consistency and normality is also investigated. The developed methods and algorithms are tested and analyzed for a number of numerical cases and then for a space robot system.

Secondly, the system considered is described by a nonlinear FOPDT model. This type of FOPDT model is an extension of the traditional FOPDT model which pre-assumes all the model parameters are constants. The nonlinearity that is defined in the model is reflected in its two categories of varying parameters, namely depending on time variable or some other variables, such as input signal etc. We refer to this type of model as a Time Varying FOPDT (TV-FOPDT) model. At first, the identifiability of the corresponding model is theoretically investigated. Then, the first concern of parameter identification of the considered systems is under assumption that the parameters of TV-FOPDT model are as time dependent. Afterwards, the input dependent parameter identification approach is considered. For these two categories of FOPDT models, the corresponding methods to make the parameters identification are proposed accordingly. Moreover, the proposed methods are further extended to make parameter identification of a kind of multiple inputs model. The proposed methods and algorithms are tested and analyzed for a number of numerical cases and finally applied to study the superheat dynamic in a Danfoss refrigeration system.

The proposed models and methods are further extended for the purpose of Fault Detection and Diagnosis (FDD). In a system where it exists possible parametric fault, if some fault happens, one or several parameters related to fault may change their values. Then the FDD procedure can be performed by identifying these fault related parameters. Afterwards, the decision whether the fault happened or how large the fault is can be made by comparison and analysis based on the estimated values.

Resume

Interessen for ulineær system-identifikation er steget betydeligt i de senere år. Det er imidlertid en hel del vanskeligere at nå frem til generelle resultater for ulineære modeller, end for lineære modeller. Årsagen er, at de ulineære model-strukturer ofte er væsentligt mere komplicerede. Følgeligt beskæftiger denne afhandling sig kun med to forskellige typer model, den ene type er en tilstands-rum-model (state space model), beskrevet ved Itô Stokastiske Differentialligninger (ISDE), den anden type er en ulineær Første Ordens Plus Død-Tid (FOPDT) model. Denne afhandling sigter mod at undersøge disse to forskellige slags ulineære modeller, samt at foreslå de tilsvarende metoder til system identifikation.

Først gennemgås den model, der knytter sig ISDE beskrivelsen. Modellen er opstået ud fra konventionelle stokastiske systemer, hvor den stokastisk varierende del af systemet ofte beskrives som en slags normal-fordelt signal overlejret signalet vedr. den deterministiske differentialligning. ISDE modellen består derfor ikke kun af en struktureret deterministisk del kaldet drifts-leddet, men også af en struktureret stokastisk varierende del kaldet diffusions-leddet. Modellen kan beskrive et system, hvor de stokastiske delsystemer er korrelerede med system tilstandene (input, output), og hvor denne relation kan beskrives eksplicit af selve modellen. Den undersøgte ulinearitet i denne model kan udtrykkes ved ulineariteten i system funktionerne. For modellerne i nærværende afhandling foretages parameter-identifikation udfra tilstands estimering, f.eks ved Extended Kalman Filtrering (EKF), eller ved Unscented Kalman Filtrering (UKF). Yderligere foreslås en ny fremgangsmåde med benyttelse af Maximum Likelihood (ML) (mest sandsynlige) teknik plus UKF, og denne metodes konvergens egenskaber undersøges m.h.t. konsistens og normalitet. De udviklede

metoder og algoritmer testes og analyseres i et antal regneeksempler samt i et system med en robot.

Derpå gennemgås det andet betragtede system, beskrevet ved en ulineær FOPDT model. Denne type FOPDT model er en udvidelse af den traditionelle FOPDT model, som forudsætter, at alle modelparametre er konstante. Det, som defineres som ulinearitet i denne betragtede model, kan henføres til to forskellige kategorier af varierende parametre, tids-varierende, eller varierende med andre variable, såsom input. Modellen kaldes Tids-Varierende FOPDT, dvs en TV-FOPDT model. Herefter tages først undersøges modellens identificerbarhed. Første skridt hen imod parameter-identifikation under antagelse af, at FOPDT modellens parametre er tidsafhængige. Derpå etableres den foreslåede identifikation af de med input varierende parametre. For disse to kategorier af FOPDT modeller foreslås de tilsvarende metoder til parameter-identifikation. Desuden udvides de foreslåede metoder til at muliggøre parameter-identifikation for en slags multi-input model. De foreslåede metoder og algoritmer testes og analyseres i en række numeriske tilfælde. Endelig bruges de til nærmere at undersøge dynamikken omkring overophedning i et Danfoss kølesystem.

De foreslåede modeller og metoder er blevet yderligere udvidet for også at dække Fejl Detektering og Diagnose (FDD). Hvis et system rummer mulighed for parametriske fejl, og der sker en fejl, da kan fejlen evt bero på, at en eller flere parametre har skiftet værdi. Her kan FDD proceduren gennemføres ved at identificere disse fejl-relaterede parametre. Derpå kan man afgøre, hvorvidt fejlen indtraf, og hvor stor den var, ved at sammenligne og analysere ud fra de estimerede værdier.

Contents

List of Figures

Chapter 1

Introduction

Interest in system identification for nonlinear systems has grown significantly in recent years. Its study demonstrates importance in process prediction, system re-configuration, fault tolerant control system and so on.

Obviously, for the nonlinear systems, different systems have their different nonlinearities. Unlike linear systems, it is difficult to find out general results for nonlinear system identification. Hence, it is required to develop different approaches to make nonlinear system identification for varies of nonlinear models. Among lots of model categories, State Space (SS) model and Input/Output (IO) model are the most popular ones. This thesis focuses on the two specific nonlinear models of these two kinds: It$\hat{o}$ Stochastic Differential Equations (ISDE) based SS model and several kinds of the extensions to First Order Plus Dead Time (FOPDT) model. It proposes two different methods to make the system identification of these two corresponding models. The proposed methods are referred to as Unscented Kalman Filter (UKF) plus Maximum Likelihood (ML) method and Mixed Integer Programming (MIP) based method in the following. Furthermore, some convergency properties and identifiability of the proposed methods are investigated. Afterwards, in order to show their advantages, these new methods are compared with some existing standard system identification methods and illustrated by some numerical examples. Finally, the proposed methods are applied to several real systems and showed their applications for the purpose of Fault Detection and Diagnosis (FDD).

1.1 Background and Motivation

Lots of engineering applications call for an accurate description of the behavior of the system under consideration, especially in the field of automatic control applications. Dynamic models that describe the system of interest can be constructed by using the first principles of physics, chemistry, biology, economy and so on. However, sometimes this kind of modeling procedure can be difficult or time consuming, because they require really much detailed specific knowledge and information, which may not be easily obtained. Nevertheless the resulting models are often very complex. In this sense, it is labor-intensive to develop the models in such a way, and hence expensive. Moreover, for a large amount of poorly understood systems, the derivation of a model setting up from the first principles is even impossible. Since the first-principle models are often complex, simulation of them may take considerable time on computers, thereby it can be challenging in the real-time applications. Moreover, Ljung proved that these constructed models are not always accurate (100). In the one hand, it is difficult to determine which elements are relevant, which effects must be included in the model, and which can be neglected. In the other hand, certain quantities needed to build the model are unknown, and have to be estimated by performing dedicated experiments. The resulting estimates often differ from the real quantities, and hence some model mismatch can occur. An alternative way of building models is to use system identification. In system identification, the aim is to estimate the dynamic model directly from observed input and output data. First principles are not directly used to model the system, but the knowledge about the system still plays an important role. Such knowledge is of great importance for setting up identification experiments to generate the required measurements, for deciding on the type of models to be used, and for determining the quality and validity of the estimated models. System identification often yields good models that are suitable for the fast on-line applications and for model-based predictive control, which has been found to be widely used in many engineering areas. Compared with the development of models set up based on first principles, it is not so labor-intensive. Moreover, at present, some steps of the identification procedure can be automated.

Nowadays, with the increasing demands for higher system performance, product quality and much more cost effective, the complexity and the automation degree of technical processes are continuously growing. This development calls for accuracy of the estimation in system and knowledge on system running. One of the most critical issues surrounding the design of automatic systems is system identification (6).

Parameter identification is one of the most important areas in the wide fields of system identification which is the procedure of using observations from a dynamic system to develop mathematical models that adequately represent the system characteristics. System identification including parameter identification generally proceeds as follows (100): Firstly, a certain type of model need to be selected that is considered to be suitable for the application at hand. Secondly, a special input signal is designed or adopted such that the respondence or output can capture the behavior of the system to be modeled. Then an identification experiment is carried out in which input and output signals are measured and collected. An identification method is selected to estimate the parameters or some functions in the model from the collected input and output measurements. Finally, the validity of the obtained model is evaluated.

The first step, also one important step in system identification, is the determination of the model type which is used to be considered. The decision is based on knowledge and information of the system under consideration, and the main properties of the candidate model. Certain types of models can be used to approximate the input-output behavior of a smooth nonlinear dynamical system in a good accuracy. These models have the so-called universal approximation capability. An example of a universal estimator is the neural network in (57). The drawback of this kind of model is that is often complicated. Hence, some other model structures have received much more attention over the years. At first, the Linear Time Invariant (LTI) model is the most popular one. It has been widely used in many engineering applications successfully, and a mature theory exists for system identification and automatic control. The authoritative guide for identification in linear system is the book written by Ljung (100). Although linear models are popular and widely used for several reasons, they still have their own limitations. In the real world, most of systems show nonlinear behavior. A linear model can only describe a limiting range of systems. With the increasing of the demanding in the world, where the performance and accurate descriptions of systems are needed, linear models are sometimes not satisfactory enough to describe the real systems anymore. Therefore, interest in nonlinear system models and nonlinear system identification methods has grown rapidly (96).

In the procedure of system identification including nonlinear system identification, the first step is model selection. For nonlinear system models, there are quite a lot of different nonlinearities and the model functions may be not time-independent during the system running, it is difficult to proposed general results for different nonlinear models. Normally, it is only possible to propose certain methods for specific models. In this sense, from the modeling point of view, two mathematical model representations, state-space and direct input-output relationship, are widely used to described the real systems. This motivates that the models discussed, including parameter identification, are within these two categories.

State-space systems are more attractive for dealing with multi-variable inputs and outputs. Just like what is stated by Rivals and Personnaz (137), state-space systems are likely to require fewer parameters, especially for multi-variable systems. Among different kinds of SS models, such as Ordinary Differential Equations (ODEs) model, Stochastic Differential Equations (SDE), discrete time model and so on, the SDE model can describe lots of real systems. It can describe system's dynamics, system noise and system disturbance. In the real world, all the measurements are often discrete time models. In this sense, the SS model considered in the thesis is SDE model with discrete time measurement, in which the nonlinearity of the system is reflected in the nonlinearity of system functions.

For the linear IO models, one popular representative model is the transfer function model, thereby First Order (FO) model was in the consideration at first. However, observed from the real industrial systems, most IO models need to take the time delay into consideration. Bearing it in the mind, the FO model is considered with time delay variable, which is generally called as First Order Plus Dead Time (FOPDT) model. Regarding the nonlinearity of this FOPDT model, another kind of nonlinearity is applied, i.e., the nonlinearity is generated by the property of time varying parameters in the systems.

For the above reasons, the thesis will consider the following two specific models:

- Nonlinear SDE model with discrete measurement
- Time Varying FOPDT (TV-FOPDT) model.

Different methods has already been proposed to handle with the system identification of these two types of models.

Nonlinear SDE model with discrete measurement:

There are some different methods to make the identification of the parameters for nonlinear SDE model with discrete measurement.

Prediction Error Method (PEM) Prediction Error Method (101) is one of the most popular methods to make the system identification, which is considered as a kind of generalized framework for it can be applied to quite arbitrary model parameterizations. It estimates model parameters by minimizing the optimally determined one step ahead output prediction error. The existing identification methods, such as Least Square Method (LSM), Ordinary Equation Method (OEM) and so on, are special cases of PEM that are proposed for different model types. The PEM for the model with Gaussian distribution is asymptotic unbiased and efficient (101). Furthermore, the use of the PEM enables an estimate of the associated uncertainties of the estimated model parameters. But it requires an explicit parameterization of the model and searches for the parameters that gives the best output prediction fit may be laborious.

Subspace Identification Method (SIM) Subspace Identification Method is first to be used for Linear Time Invariant (LTI) systems and shows good performance. Its main idea is to make the computation of the estimate of state vectors at first, then extend observable matrix from the given input-output data. However, in many cases, it provides better performance than PEM in term to the precision. Since SIM does not require a particular parameterization in the system, it is numerically attractive and suitable for multi-variable systems. In recent years subspace identification methods have been developed for certain nonlinear systems: Wiener systems (Chou and Verhaegen (21)), Hammerstein systems (Verhaegen and Westwick (163)) and so on. Although SIM is a fast, robust and convenient approach, it still has an problem with precision and few applications for closed-loop identification.

Statistical Method (SM) Statistical Method is to set up a statistical function based on the measurement and its distribution. It includes Maximum Likelihood Estimation, Least Mean Square Estimation and other different statistical methods (107). Since the estimation is only based on the measurement and does not consider the structure of the state, sometimes its performance regarding to the accuracy is not so good as expected.

Filter Based Method (FBM) Filter Based Method has became popular since the 1960s (117). FBM, in general, can be classified into two different categories. One category is referred to as direct approaches. It takes both the state variable and the unknown parameter(s) into an augmented system state. Then, the corresponding filter technique, such as KF, Extended Kalman Filter (EKF) or some other appropriate filter is used to estimate the new state. Thereby, the unknown parameter is identified. The main advantages of this kind of FBM is that it can be easily performed, so it is widely applied in the real world. However, this method only applies part of information in the system, that is the mean and variance of the state. But in most Gaussian noise system, the distribution of the state is known beforehand. In order to use it, some other statistical methods can be added to make the system identification combined with filter technique, such as Maximum Likelihood, Least Square and so on. It is another category of FBM.

Regarding the nonlinear SS models in the thesis, it is one type of SDE models with Gaussian noise. The FBM could be used to make its parameter identification. But considering the distribution given in the system, the thesis will focus on the Kalman Filter plus Maximum Likelihood method and investigate its convergence property. Moreover, in order to apply the method to the FDD procedure, this method is extended to an online manner as well.

Time Varying FOPDT (TV-FOPDT) model:
The traditional FOPDT model has three different parameters: time delay (also called Dead Time), system gain and time constant. If the system does not have time delay, it degenerates to a standard linear time invariant system which has already have a mature identification theory. Generally, there are many algorithms to estimate time delay, see (164) for some details.

Cross Correlation Method (CCM) Cross Correlation Method (10) is one of the basic method of Time Delay Estimation (TDE) problem in time series analysis. Many TDE methods are developed based on CCM. Its main idea is to cross-correlate the outputs and inputs and consider the time argument that leads to the maximum peak in the correlation series as the estimated time delay.

Maximum Likelihood (ML) Method ML method is another important method for TDE problem (70). The ML function is chosen to improve the accuracy of the estimated time delay by attenuating the signals fed into the correlator in spectral region where the Signal to Noise Ratio (SNR) is the lowest. The popularity of ML estimator stems from its relative simplicity of implementation and its optimality. For uncorrelated Gaussian signal and noise, the ML estimator of time delay is asymptotically unbiased and efficient in the limit of long observation intervals.

Average Square Difference Function (ASDF) Method The Average Square Difference Function Method (113) is based on finding the time point of the minimal error square between two received signals: no time delay signal and time delay signal, and considering this time point as the estimated time delay. Its advantage is due to the fact that it can give perfect estimation in the absence of noise while the direct correlation methods can not. Moreover, ASDF requires no multiplication, which is the most significant practical advantage over the other methods.

Least Mean Square (LMS) adaptive filter method The LMS adaptive filter is a finite impulse response (FIR) filter (65) which can automatically adapt its coefficients to minimize the mean square difference between the reference input signal and desired input signal.

From previous observation, in order to make the estimation of time delay, these methods need to perform or simulate the system several times to get enough data, i.e., different outputs signal (no time delay and time delayed outputs) or adopt different input signals. Moreover, if the other parameters rather than time delay in the models need to be identified as well, some extra procedure need to be performed after TDE. It motivates us to develop a new method which can simply identify all the parameters in the system model besides time delay.

However, from the modeling points of view, the traditional FOPDT has its own limitations, i.e., the parameters are considered as constants during the system running. But in the real world, the parameters would be changed during the system running for quite a lot of systems. The time varying property is more and more important. For these reason, in the thesis, it aims to extend the standard FOPDT model to a time varying one and find a method that

- can be cost-effectively performed to make the parameter identification of the time delay and the other parameters together only based on the measured information from system operation;

- can be used to identify the time varying parameters, so it should be an on-line method.

Application in FDD Modern systems and equipment are often subjected to some unexpected changes, such as component faults and variations in operating conditions, that tend to degrade the overall system performance. In order to design a reliable, fault-tolerant control system, or to maintain a high level of performance for complex processes, eg, spacecraft, aircraft, chemical processes and nuclear plants, etc, it is crucial that such changes are detected and diagnosed promptly so that corrective action can be taken to reconfigure the control action and accommodate the system alternation (129).

In general, a fault (63) is to be understood as an unexpected change of system function especially the parameters' change, although it may not lead to physical failure or breakdown. A technique which is used to detect and diagnose faults and identify their types or characteristics in a system is called as Fault Detection and Diagnosis (FDD) technique. The essential tasks of FDD are: Fault Detection, making a binary decision–either that something has gone wrong or that everything is fine; and Fault Diagnosis, determining the source of the fault and the fault category, eg, which sensor, actuator or component has become faulty and how is the quantitative level.

During the last three decades, the so called model-based fault detection and diagnosis approach has received increasing attention in both research and application (20; 61; 62; 63; 129; 130). This approach is based on the concept of 'analytical redundancy' as opposed to physical (hardware or parallel) redundancy, which uses measurements from redundant sensors for fault diagnosis purposes. Analytical redundancy use of signals generated by the mathematical model of the system being considered. These signals are compared with the actual measurements obtained from the system. The comparison is done using the residual quantities which give the difference between the measured signals and signals generated by the mathematical model. Hence, model based FDD can be defined as the determination of faults of a system from the comparison of available system measurements with a priori information represented by the system mathematical model through generation of residual quantities and their analysis.

In the process of FDD, residual generation can be achieved by the following methods (20):

- **Observer-based methods:** If the process parameters are known, either state observers or output observers can be applied to generate the residual. Then estimating the residual based on the knowledge, it is used to compare with the predefined threshold to make the decision of the fault (63).
- **Parity space methods:** Run the process and form an output error, then based on the error estimation to make FDD. The general process can be referred in (63) as well.
- **Parameter identification based methods:** In most practical cases the process parameters are partially not known or not known at all. Then, they can be determined with parameter identification methods by measuring input and output signals if the basic model structure is known. Then based on the results of the parameter identification and analyzed the change of the estimated value of the parameters in the system, the decision of the FDD can be made (61).

Nowadays, in quite a lot of situations, the parameter identification methods based FDD is widely used and directly performed in the fault tolerant control systems. The thesis will consider the application of system identification in the procedure of FDD.

1.2 Overview of previous work and related work

System identification handles with the problem of estimating mathematical models of systems based on the measurements of inputs and outputs in the systems. It can be originated from the work of Gauss and Fisher (39). Much of the early work was conducted within the fields of statistics, econometrics and time series analysis. Astrom and Bohlin can be marked as the starter of system identification in 1965 (100). From then on, the theory has been developed much more significantly. After four decade developing, the system identification for linear system became a field which has a relatively mature theory.

Although identification theory for nonlinear systems is almost as old as for linear systems, its progress is not so fast as that for linear systems, especially for the system which is described as a SDE model with discrete measurement, called as continuous-discrete SDE model (4). Generally, the parameter consisting in a continuous-discrete SDE model includes two different parts: parameter in the drift part and parameter in the diffusion part.

The development of SDE parameter estimation can be referred in (12). It is firstly studied to only consider the parameter in the deterministic part–drift parameter identification. Drift parameter estimation has been studied by many authors. Le Breton (88) and Dorogovcev (31) appeared to be the first persons to study estimation in discretely observed SDE model. Their model is the linear SDE model with constant diffusion coefficient. While Le Breton used Approximate Maximum Likelihood (AML) estimation, Dorogovcev used Conditional Least Squares (CLS) estimation. In 1977, Robinson used exact maximum likelihood estimation in discretely observed Ornstein-Uhlenbeck process which is a special case of SDE model. From then on, some researchers work on approximate maximum likelihood estimation (where the continuous likelihood is approximated), also called the maximum contrast estimation, such as Bellach (1983) and Yoshida (1992). All of these approaches belong to Maximum Likelihood (ML) method category.

Another category to make parameter identification of SDE model is filter based method, which was proposed a little later than ML methods. Regarding filtering and estimation theory in the discrete-discrete time framework, see e.g. (95) for more details. Similarly for the continuous-continuous framework, see e.g. (182) for more details. The latter framework is useful for design purposes, but it is argued that for filtering and estimation it is inappropriate for the true cases (66). From 1980s, the Kalman Filter (KF) technique has been more and more widely used for parameter identification in the application (94). Generally, the approaches using KF can be classified into two different categories. One category can be called as direct approaches. This kind of approach takes both the state variable and the unknown parameter(s) into an augmented system state. Then, KF, Extended Kalman Filter (EKF) or some other appropriate filter can be used to estimate the new state and thereby the estimation of unknown parameter(s). However, if the diffusion term of the SDE contains unknown parameters, this kind of approach could not be directly used. Moreover, if the system model is a nonlinear one, this method sometimes could not provide a good performance in terms of the precision.

The other kind of filter based method is to combine KF technique with some statistic methods. The scheme consists of two sequential stages. The first stage conducts the state estimation using KF, where the estimated state, both mean and variance, is a function of unknown parameters. Then, a statistic criterion, such as Maximum Likelihood (ML) and Least Mean Square (LMS), is set up in the second stage based on the estimated state. Thereby, the parameter identification problem becomes an optimization of a parameterized statistic problem. This approach can be directly applied to linear systems and an explicit solution may be found (76; 107). Nevertheless, this kind of approach needs to be extended in order to handle nonlinear cases. Then, a ML/Prediction Error Decomposition (PED) method for direct estimation of embedded parameters in SDE is proposed in (108) based on the EKF. (83) set up the scheme of parameter identification based on the EKF and ML as well as Maximum A Posteriori (MAP) estimation with software implementation. Both of these two methods can handle with the parameter identification for cases that the diffusion item consists of the unknown parameter(s). But the precision in the estimation need to be improved for some nonlinear models. In Chapter 2, a more detailed introduction will be given.

Another model, which is also widely used in application, is input/output model. FOPDT model is one of the most useful input/output models and it is well known that FOPDT model can be applied to describe many industrial processes.

The FOPDT model has three different parameters, named system gain, time constant and dead time (time delay). These parameters are often set as constants in the whole system running for the standard model. In reality, during the system running, they may not stay unchanged but vary according to the time. Thereby, in order to make up for the shortage of the standard FOPDT model, a kind of nonlinear FOPDT model in which the time varying parameters of the system can be describe is proposed in (85; 89; 123). The considered nonlinear FOPDT model is an extension of the standard FOPDT model by means that both system's gain and time constant can be changed during the system running. This nonlinear FOPDT model is generated by using a linearized method to a nonlinear model. Nearly at the same time, in (89; 123), a nonlinear FOPDT model is proposed by linearizing the nonlinear system at a number of different operating points, so the parameters of the obtained FOPDT are operating-point dependent. It is observed that some simple nonlinear FOPDT models have been already used in nonlinear control applications (15; 85; 89).

For the system identification, these were quite a lot of methods to make the estimation of the FOPDT, such as the Tangent Method, the Area Method and so on (164). Some methods have been already adopted to make the system identification for the nonlinear FOPDT model. For example, an on-line nonlinear FOPDT identification method is proposed in (85) by using the so-called longrange predictive identification method. However, the formulated system identification leads to a nonlinear optimization problem due to the unknown time-dependent time delay. Therefore, four different potential time delay scenarios are assumed, before converting the nonlinear optimization problem into a Least-Square (LS) problem using the spectral factorization technique. The assumption of time delays limits the proposed method in (85) to be applied for any other systems except for these two specific patient cases they have studied (179).

These methods have their own drawbacks, such as some ones need a special input signals, some ones only can be applied in off-line manner and so on. Moreover, all of the methods need more than two steps to make the identification of all unknown parameters in the FOPDT models. In Chapter 3, a detailed introduction of the techniques to make system identification of FOPDT is given. It describes some of the most common methods to make the parameter identification for FOPDT model.

One important application of the system identification, especially parameter identification, is in the field of FDD. The parameter identification technique based FDD is widely studied most in the reconfiguration control area. In a number of fault cases, the faults are reflected in the physical system parameters, as e.g. mass, friction, viscosity etc. It makes that the parametric faults are associated with system parameters. It is natural that system identification methods can be applied for FDD, see e.g. (41). The parameter identification techniques has been considered to use in FDD for systems with parametric faults since 1990s. Lots of researchers developed the theory of it, see e.g. (41; 44; 61; 62; 129; 150) to mention some references.

1.3 The Objectives of the Project

This thesis focuses on some issues on nonlinear parameter identification for two different kinds of nonlinear systems. It tries to use some innovative models to describe the real systems more accurately. Then based on the model formulation and the traditional system identification approaches, some methods are extended and proposed. These new methods are also compared with the traditional methods to show their advantages and difference. Furthermore, some properties of the methods are investigated. Finally, the mathematical models and their identification approaches are applied to a number of real-life relevant systems.

In order to address these objectives, the thesis contributes in the following way:

- Different state space models are discussed and compared in Chapter 2. It suggests that the Itô SDE (ISDE) model can describe dynamic systems with noise and fault much more accurately.

- A detailed review of the Kalman Filter based system identification methods for SDE model, both direct method and indirect method are given.

- Unscented Kalman Filter (UKF) plus Maximum Likelihood, to make the system identification of nonlinear SDE model is proposed. Its consistency and normality are investigated and set up the conditions under which the consistency and normality can be guaranteed for nonlinear SDE models. The method is compared with EKF based method in terms of accuracy, convergency and computation load, respectively.

- Extend the FOPDT model to a general Time-Varying FOPDT model, moreover to TV-FOPDT model with Input Depended Dead Time. The identifiability of the defined FOPDT models is discussed and some theorems are correspondingly derived.

- The methods to make the estimation of time delay are concisely described. The approaches to make the system identification of FOPDT model are given by a detailed review. Their main procedures and drawbacks/merits are discussed.

- New approach, based on the Mixed Integer Non-Linear Programming (MINLP) and Branch-and-Bound(BB) method, is proposed to make the parameter identification based on the TV-FOPDT model and the one with Input Depended Dead Time.
- All the methods proposed in the thesis are simulated for a number of numerical examples and some real-life relevant systems with some traditional methods.
- The methods are applied to the process of FDD, which are shown by several testing systems.

1.4 Thesis Outline

The thesis is organized as follows:

Chapter 2 Firstly, outlines the different system models using State Space formulations. Then, a detailed introduction of system identification for nonlinear SDE model is given. Secondly, extended from the traditional methods, the UKF plus ML method to make the nonlinear identification of ISDE model is proposed. The consistency and normality properties of the proposed method are investigated and a theorem regarding it is proved. Finally, in order to test the proposed method, a number of numerical examples are given to illustrate the properties of the methods compared with some other methods. Moreover, this approach is applied to a space robot system which is considered under some FDD scenarios.

Chapter 3 Firstly, the FOPDT model is extended to the Time Varying (TV)-FOPDT model, possibly with some input depended variables. Secondly, the identifiability analysis is performed based on the identifiability to the nonlinear systems. Thirdly, in the parameter identification, the problem is converted to a Mixed Integer Non-Linear Programming (MINL) one. The Branch and Bound (BB) method plus Least Square (LS) or Least Mean Square (LMS) method is applied to solve this optimization problem. Finally, the method is tested though a number of numerical cases and the analysis is committed based on these tests' results. The application of the model and method is illustrated by a superheat dynamic model in the supermarket refrigeration system and a FDD discussion.

Chapter 4 Finalizes the thesis by providing the conclusion and recommendations for future works.

Chapter 2

System Identification for Nonlinear SDE model and Its Application

System identification is one of the most important areas in the engineering. In case the considered models are linear, there exists well matured theory, methodology and algorithms to make the identification of them. But for nonlinear models, the situation is more complicated, and it is much more difficult than the linear ones to develop some general results.

In the nonlinear system identification, each method is only used to deal with certain kind of modeled system. In this Chapter, the system identification for a kind of SDE model is the main consideration.

These issues have been addressed throughout this chapter in the following order:

System Model Description In order to show the reason to choose Itô SDE model with discrete measurement as the concerned model in the study, some popular system models are reviewed and a brief introduction of the system identification methods is given at first. Then some basic knowledge of Itô SDE model is reviewed.

Overview of the Previous Work Different methods to make the system identification for the concerned SDE model are outlined.

UKF plus ML method An Unscented Kalman Filter (UKF) plus Maximum Likelihood (ML) method to make the system identification for the nonlinear SDE model is proposed. Then the consistency and normality of the proposed method are investigated and corresponding theorems are proved.

Numerical Test and Application Finally, some numerical tests are performed to illustrate the approach and compared with other methods. A scenario of FDD is used as an application of the work.

2.1 State Space Model

State Space (SS) Model is widely applied in different fields, such as in Control Engineering, Chemistry, Physics and so on (4). In practice, environmental disturbances, unexpected changes within the technical process under observation as well as measurement and process noises often happened in the system running. For this reason, the dynamic stochastic model is the most popular one among different kinds of SS models. The system without random features is the special case of them. In general, SS model using Stochastic Differential Equation (SDE) model can cover most of different SS model expressions (4).

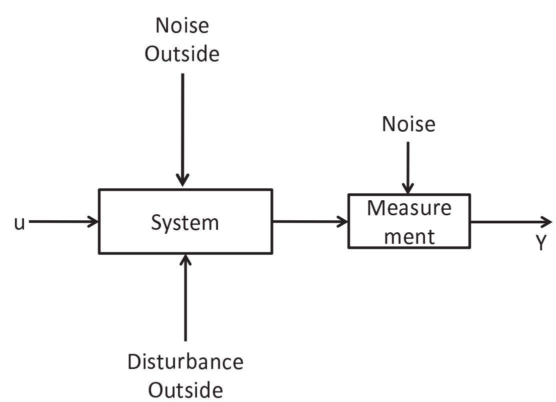

Figure 2.1: System process 1

One type of SDE model, which can describe the general process in Fig. 2.1, is expressed as:

$$\begin{cases} \dot{X} = F(X,u,t) + E_d D + \omega, \quad X(0) = X_0 \\ Y = H(X,u,t) + F_d D + \xi \end{cases} \tag{2.1}$$

Here the X is short for $X(t)$ stands for the state of the system at the time t. Y and u are the measured output and input. In case of control system concern, u is referred to as the control input, E_d, F_d are matrices of compatible dimensions, D is a deterministic but unknown input vector, ω, ξ are both the stochastic processes which are system noise and measurement noise, respectively and they are assumed to be uncorrelated with each other in most cases.

Another kind of model can describe the same system, is It$\hat{o}$ SDE (ISDE) model, which is

$$\begin{cases} dX = [F(X,u,t) + E_d D]dt + dB_t, \quad X(0) = X_0 \\ \quad Y = H(X,u,t) + F_d D + \xi \end{cases} \tag{2.2}$$

where B_t is a Brown Motion (B.M.) which will be given later. Different models (2.1) and (2.2) both can describe the same system as illustrated in Fig. 2.1.

As what we observed from most literatures, no matter how the random feature is modeled, the random part is just a simple additive stochastic process, as shown in (2.1). But in fact, in many practical situations, if random factor occurs in a complex system, it is not only related to a simple stochastic process but also to some other elements such as the state of the system. For example, a loose connection of mechanical components could lead to larger vibration influence to the relevant system comparing with normal situation, which reflects in the mathematical model as the features change of deterministic coefficients, as well as that of nondeterministic part (such as the process and measurement noises). Another example is a kind of blade distortion faulty system of wind turbine, if the system state consists of the rotation velocity, the random part in this type of system has some relation with it. According to these real systems, it is not so convincible that the random feature is only considered as a simple stochastic process. In some cases in the reality, the system is running like in Fig. 2.2 rather than Fig. 2.1.

From the modeling point of view, the SDE model has been already applied in the finance, refer (67) for more references. It is well known that the stock price could be

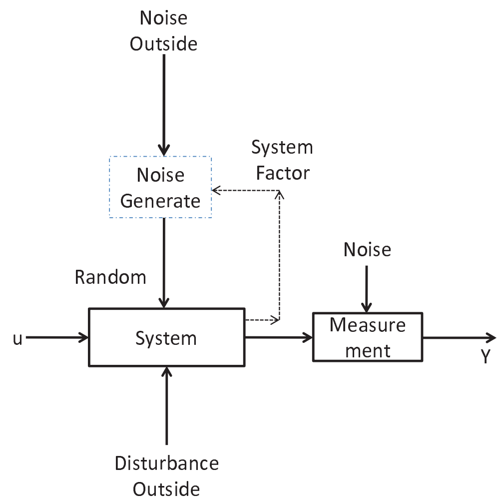

Figure 2.2: System process 2

seen as a complex stochastic process. Its random property has some relation with many factors, such as the management of the company, the profit, the economy status, the people's expectation and so on. In the financial analysis, the ISDE is just the model used to explain this complex process, which could be seen in the following. Now let S_t be the price of the stock at time t. The commonest model to describe it in the economy, decomposes the return dS_t/S_t, which is interpreted as the change rate of the stock values, into two parts. One is predictable, deterministic and anticipated return akin to the return on money invested in a risk-free bank. It gives a contribution μdt to the return dS_t/S_t, where μ is a measure of the average rate of growth of the stock price, also known as the drift item. In the simple model, μ is taken to be a constant. In more complicated model, μ can be a function of stock price S_t and time t. The second contribution to dS_t/S_t models the random change in the stock price in response to external effects, such as unexpected news, accidents and so on. It is represented by a random simple movement from a normal distribution with mean zero and reflected by a term σdB_t. Here σ is a function called the volatility, which measures the standard deviation of the returns (67). The quantity B_t is the Brown Motion (B. M.), and its definition will be given later. Putting these contributions together, the ISDE model of stock can be obtained as:

$$dS_t/S_t \ = \mu dt + \sigma dB_t \tag{2.3}$$

which is the mathematical representation for the stock price. As the generalization, the coefficients of $\mathrm{d}B_t$ and $\mathrm{d}t$ in (2.3) could be functions of S_t and t. Furthermore, the control item u_t, which is interpreted as the portfolio in the finance, can also be embedded into the functions μ and σ.

Contrasting with the above pricing of the stock, the system in the control engineering has the deterministic items and random features as well. The random features may take place in the unpredictable situations and forms. But sometimes we could get some statistic information of them, such as their distributions. Taking the normal distribution as an example, it could be interpreted as some combinations of Brown Motion since it follows the normal distribution as well. Then the random part of the system can be derived by the item of $\mathrm{d}B_t$ in the mathematics as the previous stock price model. Hence, the ISDE model could be potentially employed for the system with random feature in the engineering. According to this rule, the system in Fig. 2.2 can be reflected in mathematic model

$$\begin{cases} dX = [F(X,u,t) + E_d D]dt + G(X,u,t)dB_t, \quad X(0) = X_0 \\ Y = H(X,u,t) + F_d D + \xi \end{cases} \tag{2.4}$$

where it interprets the system noise as a structured noise. It can describe the detailed information of the noise rather than only simply using one stochastic process to describe the noise.

More generally, the system model can be written as

$$\begin{cases} dX = \tilde{F}(X,u,t)dt + G(X,u,t)dB_t, \quad X(0) = X_0 \\ Y = \tilde{H}(X,u,t) + \xi \end{cases} \tag{2.5}$$

From the theoretical point of view, the ISDE model have already been used to character some simple control systems with structured random features, such as (4; 107), and there was already a set of theory to support this kind of model in the mathematics. Some results on system identification have been obtained by extending the Kalman Filter (KF) or Extended Kalman Filter (EKF) and likelihood functions or other statistic methods in (12; 182) for linear ISDE model.

Of course, some measurable functions can also be added to the random part of the classic stochastic model (2.1). But not much theory can support this kind of model, such as the solution theory. Using the ISDE model has its own advantage to describe the random dynamic systems. It lies in:

1. ISDE offers more clear and proper description of random uncertainties in the considered systems than previous models which only describe random features as a simple class of normally distributed signals. In this sense, the ISDE model may provide a more accurate one to describe the stochastic process;

2. The structured description of random features could lead to the developed FDD algorithms based on ISDE model less conservative compared with the current results with a simple assumption of normal distributions. Since the fault could have a structure in ISDE model and it is real in some cases, it is much more convenient to deal with the fault using this model;

3. The situations that system random feature is correlated with system state, input and/or output can also be systematically handled possibly based on ISDE model, besides the situation that the system random feature is independent of system variables;

4. The ISDE model may open a proper window to go deeply to check some system random properties even for the FDD, meanwhile offer a solid platform to apply sophisticated stochastic analysis and filtering theory into control engineering.

2.2 ISDE formulation

The knowledge of ISDE is summarized in this section, referred more details in (182).

In the following, ISDE stands for It$\hat{o}$ Stochastic Differential Equation. In order to introduce this model, the definition of Brown Motion (B.M.) need to be given at first. In ISDE model, B_t stands for Brown Motion (B.M.), which is originated from the Scottish botanist Robert Brown who observed that pollen grains suspended in liquid performed an irregular motion. This motion was later explained by the random collisions with the molecules of the liquid. In this way, B_t is used to describe the motion mathematically, interpreted as a stochastic process which can describe the position of the pollen grain at time t. Strict definition is:

Definition 2.2.1 *$\{B_t\}_{t\geq 0}$ is a Brown Motion, if it satisfies*

1. *B_t is a Gaussian process,*
2. *B_t has independent increments,*
3. *B_t is continuous.*

Before giving the interpretation of ISDE, some important mathematical preliminaries are described at first in the following (182).

Definition 2.2.2 *If Ω is a given set, then a σ-algebra $\mathcal{F}$ on Ω is a family $\mathcal{F}$ of subsets of Ω with the following properties:*

1. *The null set $\phi \in \mathcal{F}$*
2. *If $F \in \mathcal{F}$, then $F^C \in \mathcal{F}$, where $F^C = \Omega \setminus F$ is the complement of F in Ω*
3. *If $A_1, A_2 \cdots \in \mathcal{F}$, then $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$*

Definition 2.2.3 *The triple $(\Omega, \mathcal{F}, P)$ is called a probability space if Ω is a given set, $\mathcal{F}$ is the σ-algebra in Ω and P is the probability measure.*

Definition 2.2.4 *If $(\Omega, \mathcal{F}, P)$ is a given probability space, then a function $Y : \Omega \to \mathbb{R}^n$ is called $\mathcal{F}$-measurable if*

$$Y^{-1}(U) := \{\omega \in \Omega; Y(\omega) \in U\} \in \mathcal{F}$$

for all open sets $U \in \mathbb{R}^n$.

Noted that in the following, the $(\Omega, \mathcal{F}, P)$ is a given probability space in which the SDE is defined. Then the SDE is given in the following:

Definition 2.2.5 *A equation is called an Stochastic Differential Equation (SDE) if it has the format*

$$dX_t = f(t, X_t)dt + g(t, X_t)dB_t, \tag{2.6}$$

where t is the time invariable and X_t is shot for $X(t)$, which is a real-valued function of t, B_t is a Brown Motion (B.M.). $f(t, X_t)$ is called as the drift coefficient and $g(t, X_t)$ is called as the diffusion coefficient.

The notation used in (2.6) is shorthand for the corresponding integral interpretation and is therefore ambiguous unless a specific integral interpretation is given. SDEs may be interpreted both in the sense of Stratonovich formulation (67) and in the sense of Itô formulation (182).

Definition 2.2.6 *Suppose that B_t is a Brown Motion and X_t is a measurable function regarding the σ-algebra $\mathcal{F}$ generated by B_t, then the Itô integral*

$$\int_0^T X_t dB_t : \Omega \to \mathbb{R}$$

is defined to be the limit of

$$\sum_{i=0}^{k-1} X_{t_i}(B_{t_{i+1}} - B_{t_i}) \tag{2.7}$$

as the mesh of the partition $0 = t_0 < t_1 < \cdot < t_k = T$ of $[0,T]$ tends to 0 (in the style of a Riemann C Stieltjes integral).

It can be seen that the Itô integral uses the left endpoint of each subinterval to make a sum, but Stratonovich formulation just applies the value of the process X_t at the meddle point of each subinterval: i.e., $\frac{X_{t_{i+1}}+X_{t_i}}{2}$ in place of X_{t_i} in (2.7). In most cases, the system identification need to be performed based on the information obtained until sampling time points. For this reason, the Stratonovich interpretation is unsuitable for system identification, then the Itô interpretation of the integral is adapted in the system identification.

Now existence and uniqueness result for the solution of the ISDE is given (182):

Theorem 2.2.7 *Let $T > 0$ and $f(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $g(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable functions satisfying*

$$|f(t,x)| + |g(t,x)| \leq C(1+|x|); \quad x \in \mathbb{R}^n, \ t \in [0,T], \tag{2.8}$$

for some constant C, (where $|g(t,x)|^2 = \sum\limits_{i,j} |g(t,x)_{ij}|^2$) and such that

$$|f(t,x) - f(t,y)| + |g(t,x) - g(t,y)| \leq D|x-y|; \quad x,y \in \mathbb{R}^n, \ t \in [0,T], \tag{2.9}$$

for some constant D. Let Z be a random variable which is independent of the σ-algebra generated by $B_s(\cdot)$, $s \geq 0$ and such that

$$E[|Z|^2] < \infty.$$

Then the stochastic differential equation

$$dX_t = f(t, X_t)dt + g(t, X_t)dB_t, \quad 0 \leq t \leq T, \;\; X_0 = Z \tag{2.10}$$

has a unique t-continuous solution X_t with the property that

$$E[\int_0^T |X_t|^2 dt] < \infty. \tag{2.11}$$

In the thesis, all the ISDE models are chosen to satisfy these conditions in order to guarantee the existence and uniqueness of the solution. For the implicity, this condition will not be checked again when dealing with the ISDE models.

From the general ISDE model (2.6), it can be observed that the random part in the system can depend not only on the time but also on the state variable. For some special kinds of state dependent noised ISDE model, a mathematical tool, called as Itô Formula (182), can be applied to simplify the model.

Theorem 2.2.8 *Consider the ISDE (2.6), suppose $F(t,x)$, a real-valued function, defined for $x \in \mathbb{R}^n$ and $t \in [a,b], 0 \leq a \leq b$, with continuous partial derivatives, $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x}$ and $\frac{\partial^2 F}{\partial x^2}$, then it can be obtained that*

$$dF(t, X_t) = \tilde{f}(t, X_t)dt + \tilde{g}(t, X_t)dB_t, \tag{2.12}$$

where

$$\tilde{f}(t, X_t) = \frac{\partial F(t, X_t)}{\partial t} + f(t, X_t)\frac{\partial F(t, X_t)}{\partial X_t} + \frac{1}{2}g^2(t, X_t)\frac{\partial^2 F(t, X_t)}{\partial X_t^2} \tag{2.13}$$

and

$$\tilde{g}(t, X_t) = g(t, X_t)\frac{\partial F(t, X_t)}{\partial X_t}. \tag{2.14}$$

For example, consider the ISDE

$$dX_t = f(t, X_t)dt + \alpha X_t^2 dB_t. \tag{2.15}$$

If $F(t, X_t)$ is chosen as $\frac{1}{X_t}$, then according to (2.12)

$$dF(t, X_t) = \tilde{f}(t, X_t)dt - \alpha dB_t. \tag{2.16}$$

From the former theorem 2.2.8 and the example, it can be seen that for some kinds of state dependent noised ISDE models, Itô Formula can really simplify them to ones without state dependent noise. This is also an advantage to use ISDE model to describe the complex process.

2.3 System Identification for ISDE Model

In this section, firstly, the system identification problem is formulated, and then the approaches to make system identification for ISDE model are considered.

2.3.1 Parameterized ISDE Model

The considered system is described by the following ISDE:

$$dX(t) = g_1(X(t), u(t), t, \theta)dt + g_2(X(t), u(t), t, \theta)dB_t, \tag{2.17}$$

where $t \in \mathbb{R}$ is the time variable, $X(t) \in \mathscr{X} \subset \mathbb{R}^n$ is a vector of state variables, $u(t) \in \mathscr{U} \subset \mathbb{R}^m$ is a vector of input variables, $g_1(\cdot) \in \mathbb{R}^n$, $g_2(\cdot) \in \mathbb{R}^{n\times n}$ are nonlinear or linear functions and $\{B_t\}$ is an n-dimensional Brown Motion. $\theta \in \Theta$ is a stack consisting of all unknown parameters. For simplicity purpose, $X(t), u(t)$ are denoted as X, u respectively in the following.

The measurement of the considered system is described by

$$Y_k = h(X_k, u_k, t_k) + \varepsilon_k, \tag{2.18}$$

where $Y_k \in \mathscr{Y} \subset \mathbb{R}^l$ is a vector of output variables, $h(\cdot,\cdot,\cdot) \in \mathbb{R}^l$, t_k, $k = 0, 1, \ldots, N$ are sampling instants, $\{\varepsilon_k\}$ is an l-dimensional noise process with $\varepsilon_k \sim \mathscr{N}(0, R)$ (R is an $l \times l$ matrix) and X_k is the state value at time t_k.

Noted that in order to make the identification of the parameter, an assumption of the measurement need to be made as the prerequisite, i.e., the frequency of sample points obtained from the measurement should be much larger than the frequency of parameter variation. Although it is impossible to get the true parameter value before the identification, this assumption can be satisfied by reducing the sampling interval as little as possible. For the simplicity, in the thesis, it is chosen that all the measurements (sample frequency) satisfy this condition.

In (2.18), the measurement of the system is considered as discrete one. Since the diffusion coefficient is almost surely determined by the process, i.e., it can be estimated without any error if observed continuously throughout a time interval for the linear models (43). In the other hand, parameter estimation in diffusion processes based on measurement at discrete time points is of much more practical importance due to the difficulty of observing diffusions continuously throughout a time interval. Hence, the thesis focuses on the parameter identification for continuous system with discrete measurement model.

The considered parameter identification problem could be described as:

(P): **Estimate the unknown parameter θ in the system (2.17) based on a set of data which consists of some measured output signals Y_k generated by (2.18) and corresponding input signals u_k.**

2.3.2 Conventional Methods

The parameter identification for continuous-discrete ISDE model could split into two parts: parameter estimation in the drift coefficient and diffusion coefficient.

In system identification for the parameterized ISDE model, the Least Square (LS) method was firstly applied to cope with the linear ISDE model. Considering the ISDE model (2.17), the idea of LS method (52) is to minimize the quadratic cost function:

$$Q(\theta) = \sum_{N}^{k=1} \frac{[X_k - X_{k-1} - g_1(X_{k-1}, u_{k-1}, t_{k-1}, \theta)(t_k - t_{k-1})]^2}{g_2^2(X_{k-1}, u_{k-1}, t_{k-1}, \theta)(t_k - t_{k-1})}. \tag{2.19}$$

But since the model (2.17) implies the distribution of the state variable, the Maximum Likelihood (ML) method, which considers the distribution in the system, was proposed and widely used in the system identification for the linear SDE model. It is firstly studied to consider only the parameter in the deterministic part–drift parameter identification. Drift parameter estimation in stochastic processes based on discrete measurement has been studied by many authors since 1970s. Le Breton (88) appeared to be the first person to study the estimation in discretely observed ISDE model. His models are the linear ISDE models with constant diffusion coefficients. And Le Breton used Approximate Maximum Likelihood (AML) estimation. In 1977, Robinson studied exact maximum likelihood estimation in discretely observed Ornstein-Uhlenbeck process which is a special ISDE model. This can be referred in (116). From then on, some researchers worked on approximate maximum likelihood estimation (where the continuous likelihood function is approximated), also called the maximum contrast estimation, such as in the works of (5; 46; 78) and so on. These kinds of approaches can all be referred to as Maximum Likelihood (ML) method.

The main idea of the ML method is as follows. Since the model (2.17) has the Markov property, it is possible to express the likelihood function of a given sequence of measurements $Y_0, Y_1, \ldots, Y_k, \ldots, Y_N$ solely in terms of the transition Probability Density Functions (pdfs), i.e.

$$L(\theta) = \prod_{k=1}^{N} p(Y_0, Y_1, \ldots, Y_k, \ldots, Y_N; \theta) \tag{2.20}$$

with θ is the unknown parameter vector, p is the probability density of corresponding measurements based on the parameter θ. The ML estimates are given by

$$\theta_{ML} = \arg\max_{\theta \in \Theta} L(\theta). \tag{2.21}$$

Then take θ_{ML} as the result of the parameter identification. In order to get the conditional probability density function in (2.20), the distribution of the state variable in the system need to be applied before the estimation.

The ML method can belong to the category of the statistic method. The main idea of the statistic method is to present a suitable statistic function, then optimize the proposed function and get the optimal value as the estimate. From another point of view, the ML method and LS method could also be seen as special cases of the Method of Moments (MM) (55). This method was originally developed for discrete time stochastic models, yet it may be applied to ISDE by computing moment conditions from a discrete version of the ISDE. It only used the certain moment condition to form a function and provide an estimate by minimizing the corresponding function. Its main advantage is that it requires specification only of certain moment conditions rather than the entire pdfs. This can also be a drawback, for it does not make efficient use of all the information in the samples, only applies the first moment (mean) or second moment (variance), which may lead to a loss of efficiency. For the model, the MM can not deal with the cases in which the state variable is unmeasured.

Another category to make parameter identification is filter based method. The ISDE model (2.17) plus (2.18) is considered as well. This kind of method is based on the filter technique, especially Kalman Filter (KF) technique. It has been more and more widely used for parameter identification in the application (94) since last century. Generally, the approaches using KF can be classified into two different categories. One category is referred as direct approaches. This kind of approach firstly takes both of the state variable and the unknown parameter(s) into a new augmented system state. Then, KF, Extended Kalman Filter (EKF) or some other suitable filter is used to make the estimation of the new state and thereby the estimation of unknown parameter(s) is obtained from this new state estimate. However, if the diffusion term of the ISDE model contains unknown parameters, this kind of approach could not show a good performance regarding the precision. Moreover, if the system model is a nonlinear one, this method sometimes could not provide a good performance with regard to the accuracy because of the nonlinearity of the system.

The other kind of KF based method is to combine KF technique with some statistic methods. This scheme mainly consists of two sequential stages. The first stage conducts the state estimation using KF, where the estimated state is a function of unknown parameters. Then, a statistic criterion, such as Maximum Likelihood (ML) and Least Mean Square (LMS), is set up in the second stage based on the estimated state. Thereby, the parameter identification problem becomes an optimization of a parameterized statistic problem. This approach can be directly applied to linear systems and explicit solutions may be found in (76; 107) and so on. Nevertheless, this kind of approach needs to be extended in order to handle nonlinear cases. Then a ML/Prediction Error Decomposition (PED) method for direct estimation of parameters in ISDE is proposed in (108) based on the EKF. Kristensen, Madsen and Jørgensen, in (83), set up the scheme of parameter identification based on the EKF and ML as well as Maximum A Posteriori (MAP) estimation with software implementation. Both of the two methods can handle parameter identification for cases that the diffusion item consists of the unknown parameter(s). But the precision of estimation need to be improved for some nonlinear models (96).

For the parameter identification using the filter techniques to estimate the state, since it is based on the state space model, this method is referred as state estimation based method in the thesis. In the next section, we will explain it in detail. In the thesis, EKF plus ML method is firstly considered to make the parameter identification for the parameters both in drift and diffusion items. Then this method is improved and extended to Unscented Kalman Filter (UKF) plus ML method to deal with some other nonlinear cases. It shows that using UKF plus ML method has its own advantages than the EKF plus ML method in terms of accuracy and convergency.

2.4 State Estimation Based Methods

At first, the detailed parameter identification methods are introduced for a general state space model, which is described by parameterized ISDE model given by (2.17) plus (2.18). Since the thesis focuses on the KF based methods, before the introduction of the identification methods, the two kinds of Kalman Filters are summarized in the following.

2.4.1 Discretized Model

In order to apply the EKF and UKF, the model (2.17) need to be discretized beforehand. According to Euler discretization, the (2.17) can be discretized as:

$$X_k = X_{k-1} + g_1(X_{k-1}, u(k-1), t_{k-1}, \theta)(t_k - t_{k-1}) + g_2(X_{k-1}, u(k-1), t_{k-1}, \theta)(B_k - B_{k-1}). \tag{2.22}$$

Based on the discretized model (2.22), the EKF or UKF can be performed in the following.

2.4.2 Extended Kalman Filter (EKF)

Based on the system (2.22) with (2.18), the EKF can be performed according to the following procedure (18).

Initialization with: original state estimation X_0 and variance estimation P_0.

Time-updated (Prediction):

$$\begin{aligned}
&\hat{X}_{k|k-1} = \hat{X}_{k-1|k-1} + g_1(\hat{X}_{k-1|k-1}, u_{k-1}, t_{k-1}, \theta)(t_k - t_{k-1}),\\
&P_{k|k-1} = \Phi_{k-1} P_{k-1|k-1} \Phi_{k-1}^T + g_2 g_2^T(\hat{X}_{k-1|k-1}, u_{k-1}, t_k, \theta)(t_k - t_{k-1}),\\
&S_k = H_k P_{k|k-1} H_k^T + R,\\
&K_k = P_{k|k-1} H_k^T S_k^{-1},
\end{aligned}$$

where $\hat{X}_{k|k-1}$ and $P_{k|k-1}$ are the estimates of state and variance of state at time t_k conditionally on all the information available at time t_{k-1}, S_k is the estimate of the variance of measurement at time t_k and K_k is Kalman gain.

Measurement-updated (Update):

$$
\begin{aligned}
r_k &= Y_k - h(\hat{X}_{k|k-1}, u_k, t_k),\\
\hat{X}_{k|k} &= \hat{X}_{k|k-1} + K_k r_k,\\
P_{k|k} &= (I - K_k H_k) P_{k|k-1},
\end{aligned}
$$

where $k \geq 1$, $g_2^T(\cdot)$ stands for the transpose of $g_2(\cdot)$, r_k is the error between the measurement and estimated measurement, and

$$
\begin{aligned}
\Phi_k &= \frac{\partial(X_k + g_1(X_k, u_k, t_k, \theta)(t_k - t_{k-1}))}{\partial X_k}\Big|_{X_k = \hat{X}_{k|k}},\\
H_k &= \frac{\partial(h(X_k, u_k, t_k))}{\partial X_k}\Big|_{X_k = \hat{X}_{k|k-1}}
\end{aligned}
$$

Note that in the variance prediction stage, a property of standard Brown Motion is applied (182), i.e,

$$E(B_k - B_{k-1})^2 = (t_k - t_{k-1})I.$$

2.4.3 Unscented Kalman Filter (UKF)

The same to the EKF, during the first stage, the state estimation can be accomplished by UKF (35) as well. Its procedure is as follows:

Initialization with: original state estimation X_0 and variance estimation P_0.

The first step is to create $2n+1$ sigma-points in such a way that these points together can capture both the mean and covariance of the state. Then, the matrix χ is formulated to contain these points, and its columns are calculated as follows:

$$
\begin{aligned}
\chi_{i,k-1} &= X_{k-1}, && i = 0\\
\chi_{i,k-1} &= X_{k-1} + (\sqrt{(n+\lambda)P_{k-1}})_i, && i = 1, \ldots, n\\
\chi_{i,k-1} &= X_{k-1} - (\sqrt{(n+\lambda)P_{k-1}})_{i-n}, && i = n+1, \ldots, 2n
\end{aligned}
$$

where i in the subscript means the i-th column, $k \geq 1$, $\lambda = \alpha^2(n+\kappa) - n$ is a scaling parameter, α determines the spread range of the sigma points around the state X_{k-1} and is usually set as a small positive value in order to avoid non-local effects (in the examples of thesis, α is chosen as 0.001), κ is called as secondary scaling parameter which is usually set as 0.

Each sigma-point is combined with a weight. These weights are calculated by comparing the moments of these sigma-points with Taylor series expansion of the models

(35). As a result, the weights for calculating mean and covariance estimates are given as:

$$
\begin{aligned}
&W_0^{(m)} = \tfrac{\lambda}{(n+\lambda)}, \\
&W_0^{(c)} = \tfrac{\lambda}{(n+\lambda)} + (1-\alpha^2+\beta), \\
&W_i^{(m)} = W_i^{(c)} = \tfrac{1}{2(n+\lambda)}, \qquad\qquad i = 1,\ldots,2n
\end{aligned}
$$

where parameter β is used to incorporate prior knowledge of the distribution of X, for Gaussian distributions $\beta = 2$ is optimal in general cases. The superscripts m and c stand for that the corresponding weights are used to calculate the mean and covariance of the state respectively.

The filter then predicts the state in the following step by propagating sigma-points through the state and measurement models, and calculating weighted averages and covariance matrices of the states:

$$
\begin{aligned}
&\chi_{i,k|k-1} = \chi_{i,k-1} + g_1(\chi_{i,k-1}, u_{k-1}, t_{k-1}, \theta)(t_k - t_{k-1}) \\
&\hat{X}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} \chi_{i,k|k-1} \\
&P_{k|k-1} = \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{X}_{k|k-1}][\chi_{i,k|k-1} - \hat{X}_{k|k-1}]^T \\
&Y_{k|k-1} = h(\chi_{k|k-1}, u_{k-1}, t_{k-1}) \\
&\hat{Y}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} Y_{i,k|k-1}
\end{aligned}
$$

The predictions are updated by: first, calculating the measurement covariance and state-measurement cross correlation matrices, and then, determining the Kalman gain, at last the updated estimation of state and variance is obtained:

$$
\begin{aligned}
&P_{YY} = \sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \hat{Y}_{k|k-1}][Y_{i,k|k-1} - \hat{Y}_{k|k-1}]^T \\
&P_{XY} = \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{X}_{k|k-1}][Y_{i,k|k-1} - \hat{Y}_{k|k-1}]^T \\
&K_k = P_{XY} P_{YY}^{-1} \\
&r_k = (Y_k - \hat{Y}_{k|k-1}) \\
&\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k r_k \\
&P_{k|k} = P_{k|k-1} - K_k P_{YY} K_k^T
\end{aligned}
$$

In stead of linearizing a nonlinear function, UKF generates $2N+1$ sigma points for states estimation which are propagated through the actual non-linear function, eliminating linearization. The points are chosen such that their mean, covariance as well as other higher order moments can be easily caught up. These propagated points help in recalculating the mean and covariance, then yielding more accurate results compared to ordinary function linearization. The underlying idea is to approximate the probability distribution instead of the function (35). This strategy results in decrement in computational complexities at the same time increasing estimation accuracy, gaining faster and more accurate results.

The unscented transform approach provides another advantage of treating noise in a nonlinear system to account for non-Gaussian or non-additive noises. For doing so firstly noise is propagated through the functions by augmenting the state vector including the noise sources. This technique was first introduced by Julier (72) and later developed by Merwe (109). Sigma point samples are then selected from the augmented state, which includes the noise values. This technique results in the accuracy of process and measurement noise captured with same accuracy as that of the state, which in turn increases the accuracy of the estimation for non-additive noise systems (35).

2.4.4 Parameter Identification Based on the KF Methods

The scheme of the classic method based on the KF to solve the problem **(P)** is given:

Direct approach–only using the KF technology

- Initialization with state X_0 and variance P_0,
- Take the unknown parameter as the augment state to the system, rewrite the system model using the new state as:

$$\begin{cases} \tilde{X}_k = [X_k, \theta]' \\ \tilde{X}_k = \tilde{X}_{k-1} + \begin{pmatrix} g_1(\tilde{X}_{k-1}, u_{k-1}, t_{k-1}) \\ 0 \end{pmatrix} (t_k - _{k-1}) + \begin{pmatrix} g_2(\tilde{X}_{k-1}, u_{k-1}, t_k) \\ 0 \end{pmatrix} (B_k - B_{k-1}) \end{cases} \tag{2.23}$$

- Use KF technique, like EKF and UKF, to estimate the state, and take last part of the estimation of the augment state as the result for parameter identification.

KF plus ML method

- Initialization with state X_0 and variance P_0,
- Use KF technique to estimate the state which is parameterized by the unknown system parameters,
- Form the Maximum Likelihood function using the parameterized state estimation,
- Solve the optimization problem of the parameterized Maximum Likelihood function, then get optimal solution as the result of the parameter identification.

The original KF plus ML method applied linear Kalman Filter firstly for linear state space model (144; 148). This method is developed to use EKF to handle the nonlinear system in (83). In order to improve the performance for some kinds of nonlinear system, this thesis applied UKF instead of EKF in the estimation. The scheme of the new method based on the UKF to solve the problem **(P)** adopts UKF in the second state estimation step and repeats the same procedure as that in the KF plus ML method.

The first two steps of the KF plus ML method can be followed by the previous procedures of EKF and UKF. As soon as the state estimation is obtained, the last two steps will begin in the following.

Introducing the notation

$$\mathscr{Y}_k = [Y_k, Y_{k-1}, \ldots, Y_1, Y_0],$$

then, the likelihood function becomes the joint probability density, i.e.,

$$L(\theta; \mathscr{Y}_N) = p(\mathscr{Y}_N \mid \theta), \tag{2.24}$$

or equivalently

$$L(\theta; \mathscr{Y}_N) = \left(\prod_{k=1}^{N} p(Y_k \mid \mathscr{Y}_{k-1}, \theta) \right) p(Y_0 \mid \theta). \tag{2.25}$$

In order to carry out the optimization of the likelihood function, the state estimation needs to be solved beforehand in order to obtain the estimated outputs. For the ISDE in (2.17) is driven by a Brown Motion which can be seen as a Wiener process, and the increments of a Wiener process are still Gaussian, it is reasonable to assume the conditional densities can be well approximated by Gaussian densities, which have two parameters, i.e., the means and covariances. Based on this fact and the previous results of state estimation, the parameterized likelihood function can be formulated as:

$$L(\theta;\mathscr{Y}_N) = \left(\prod_{k=1}^{N} \frac{\exp(-\frac{1}{2}r_k^T P_{YY}^{-1} r_k)}{\sqrt{\det(P_{YY})}(\sqrt{2\pi})^n} p(Y_0 \mid \theta)\right), \tag{2.26}$$

where P_{YY} is the covariance matrix of the measurement Y, while the same matrix is represented as S_k in EKF, and superscript -1 stands for the inverse of the corresponding matrix.

Then, the previous identification problem **(P)** can be converted to an optimization problem which may be described as:

(P') Given a set of measured output $\mathscr{Y}_k$ and input signals $u(t_k) \in \mathscr{U}$, find θ by solving the optimization problem which is defined in the following

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\{-\ln(L(\theta;\mathscr{Y}_N \mid Y_0))\}. \tag{2.27}$$

2.4.5 Optimization Computing Method

Some optimization algorithms which are used to make computation are needed to solve the optimization problem of the ML function (2.27). Generally, in order to get the solution of it, the convex property of the formulated optimization problem (2.27) needs to be explored firstly. Even though it might be a non-convex problem sometimes and hardly to obtain the global solution. But a better initial value could possibly lead to a global optimal solution. Especially, for some special cases such as linear systems, the global optimal solution could be obtained. The standard optimization method to solve ML problem could be seen see e.g. (14). Another popular method to solve this optimization problem of ML function is the Expectation Maximization (EM) algorithm. It was originally proposed in (26). The EM algorithm includes two steps: expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and maximization (M) step, which computes to find parameters maximizing the expected log-likelihood. It shows good performance especially to find the maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. The details for the EM algorithm can be found in (106), see also from (47) robust estimation of Linear Time Invariant (LTI) state space models. In (139), a solution of more complicated problem to estimate non-linear state space models is given, using particle filter.

In the case of this thesis, since the KF based methods are concerned and the explicit system model is known beforehand, EM approach is not adopted, here the quasi-Newton using BFGS update method (14) is adopted, which is summarized as (for simplicity, assume $F(\theta) = -\ln(L(\theta; \mathscr{Y}_N \mid Y_0))$):
Starting with an initial guess θ_0 and an approximate Hessian matrix C_0, the following steps are repeated until θ_k converges to the solution.

1. Obtain a direction d_k by solving: $C_k d_k = -\bigtriangledown F(\theta)$, where $\bigtriangledown F(\theta_k)$ means the gradient of the function F at point θ_k.

2. Perform a line search to find an acceptable step size α_k by minimizing $F(\theta_k + \alpha d_k)$ over $\alpha \geq 0$, then update $\theta_{k+1} = \theta_k + \alpha_k d_k$.

3. Set $S_k = \alpha_k d_k$.

4. $\delta_k = \bigtriangledown F(\theta_{k+1}) - \bigtriangledown F(\theta_k)$.

5. $C_{k+1} = C_k + \frac{\delta_k \delta_k^T}{\delta_k^T S_k} - \frac{C_k S_k (C_k S_k)^T}{S_k^T C_k S_k}$.

Convergence can be checked by observing the norm of the gradient, $\mid \bigtriangledown F(\theta_k) \mid$. If its value less than a predefined threshold, the process will be terminated. Take the value θ_k at the recent step as the optimal solution of the original problem, denoted as $\hat{\theta}$. It is also severed as the estimated value of the unknown parameter in the system.

2.5 Consistency and Normality

Before the identification method is used, the property of the estimation should be investigated firstly.

2.5.1 Pre-knowledge

In order to make the content clear, several definitions need to be given according to the statistical theory beforehand, (24).

- **Consistency:** $\hat{\theta}_N$ is said to be consistent if $\hat{\theta}_N \to \theta_0$ in probability 1 as $N \to \infty$, i.e.,

$$P\{\lim_{N\to\infty} \hat{\theta}_N = \theta_0\} = 1,$$

  where θ_0 is the true unknown parameter value. In the following, if $\hat{\theta}_N$ is consistent, it is noted as $\hat{\theta}_N \xrightarrow{p} \theta_0$.

- **Asymptotic Normality:** $\hat{\theta}_N$ is said to be asymptotic normality if there exists a function $d(N)$ such that when $N \to \infty$ the limiting distribution of $d(N)(\hat{\theta}_N - \theta_0)$ is normal distribution with 0 mean and variance $\sigma^2_{\theta_0}$, noted as $d(N)(\hat{\theta}_N - \theta_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\theta_0})$. Here $\sigma^2_{\theta_0}$ is called the asymptotic variance of the estimate θ_0.

- **Fisher Information Matrix (FIM):** Fisher information matrix of measurement Y with regard to the parameter θ_0 is defined by

$$\varphi_N(\theta_0) = E[\frac{\partial l_N(\theta_0)}{\partial \theta} \frac{\partial l_N^T(\theta_0)}{\partial \theta}].$$

  Note that in the calculation in (2.24), all the sample points are taken as known information, it is not necessary to consider the expectation. But if the sampled point is taken as a variable, it will have the stochastic property. Then the fisher information considers the expectation.

  A property of fisher information matrix is needed in the following, it is summarized as the following lemma:

  Lemma 2.5.1

$$E(\frac{\partial^2 l_N(\theta_0)}{\partial \theta \partial \theta^T}) = -\varphi_N(\theta_0).$$

  This Lemma and proof can be found in (91). Noted that from the definition of FIM, it is a positive semidefinite symmetric matrix. If in the non positive definite case, its inverse means Moore-Penrose inverse. For the simplicity, in the thesis, $\varphi_N(\theta)$ will be noted as $\varphi(\theta)$.

2.5.2 State of the Art

The KF technique plus ML methods may belong to the category of Maximum Likelihood Estimation. Crowder (22), who was firstly proposed ML estimation, showed MLE has the weak consistency and asymptotic normality for independent observations if the following conditions are satisfied:

- $\lambda_{\min}\{\varphi(\theta_0)\} \to \infty$ as $N \to \infty$, where $\lambda_{\min}\{\varphi(\theta_0)\}$ is the minimum characteristic root of the matrix $\varphi(\theta_0)$ and $\varphi(\theta_0)$ is the information matrix $E(-\frac{\partial L}{\partial\theta\partial\theta^T}) = \varphi$ evaluated at θ_0 which is the true value of the parameter vector.
- $-\varphi(\theta_0)^{-1}(\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta^T}) \to I$.
- Given $\varepsilon > 0$, $\exists\ \delta(\varepsilon) > 0$ subject to

$$P[|\{\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta^T}(\theta_0)\}^{-1}\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta^T}(\theta)| < \varepsilon] \to 1, \quad as \quad N \to \infty, \quad when \ |\theta - \theta_0| \leq \delta. \tag{2.28}$$

And in 1980, Adrian Pagan (125) proved that:

Theorem 2.5.2 *For a linear stochastic differential model which can be described by* (2.17) *and* (2.18) *with that all functions are linear, if*

A(i) *all random features are stationary,*

A(ii) $\theta_0 \in$ *interior of* Θ *compact in Euclidean space*

are satisfied, then

$$\varphi^{\frac{1}{2}}(\hat{\theta}_{ML} - \theta_0) \to \mathcal{N}(0, I). \tag{2.29}$$

Here the value $\hat{\theta}_{ML}$ is the estimate using Kalman Filter plus Maximum Likelihood method. This theorem showed that the estimation based on the KF plus ML method tends to a normal distribution when the sampling number tends to infinity. Moreover, the mean of the estimation tends to the true value of the parameter and its variance tends to the inverse of the information matrix for the maximum likelihood function.

At the same time, some condition which is called as non-local minimum have been developed to guarantee global convergence for certain types of models, such as (5) for ARMA model, (48) for ARMAX model and so on. As is observed, regarding ML method based on KF technique, in (125), (46), Pagan and Ghosh considered KF plus ML parameter identification for different linear models and they both showed under some conditions, the consistency and normality of the estimate are hold. But so far, the investigation for the convergence property of ML methods was only based on the linear models or some certain input-output models. Bearing them in the mind, in the thesis, the consistency and normality with regard to the UKF plus ML method to handle some nonlinear systems are derived.

In order to make the content more simplicity, the convergence property is investigated for the nonlinear discrete systems. In the following parts, the system model and the identification approach are rewritten by their discrete version. This work is based on the UKF plus ML method to identify some kinds of the nonlinear SS models.

2.5.3 UKF plus ML Method

For the application convenience, since most models or systems are performed in the discrete version, the model considered here is a discrete one with noise. It is described by the following discrete time model:

$$x_k = F(x_{k-1}, u_{k-1}, \theta) + \omega_{k-1}, \tag{2.30}$$

where $k \in \mathbb{Z}$ is the discrete time variable, $x_k \in \mathscr{X} \subset \mathbb{R}^n$ is a vector of state variables, $u_k \in \mathscr{U} \subset \mathbb{R}^m$ is a vector of input variables, $\{\omega_k\}$ is an n-dimensional standard Wiener process. $\theta \in \Theta$ is the unknown parameter vector.

The measurement of the considered system is described by

$$y_k = h(x_k) + \varepsilon_k, \tag{2.31}$$

where $y_k \in \mathscr{Y} \subset \mathbb{R}^m$ is a vector of output variables, $h(\cdot) \in \mathbb{R}^m$, $\{\varepsilon_k\}$ is an m-dimensional white noise process.

In order to make the parameter identification to the nonlinear system described by (2.30) and (2.31), the UKF plus ML method is applied. It can be summarized in the following steps.

2.5.3.1 Parameterized State Estimation

Unscented Kalman Filter (UKF):

Initialization with: state estimation x_0 and variance estimation P_0.

Creating sigma-points

$$\begin{array}{ll} \chi_{i,k-1} = x_{k-1}, & i = 0 \\ \chi_{i,k-1} = x_{k-1} + (\sqrt{(n+\lambda)P_{k-1}})_i, & i = 1, \ldots, n \\ \chi_{i,k-1} = x_{k-1} - (\sqrt{(n+\lambda)P_{k-1}})_{i-n}, & i = n+1, \ldots, 2n \end{array}$$

where the notation follows the previous equations.

Define the weight function

$$\begin{array}{ll} W_0^{(m)} = \frac{\lambda}{(n+\lambda)}, & \\ W_0^{(c)} = \frac{\lambda}{(n+\lambda)} + (1-\alpha^2+\beta), & \\ W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n+\lambda)}, & i = 1, \ldots, 2n \end{array}$$

.

The prediction procedure of the UKF is

$$\begin{array}{l} \chi_{i,k|k-1} = F(\chi_{i,k-1}, u_{k-1}, \theta) \\ \hat{x}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} \chi_{i,k|k-1} \\ P_{k|k-1} = \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{X}_{k|k-1}][\chi_{i,k|k-1} - \hat{X}_{k|k-1}]^T \\ Y_{i,k|k-1} = h(\chi_{i,k|k-1}) \\ \hat{y}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} Y_{i,k|k-1} \end{array}$$

.

Then updating:

$$\begin{array}{l} P_{yy,k} = \sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \hat{y}_{k|k-1}][Y_{i,k|k-1} - \hat{y}_{k|k-1}]^T \\ P_{xy,k} = \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_{k|k-1}][Y_{i,k|k-1} - \hat{x}_{k|k-1}]^T \\ K_k = P_{xy,k} P_{yy,k}^{-1} \\ r_k = (y_k - \hat{y}_{k|k-1}) \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k r_k \\ P_{k|k} = P_{k|k-1} - K_k P_{yy,k} K_k^T \end{array}$$

. Here the estimated variable has the following interpretation in probability.

$$\begin{aligned}\hat{x}_{k|k-1} &= E(x_k \mid \mathscr{Y}_{k-1})\\ \hat{x}_{k|k} &= E(x_k \mid \mathscr{Y}_k)\\ P_{k|k-1} &= E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T \mid \mathscr{Y}_{k-1})\\ P_{k|k} &= E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T \mid \mathscr{Y}_k)\end{aligned}$$

2.5.3.2 ML Optimization

Formulating the ML function based on the previous state estimation,

$$L(\theta;\mathscr{Y}_N) = \left(\prod_{k=1}^{N} \frac{\exp(-\frac{1}{2} r_k^T P_{yy,k}^{-1} r_k)}{\sqrt{\det(P_{yy,k})}(\sqrt{2\pi})^n}\right) p(y_0 \mid \theta), \tag{2.32}$$

then the identification problem is to solve the following optimization problem

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\{\ln(L(\theta;\mathscr{Y}_N)\}. \tag{2.33}$$

In order for the convenience, in the following, let

$$\begin{aligned}l_N(\theta) &= \ln(L_N(\theta))\\ &= \ln(P(y_0 \mid \theta)) - \sum_{k=1}^{N} \frac{1}{2}\{\ln(\det(P_{yy,k})) + nN\ln(2\pi) + r_k^T P_{yy,k}^{-1} r_k\}.\end{aligned} \tag{2.34}$$

2.5.4 Properties of UKF Plus ML Method

In this section, the boundness is considered as element boundness for vectors and element boundness for matrices with ignoring the influence of the noise. Moreover, the boundness means the element has both lower and upper bounds.

The main theorem for the convergence of UKF plus ML method is described as:

Theorem 2.5.3 *For the stochastic parameterized system which is described by* (2.30) *and* (2.31)*, if the following conditions are satisfied*

(a) *Function $F(\cdot,\cdot,\cdot)$ and its derivatives up to second order with regard to θ are bounded with different lower and upper bounds and θ-continuous, $h(\cdot)$ and its derivatives up to second order are bounded with different lower and upper bounds and continuous.*

(b) *For the composite function $h(F)$, there exists θ-continuous functions g_1 and g_2 such that*

$$h(F(x_{k-1},u_{k-1},\theta)+\omega_{k-1}) = g_1(x_{k-1},u_{k-1},\theta) + g_2(u_{k-1},\theta)\omega_{k-1}, \tag{2.35}$$

and

$$h(F(x_{k-1},u_{k-1},\theta)) = g_1(x_{k-1},u_{k-1},\theta). \tag{2.36}$$

(c) *The true value θ_0 of θ is an interior point of a compact set Θ.*

then the estimation value $\hat{\theta}$ using UKF plus ML method is consistent and asymptotic normal, i.e., when $N \to \infty$, there is

$$\hat{\theta}_N \xrightarrow{p} \theta_0, \tag{2.37}$$

and furthermore

$$\hat{\theta}_N - \theta_0 \xrightarrow{d} \mathcal{N}(0, \varphi^{-1}(\theta_0)). \tag{2.38}$$

The proof of Theorem 2.5.3 is based on the Crowder's theorem in (22) which showed that the ML for dependent observations is consistent if three conditions are satisfied and these conditions can be seen in previous section.

In order to make the analysis clear, boundness condition (a) is rewritten in mathematics. The condition tells that there exists some couples of known boundaries (m_{Fi}, M_{Fi}) and (m_{hi}, M_{hi}) with $m_{Fi} < M_{Fi}$ and $m_{hi} < M_{hi}$ as $i = 0,1,2$ such that

$$m_{Fi} \le element\{F_\theta^i(x,u,\theta)\} \le M_{Fi}, \tag{2.39}$$

and

$$m_{hi} \le element\{h^i(\cdot)\} \le M_{hi}. \tag{2.40}$$

Here the $element\{A\}$ means each element of A no matter A is a vector or matrix. In the following, for simplicity, (2.39) and (2.40) are noted as $m_{Fi} \le F_\theta^i \le M_{Fi}$ and $m_{hi} \le h^i \le M_{hi}$.

In order to prove Theorem 2.5.3, firstly the optimization function need to be checked for its derivatives up to second order with regard to the unknown parameter. Differentiating $l_N(\theta)$ in (2.34) with respect to one parameter θ_i, it gives:

$$\frac{\partial l_N(\theta)}{\partial \theta_i} = \sum_{k=1}^{N}\{-\frac{1}{2}tr(P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial \theta_i}) - \frac{\partial r_k^T}{\partial \theta_i}P_{yy,k}^{-1}r_k + \frac{1}{2}r_k^T P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial \theta_i}P_{yy,k}^{-1}r_k\}. \tag{2.41}$$

Here, the expressions for derivatives of a symmetric matrix $A(x)$ are used (104)

$$\frac{\partial(\det(A(x)))}{\partial x} = \det(A(x))tr(A^{-1}(x)\frac{\partial A(x)}{\partial x}),$$
$$\frac{\partial A^{-1}(x)}{\partial x} = -A^{-1}(x)\frac{\partial A(x)}{\partial x}A^{-1}(x).$$

Since the last term of (2.41) is a scale, it equals to its trace, then

$$\frac{\partial l_N(\theta)}{\partial \theta_i} = \sum_{k=1}^{N}\{-\frac{1}{2}tr(P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial \theta_i})(I-P_{yy,k}^{-1}r_k r_k^T) - \frac{\partial r_k^T}{\partial \theta_i}P_{yy,k}^{-1}r_k\}. \tag{2.42}$$

The second order derivative is

$$\begin{aligned}
\frac{\partial^2 l_N(\theta)}{\partial \theta_i \partial \theta_j} = &-\frac{1}{2}\sum_{k=1}^{N} tr[\partial(P_{yy,k}^{-1}\partial P_{yy,k}/\partial\theta_i)/\partial\theta_j(I-P_{yy,k}^{-1}r_k r_k^T)] \\
&-\frac{1}{2}\sum_{k=1}^{N} tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}r_k r_k^T] \\
&+\frac{1}{2}\sum_{k=1}^{N} tr\{P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}(\frac{\partial r_k}{\partial\theta_j}r_k^T + r_k\frac{\partial r_k^T}{\partial\theta_j})\} \\
&-\sum_{k=1}^{N}\frac{\partial^2 r_k}{\partial\theta_i\partial\theta_j}P_{yy,k}^{-1}r_k + \sum_{k=1}^{N}\frac{\partial r_k^T}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}r_k - \sum_{k=1}^{N}\frac{\partial r_k^T}{\partial\theta_i}\cdot P_{yy,k}^{-1}\cdot\frac{\partial r_k}{\partial\theta_j}.
\end{aligned} \tag{2.43}$$

The only random variable in (2.43) is r_k and from the algorithm of UKF, (2.35) and (2.36), it can be obtained that $E(r_k r_k^T) = P_{yy,k}$, and

$$\begin{aligned}
Er_k &= E(y_k - \hat{y}_{k|k-1}) \\
&= E\{h(x_k) + \varepsilon_k - \sum_{i=0}^{2n} W_i^{(M)} Y_{i,k|k-1}\} \\
&= E\{h(x_k) - \sum_{i=0}^{2n} W_i^{(M)} Y_{i,k|k-1}\} \\
&= E\{h(F(x_{k-1},u_{k-1},\theta)+\omega_{k-1}) - \sum_{i=0}^{2n} W_i^{(M)} h(F(\chi_{i,k-1},u_{k-1},\theta))\} \\
&= E\{g_1(x_{k-1},u_{k-1},\theta) - \sum_{i=0}^{2n} W_i^{(M)} g_1(\chi_{i,k-1},u_{k-1},\theta)\} \\
&= 0.
\end{aligned} \tag{2.44}$$

and if the measurement y_k is taken as a known scale, then

$$\begin{aligned}
\frac{\partial r_k}{\partial \theta_i} &= \frac{\partial (y_k - \hat{y}_{k|k-1})}{\partial \theta_i} \\
&= \frac{\partial y_k}{\partial \theta_i} - \frac{\partial \hat{y}_{k|k-1}}{\partial \theta_i} \\
&= 0 - \frac{\partial \hat{y}_{k|k-1}}{\partial \theta_i} \\
&= -\frac{\partial (\sum_{i=0}^{2n} W_i^{(M)} Y_{i,k|k-1})}{\partial \theta_i} \\
&= -\sum_{i=0}^{2n} \frac{\partial [\sum_{i=0}^{2n} W_i^{(M)} h(\chi_{i,k|k-1})]}{\partial \theta_i} \\
&= -\sum_{i=0}^{2n} \frac{\partial [\sum_{i=0}^{2n} W_i^{(M)} h(F(\chi_{i,k-1}, u_{k-1}, \theta))]}{\partial \theta_i} \\
&= -\sum_{i=0}^{2n} [\sum_{i=0}^{2n} W_i^{(M)} h'_F \cdot \frac{\partial F(\chi_{i,k-1}, u_{k-1}, \theta)}{\partial \theta_i}].
\end{aligned} \tag{2.45}$$

It can be observed that the random property of the first order derivative of r_k with regard to θ_i only comes from $\chi_{i,k-1}$. As a result, from the observation of kth step, the first order derivative of r_k with regard to θ_i only depends on past innovations and the known input signal (control variable). It is the same to $\frac{\partial^2 r_k}{\partial \theta_i \partial \theta_j}$. Then it can be concluded that the first order and second order derivatives of r_k with regard to θ_i are independent with r_k.

From (2.45) and the previous independency interpretation, taking the expectation of (2.43),

$$
\begin{aligned}
E(\frac{\partial^2 l_N(\theta)}{\partial\theta_i\partial\theta_j}) = & -\frac{1}{2}\sum_{k=1}^{N} tr\{\partial(P_{yy,k}^{-1}\partial P_{yy,k}/\partial\theta_i)/\partial\theta_j[I-P_{yy,k}^{-1}E(r_k r_k^T)]\} \\
& -\frac{1}{2}\sum_{k=1}^{N} tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}E(r_k r_k^T)] \\
& +\frac{1}{2}\sum_{k=1}^{N} tr\{P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}[E(\frac{\partial r_k}{\partial\theta_j})E(r_k^T)+E(r_k)E(\frac{\partial r_k^T}{\partial\theta_j})]\} \\
& -\sum_{k=1}^{N} E(\frac{\partial^2 r_k}{\partial\theta_i\partial\theta_j})P_{yy,k}^{-1}E(r_k)+\sum_{k=1}^{N} E(\frac{\partial r_k^T}{\partial\theta_i})P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}E(r_k) \\
& -\sum_{k=1}^{N} E(\frac{\partial r_k^T}{\partial\theta_i}\cdot P_{yy,k}^{-1}\cdot\frac{\partial r_k}{\partial\theta_j}) \\
= & \, 0-\frac{1}{2}\sum_{k=1}^{N} tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}]+0-0+0-\sum_{k=1}^{N} E(\frac{\partial r_k^T}{\partial\theta_i}\cdot P_{yy,k}^{-1}\cdot\frac{\partial r_k}{\partial\theta_j}).
\end{aligned}
\tag{2.46}
$$

Then, using the Lemma 2.5.1 it can be obtained that the ijth element of the information matrix is

$$
\begin{aligned}
\varphi_{ij} &= -E(\frac{\partial^2 l_N(\theta)}{\partial\theta_i\partial\theta_j}) \\
&= \sum_{k=1}^{N}\frac{1}{2}tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}]+\sum_{k=1}^{N} E(\frac{\partial r_k^T}{\partial\theta_i}\cdot P_{yy,k}^{-1}\cdot\frac{\partial r_k}{\partial\theta_j}).
\end{aligned}
\tag{2.47}
$$

Moreover, in order to prove the Theorem 2.5.3, the following Lemmas need to be applied.

Lemma 2.5.4 *If functions $F(\cdot,\cdot,\cdot)$ and $h(\cdot)$ are uniformly bounded, then $P_{k|k}$ is uniformly bounded.*

Proof: If $F(\cdot,\cdot,\cdot)$ and $h(\cdot)$ are uniformly bounded, from (2.30) and (2.31), all of the state, measurement and their one step estimation are uniformly bounded. According to the UKF algorithm, $\chi_{i,k|k-1}$ and $Y_{i,k|k-1}$ are uniformly bounded as well.

$$\begin{aligned}
P_{k|k} &= P_{k|k-1} - K_k P_{yy,k} K_k^T \\
&= P_{k|k-1} - P_{xy,k} P_{yy,k}^{-1} P_{xy,k}^T \\
&= \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{X}_{k|k-1}][\chi_{i,k|k-1} - \hat{X}_{k|k-1}]^T \\
&+ \{\sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_{k|k-1}][Y_{i,k|k-1} - \hat{x}_{k|k-1}]^T\} \\
&\quad \{\sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \hat{y}_{k|k-1}][Y_{i,k|k-1} - \hat{y}_{k|k-1}]^T\}^{-1} \\
&\quad \{\sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_{k|k-1}][Y_{i,k|k-1} - \hat{x}_{k|k-1}]^T\}^T.
\end{aligned} \tag{2.48}$$

From (2.48), it can be seen that $P_{k|k}$ is a function of the state, measurement and their one step estimation. Since all of them are uniformly bounded, then $P_{k|k}$ is uniformly bounded.♯

Lemma 2.5.5 *If condition (a) is satisfied, then $P_{yy,k}$ and its derivatives up to second order with regard to θ are bounded.*

Proof:

$$\begin{aligned}
P_{yy,k} &= \sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \hat{y}_{k|k-1}][Y_{i,k|k-1} - \hat{y}_{k|k-1}]^T \\
&= \sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \sum_{i=0}^{2n} W_i^{(m)} Y_{i,k|k-1}][Y_{i,k|k-1} - \sum_{i=0}^{2n} W_i^{(m)} Y_{i,k|k-1}]^T \\
&= \sum_{i=0}^{2n} W_i^{(c)} [h(\chi_{i,k|k-1}) - \sum_{i=0}^{2n} W_i^{(m)} h(\chi_{i,k|k-1})][h(\chi_{i,k|k-1}) - \sum_{i=0}^{2n} W_i^{(m)} h(\chi_{i,k|k-1})]^T \\
&= \sum_{i=0}^{2n} W_i^{(c)} [h(F(\chi_{i,k-1}, u_{k-1}, \theta)) - \sum_{i=0}^{2n} W_i^{(m)} h(F(\chi_{i,k-1}, u_{k-1}, \theta))] \\
&\qquad [h(F(\chi_{i,k-1}, u_{k-1}, \theta)) - \sum_{i=0}^{2n} W_i^{(m)} h(F(\chi_{i,k-1}, u_{k-1}, \theta))]^T.
\end{aligned} \tag{2.49}$$

Since condition (a) is satisfied, from (2.40)

$$m_{h0} \leq h(F(\chi_{i,k-1}, u_{k-1}, \theta)) \leq M_{h0}, \tag{2.50}$$

it can be obtained

$$element\{P_{yy,k}\} \leq (M_{h0} - m_{h0})^2. \tag{2.51}$$

(2.51) means $P_{yy,k}$ is bounded. For the derivatives of $P_{yy,k}$ with regard to θ, from (2.49), it can be seen that it only depends on the derivative of $h(F(\chi_{i,k-1}, u_{k-1}, \theta))$ with regard to θ.

$$\frac{\partial h(F(\chi_{i,k-1}, u_{k-1}, \theta)}{\partial \theta} = \frac{\partial h(F(\chi_{i,k-1}, u_{k-1}, \theta))}{\partial F(\chi_{i,k-1}, u_{k-1}, \theta)} \cdot F'_{\theta}(\chi_{i,k-1}, u_{k-1}, \theta). \tag{2.52}$$

According to condition (a), (2.39) and (2.40), there exists lower and upper bounds for the right part in (2.52), noted as m_{hf1} and M_{hf1}. Performing the same procedure as the proof for the $P_{yy,k}$, it can be concluded that the first order derivative of $P_{yy,k}$ with regard to θ is bounded. Similarly, the second order derivative of $P_{yy,k}$ with regard to θ can be proved bounded as well. ♯

Lemma 2.5.6 *If condition (b) is satisfied, then*

$$E(\frac{\partial \hat{x}_{k|k}}{\partial \theta} \cdot \frac{\partial \hat{x}_{k|k}^T}{\partial \theta}) < \infty, \tag{2.53}$$

and

$$E(\hat{x}_{k|k}\hat{x}_{k|k}^T) < \infty. \tag{2.54}$$

Proof: Using the condition (b), the conclusions can be obtained directly by the system model equation and the definition of $\hat{x}_{k|k}$ in the UKF. ♯

Lemma 2.5.7 *If condition (b) is satisfied,*

$$0 < E(\frac{\partial r_k^T}{\partial \theta_i} \cdot P_{yy,k}^{-1} \cdot \frac{\partial r_k}{\partial \theta_j}) < \infty. \tag{2.55}$$

Proof: Since $P_{yy,k}$ is positive definite, its inverse $P_{yy,k}^{-1}$ is positive definite as well. As a result, $\frac{\partial r_k^T}{\partial \theta_i} \cdot P_{yy,k}^{-1} \cdot \frac{\partial r_k}{\partial \theta_j} > 0$. Taking the expectation, the left inequality is proved. To prove the right part, firstly from the definition of r_k, it can be obtained that

$$
\begin{aligned}
& E(\frac{\partial r_k^T}{\partial \theta_i} \cdot P_{yy,k}^{-1} \cdot \frac{\partial r_k}{\partial \theta_j}) \\
&= E(\frac{\partial \hat{y}_{k|k-1}^T}{\partial \theta_i} \cdot P_{yy,k}^{-1} \cdot \frac{\partial \hat{y}_{k|k-1}}{\partial \theta_j}) \\
&= E\{\frac{\partial [\sum_{i=0}^{2n} W_i^{(m)} h(F(\chi_{i,k-1}, u_{k-1}, \theta))]^T}{\partial \theta_i} \cdot P_{yy,k}^{-1} \cdot \frac{\partial [\sum_{i=0}^{2n} W_i^{(m)} h(F(\chi_{i,k-1}, u_{k-1}, \theta))]}{\partial \theta_j}\}
\end{aligned} \tag{2.56}
$$

$P_{yy,k}$ is positive definite and bounded, then $P_{yy,k}^{-1}$ is also bounded. Combining with (2.52), the three parts in the product of right part in (2.56) are all bounded, then the right part of (2.56) is bounded. Then, the right part of (2.55) is proved. In a whole, (2.55) is established. ♯

Here a corollary can be obtained.

Corollary 2.5.8 *If condition (b) is satisfied, then* $\frac{\partial^2 P_{k|k}}{\partial \theta_i \partial \theta_j}$ *is uniformly bounded*

Proof: According to the algorithm of the UKF, it can be seen that

$$
\begin{aligned}
P_{k|k} &= P_{k|k-1} - K_k P_{yy,k} K_k^T \\
&= P_{k|k-1} - P_{xy,k} P_{yy,k}^{-1} P_{xy,k}^T \\
&= \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{X}_{k|k-1}][\chi_{i,k|k-1} - \hat{X}_{k|k-1}]^T \\
&\quad - \{\sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_{k|k-1}][Y_{i,k|k-1} - \hat{x}_{k|k-1}]^T\} \\
&\quad \{\sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \hat{y}_{k|k-1}][Y_{i,k|k-1} - \hat{y}_{k|k-1}]^T\}^{-1} \\
&\quad \{\sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_{k|k-1}][Y_{i,k|k-1} - \hat{x}_{k|k-1}]^T\}^T
\end{aligned} \tag{2.57}
$$

$\chi_{i,k|k-1}$, $\hat{x}_{k|k-1}$ only depend on function $F(\cdot,\cdot,\cdot)$ and $Y_{i,k|k-1}$, $\hat{y}_{k|k-1}$ only depend on functions $h(\cdot)$ and $F(\cdot,\cdot,\cdot)$. If differentiating $P_{k|k}$ by second order with regard to θ, the derivative only depends on second order derivative of $h(\cdot)$, $F(\cdot,\cdot,\cdot)$ and its derivatives up to second order with regard to θ. Condition (b) guarantees that all of them are bounded. As a result, $\frac{\partial^2 P_{k|k}}{\partial\theta_i\partial\theta_j}$ is bounded. ♯

This corollary shows that under the condition of the Theorem 2.5.3, the second derivative of the updated variance for the state is also bounded.

Theorem 2.5.9 *Under the condition of Theorem 2.5.3,*

$$\frac{1}{N}[\frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'} + \varphi(\theta_0)] \xrightarrow{p} 0. \tag{2.58}$$

Proof: From (2.43) and (2.47),

$$\begin{aligned}
&\frac{1}{N}[\frac{\partial^2 l_N(\theta_0)}{\partial\theta_i\partial\theta_j} + \varphi_{N,ij}(\theta_0)] \\
&= \frac{1}{2N}\sum_{k=1}^{N} tr[P_{yy,k}^{-1}\frac{\partial^2 P_{yy,k}}{\partial\theta_i\partial\theta_j}(P_{yy,k}^{-1}r_k r_k^T - I)] \\
&- \frac{1}{2N}\sum_{k=1}^{N} tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}r_k r_k^T] \\
&- \frac{1}{2N}\sum_{k=1}^{N} tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}r_k r_k^T] \\
&- \frac{1}{N}\sum_{k=1}^{N}\frac{\partial^2 r_k}{\partial\theta_i\partial\theta_j}P_{yy,k}^{-1}r_k \\
&+ \frac{1}{N}\sum_{k=1}^{N}\frac{\partial r_k^T}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}r_k \\
&+ \frac{1}{2N}\sum_{k=1}^{N} tr\{P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}(\frac{\partial r_k}{\partial\theta_j}r_k^T + r_k\frac{\partial r_k^T}{\partial\theta_j})\} \\
&+ \frac{1}{2N}\sum_{k=1}^{N} tr(P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}) \\
&+ \frac{1}{2N}\sum_{k=1}^{N} tr(P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}).
\end{aligned} \tag{2.59}$$

Since

$$
\begin{aligned}
& E(tr[P_{yy,k}^{-1}\frac{\partial^2 P_{yy,k}^{-1}}{\partial\theta_i\partial\theta_j}(P_{yy,k}^{-1}r_k r_k^T - I)]) \\
&= tr[P_{yy,k}^{-1}\frac{\partial^2 P_{yy,k}^{-1}}{\partial\theta_i\partial\theta_j}(P_{yy,k}^{-1}E(r_k r_k^T) - I)] \\
&= tr[P_{yy,k}^{-1}\frac{\partial^2 P_{yy,k}^{-1}}{\partial\theta_i\partial\theta_j}(P_{yy,k}^{-1}P_{yy,k} - I)] \\
&= 0
\end{aligned}
\tag{2.60}
$$

the expectation of the first term in the bracket of the right part in (2.59) is zero, then applying Kolmogorov's law of large numbers, the first term in (2.59) converges to zero in probability one when $N \to \infty$. Similarly, in (2.59) the second and the seventh term together, the third and the eighth terms together converge to zero as well.

Applying the independency between r_k and its derivatives up to second order with regard to θ, there are

$$
E(\frac{\partial^2 r_k}{\partial\theta_i\partial\theta_j}P_{yy,k}^{-1}r_k) = E(\frac{\partial^2 r_k}{\partial\theta_i\partial\theta_j})P_{yy,k}^{-1}E(r_k) = 0, \tag{2.61}
$$

$$
E(\frac{\partial r_k^T}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}r_k) = E(\frac{\partial r_k^T}{\partial\theta_i})P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_j}P_{yy,k}^{-1}E(r_k) = 0, \tag{2.62}
$$

and

$$
\begin{aligned}
& E\{tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}(\frac{\partial r_k}{\partial\theta_j}r_k^T + r_k\frac{\partial r_k^T}{\partial\theta_j})]\} \\
&= tr\{P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}[E(\frac{\partial r_k}{\partial\theta_j})E(r_k^T) + E(r_k)E(\frac{\partial r_k^T}{\partial\theta_j})]\} = 0.
\end{aligned}
\tag{2.63}
$$

From (2.61), (2.62) and (2.63), the expectation of fourth, fifth and sixth terms in the bracket of the right part in (2.59) are zero as well. Then according to the Kolmogorov's law of large numbers, these terms converge to zero in probability one when $N \to \infty$.

In all,

$$
P(\lim_{N\to\infty}\frac{1}{N}[\frac{\partial^2 l_N(\theta_0)}{\partial\theta_i\partial\theta_j} + \varphi(\theta_0)] = 0) = 1. \tag{2.64}
$$

Then (2.58) is satisfied. $\sharp$

Now we turn to prove the main Theorem 2.5.3

Proof of Theorem 2.5.3:

1. From (2.47), it can be observed that

$$\begin{aligned}\varphi_{ii}(\theta_0) &= -E(\frac{\partial^2 l_N(\theta_0)}{\partial\theta_i\partial\theta_i}) \\ &= \sum_{k=1}^{N}\{\frac{1}{2}tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}] + E(\frac{\partial r_k^T}{\partial\theta_i}\cdot P_{yy,k}^{-1}\cdot\frac{\partial r_k}{\partial\theta_i})\}\end{aligned} \tag{2.65}$$

Since $(P_{yy,k})^{-1}P_{yy,k} = P_{yy,k}(P_{yy,k})^{-1} = I$, then for $i = 1, 2, \ldots, n$, there exists

$$\sum_{l=1}^{n}(P_{yy,k}^{-1})_{il}(P_{yy,k})_{li} = \sum_{l=1}^{n}(P_{yy,k})_{il}(P_{yy,k}^{-1})_{li} = 1$$

According to (2.51), for i=1,2,...,n,

$$\begin{cases}\sum_{l=1}^{n}(P_{yy,k}^{-1})_{li} \geq 1/(M_{h0}-m_{h0})^2 \\ \sum_{l=1}^{n}(P_{yy,k}^{-1})_{il} \geq 1/(M_{h0}-m_{h0})^2\end{cases} \tag{2.66}$$

From Lemma 2.5.5, since $\frac{\partial P_{yy,k}}{\partial\theta_i}$ is a symmetric matrix, it can be derived that there exits $m_{p1} > 0$ such that

$$element[\frac{\partial P_{yy,k}}{\partial\theta_i}]^2 > m_{p1}. \tag{2.67}$$

Combine (2.66) and (2.67), and since $P_{yy,k}$ and $\frac{\partial P_{yy,k}}{\partial\theta_i}$ are both symmetric matrices, any permutation is allowed for their products, there exists:

$$\begin{aligned}& tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}] \\ &= tr[P_{yy,k}^{-1}\frac{\partial P_{yy,k}}{\partial\theta_i}\frac{\partial P_{yy,k}}{\partial\theta_i}P_{yy,k}^{-1}] \\ &\geq n[\frac{1}{(M_{h0}-m_{h0})^2}\cdot m_{p1}\cdot\frac{1}{(M_{h0}-m_{h0})^2}] > 0\end{aligned} \tag{2.68}$$

Followed by Lemma 2.5.7 and (2.68), the item in the brackets of the right part in (2.65) is positive and has a positive lower bound. Then, the whole right part of (2.65) is tend to infinity as $N \to \infty$. It shows that all the diagonal elements of $\varphi(\theta_0)$ tend to infinity as $N \to \infty$. $\varphi(\theta_0)$ is a symmetric positive matrix, then all of its eigenvalues tend to ∞ as well. The first condition of Crowder's theorem is satisfied.

2. If condition (a) and (b) are hold, from Theorem 2.2, it can be obtained that when $N \to \infty$,
$$-\varphi(\theta_0)^{-1}(\frac{\partial^2 l_N(\theta_0)}{\partial\theta\partial\theta^T}) \xrightarrow{p} I. \tag{2.69}$$
The second condition of Crowder's theorem is satisfied.

3. Considering the continuity of θ, if all the functions in the system are θ-continuous and θ is in its compact set Θ, then for a given $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$, such that $\mid l_N(\theta) - l_N(\theta_0) \mid < \varepsilon$. From the condition (a), all the derivatives with regard to θ are continuous. Then it can be obtained that function $\frac{\partial^2 l_N(\theta)}{\partial\theta\partial\theta^T}$ is continuous as well. (2.28) can be directly got by letting $N \to \infty$.

From the above, if the system model satisfies the conditions (a)-(c), the conditions of Crowder's theorem can be proved to be hold as well. Following the result of Crowder's theorem, it can be claimed that
$$\hat{\theta} \xrightarrow{p} \theta_0. \tag{2.70}$$
Next in order to prove the asymptotic normality, the Mean Value Theorem
$$\frac{f(b)-f(a)}{b-a} = f'(c) \quad for \;\; c \in [a,\; b] \tag{2.71}$$
is applied with $f(\theta) = l'_N(\theta)$, $b = \hat{\theta}$ and $a = \theta_0$. Then
$$l'_N(\hat{\theta}) = l'_N(\theta_0) + \frac{\partial^2 l_N(\hat{\theta}_1)}{\partial\theta\partial\theta^T}(\hat{\theta} - \theta_0) \tag{2.72}$$
for some $\hat{\theta}_1 \in [\hat{\theta},\; \theta_0]$.

Since the likelihood function $l_N(\theta)$ is continuous, the maximum solution satisfies that $l'_N(\theta) = 0$. Moreover, when $N \to \infty$, since $\hat{\theta}_1 \in [\hat{\theta},\; \theta_0]$ and (2.70), it can be obtained that $\hat{\theta}_1 \to \theta_0$ and
$$\hat{\theta} - \theta_0 = -\{\frac{\partial^2 l_N(\theta_0)}{\partial\theta\partial\theta^T}\}^{-1}\varphi(\theta_0)\varphi^{-1}(\theta_0)l'_N(\theta_0). \tag{2.73}$$
From (2.69), there exists $-\{\frac{\partial^2 l_N(\theta_0)}{\partial\theta\partial\theta^T}\}^{-1}\varphi(\theta_0) \xrightarrow{p} I$ as well. The other part in the right side of (2.73) tends to the normal distribution $\mathcal{N}(0, \varphi^{-1}(\theta_0))$ according to central limit theorem in (140) and the definition of $\varphi(\theta_0)$. In all, it can be obtained that $\hat{\theta} - \theta_0 \xrightarrow{d} \mathcal{N}(0, \varphi^{-1}(\theta_0))$. ♯

In order to show how to apply the Theorem 2.5.3 to system models, an example is adopted in this part.

Example:

Consider a system model which is described by

$$\begin{cases} x_k = \dfrac{\theta x_{k-1}}{1+x_{k-1}^2} + \omega_{k-1} \\ y_k = x_k + \varepsilon_k \end{cases} \tag{2.74}$$

with ω_k and ε_k are both one dimensional standard Gaussian noise, which means that ω_k, $\varepsilon_k \sim \mathcal{N}(0,1)$. In the system, the true value of θ is set as 1 and the initial value of the system state is $x_0 = 0.1$.

First, the condition of theorem 2.5.3 is checked.

Condition (a) Rewrite the nonlinear function in the system (2.74),

$$\frac{\theta x_{k-1}}{1+x_{k-1}^2} = \frac{1}{1/x_{k-1} + x_{k-1}}\theta. \tag{2.75}$$

According to the inequality of arithmetic and geometric means, we get

$$1/x_{k-1} + x_{k-1} \leq -2 \quad or \;\; 1/x_{k-1} + x_{k-1} \geq 2. \tag{2.76}$$

Then

$$-\frac{1}{2} \mid \theta \mid \leq \frac{\theta x_{k-1}}{1+x_{k-1}^2} \leq \frac{1}{2} \mid \theta \mid . \tag{2.77}$$

As a result of (2.77), if neglecting the effect of the system noise and define $\theta \in \Theta$ where Θ is a bounded, the state, the system function and function in the measurement (unit function) are all bounded. Regarding the derivatives of functions up to second order with regard to θ, their boundness can be obtained by calculating the deviations of the functions and applying the boundaries of the states.

Condition (b) For the system model (2.74), this condition is naturally satisfied if replacing x_k in the measurement by using the system equation.

Condition (c) If the possible set Θ is chosen as a bounded compact set in $\mathbb{R}$, the condition can be satisfied.

From the above analysis, it can be seen that the parameter in (2.74) can be estimated consistently using UKF plus ML method. The combined UKF and ML method is simulated to make the parameter identification of the above system model. It is assumed that the parameter θ is between 0 and 10. Here the initial value of the estimation is chosen as 0.1. Fig. 2.3 shows 300 estimations started at 1th samples and ended at 300 samples. For each estimation, it applied all the data obtained to make the parameter identification of the system. The horizontal axis is the number of the identification, started at 1 and ended at 300. The vertical axis stands for the estimated value. From the result, it displays that the estimated values gradually converge to the true value 1. Moreover, it is obvious that when the number of samples that is used for estimation tends to infinity, the estimated value will much closer to the true value of the parameter.

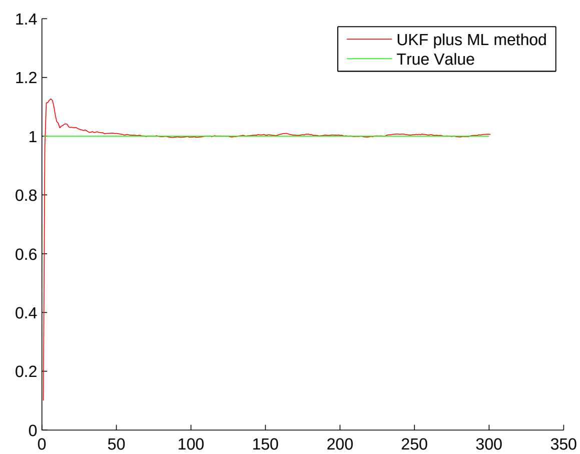

Figure 2.3: Estimations result for *example*

Furthermore, according to the Theorem 2.5.3, the variance of the estimation can be predicted using $\varphi_N^{-1}(\theta_0)$ as its approximation. In this case, $\varphi_N^{-1}(\theta_0) = 0.02789$ for the estimation using 100 sampling points. If using the 50 sampling points to make the estimation, the variance of the identification value is 0.0352. The difference between the variances of these two different identifications is with about 30 percent. However, the theorem only shows asymptotic property. The more points applied, the more accurate the result is compared with the true value and the more closer the variance is to zero.

This is the real case, if different large N, for example using hundreds or thousands sampling points, is adopted to make the estimation, the identification values are nearly the same, and the variance are closer to zero.

2.5.5 Case Studies Using KF Based Methods

In this part, two kinds of nonlinear system models are considered. During the simulations, four KF technique based methods are implemented. Since the direct filter methods can not lead to the good results to estimate the parameter in the diffusion item, the system in simulation one only considers the unknown parameter in the drift item. The simulation two will consider two parts (both drift item and diffusion item) parameter identification problem.

2.5.5.1 Simulation One: Classic Model

The system considered in this part can be seen as the classic model in which system noise is considered as one simple Gaussian process, but the system model is rewritten as the ISDE model formulation. The objective of these simulations is to make the comparison of different KF based methods.

The system is described as:

$$\begin{cases} dX = f(X,U,\theta)dt + \sigma dB_t \\ Y(k) = h(X(k)) + \varepsilon_k \end{cases} \tag{2.78}$$

where X is the system state, and it is two dimensional vector rewritten as $(X_1, X_2)^T$, U is the input signal, θ is the system unknown parameter, σ is a constant related to process noise variance, and $Y(k) = (Y_1(k), Y_2(k))^T$ is the measurement vector. Functions $f(\cdot) \in \mathbb{R}^2$ and $h(\cdot) \in \mathbb{R}^2$ are some specific nonlinear or linear functions as below:

$$\begin{cases} f(X,U,\theta) = \begin{pmatrix} X_2 + U \\ 2(1-\theta X_1^2)X_2 - X_1 U \end{pmatrix}, \\ h(X(k)) = \begin{pmatrix} X_1(k) \\ X_2(k) \end{pmatrix}. \end{cases}$$

Here B_t is a two dimensional Brown Motion, and noise in the measurement $\varepsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ with

$$\mathbf{S} = \begin{pmatrix} 0.0001 & 0 \\ 0 & 0.0001 \end{pmatrix}$$

.

In the simulation, parameters are set as $\theta = 0.5$, $\sigma = 0.1$, and the initial state is $(1,1)^T$. The input signal U is a kind of sweeping signal which is plotted in the Fig. 2.4. The outputs is generated by simulating the predefined system and the data is plotted in Fig. 2.5.

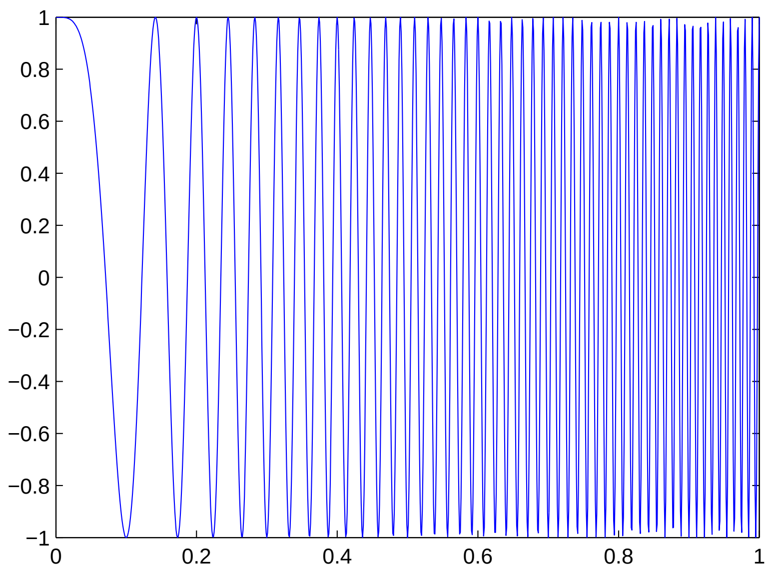

Figure 2.4: The input U for 2.5.5.1

The Data used in the identification process is chosen as the continuous 100 couples with sample interval 0.01 seconds.

For the direct methods using EKF and UKF, it need to generate an augment state to the system. Since in the system θ is set as a constant, the new system with the augment state could be rewritten as:

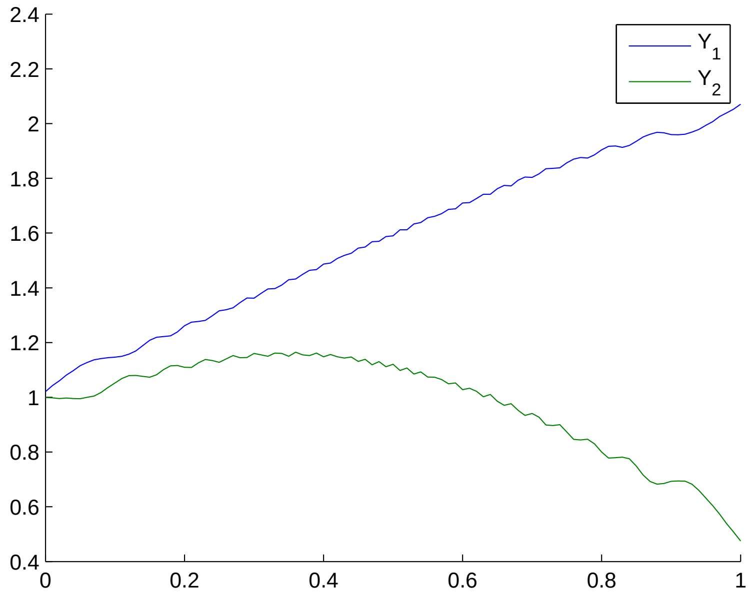

Figure 2.5: The output Y for 2.5.5.1

$$\begin{cases} \tilde{X} = [X, \theta]' \\ \dot{\tilde{X}} = \begin{pmatrix} \tilde{X}_2 + U \\ 2(1 - \theta \tilde{X}_1^2)\tilde{X}_2 - U\tilde{X}_1 \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \\ 0 & 0 \end{pmatrix} dB_t \\ h(\tilde{X}(k)) = \begin{pmatrix} \tilde{X}_1(k) \\ \tilde{X}_2(k) \end{pmatrix}. \end{cases} \tag{2.79}$$

Applying EKF and UKF using the data as the previous parts to the system model, the estimation of the augment state can be obtained. Then the last component $\tilde{X}_3(k)$ can be taken as the result of the parameter θ identification.

- The estimation results using EKF/UKF directly for 2.5.5.1 could be seen in Table 2.1.

- The two approaches are implemented under the same computational condition (cpu: Intel Core2 Duo CPU T5900. Memory: 3GB.). The EKF based method needs 0.033732 seconds while UKF based method needs 0.081804 seconds.

Table 2.1: The estimation results using EKF/UKF for 2.5.5.1

Approach	EKF	UKF
θ	0.4968	0.4989

Table 2.2: The estimation results using EKF+ML/UKF+ML for 2.5.5.1

Approach	EKF+ML	UKF+ML
θ	0.4987	0.5003
Computing Time	4.689910 seconds	7.365839 seconds

.

The results using EKF plus ML and UKF plus ML methods are listed in Table. 2.2 and Figure. 2.6.

The comparison with regard to convergence of the two methods can be judged according to the number of iterations in solving the optimization problem required to reach the same tolerant criteria. In the simulation, it made 100 iteration steps in the optimization and results can be seen in Figure. 2.6. If the tolerant level is selected as $1.0000e-004$ in the concern, and Table. 2.3 shows iteration numbers of these two approaches.

Comparison of the Four Methods:

From the simulation tests, the following discussions could be made:

- Precision: It can be observed that the parameter estimated using UKF+ML based method is the closest one to the true value than the results using the other methods. In all, regarding the precision, the order is: UKF+ML, UKF, EKF+ML and

Table 2.3: The number of required iterations for 2.5.5.1

Approach	EKF MLE	UKF MLE
The number of the iteration	66	32

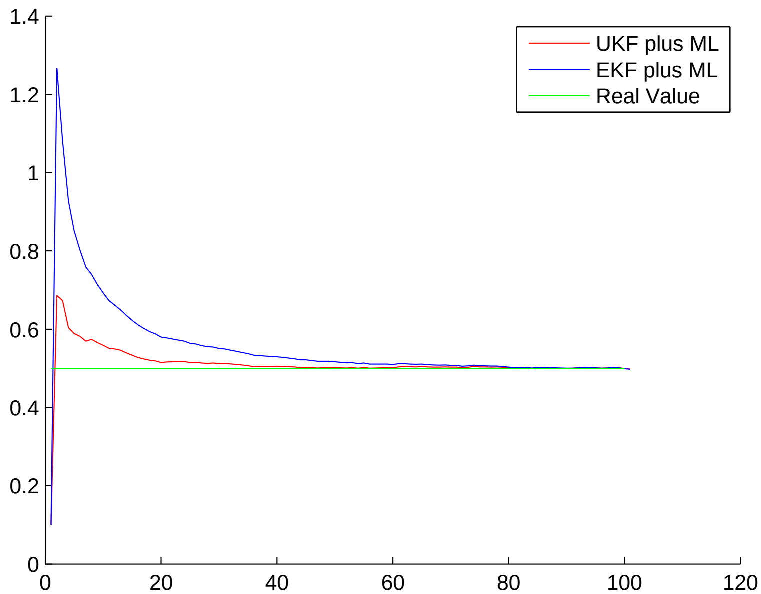

Figure 2.6: Optimization in parameter identification using EKF plus ML and UKF plus ML for 2.5.5.1

EKF methods. Since the model is a nonlinear one, the most important factor to influence the precision is the choice of the Filter–no doubt that using UKF could provide a more accurate state estimation. Under the condition using the same filter technique, the method with maximum likelihood can use more information of the system (distribution of the noise or disturbance) than only applying KF technique, hence it is better to use Kalman Filter method with the ML approach to make parameter identification for some parameterized nonlinear models.

- Computation load: From the view of the computational time, under the same conditions, it is clear that UKF based method needs more calculation power than EKF based method does. It is because in the state estimation stage, UKF uses a number of sigma-points which need to be generated and the Cholesky decomposition of the covariance matrix needs to be carried out as well.

- Convergence: Regarding the convergence for ML methods, UKF+ML method have the faster convergence property than EKF+ML. It is due to that in the state estimation stage, UKF does not make the linearization to the nonlinear system, while EKF makes the linearization to the original system. In the sense, UKF based method can catch more properties of the system than EKF based method. Then, it can find the optimal solution much more quickly.

As discussed, if the Filter technique is used to deal with the case where the diffusion item have the unknown parameter without any other tool, the performance with regard to the precision is too bad. Hence, in the following, the KF+ML is the main method to handle the parameter identification problem for the ISDE model in which the diffusion part contains the unknown parameters.

2.5.5.2 Simulation Two (Case (A) and (B)): Parameter Identification Using KF plus ML Methods

This part, three different cases are shown to make the comparison of EKF+ML and UKF+ML methods.

First, it will be shown that nonlinear ISDE model can deal with some kinds of the system where the random feature can be affected by the state of the system.

Case A0:

This system is chosen using the so-called Cox-Ingersoll-Ross (CIR) model (124). It is used to describe the term structure of interest rates.

$$\begin{cases} dX = a(b-X)dt + \sigma X^{\frac{1}{2}} dB_t, & X(0) = 0.1 \\ Y = X + \omega_t \end{cases} \tag{2.80}$$

where X is the continuous-time short-term interest rate. Many structure models can be found using this kind of model class by setting appropriate parameter constraints (see Chan et al., 1992) for a survey. In the test, the true parameters are set as $a = 0.5$, $b = 1$, $\sigma = 0.5$.

According to the It$\hat{o}$ Formula, let a new variable $Z = 2X^{\frac{1}{2}}$, it can be obtained that

$$\begin{aligned} dZ &= \frac{\partial 2X^{\frac{1}{2}}}{\partial X} dX + \frac{1}{2}\frac{\partial^2 2X^{\frac{1}{2}}}{\partial X^2} \cdot (\sigma X^{\frac{1}{2}})^2 dt \\ &= X^{-\frac{1}{2}} dX + \frac{1}{2} \cdot (-\frac{1}{2} X^{-\frac{3}{2}}) \cdot \sigma^2 X dt \\ &= (abX^{-\frac{1}{2}} - aX^{\frac{1}{2}})dt + \sigma dB_t - \frac{1}{4}\sigma^2 X^{-\frac{1}{2}} dt \\ &= (abX^{-\frac{1}{2}} - aX^{\frac{1}{2}} - \frac{1}{4}\sigma^2 X^{-\frac{1}{2}})dt + \sigma dB_t. \end{aligned} \tag{2.81}$$

From the relationship between Z and X, $X = \frac{1}{4}Z^2$, take place of X in (2.81) and the measurement in (2.80), then the transformed system model can be described using the new state variable Z as:

$$\begin{cases} dZ = (2abZ^{-1} - \frac{1}{2}aZ - \frac{1}{2}\sigma^2 Z^{-1})dt + \sigma dB_t, & Z(0) = 2(X(0))^{\frac{1}{2}} \\ Y = \frac{1}{4}Z^2 + \omega_t \end{cases} \tag{2.82}$$

It can be seen that the diffusion item (stochastic part) of new model (2.82) does not depend on the new system variable Z but all the unknown parameters are not changed. It means that to make the system identification of the original system can be accomplished by estimating the transformed system (2.82). Moreover, the system model (2.82) is the common model where the state variable only affect the deterministic part of the system. Then different methods can be adopted to make its system identification. This example can also be taken as one to show the merit of using ISDE model that the ISDE model can deal with some systems with state dependent random feature or noise.

In the thesis, the proposed UKF plus ML method is applied to make the estimation of (2.82). Here, the data adopted is 200 points from the beginning, the step interval is set as 0.1 second. The result is $\hat{a} = 0.4978$, $\hat{b} = 1.012$, $\hat{\sigma} = 0.4887$. The performance of the estimation is really nice, but the estimation for the parameter related to the random feature has a relatively large bias due to the random noise generation.

Case A:

This example we use is the same to the example 1 in the paper (83), in which it proposed a detailed algorithm using EKF+ML/MAL method to make the system identification for the ISDE equations. The system is described as

$$d\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} VX_1 - \frac{UX_1}{X_3} \\ -\frac{VX_1}{Y} + \frac{U(10-X_2)}{X_3} \\ U \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} dB_t,$$

where $(X_1, X_2, X_3)^T$ is the state of the system, and

$$V = \theta \frac{X_1}{0.5X_2^2 + X_2 + 0.03},$$

θ is the system parameter in the drift term of the SDE, U is the input variable. $\sigma_1, \sigma_2, \sigma_3$ are unknown parameters in the diffusion term.

The measurement equation is given as

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}_k = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}_k + \varepsilon_k,$$

where $(Y_1, Y_2, Y_3)^T$ is the measurement of the state, and $\varepsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ with

$$\mathbf{S} = \begin{pmatrix} S_{11} & 0 & 0 \\ 0 & S_{22} & 0 \\ 0 & 0 & S_{33} \end{pmatrix}$$

and $S_{11} = 0.01, S_{22} = 0.001, S_{33} = 0.01$.

The true parameters are assumed as $\theta = 1$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma = 0.1$, sampling interval is chosen as 0.01s, and the initial state is $(1, 0.24495, 1)^T$. The U is a kind

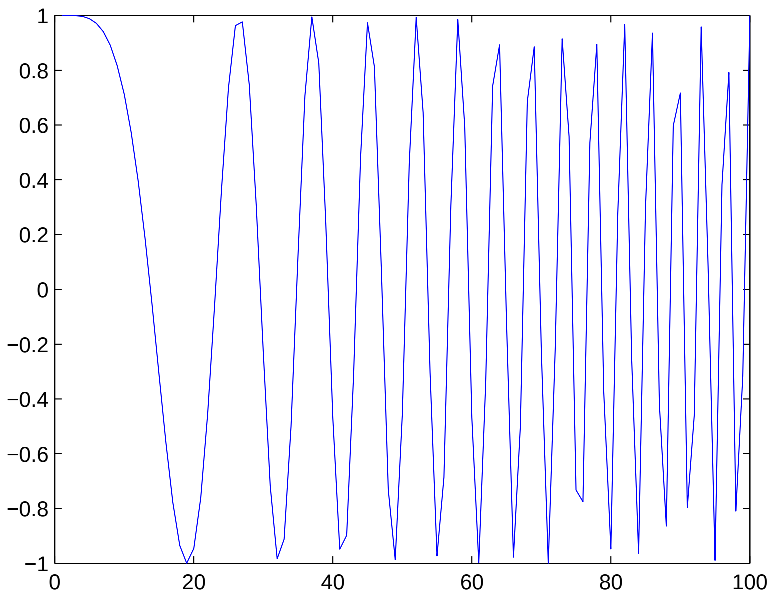

Figure 2.7: The input for the case **A**

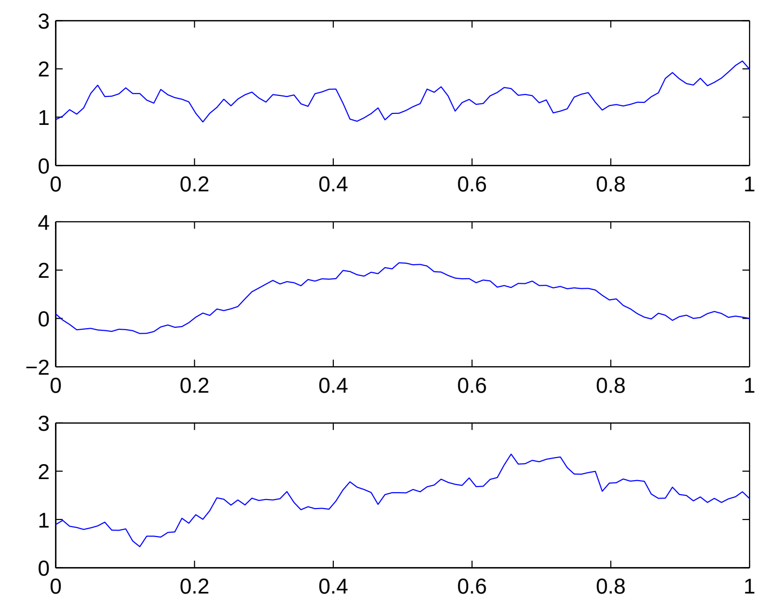

Figure 2.8: The measurement $(Y^1, Y^2, Y^3)^T$ for the case **A**.

Table 2.4: The estimation results for case **A**

Approach	EKF MLE	UKF MLE
θ	1.0422	0.9983
σ	0.0935	0.0984

Table 2.5: The number of required iterations for case **A**

Approach	EKF MLE	UKF MLE
The number of the iteration	73	53

of sweeping signal which is plotted in the Fig. 2.7. A set of outputs (100 samples) is generated by simulating the predefined system and the data is plotted in Fig. 2.8.

Both the EKF and UKF plus ML methods are examined and compared in the following two scenarios.

1. **Normal test**, i.e., the data used for identification is generated from the true system, which is plotted in Fig. 2.7 and Fig. 2.8.

   - Precision:

     The estimation results are shown in Table. 2.4. It can be observed that the parameter estimated using UKF based method is closer to the true value than the situation using EKF based method. This is because UKF does not apply linearization during the state estimation stage. Some experimental results indicate that UKF could yield results comparable to a third order Taylor series expansion of the state-model, while EKF of course only is accurate to a first order linearization.

   - Convergence issue.

     The tolerant level is selected as $1.0000e-004$ in our concern, and Table 2.5 shows iteration numbers of these two approaches. It can be noticed that the UKF plus ML method converges faster than EKF based method does for this example. The fast convergence also comes from the fact that UKF based method does not make linearization in the state estimation. It can catch more properties of the system than EKF based method.

Table 2.6: The estimated parameter for robustness test in case **A**

Approach	EKF+ML	UKF+ML
θ	1.1325	1.2578
σ	0.1082	0.1115

- Computation load.

  The two approaches are implemented under the same computational condition (cpu: Intel Core2 Duo CPU T5900. Memory: 3GB.). The EKF based method needs 4.272164 seconds while UKF based needs 8.672853 seconds. From the computation point of view, it is clear that UKF based method needs more calculation power than EKF based method does. The most computationally demanding part of UKF is the matrix square-root used to calculate sigma points. Matrix diagonalization or Cholesky factorization of the covariance matrix can be used to solve this problem, but still need heavier computation load. A more direct square root approach, propagating only the square-roots of the covariance matrices, may offer higher computationally efficiency. Merwe proposed an approach for doing this in (109).

2. **Robustness test**, i.e., the data are generated from the system in which there exists the modeling error.

   Here the modeling error concerned only happens in variable V. The data is generated according to the new V, noted as V_1,

   $$V_1 = \theta \frac{X_1}{0.55X_2^2 + X_2 + 0.03}.$$

   However, the following estimation still uses the original system model. The convergent values are listed in Table. 2.6.

   It can be observed both results have some deviations compared with the ”true” identification. Here the criterion to evaluate the robustness is made as:

   $$l_a = \frac{|\hat{a} - \hat{a}_e|}{\hat{a}} \times 100\%,$$

where $\hat{a}$ is the nominal result of the identification while $\hat{a}_e$ is the result based on the modeling error data (assume a is an unknown parameter of the system). The less l_a is, the more robust the method is.

According to this criterion,

- For the estimation of θ,

$$\begin{cases} l_\theta = 0.0866, & for\ \ EKF+ML\ \ method. \\ l_\theta = 0.2599, & for\ \ UKF+ML\ \ method. \end{cases}$$

- For the estimation of σ,

$$\begin{cases} l_\sigma = 0.1572, & for\ \ EKF+ML\ \ method \\ l_\sigma = 0.0853, & for\ \ UKF+ML\ \ method. \end{cases}$$

The results evidently show that UKF based method has larger deviations than EKF based method. This means that the UKF based method is more sensitive than EKF based method in the deterministic parameter identification regarding the modeling error. But regarding the random part, the EKF based method is more sensitive. This is because the model error only happened in the deterministic feature of the system without in random feature. Since the UKF based method can catch more information of the systems, it caused different comparisons for the two parts parameter identification.

Case B:

Two scenarios are investigated in this part: nonlinear systems described as a polynomial format and a division format. For simplicity, all the systems are simulated in one time unit and the parameter identification is based on 50 continuous sampling points with uniform time intervals of 0.01.

The system is generally described as:

$$dX = f(X,U,\theta)dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} dB_t$$
$$Y_k = h(X_k) + \varepsilon_k$$

Table 2.7: The estimation result for case **B-1**

Approach	EKF+ML	UKF+ML
θ	0.7729	0.8012
σ	0.1056	0.1045

Table 2.8: The number of required iterations for case **B-1**

Approach	EKF MLE	UKF MLE
The number of the iteration	53	71

where X is the system state, and it is rewritten as $(X_1, X_2, X_3)^T$, U is the input variable. θ is the system unknown parameter, and there is $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$. $Y(k) = (Y_1(k), Y_2(k), Y_3(k))^T$ is the measurement, $f(\cdot) \in \mathbb{R}^3$, $h(\cdot) \in \mathbb{R}^l, l \leq 3$ are some specific nonlinear or linear functions.

B-1: The function $f(\cdot)$ is a nonlinear polynomial:

$$f(X, U, \theta) = \begin{pmatrix} X_2^2 X_1 + U X_1 \\ X_3 + U X_2 \\ \theta X_1 (X_2 + X_3) + U \end{pmatrix}$$

and the measurement equation is

$$Y(k) = X_1(k) + \varepsilon_k$$

with $\varepsilon_k \sim \mathcal{N}(0, 0.1)$. Here the true values are that $\theta = 0.8$, $\sigma = 0.1$, and the initial state is $(1, 0, 1)^T$. It should be remarked that the system states become partially measurable, i.e., only X_1 is measured, while in the previous cases, all system states are directly measured. In this case, the input variable U is set as $U(t) = 0.5 sin(8t)$ and the output signal is obtained by simulating the system.

Similarly as what we do for the former cases, the estimated and computing results are listed in Table. 2.7 and Table. 2.8.

- Parameters estimation (Table. 2.7)

- Number of iteration (Table. 2.8)

Table 2.9: The estimation result for case **B-2**

method	EKF	UKF
θ	0.7881	0.4883
σ	0.0950	0.0973

Table 2.10: The number of required iterations for case **B-2**

Approach	EKF MLE	UKF MLE
The number of the iteration	82	49

- Computation load.

  Here the condition of the computation is the same to the previous case. The EKF method need 2.376364 seconds while UKF need 6.419088 seconds.

B-2: The function $f(\cdot)$ has simple divisions. The only difference to the case **B-1** is that the function $f(\cdot)$ converts to the following function which has simple divisions.

$$f(X,U,\theta) = \begin{pmatrix} X_2^2/X_3 + UX_1/X_3 \\ \theta X_3/X_2 \\ X_1 + U \end{pmatrix}$$

Here the true value of θ is 0.5, initial state is $(1,1,1)^T$ and other variables are just the same to **B-1**.

Repeat the same process and the results are shown in the below tables (Table. 2.9 and 2.10).

- Parameters estimation (Table. 2.9)

- Number of iteration (Table. 2.10)

- Computation load.

  The EKF based method needs 3.666436 seconds while UKF based method needs 7.084774 seconds under the same computing condition.

In the case **B**, the robustness test is not listed because the results have the same conclusion with case **A**.

discussion:

The two case studies of case **B** showed almost the same results as situation with case **A**. In the polynomial case, the two estimation results illustrate that UKF based method has a little better performance than EKF based method, if the computation load is not a concern. Regarding to the converging property, the EKF based method is a slightly better than UKF based method. However, regarding the division case, the UKF based method is obviously better than EKF based method without concerning of computational loads. And UKF based method converges much faster than EKF based method as well. It could be concluded that the UKF based method is better than EKF based method for systems with rather complex nonlinearity.

2.5.5.3 Conclusion for Studies

Through the above studies, the characteristics of both EKF and UKF based methods are illustrated. In general, the UKF based method can provide more accurate result than EKF based method. Meanwhile, the UKF based method also provides faster converging rate than EKF based method although there are some special cases. It is due to that the EKF just picks up the first order term through linearization of the nonlinear system and drops all items higher than the first order. If the influence of the higher order items can not be ignored in the system, the EKF may provide a poor performance in terms of the accuracy. In contrast, the UKF uses sigma-points that are dedicatedly chosen. S. Julier indicated that UKF yields results comparable to a third order approximation of Taylor expansion (73). As a result, it can provide a better estimation to the state of the system. That could be the reason why UKF based method is generally better in parameter estimation. Furthermore, the studies suggest to use UKF plus ML method to make the parameter identification for some nonlinear systems. The payoff for better performance of the UKF based method , including combining with ML method, is the computational load. The UKF needs to handle the Cholesky decomposition and calculation based on double-sized sigma-points. Moreover, it has been found that the UKF based method is more sensitive than EKF based method regarding to potential modeling errors.

2.6 FDD Application

In this section, the previously proposed model and the method are applied to Fault Detection and Diagnosis (FDD) procedure based on the ISDE model formulations of the systems.

2.6.1 The ISDE Model with Parametric Fault

Consider the following nominal ISDE model, which is a parametric one

$$dX(t) = g_1(X(t), u(t), t, \theta)dt + g_2(t, \theta)dB_t, \tag{2.83}$$

with the measurement

$$Y(k) = h(X(k), t(k)) + \varepsilon_k, \tag{2.84}$$

where the definitions of the variables are the same as previous sections. Here assuming that the fault is a parametric fault, i.e., if the fault happens, it only influence the parameter θ of the systems. Suppose θ changes from the normal value θ_0 to the faulty value θ_1 if fault happens.

According to the fault characteristics, if the fault happens, the change of the system could be described as:

$$df(t) = [g_1(X(t), u(t), t, \theta_1) - g_1(X(t), u(t), t, \theta_0)]dt + [g_2(t, \theta_1) - g_2(t, \theta_0)]dB_t.$$

where $df(t)$ describes the change of the system, and $df(t) = 0$ when no fault happens. Note that here the diffusion coefficients of the system and the system change considered when the fault happens, is independent on the state. Because if considering the state depended diffusion item, It$\hat{o}$ Formula can simplify the model to one without state depended diffusion model.

2.6.2 FDD Methods

FDD methods for the system need to consider the following problems:

1. How to make fault detection?

2. If the fault happens, how to evaluate the fault?

Generally, for the system with possible parametric fault, the previous two problems can be considered and accomplished together, if the parameter identification of the fault parameter θ can be made in an online manner. In this way, if the FDD procedure has detected that the parameter changed deviating from the normal value at some time, it can be claimed that the fault happened and of course, the change between the normal value and the current value can be used to evaluate the fault. Sometimes in order to deal with the fault and maintain the system running not too bad, the information of the states is also quite important for system reconfiguration. For this reason, the fault estimation often accompanies with the state estimation (64). Thereby, in the thesis, the Joint Parameter Identification and State Estimation (JPISE) technique for a FDD design for a class of ISDE modeled systems is considered. The considered faults are types of abrupt parametric faults, which indicates that some system parameters will immediately deviate from their normal values if faults happen. The concerned system parameters consist of deterministic parts as well as those describing the stochastic features in the system, such as the new covariances of the process noise and measurement noise.

The JPISE problem is a nonlinear problem, no matter the considered system is a linear one or not (59). In general, the techniques to solve a JPISE problem can be classified into two categories. The basic idea of one category, it is named as state estimation approaches, which is to extend the unknown system parameters as additional system states, so that an augmented state space model can be achieved. Then the Extended Kalman Filter (EKF) technique is used to estimated the augmented system states, which includes the original system states and the unknown system parameters (18). The parameter identification and state estimation can be simultaneously obtained at each sampling step. However, this kind of approach gives rise to explicit multiplication of states by other states, meanwhile it is well known that the EKF is a kind of first-order approximation and no guarantee for global convergency (59). Another category for solving JPISE problem is named as ”bootstrap” methods by (59). Within this type of method, the parameter identification and state estimation are carried out sequently. Either the (parameterized) state estimation is first obtained and then substituted into a parameter identification process, or *vice versa*. The Kalman Filter with Maximum Likelihood (KF-ML) method (83; 152) is a typical approach in this category. However, this category also suffers some potential drawbacks, such as non-convex optimization problem and suboptimal solution. Nevertheless, this kind of bootstrap method seems more flexible than the state estimation methods, e.g., being able to directly deal with identification of nonlinear system with unknown stochastic characteristics. Thereby in the following, the previous KF-ML method is applied to deal with a JPISE problem for a fault-tolerant space robot system which is the same to the system in (169).

The parameter identification using EKF plus ML and UKF plus ML methods to the ISDE model is first to use the input and output data to make state estimation, then form a ML function based on the result of the state estimation and solve the optimization of the ML function. It takes the optimal solution as the parameter estimate. It is only a off-line method. If the KF plus ML methods can be applied to FDD, it need to extend to an on-line version because the fault parameter must be investigated all the time to grantee that if the fault happened, it could be detected immediately.

In the following, the moving windows technique is adopted to extend the KF plus ML methods to an online manner to fit for the FDD demanding.

2.6.3 Entire FDD Procedure

Before the process is up to run, the length of one moving window N need to be chosen at first. Then based on the input and output data, the estimation procedure can be performed in an on-line way. When a new couple of data is collected, the latest N couples of input and output data are used to make the parameter identification of the system. The result of the identification is taken as the estimation of the fault parameter.

Take $\hat{\theta}$ as estimation of θ, then the predefined threshold method or some statistical methods such as cusum method can be used to determine whether the fault happened or not (178). Here, for the simplicity, the deterministic threshold method is applied, i.e., if the value of $\hat{\theta}$ is within 10% deviation to the normal value θ_0, the system is claimed running normally. Otherwise, it will be claimed that a fault has happened.

Associated with the fault detection, state estimation can be obtained as well as a by-product of the FDD procedure. It can be got by substituting the estimated parameter value $\hat{\theta}$ to the parameterized state estimation, the state estimation $\hat{x}_k(\hat{\theta})$ and $P_k(\hat{\theta})$ using KF method are then obtained. Sometimes if the state estimation is not smooth enough, Kalman Smoother technique can be adopted to make the estimation more

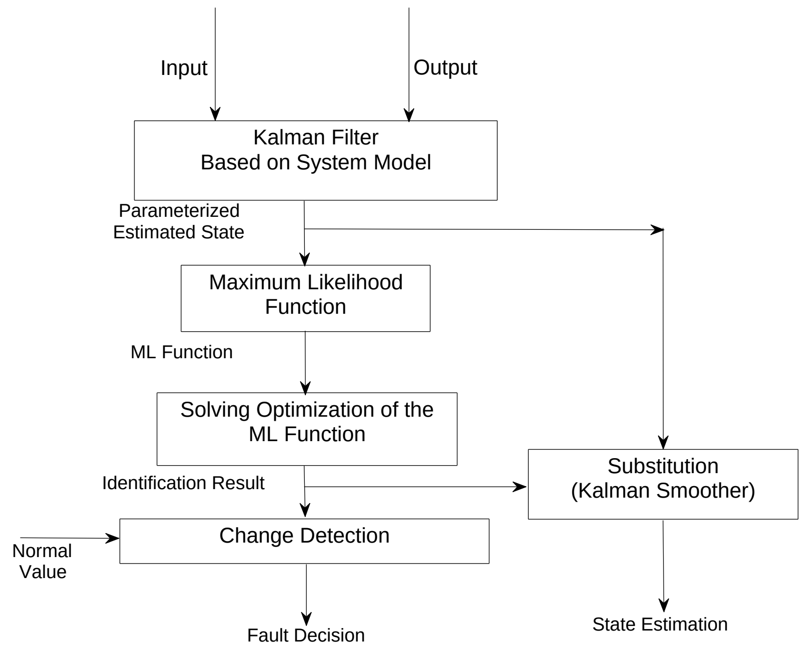

Figure 2.9: The scheme using KL and ML method

accurate. The Kalman Smoother proceeds backward in time (18) and it is summarized as:

Initial with $\hat{x}_k(\hat{\theta})$ and $P_k(\hat{\theta})$, and let $j = k-1, k-2, \ldots, k-N+1$, there is:

$$\begin{aligned} L_j(\hat{\theta}) &= P_j(\hat{\theta})\overline{A}^T(\hat{\theta})P_{j+1}^{-}(\hat{\theta}), \\ \hat{x}_{j|k}(\hat{\theta}) &= \hat{x}_j(\hat{\theta}) + L_j(\hat{\theta})(\hat{x}_{j+1|k}(\hat{\theta}) - \hat{x}_{j+1}^{-}(\hat{\theta})), \\ P_{j|k}(\hat{\theta}) &= P_j(\hat{\theta}) + L_j(\hat{\theta})(P_{j+1|k}(\hat{\theta}) - P_{j+1}^{-}(\hat{\theta}))L_j^T(\theta). \end{aligned} \tag{2.85}$$

Summarizing the above steps for FDD, the entire scheme is illustrated in Fig. 2.9. Suppose the procedure begins at k-th samples:

- Employ KF technique to make state estimation (mean and covariance), it need to determine the certain specific KF format according to the specific form of the system, such as KF for linear systems, EKF or other nonlinear filters for nonlinear systems.

- Form the parameterized ML function based on the results from previous state estimation.

- Solve the ML optimization problem, and obtain the optimal solution $\hat{\theta}$ as the estimation of fault parameter.
- Compare the identification result with the value under the normal situation system and make the fault detection decision using the predefined deterministic threshold.
- Substitute the identified parameter into parameterized KF solution, and then obtain the state estimation. If necessary, apply KS for smoothing purpose.
- Repeat the former steps when the new couple of input and output data is obtained.

Note that the first three steps are just the parameter identification using KF technique plus ML method.

2.7 Cases Study for a Space Robot System

In order to show the performance of the proposed method, a case of space robot is studied with different fault scenarios.

2.7.1 The ISDE Model Formulation

The space robot system used in (169) is considered here, the process could be seen in Fig. 2.10. In the normal situation, system parameters are listed in Table 2.11, and the dynamic of the normal system is described by:

$$N^2 I_m \ddot{\Omega} + I_{son}(\ddot{\Omega} + \ddot{\varepsilon}) + \beta(\dot{\Omega} + \dot{\varepsilon}) = T_j^{eff}, \tag{2.86}$$

$$I_{son}(\ddot{\Omega} + \ddot{\varepsilon}) + \beta(\dot{\Omega} + \dot{\varepsilon}) = -T_{def}. \tag{2.87}$$

The actuator part including a DC-motor and a gear box is simplified as $T_j^{eff} = NT_m$ and $T_m = k_t i_c$. Torque T_{def} due to the deformed spring is described by $T_{def} = c\varepsilon$.

In the actual system, the controllable input is the motor current i_c, and the measured signals are encoder output $\Theta = \Omega + \varepsilon$ and tachometer output $N\dot{\Omega}$. The original system

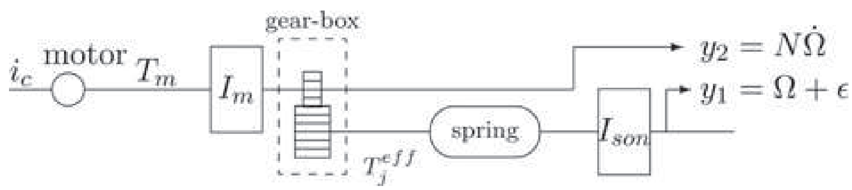

Figure 2.10: The process of a robot space referred (169)

was a SIMO system. Define state vector $X = [\Omega, \dot{\Omega}, \varepsilon, \dot{\varepsilon}]^T$, output vector $Y = [\Omega + \varepsilon, N\dot{\Omega}]^T$.

The state-space model of the system is obtained as follows:

$$\begin{cases} dX(t) = [AX(t) + BU(t)]dt + a\sigma dB_t \\ Y(t) = CX(t) + \omega_t \end{cases} \tag{2.88}$$

where the system matrices are:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{c}{N^2 I_m} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{\beta}{I_{son}} & -(\frac{c}{N^2 I_m} + \frac{\beta}{I_{son}}) & -\frac{\beta}{I_{son}} \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ \frac{k_t}{NI_m} \\ 0 \\ -\frac{k_t}{NI_m} \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & N & 0 & 0 \end{bmatrix},.$$

Here B_t is a two dimensional Brown Motion, $a\sigma$ is the item related to the covariance of the noise in the process, where $a = 0.001$ and σ is the parameter in the diffusion item. The noise in the measurement ω_t is the two dimensional Gaussian processes with means $\mathbf{0}$ and covariances R. Here,

$$R = \begin{bmatrix} 0.001^2 & 0 \\ 0 & 0.001^2 \end{bmatrix}.$$

2.7.2 Test Conditions

In the following situation, the different scenarios are considered to the fault detection and state estimation. In the model, the whole time the system running is 31.4 seconds and at the 10th second the fault happened. The initial condition for the system is

Table 2.11: System parameters of the space robot system referred in (169)

Symbol	Description	Unit
N=-260.6	Gear-box ratio	–
$I_m = 0.0011$	Inertia of the input axis	kg m^2
Ω	Joint angle of the internal axis	rad
$I_{son} = 400$	Inertia of the output axis	kg m^2
T_j^{eff}	Torque of effective joint input	Nm
ε	Joint angle of output axis	rad
$k_t = 0.6$	Motor torque constant	N/%
i_c	Motor current	Am
$\beta = 0.4$	Damping coefficient	N/%
$c = 130\,000$	Spring coefficient	N/%
T_{def}	Deformation torque of gear box	Nm
T_m	Motor torque	Nm

$x(0) = [0.01, 0, 0, 0]^T$. The test is made by using two different kinds of input signals, i.e., piecewise constant input and sinusoid input with different frequency:

$$u_1(t) = \begin{cases} 0.1, & for \;\; t < 5s \\ -0.5, & for \;\; 5s \leq t < 20s \\ 0.2, & for \;\; others. \end{cases}$$

$$u_2(t) = 0.5sin(0.8t).$$

$$u_3(t) = sin(0.2t).$$

$$u_4(t) = 0.8sin(1.5t).$$

2.7.3 Fault in the Deterministic Part

2.7.3.1 Case (C): Fault In the Actuator Part

In the first, the fault considered only takes place in the deterministic input item. When the fault happens, it is assumed to only disturb the motor constant. Fault parameter is assumed as θ and motor toque constant is taken as θk_t.

In the case, the normal system and faulty system can be written as:

$$\begin{cases} dX(t) = [AX(t) + B(\theta)U(t)]dt + a\sigma dB_t \\ \quad Y(t) = CX(t) + \omega_t \end{cases} \tag{2.89}$$

with $B(\theta) = \theta B$ and

$$\theta = \begin{cases} \theta_0, & Normal\ \ system, \\ \theta_f, & Faulty\ \ system. \end{cases} \tag{2.90}$$

In the simulation, the data is generated by setting $\sigma = 1$ is a constant, the values of the parameter are $\theta_0 = 1$ and $\theta_f = 1.5$, choose the sample interval as 0.1 seconds and the initial value of the θ is 0.9. The simulated output is plotted in Fig. 2.11

The FDD procedure and state estimation are performed using the proposed KF+ML method. For this case, the system is a linear one, so Kalman Filter is applied for KF stage. The estimation is implemented with different detection windows (N) and different inputs. The identification of the fault parameter need to wait for the first N outputs at the beginning. Before the time when enough data is collected, there are two methods to cope with the estimation. One is to make the estimation based on all the data obtained at sampling time. The other one is to set the fault parameter estimated as the initial value and state estimation is based on all the sample points before reaching N-th point. In the thesis, the later one is applied in order to show the performance in detail. As soon as N sample points are collected, a moving window with the length of N is used to on-line update the estimation according to the previous KF plus ML algorithm.

Piecewise Constant Input

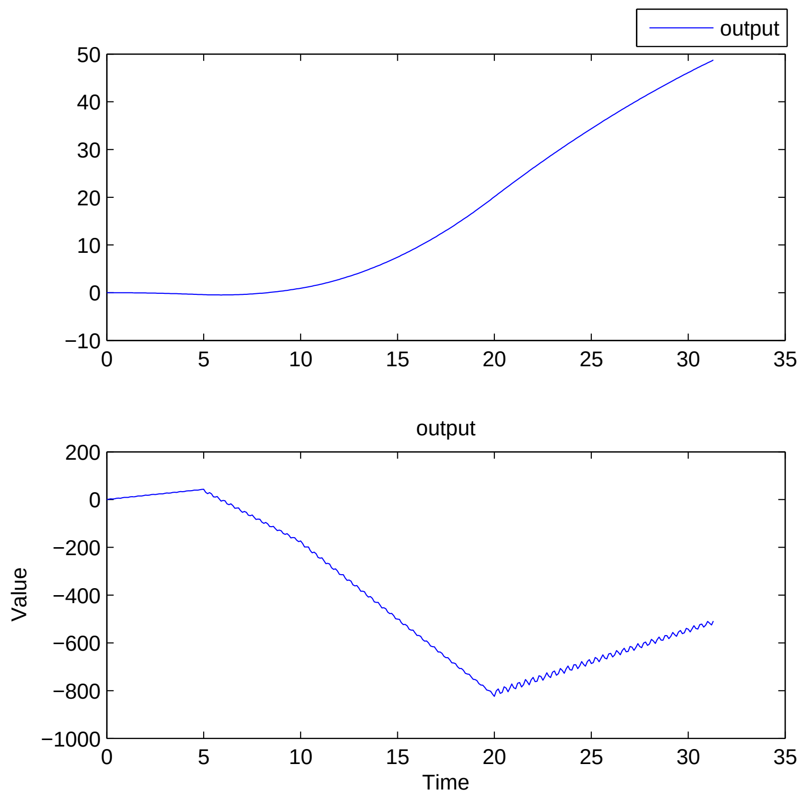

Figure 2.11: Output for case (**C**) with u_1

Using the input u_1 and different windows lengths, the simulated output is plotted in Fig 2.11 and the estimation results could be seen in Fig. 2.12, Fig. 2.13, Fig. 2.14, and Fig. 2.15.

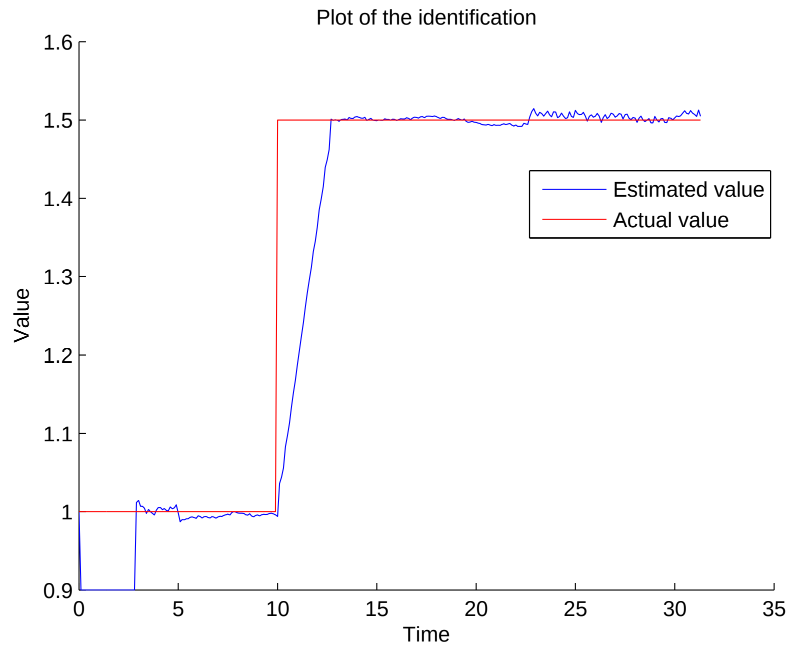

Figure 2.12: Parameter identification for case (**C**) with u_1 and $N = 30$

From the tests, we could get the following results:

- *Fault detection:*
  As shown in Fig. 2.12 and Fig. 2.14, the algorithm needs to wait for the first N points, thereby the estimated parameter just remains at the initial value in the beginning. Before 10th second, the estimated fault parameter is much close to 1 which is the normal value of the system. At the 5th second, the estimated value has a small deviation to 1 since the effect of the input signal is changed to the different direction. When the fault happened at 10th second, the estimated value has a large jump or deviation at the beginning period. After a while, the estimated values converge to some steady-state values which close to the real system values. At the moment, it is believed that the fault has happened and

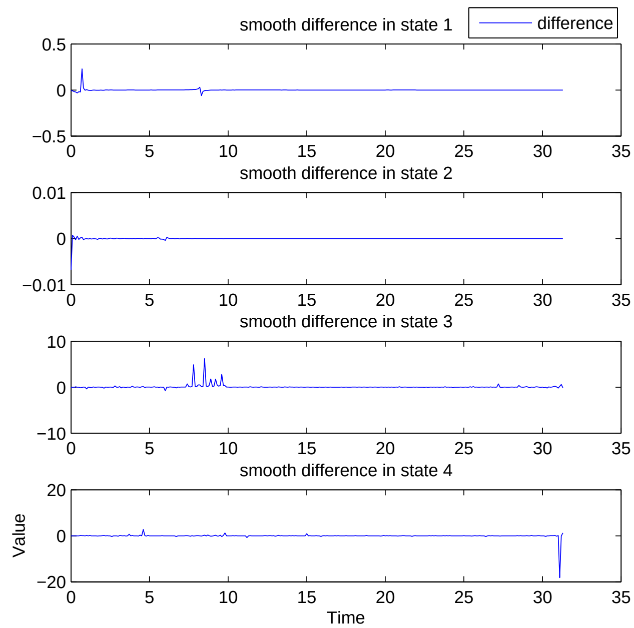

Figure 2.13: State estimation error for case (**C**) with u_1 and $N = 30$

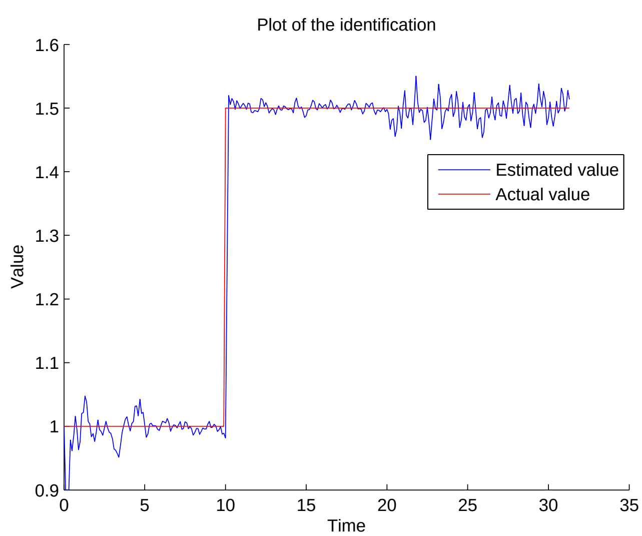

Figure 2.14: Parameter identification for case (**C**) with u_1 and $N = 5$

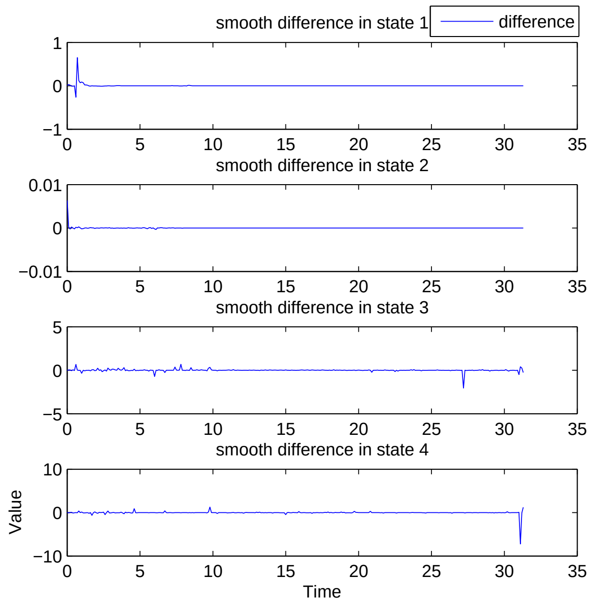

Figure 2.15: State estimation error for case (**C**) with u_1 and $N = 5$

its magnitude is obtained as well. Around the 20th second, a relatively large deviation can be seen and it is due to the change of the input signal. During almost all the time, the error for the fault identification is within 4%. Regarding the identification results, if the predefined threshold is already set before the detection, such as 10% from the normal value, it can be claimed that the fault has happened after approximately 11th second.

- *State estimation:*
  Fig. 2.13 and Fig. 2.15 show the errors of the state estimation and they are expressed in percentage. Most of state estimation errors are within 1%. But as same as the phenomenon observed in the fault detection results, at those intervals and times when the condition of the system/input is changed, the estimations may have a relatively large temporal oscillations. Moreover, when the system tends to stop, sometimes there may be a large deviation to the estimated value.

- *Length of Moving Windows:*
  In the test, moving windows with 5 points and 30 points are considered. From the

parameter identification results, 30 points could give us a much better estimation of the parameter at the cost of the detection time. But 5 points, not better than the 30 points in the accuracy, but it saves much time in the procedure. Further, it can react to the fault much more quickly. Meanwhile, for the state estimation the performance of the two different kinds of sample points is nearly the same.

Sinusoid Input

In this part, the input variable is changed, 3 sinusoid inputs with different frequencies and amplitudes are adopted, that are the previous inputs u_2, u_3 and u_4. Here we also used two different lengths of moving windows, 5 and 30.

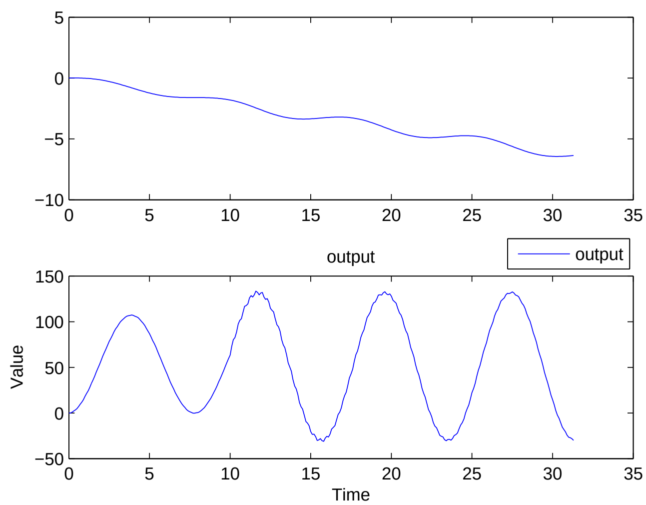

Figure 2.16: Output for case (**C**) with u_2

From the output figures, the difference between 3 inputs is evidence since the frequencies of these inputs are different. Even it could be guessed from the output which input it used.

The comparison : It could hardly see the difference from the identification using 30 points, all of them are quite fine. But from the identification using 5 points, the difference is obviously displayed. We could see the periodical small bias from the results. It could be seen that these small biases emerge nearly 4 seconds

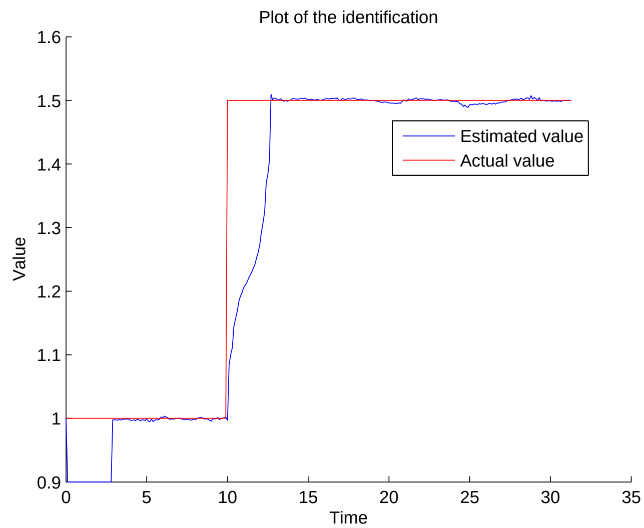

Figure 2.17: Parameter identification for case (**C**) with u_2 and $N = 30$

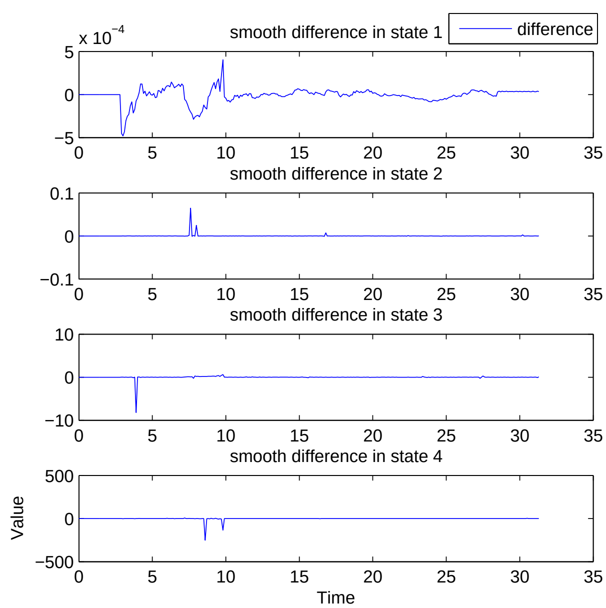

Figure 2.18: State estimation error for case (**C**) with u_2 and $N = 30$

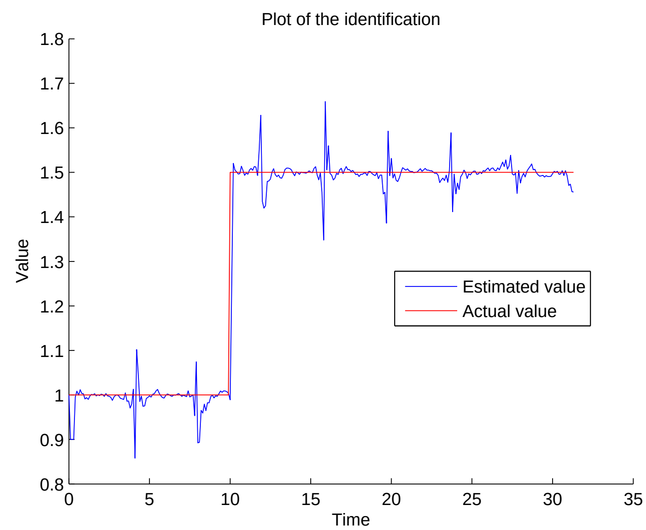

Figure 2.19: Parameter identification for case (**C**) with u_2 and $N = 5$

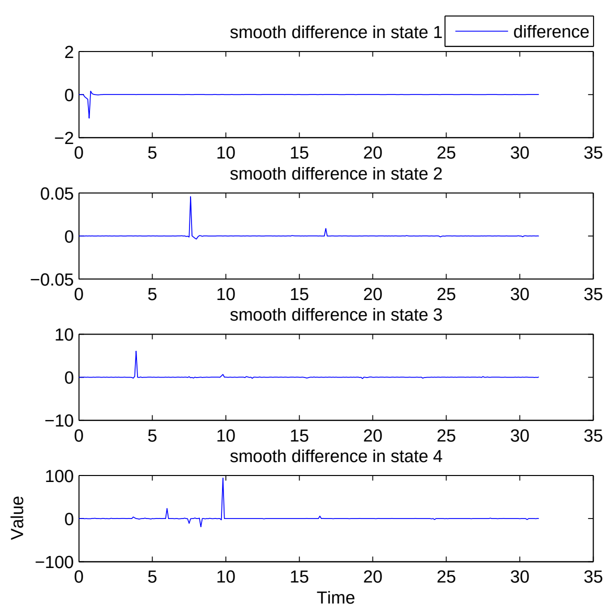

Figure 2.20: State estimation error for case (**C**) with u_2 and $N = 5$

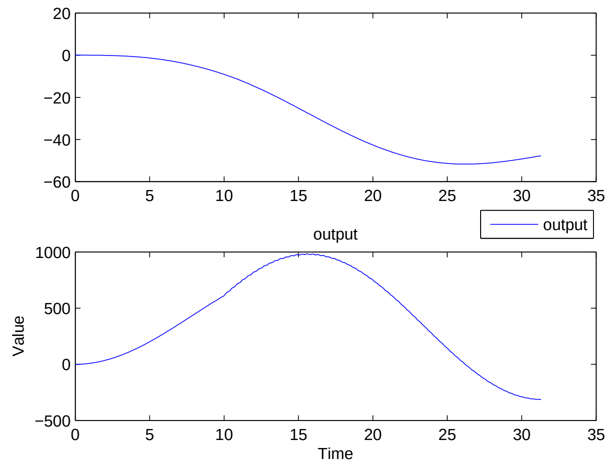

Figure 2.21: Output for case (**C**) with u_3

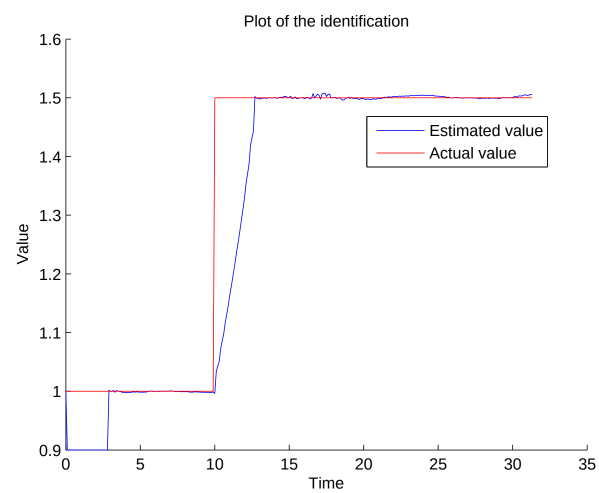

Figure 2.22: Parameter identification for case (**C**) with u_3 and $N = 30$

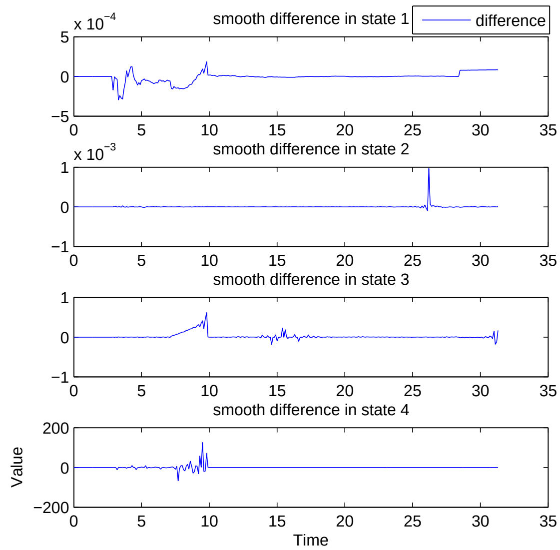

Figure 2.23: State estimation error for case (**C**) with u_3 and $N = 30$

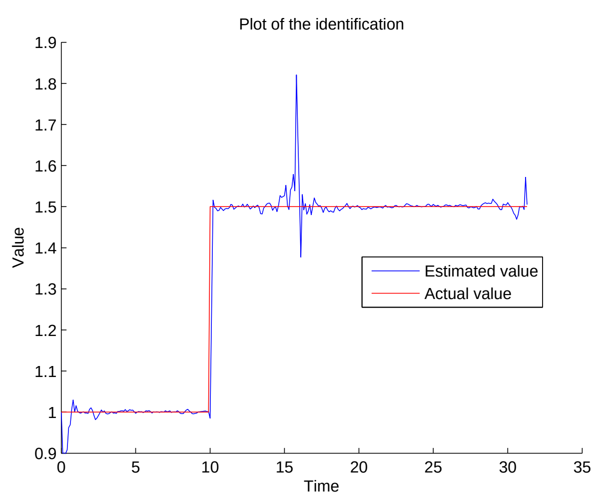

Figure 2.24: Parameter identification for case (**C**) with u_3 and $N = 5$

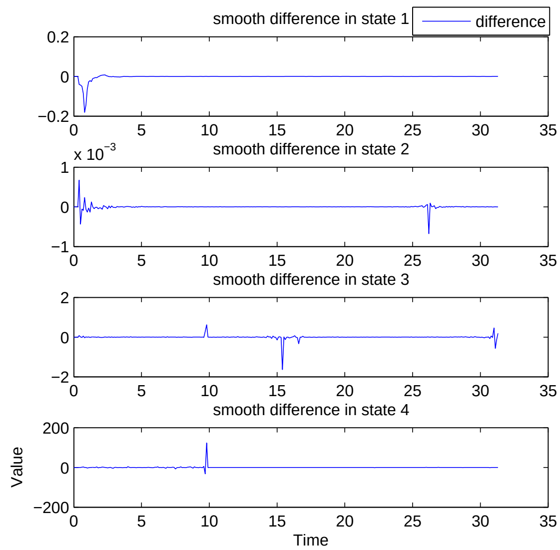

Figure 2.25: State estimation error for case (**C**) with u_3 and $N = 5$

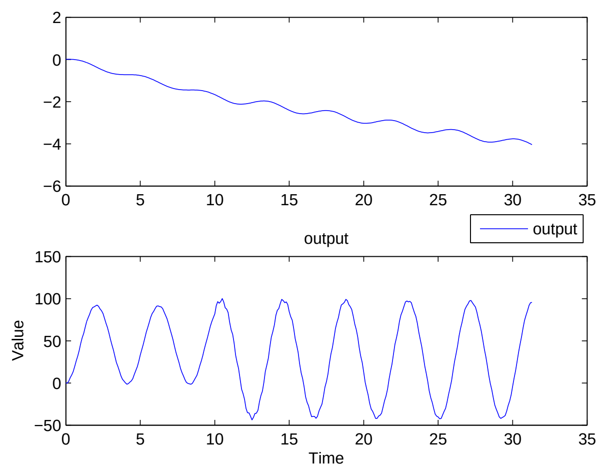

Figure 2.26: Output for case (**C**) with u_4

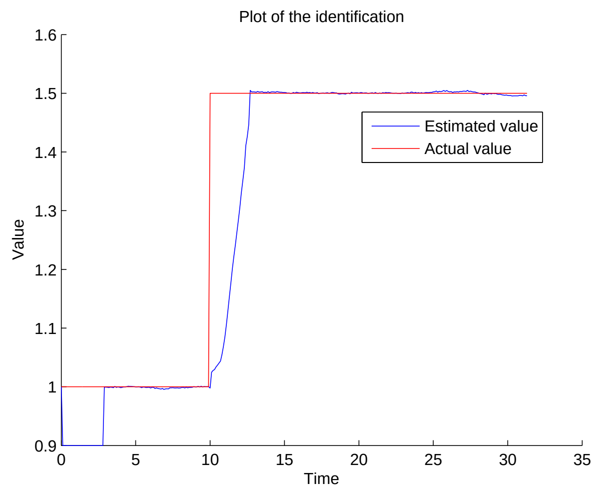

Figure 2.27: Parameter identification for case (**C**) with u_4 and $N = 30$

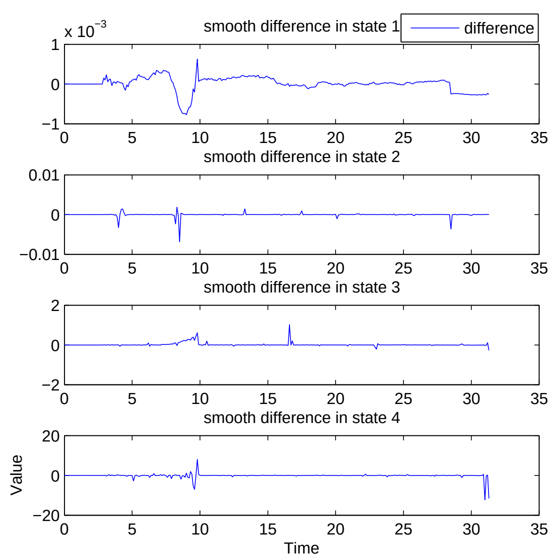

Figure 2.28: State estimation error for case (**C**) with u_4 and $N = 30$

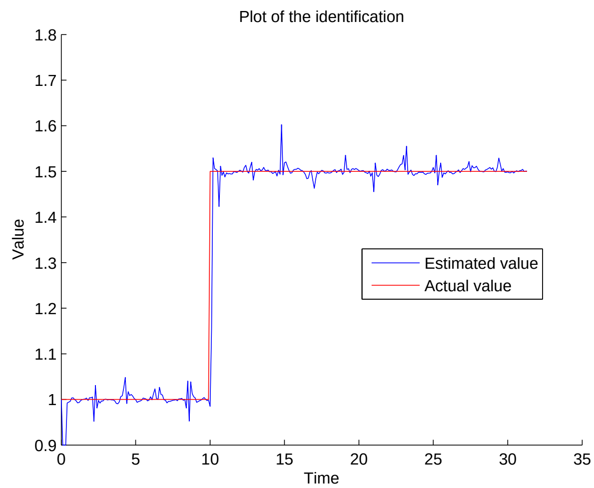

Figure 2.29: Parameter identification for case (**C**) with u_4 and $N = 5$

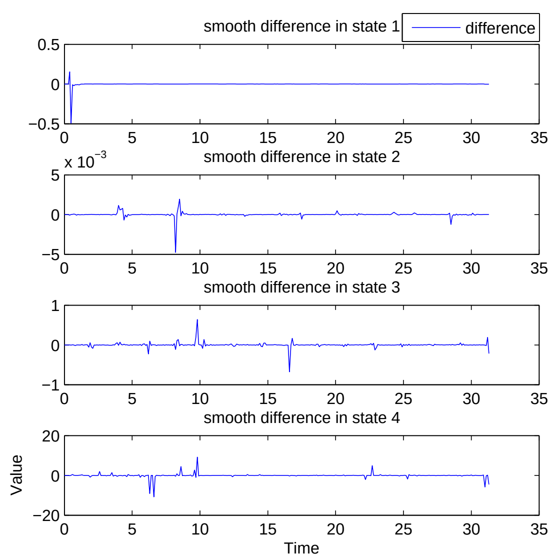

Figure 2.30: State estimation error for case (**C**) with u_4 and $N = 5$

one time for u_2 case, only one time for u_3 case and more frequently for u_4 in the whole procedure. From the results for the cases using u_2 and u_3, the bias always happens at peak time of the input. It can be concluded that the input frequency could influence the parameter identification periodically. But the amplitude could not disturb the estimation. Since the influence is not huge, the state estimations are nearly the same here and the performance is quite good in terms of the precision.

2.7.3.2 Case (D): Fault in the state item

In this part, we consider the fault only happens in the state item, that means it could change the part of system which explicitly has the states. In the model, it reflects in the model that the matrix A changes if the fault happens. Follow the same procedure as in case $(\mathbf{C})$, the normal and faulty system can adopt the model:

$$\begin{cases} dX = [A(\theta)X + BU]dt + a\sigma dB_t \\ Y = Cx + \omega_t \end{cases} \tag{2.91}$$

with

$$\theta = \begin{cases} \theta_0, & Normal \ \ system, \\ \theta_f, & Faulty \ \ system. \end{cases}$$

where $A(\theta) = \theta A$. Other variable is defined as the model (2.74).

The test is using time interval as 0.05 second, input u_1 and the length of moving windows with 30. The fault variable is making as $\theta_0 = 1$ and $\theta_f = 0.8$ and the initial value of the estimation is made as 0.9.

The output is seen in Fig. 2.31.

The fault variable identification could be seen in the following Fig. 2.32. Error for the state estimation is plotted in Fig. 2.33. From the figures, it can be seen that if the fault happened in the state part, it can affect the system much more hugely than the fault happened in the input item. But the estimation result shows nearly the same comparison.

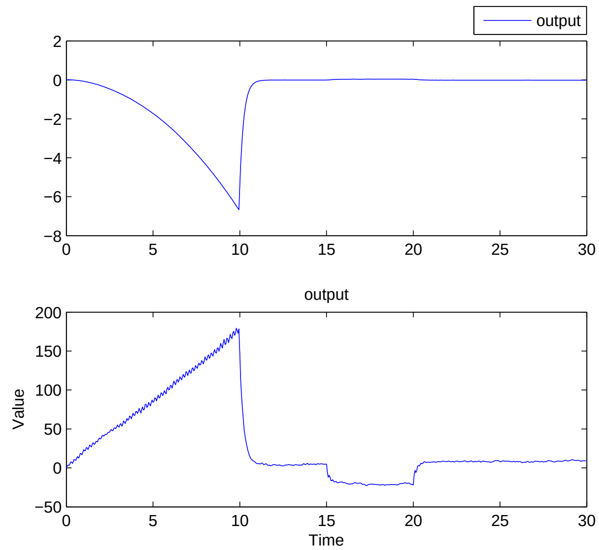

Figure 2.31: Output for case (**D**) with u_1

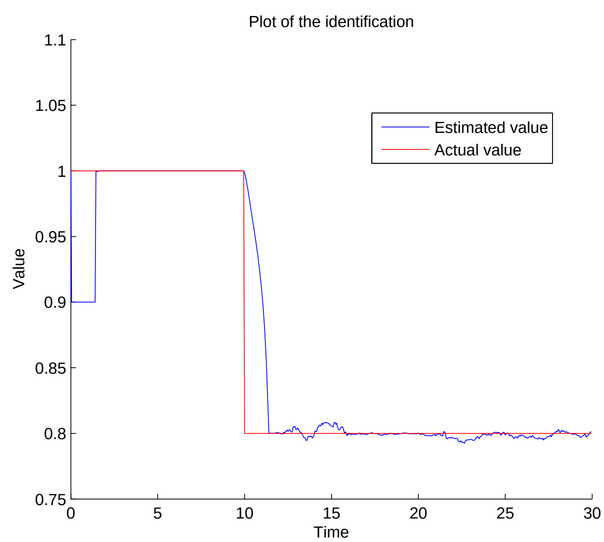

Figure 2.32: Fault detection for case (**D**) with u_1 and $N = 30$

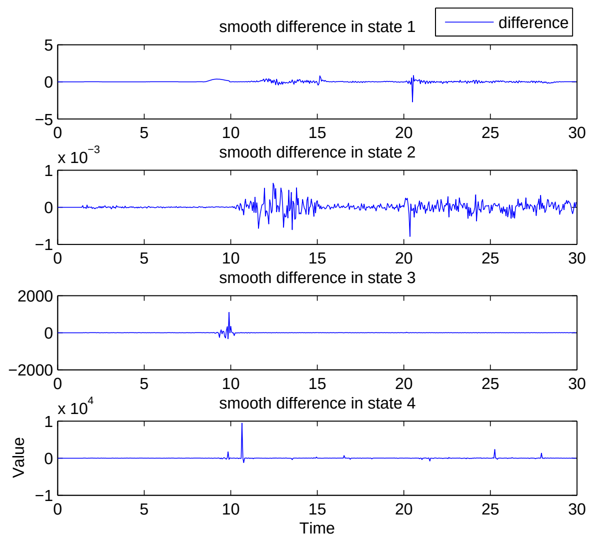

Figure 2.33: State estimation error for case (**D**) with u_1 and $N = 30$

2.7.3.3 Case (E): Fault in the all deterministic items

This part we combine the former two deterministic fault scenarios, that is the fault could influence both the input item and the state item. The model can be described as:

$$\begin{cases} dX = [A(\theta)X + B(\theta)U]dt + a\sigma dB_t \\ Y = CX + \omega_k \end{cases} \tag{2.92}$$

where $A(\theta) = \theta_A A$ and $B(\theta) = \theta_B B$, and let

$$\theta = [\theta_A \quad \theta_B]^T = \begin{cases} [\theta_{A0} \quad \theta_{B0}]^T, & Normal \quad system, \\ [\theta_{Af} \quad \theta_{Bf}]^T, & Faulty \quad system. \end{cases} \tag{2.93}$$

When the system is running without fault, all the system matrices are $A(\theta_{A0}) = A$ and $B(\theta_{B0}) = B$. If the fault happens, $\theta_{Af} = 0.9$ *or* 0.8 and $\theta_{Bf} = 0.2$ *or* 0.5.

We make tests for the following 4 situations.

E-a. u_1, $\theta_{Af} = 0.8$, $\theta_{Bf} = 0.5$ and 30 sample points.

E-b. u_2, $\theta_{Af} = 0.9$, $\theta_{Bf} = 0.2$ and 30 sample points.

E-c. u_3, $\theta_{Af} = 0.9$, $\theta_{Bf} = 0.2$ and 5 sample points.

E-d. u_4, $\theta_{Af} = 0.9$, $\theta_{Bf} = 0.2$ and 10 sample points.

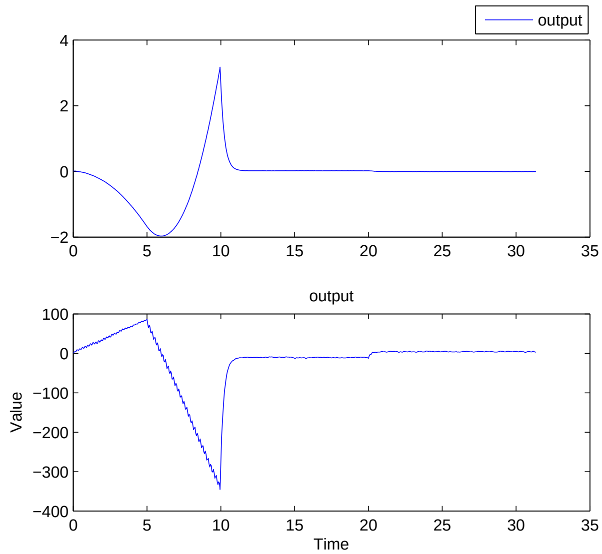

Figure 2.34: Output for case E-a.

The output data and estimation results for Case E are showed in Fig. 2.34–Fig. 2.45.

2.7.3.4 Results Analysis

Fault detection and state estimation are implemented using Kalman Filter technique plus Maximum Likelihood method to a class of control system which is modeled by the ISDE equation. In the system, the fault is considered to be parametric one, that is if the fault happens, some of the system parameters will be changed. When the fault only affect the deterministic part of the system (drift item for the ISDE equation), the following properties can be obtained for the method.

- **Precision:**
  Regarding the accuracy performance of the method, both the identification of the fault parameter and state estimation are quite fine except for the data collected period. When the fault happens, it also need some time to recover to the steady identification. The method can accurately make the estimation of the fault parameter and state.

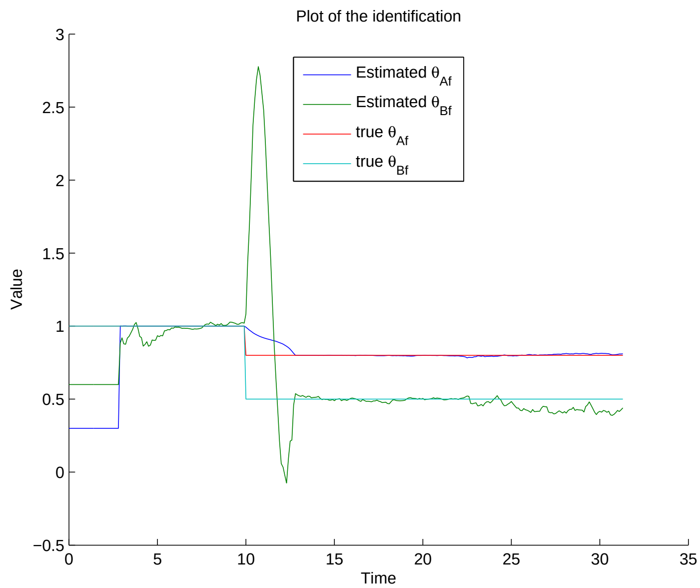

Figure 2.35: Fault detection for case E-a.

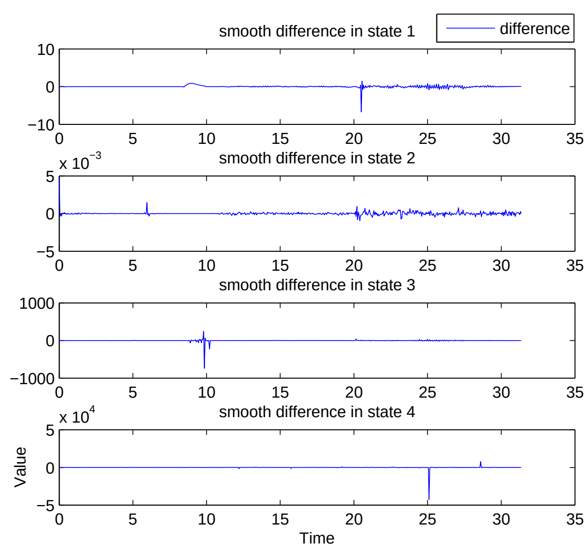

Figure 2.36: State estimation error for case E-a.

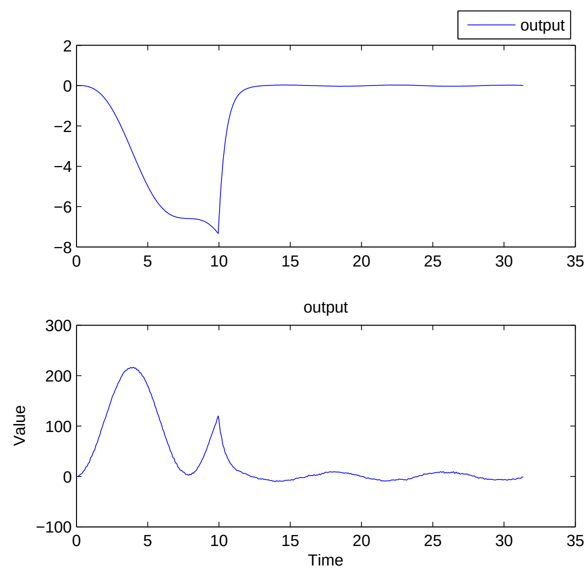

Figure 2.37: Output for case E-b.

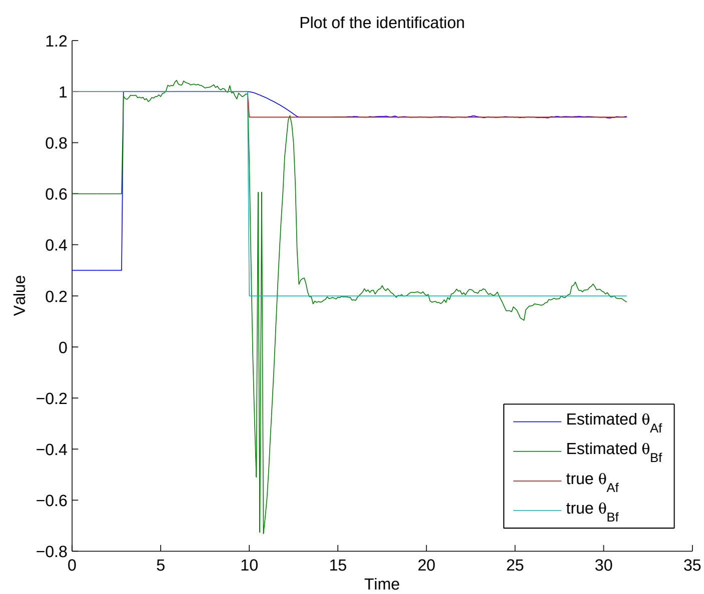

Figure 2.38: Fault detection with case E-b.

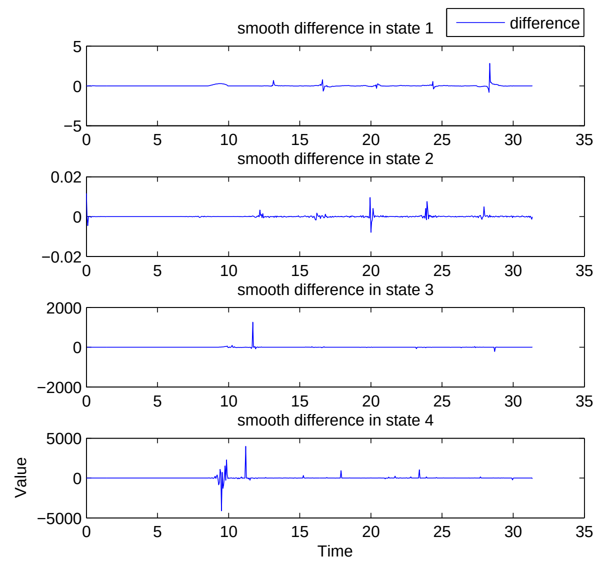

Figure 2.39: State estimation error for case E-b.

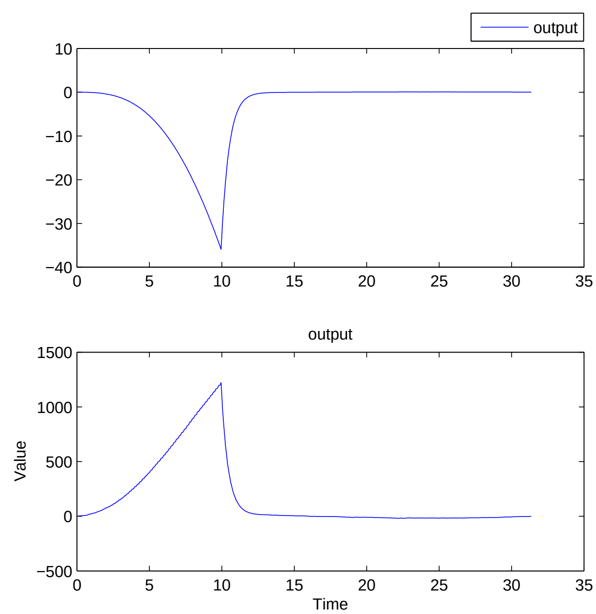

Figure 2.40: Output for case E-c.

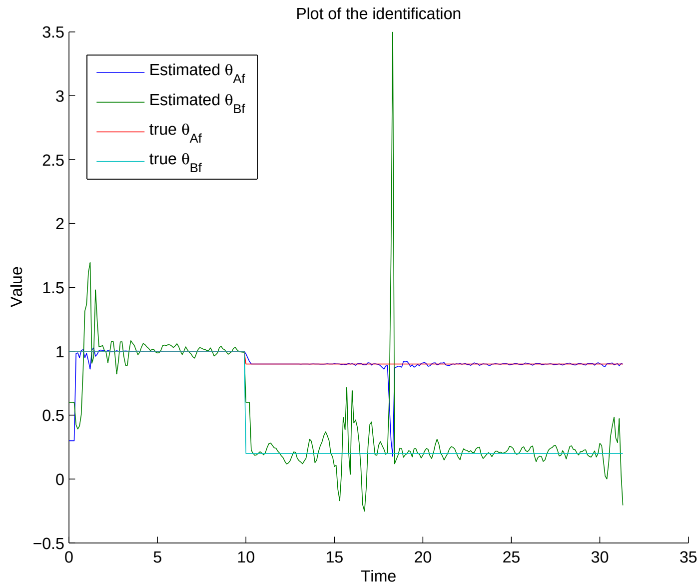

Figure 2.41: Fault detection with case E-b.

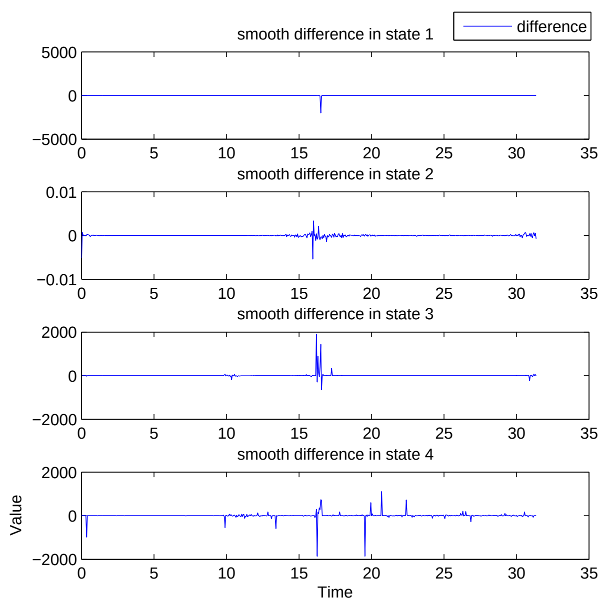

Figure 2.42: State estimation error for case E-b.

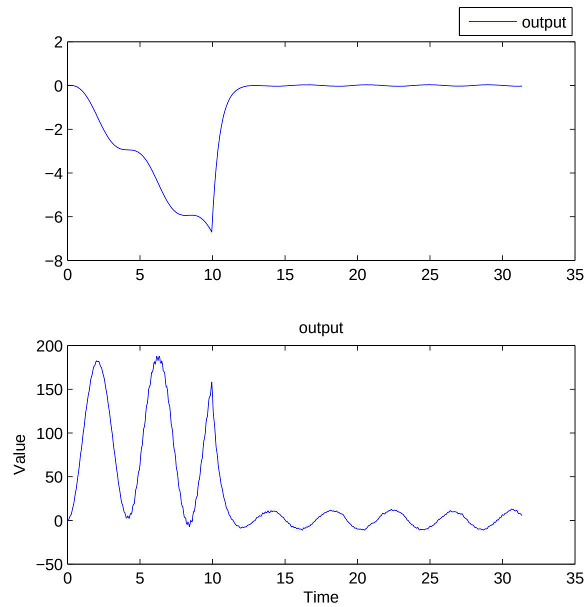

Figure 2.43: Output for case E-d.

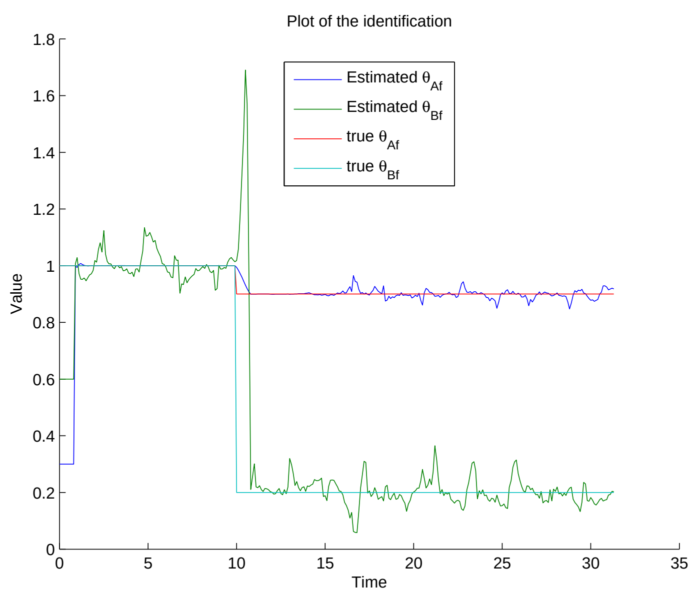

Figure 2.44: Fault detection for case E-d.

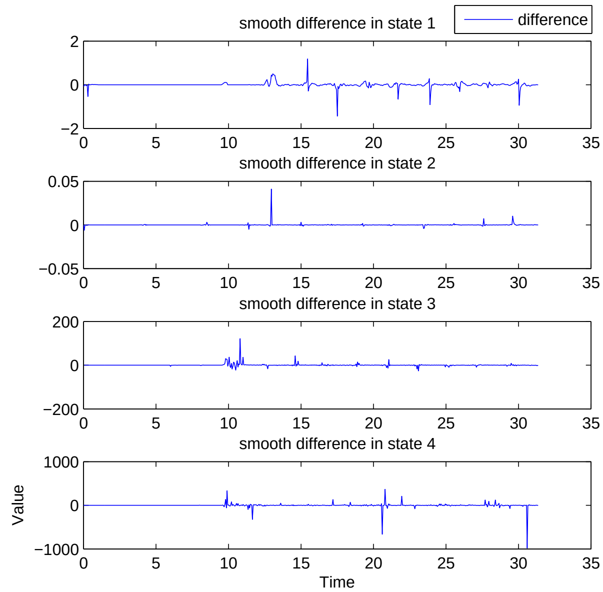

Figure 2.45: State estimation error for case E-b.

- **Length of Moving Windows:**
  In the simulation, only two kinds of length for moving windows are adopted, 5 and 30. In general, the length of moving windows should have a lower limit used in the first stage of KF technique, if the used data is too little the state estimation will be bad. If the windows length is beyond the low limit, the more it is, the better the estimation is for the time invariant system. But for the fault control system, it is a time varying system, the estimation using much data (long windows) would lead to the results losing time varying property. Moreover, without considering the accuracy in the fault estimation, one long windows means large time delay for the fault detection. It is a dilemma that the longer windows we use, the more continuous estimation is obtained, but the more time delay in the fault decision. It is important to find a balance to determine the length of moving windows.

- **Input Signals:**
  The input signals can also affect the estimation results for the system. From the estimation using 5 points as the length of moving windows, it is more obvious, if

the input have seriously changed, such as direction change, sudden jump, etc. or periodically changed related with the frequency, the estimation show the same properties with it, i.e., the performance of the estimation can be affected by the characteristic and frequency of the input signals.

2.7.4 Case (F): Fault in both deterministic part and random part

In this part, the system considered that if the fault happens, it can affect both deterministic part and random part. For the simplicity, here the fault in deterministic part is only considered to happen in the input item.

In the case, the normal system and faulty system can be written as:

$$\begin{cases} dX(t) = [AX(t) + B(\theta)U(t)]dt + a\sigma dB_t \\ Y(t) = CX(t) + \omega_t \end{cases} \tag{2.94}$$

and

$$\Theta = [\theta \;\; \sigma]^T = \begin{cases} [\theta_0 \;\; \sigma_0]^T = [1 \;\; 1]^T, & Normal \;\; system, \\ [\theta_f \;\; \sigma_f]^T, & Faulty \;\; system. \end{cases} \tag{2.95}$$

Note that in this model, the parameter σ in the diffusion part is not a constant but an unknown parameter related to the fault like θ in the drift part need to be identified. The fault detection need to be performed by identifying $\Theta = [\theta \;\; \sigma]^T$.

Two different fault scenarios are considered in the following.

Same fault in both deterministic part and random part:
At first, the fault is considered the same in both deterministic part and random part. The values in the data generation are set as $\theta_f = 1.5$, $\sigma_f = 1.5$.

F-a. One parameter method: In this part, since the fault influence the deterministic part and random part in the same way, the two parameters in them can be handled by only one parameter. Hence, the fault variable both in the deterministic part and random part could be considered as only one parameter β, where $\beta_0 = 1$ and $\beta_f = 1.5$. In the simulation, the input variable is using u_2. The length of moving windows is set as 30 samples. The output can be seen in Fig. 2.46 and results of the fault identification and state estimation are plotted by corresponding figures.

F-b. Multi-parameters method: Considering the system model in case F-a., it is to make the parameter identification of (θ, σ) although these two parameters change in the same way. The results are not listed but there is a little difference to the case F-a that the performance is not so better than it.

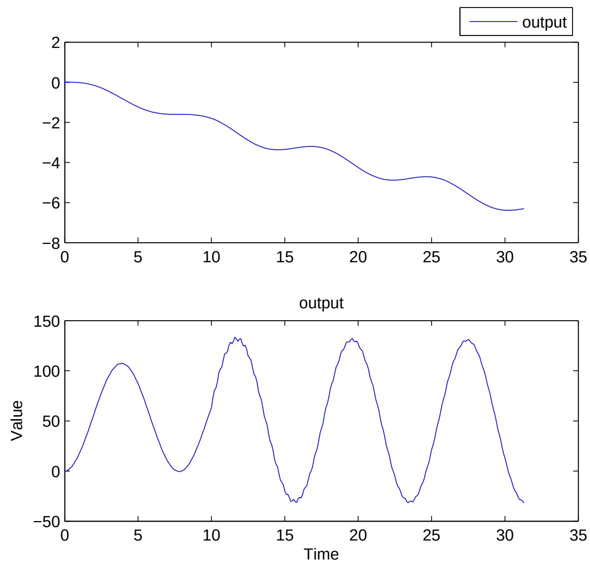

Figure 2.46: Output for case F-a.

Different fault in both deterministic part and random part:

3 different tests are made in this part. The input variable is using u_1.

F-c. $\theta_B = 0.5$, $\sigma_f = 1.5$.

F-d. $\theta_B = 0.5$, $\sigma_f = 10$.

F-e. Another test–using the real system (2.94) with (2.95) to generate the data but for the detection using deterministic fault model (2.89) with (2.90), $\theta_B = 0.5$, $\sigma_f = 10$.

The output signal and estimation results of Case F-c, F-d, F-e could be seen in the following figures Fig. 2.49–Fig. 2.57.

Results analysis:

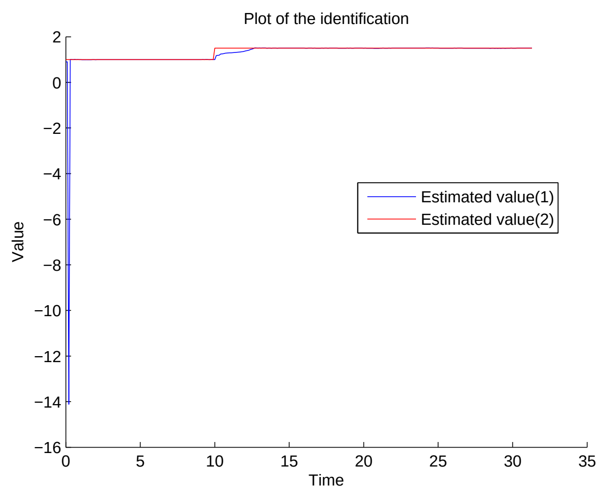

Figure 2.47: Parameter identification for case F-a.

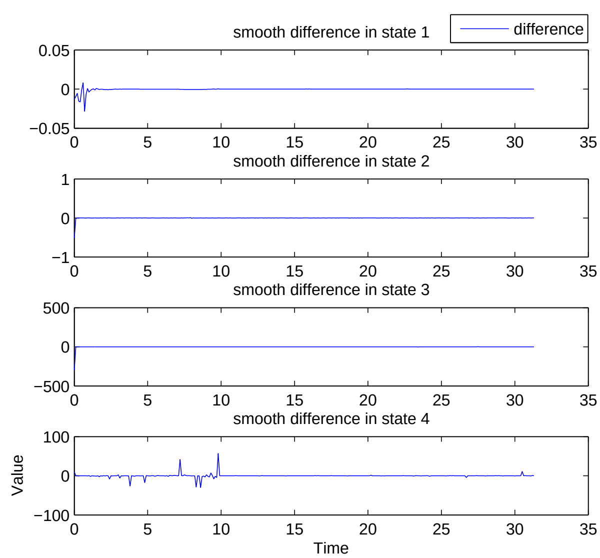

Figure 2.48: State estimation error for case F-a.

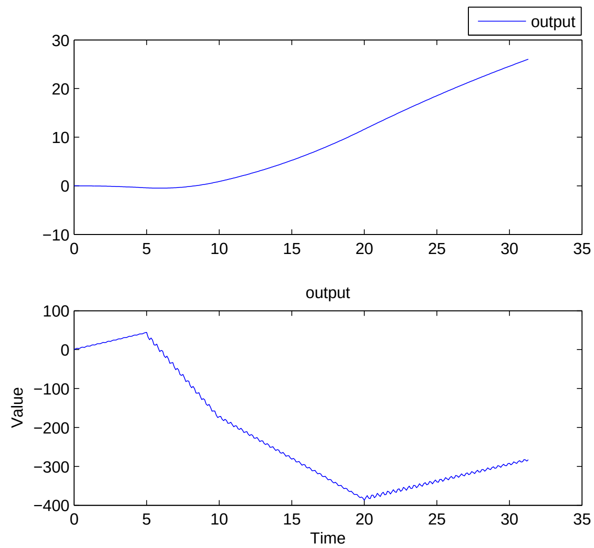

Figure 2.49: Output with case F-c.

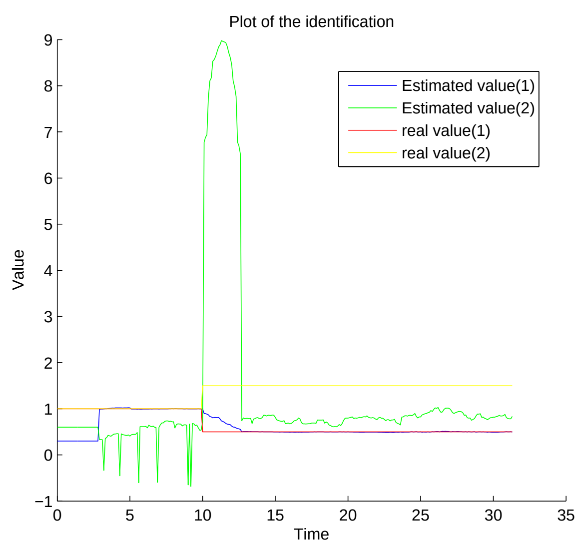

Figure 2.50: Fault detection with case F-c.

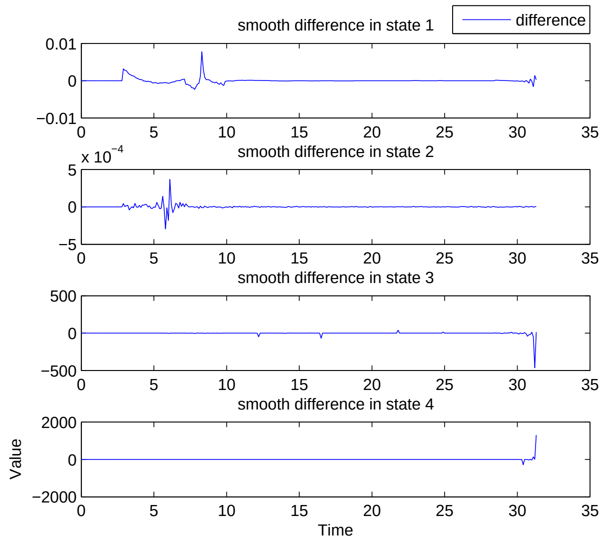

Figure 2.51: State estimation error for case F-c

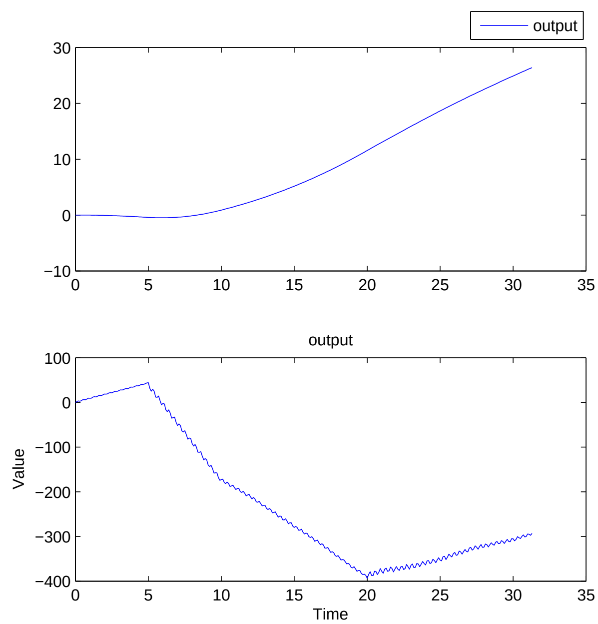

Figure 2.52: Output with case F-d

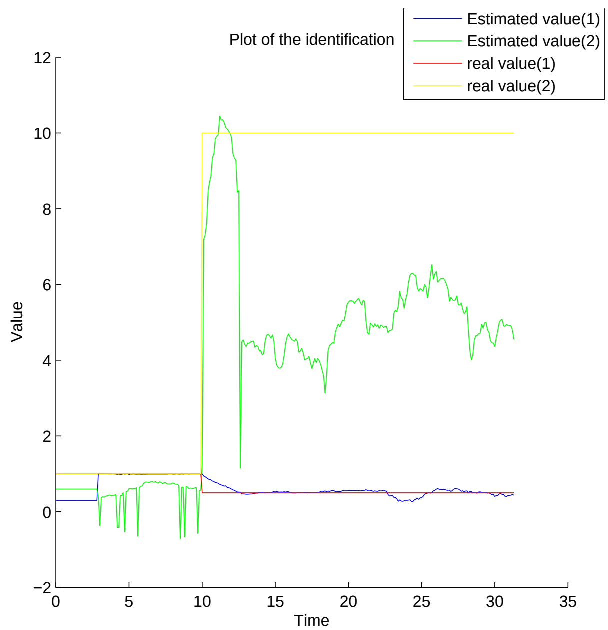

Figure 2.53: Fault detection with case f-d

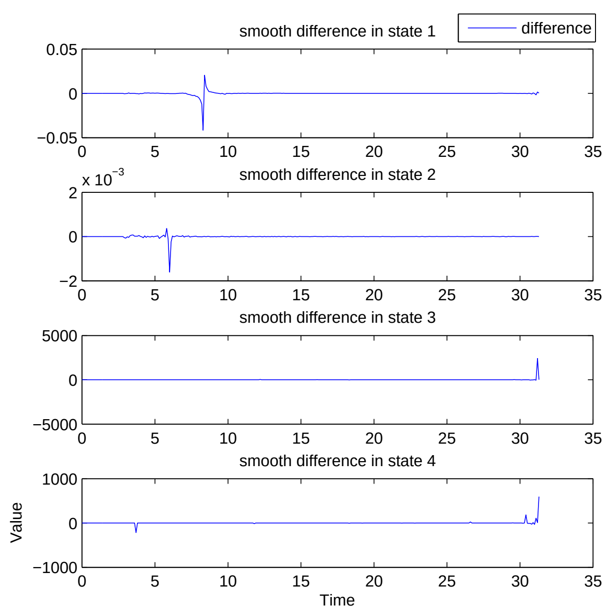

Figure 2.54: State estimation error for case F-d

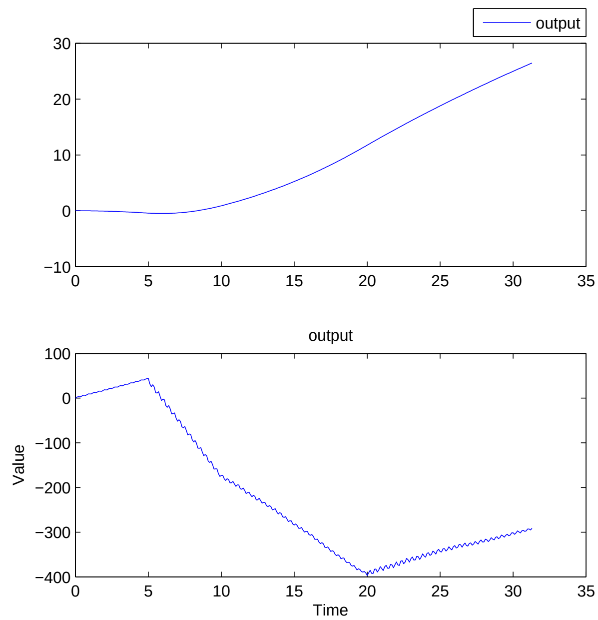

Figure 2.55: Output with case F-e

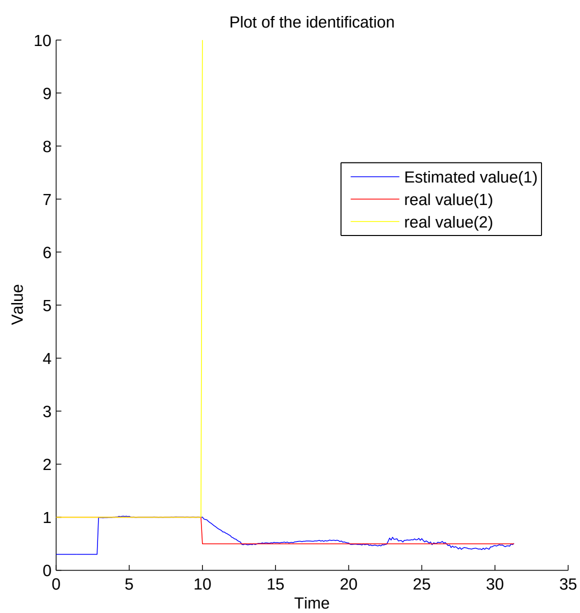

Figure 2.56: Fault detection with case F-e

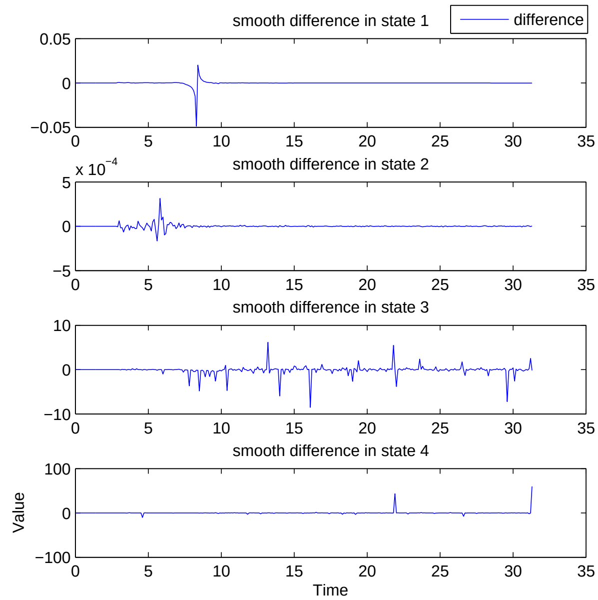

Figure 2.57: State estimation error for case F-e

Parameter identification: The results of the parameter identification show that only the deterministic part of the fault are estimated accurately for all the cases. For the random part of the fault, the precisions of case F-a and F-b are better. Estimations under the cases F-c and F-d are bias from the true values. Case F-e is the worst, the result is even not convergent. From the results, it can be observed that if the parameter in the random part is considered to the identification, the accuracy of the identification of the deterministic will be better than that without considering the parameter in the random part. But for the estimated parameter in random part, the performance will depend on the system itself. From the fault detection, the method with the model can detect both part of the fault on time without considering the sample delay. This phenomenon is generated since the random property of the noise sometimes destroys its main property. However, it can still diagnosis the fault accurately in some senses with a little bias for the parameter estimated although estimation for the random part is not good for some cases.

Sample points: In this section, the cases using different sample points (moving windows) are not listed in the thesis. This is because its effect shows the same phenomenon as the previous tests. For example, 30 points estimation is much better than the 5 points estimation with regards to the stability, continuous property. It is really the case that the more points applied, the more stable and continuous the estimation was. However, the cost of more sample points used is more time delay to detect whether the fault happened.

State estimation: It is obvious that for all the cases, the state estimations are quite good. But for any of these case, at some initial time the estimation is bad, which is because the process need some time to catch up the property of the system. After a short while, the estimation shows good performance. Then when the fault happens, the estimation is destroyed. It need some time to recover to the good level for the estimation. Sometimes, at the end of the running time, there may be several estimation which is not good.

2.8 Conclusion

A system identification method with state estimation using UKF-ML technique for ISDE model is proposed. The KF technique is firstly applied to get a parameterized state estimation. Secondly, the ML function is formed using the parameterized state estimation and the noise distribution knowledge. Then, an optimization problem of the ML function needs to be solved and the optimal value is taken as the estimated system parameter. The method is proved to be consistency and normality for the considering systems. And it can be extended to an online manner.

A large amount of numerical simulation showed that it could provide a better performance than traditional methods, such as EKF, UKF and EKF plus ML methods, in terms of the accuracy and the convergency at the cost of more computation load. But with the increasing of computer, this is not the concerned problem as before. If the approach is using in an online manner, it can be seen obvious that several factors can affect the performance of the estimation, such as the length of the moving windows, input signals.

The model and methods can be also applied to the FDD process. Based on the predefined threshold method, the fault decision can be made based on the system identification. Meanwhile, the estimated parameter related with fault is substituted into the parameterized state estimation and the Kalman Smoother is applied for the state estimation. Thereby the state estimation can be obtained as another byproduct.

The simulation results based on a robot system showed a promising performance of the proposed method in terms of providing a quick, accurate and robust fault parameter identification and state estimation.

Chapter 3

System Identification Method for TV-FOPDT Model and Its Application

In this chapter, we will discuss the system identification of a nonlinear FOPDT model and its application in a real-life relevant system. Extending the standard FOPDT model, this Chapter proposed some new FOPDT models and corresponding methods to make parameter identification of them.

The content in this chapter is as follows:

Overview of the Previous Work In order to show the motivation, the model development is shortly described. Furthermore, the methods to make the parameter identification of the standard FOPDT model are summarized.

Model Extension and Identification Methods Several different models are extended based on the standard FOPDT model. According to the characteristic of the new models, the identifiability is defined and investigated. Corresponding theorems regarding the identifiability are proved. Then some new methods based on a kind of nonlinear programming problem are proposed to make parameter identification of these different models.

Numerical Test and Application Finally, a number of numerical tests are performed to illustrate the approach and compared with other methods. A scenario of the application in superheat dynamic modeling is used as an application of the work.

3.1 Motivation and Purpose

In last chapter, the state space model is discussed. It extends the stochastic state space model using ISDE model. There is another kind of model which is also widely used in application, that is input/output model. First Order Plus Dead Time (FOPDT) model is the famous one in this category which is widely applied and it can model many industrial processes.

The FOPDT model has three different parameters, named system gain, time constant and dead time (time delay). These parameters are often set as constants in the whole system running for the standard model. In reality, during the system running, the system may not stay unchanged but vary according to the time. Thereby, in order to make up for the shortage of the standard FOPDT model, a kind of nonlinear FOPDT model in which the time varying parameters of the system can be describe is proposed in (85; 89; 123). The considered nonlinear FOPDT model is an extension of the standard FOPDT model by means that both system's gain and time constant can be changed during the system running. This nonlinear FOPDT model is generated by using a linearized method to a nonlinear model. In the thesis, a new type of explicit nonlinear FOPDT model is proposed as well, named Time-Varying FOPDT (TV-FOPDT) model, which is used to model the superheat dynamic in a supermarket refrigeration system. The TV-FOPDT model is an extension of the standard FOPDT by allowing the system parameters (system gain, time constant and time delay) to be time dependent variables.

Sometimes, in the practical system, the parameters may depend on the other variables besides the time. For example, considering the system of the superheat in a refrigeration system, the time that used in the process of changing the evaporation temperature depends on some conditions, such as the refringent filling of the evaporator. The more the refringent is filled in the evaporator, the more time needed. If this refringent filling is taken as the system input, the time that used in the temperature change process, which could be taken as the time delay for the system changing, may depend on this input refringent filling (87; 135; 171). In order to express this property of system, this thesis will extend the TV-FOPDT model to a more general one with assuming that the dead time (time delay) can be also input dependent and the model can be called as a kind of TV-FOPDT model with input dependent dead time. Furthermore, the model is extended to Multiple Input (MI) systems.

For the different proposed models, the corresponding methods to make parameter identification are developed based on some nonlinear programming techniques. In the beginning, the traditional methods to make parameter identification of the standard FOPDT model are reviewed.

3.2 FOPDT Identification

A standard FOPDT model can be expressed by the following equation and transfer function:

$$Y(s) = G(s)U(s), \tag{3.1}$$

with transfer function

$$G(s) = \frac{K}{T_p s + 1} \exp^{-T_d s}. \tag{3.2}$$

where $Y(s)$ and $U(s)$ are the Laplace-transform of system output and system input, K is the system gain, T_p is the time constant and T_d is the (apparent) time delay (dead time).

Different methods (164) have been already proposed to estimate these three parameters in the FOPDT model (3.1) with (3.2) by performing a simple experiment on the plant. This is motivated by the fact that many processes can be described effectively by this dynamic model and it suits well with the simple structure of some kinds of controller.

Tangent Method referred in (17), firstly draws the tangent of the system response at the inflection point. Then, the method determines the system gain by dividing the steady-state change in the system output y using the amplitude of the step in input. And the dead time T_d can be determined as the time interval between the application of the step input and the intersection of the tangent line with the time axis. Finally, the value of $T_p + T_d$ is estimated as the time interval between the application of the step input and the intersection of the tangent line with line $y = y_\infty$ where y_∞ is the final steady state value of the system output. And the time constant T_p can be calculated by subtracting the previously estimated value of the time delay T_d. This method can provide exact results for a true FOPDT system. But its main drawback is that it only depends on one single point of the reaction curve (i.e., the inflection point) and for this reason, it is much sensible to the measurement noise. In fact, the measurement noise may cause large errors in the estimation of inflection point and of time derivative of the system output (164).

Area Method is an approach that is more robust to the measurement noise (164). Considering that the system gain K can be determined the same as for the tangent method, it firstly calculates the area between the system output and line $y = y_\infty$. Then, $T_p + T_d$ is to be determined by the division between this area and estimated K. Subsequently, the area between the system output and the time axis in the time interval from initial time to $T_p + T_d$ is calculated. Finally, T_p and T_d are determined by the combination of these two area calculation and estimated K. Since it need to calculate some integrals, it is more relevant from the computational view, but it has advantage that it is more robust to the noise in the measurement than the tangent method. However, it has a drawback in the possible determination of a negative value of the time delay T_d when the process exhibits a nonlinear lag-dominant dynamics (164).

Two-points-based Method is based on the estimation of two time instants of the reaction curve, which has been proposed in (155) (it is also reported in (141)). It consists in determining two time instants when the process output attains 35.3% and 85.3% of its final steady state respectively. Then, the dead time and the time constant are calculated by the combination of these two instants. The gain of the process is determined as in the area method. This approach is very simple and it can be applied by hand easily. This technique, in addition to the problem of being sensible to the measurement noise in the estimation of the two times, suffers from the same problem as the area method (164).

Optimization-based Method Optimization based method is to estimate the three transfer function parameters K, T_d and T_p by minimizing the integral of difference between the experimental step response and the model step response (132). The major drawback of this method is the computation load.

Least Square (LS) Method LS method, referred in (154), is widely used to make the identification of FOPDT model. This method firstly applies moving covariance to find the dead time of the system. And then uses least square method to identify the other two parameters.

Prediction Error Method (PEM) PEM is a more general method that can identify many models (100). It is already being implemented in software matlab. But for the certain model, it is not the best method with regarding to the accuracy and computation load (179).

3.3 TV-FOPDT System Identification

Based on the standard FOPDT model, a TV-FOPDT model is proposed. Now, the new TV-FOPDT model is the main concern in this section.

3.3.1 TV-FOPDT Formulation and Its Identification Problem

In the following, a kind of First-Order Plus Dead-Time (FOPDT) process model by:

$$y(s) = G^t(s)u(s), \tag{3.3}$$

with transfer function

$$G^t(s) = \frac{K^t}{T_p^t s + 1} \exp^{-T_d^t s}. \tag{3.4}$$

Here $y(s)$ is the system output, $u(s)$ is the system input, K^t is the process gain, T_p^t is the system time constant and T_d^t is the time delay in the system. Note that the superscript t means that the corresponding variable may have the alteration during the whole running time of the system. In order to recognize this model from the standard FOPDT model, it is called as Time Varying FOPDT (TV-FOPDT) model.

Then the corresponding system identification could be described as follows, which is the main problem concerned in this Chapter as well.

(P): **Estimate the parameters including K^t, T_p^t and T_d^t in the system modeled by (3.3) with (3.4) based on a set of input and output data.**

3.3.2 Model Discretization

System model (3.3) with (3.4) is firstly approximated by its discrete version. The transfer function (3.4) is discretized as

$$G^t(z) = \frac{K^t(1-\alpha^t)}{z^l(z-\alpha^t)}, \tag{3.5}$$

where $\alpha^t \triangleq \exp^{-\frac{T_s}{T_p^t}}$, and T_s is the sample interval. Here, l is the simplicity of l^t, which is the discrete approximation of system delay T_d^t, and it is an integer number with property: $lT_s \leq T_d^t \leq (l+1)T_s$.

Now define $\beta^t \triangleq K^t(1-\alpha^t)$, the TV-FOPDT model (3.3) with (3.4) can be further transferred into a description using difference equation as:

$$y(k) = \alpha^t y(k-1) + \beta^t u(k-l-1), \tag{3.6}$$

for $k = l+1, l+2, \cdots\infty$.

Then the model identification problem **(P)** with parameters K^t, T_p^t and T_d^t is converted to estimate the new parameters α^t, β^t and l for the discrete model version (3.6).

3.3.3 Identifiability Analysis

Before the identification procedure is performed to the system models, the identifiability of the corresponding models should be firstly checked. In (99), the identifiability of parameterized model was given. It proposed to express the identifiability of the parameterized model as that the identified value is the same to the true value of the model. It is described in (100) for some kinds of models such as SISO transfer function model and state space model. But sometimes, it is hard to know the true values of the system parameters beforehand. The comparison between the estimation and the true value can not be accomplished.

In this chapter, the model considered is a nonlinear FOPDT model, in which the nonlinearity is expressed by its time varying property and time delay, no identifiability analysis could be found for this kind of system in the previous work. Here the thesis tends to set up the definition of the identifiability based on the time varying nonlinear FOPDT and then prove a corresponding theorem.

Definition 3.3.1 *Suppose the nonlinear model $\mathcal{M}$ with discrete measurement, Θ is the parameters in the model, consider the identification method $\mathcal{I}$, if there exists an integer N, based on any given N couples of data points $\{y_i\}_{i=t}^{t+N-1}$ and corresponding input signal, the identification results $\hat{\Theta}_{t+N-1}$ using $\mathcal{I}$ is unique for any time t, then the identification method $\mathcal{I}$ is said to be globally N identifiable for model $\mathcal{M}$.*

This definition of globally identifiable is based on the sample points and the identification method. It is reasonable, because for the nonlinear identification, in general, the method can only be applied for some certain models and the accuracy performance is affected by the number of the sample points much hugely. And it is set up on the uniqueness of the estimation. In some sense, it would be equivalent to the definition of the identifiability in (100). However, this definition may be more practical because it do not need to know the true value before the identification.

In the following, if the identification method $\mathcal{I}$ is globally N identifiable for model $\mathcal{M}$, it is noted as $\mathcal{M}$-N globally identifiable for simplicity.

Proposition 3.3.2 *Suppose model $\mathcal{M}$ is time invariant system, if the identification method $\mathcal{I}$ is $\mathcal{M}$-N_0 globally identifiable, then for any $N \geq N_0$, $\mathcal{I}$ is $\mathcal{M}$-N globally identifiable as well.*

Proof: This proposition can be easily proved by the method of contradiction. It will be omitted here.

For the time invariant system, the proposition shows that if a method is globally identifiable, it will be globally identifiable when the number of data points used for the identification exceeds a fixed number. That means the sample points should be sufficiently enough to get the right estimation of the system. But for the time varying system, the proposition 3.3.2 is no longer hold. Since the time varying property, too many data will lose the time varying of the parameter so that the identification only shows the average level which would lead to bad result. However, less samples can maintain the time varying property, but can not grab all the information of system so it would not provide an accurate identification. From these two angles, there should be a balance to choose the number of the sample points to make the identification. That is to say, for the time varying system, there may exist an optimal sample number to make system identification. For this sense, it need to re-consider the property of identifiability.

Now, considering the model (3.6), which is the one in the linear manner and the parameters are time varying. In order not to loss the generality, in the following, the more general time varying linear model is studied. First, we will prove the following theorem regarding to the identifiability.

Theorem 3.3.3 *Consider a time varying transfer function system*

$$Y(s) = G(s, \Theta^t)U(s) \tag{3.7}$$

here the superscript t means the corresponding variable is time varying. If

1. *The system (3.7) can be rewritten equivalently as*

$$y(t) = \phi^T(t)\tilde{\Theta}^t, \tag{3.8}$$

with $\phi(t)$ is the information vector consisting of some observations up to time $(t-1)$, and $\tilde{\Theta}^t$ is corresponding parameter vector converted by Θ at time t. And the frequency of the observation is much larger than the frequency of parameter changing.

2. *The input $u(t)$ is persistently excited.*

are satisfied, then there exists an integer N such that the Least Square (LS) estimator is (3.8)-N globally identifiable, then LS is (3.7)-N globally identifiable.

Note that the first condition requires that the system should be equal to a time varying linear system. And there is frequency requirement on the observation. It is really natural to guarantee the estimation can track the change of the parameters. Although it is difficult to know the frequency of parameter changing beforehand, the frequency demanding can still be satisfied by decreasing the samples interval as little as possible. The second condition is the requirement to the input signal.

Proof: If the input $u(t)$ is persistently excited, it means that there exists α, β satisfying $0 < \alpha \leq \beta < \infty$ and a positive integer N, such that for the successive N samples

$$\alpha I \leq \sum_{i=t+1}^{t+N} \phi(i)\phi^T(i) \leq \beta I, \quad a.s. \;\; for \;\; any \;\; t > 0. \tag{3.9}$$

Also, for the equivalent system (3.8), there is

$$y(t+i) = \phi^T(t+i)\tilde{\Theta}^t, \tag{3.10}$$

for any $t > 0$ and $i = 1, 2, 3, \cdots, N$. It can be formalized by matrices

$$Y = \Phi\tilde{\Theta}^t, \tag{3.11}$$

where $Y = [y(t+1)\ y(t+2)\ \cdots\ y(t+N)]^T$, $\Phi = [\phi(t+1)\ \phi(t+2)\ \cdots\ \phi(t+N)]^T$. From (3.9), the matrix $\sum_{i=t+1}^{t+N} \phi(i)\phi^T(i)$ is a full rank matrix, noted its rank as R, where $R = dim(\tilde{\Theta}^t)$, and

$$\sum_{i=t+1}^{t+N} \phi(i)\phi^T(i) = \Phi\Phi^T. \tag{3.12}$$

As a result, $\Phi\Phi^T$ is nonsingular and its inverse exists. According to the algorithm of Least Square Estimator, the former equation (3.11) has a unique solution. It is also the solution of the system identification of (3.8).

From the equivalence of the model (3.8) and (3.7), the parameter identification can be obtained uniquely based on the estimation of $\tilde{\Theta}^t$ in the model (3.8). At last, it is proved that there exists N, the Least Square estimator is (3.8)-N globally identifiable and hence (3.7)-N globally identifiable. $\sharp$

Now the LS estimator is proved to be (3.8)-N globally identifiable and (3.7)-N globally identifiable. For time varying system, Theorem 3.3.3 only proved the existence of N to make LS estimator identifiable. There should be many choices of N. But since the model is time varying one, different choices of N will lead to different estimation performance. It need us to select an optimal one to get the best estimation.

Theorem 3.3.4 *Consider the model,*

$$y(t) = \phi^T(u(t),t)\Theta^t, \tag{3.13}$$

the variables are defined the same to the previous Theorem 3.3.3, suppose

1. *$u(t)$ is persistently excited, i.e., there exists α, β satisfying $0 < \alpha \leq \beta < \infty$ and a positive integer N, such that for the continuous N samples, the (3.9) is satisfied.*

2. *parameter changing rate is bounded, i.e., $\Delta^t = \Theta^t - \Theta^{t-1} \leq M$.*

then the best choice of data sample number in the system identification using LS estimator is N, i.e., LS estimator with N samples will get the estimation with the least upper bound of the error.

Proof: Define the estimation error

$$\varepsilon_t = \hat{\Theta}^t - \Theta^t. \tag{3.14}$$

where the recursive LS algorithm with fixed windows length q gives the estimation

$$\begin{aligned}\hat{\Theta}^t &= \hat{\Theta}^{t-1} + P_t\phi_t[y(t) - \phi_t^T\hat{\Theta}^t] \\ P_t^{-1} &= P_{t-1}^{-1} + \phi_t\phi_t^T - \phi_{t-N+1}\phi_{t-N+1}^T.\end{aligned} \tag{3.15}$$

Then,

$$\begin{aligned}\varepsilon_t &= \hat{\Theta}^t - (\Theta^{t-1} + \Delta^t) \\ &= \hat{\Theta}^t - (\hat{\Theta}^{t-1} - \varepsilon_t + \Delta^t) \\ &= P_t\phi_t[y(t) - \phi_t^T\hat{\Theta}^t] + \varepsilon_t - \Delta^t \\ &= [I - P_t\phi_t\phi_t^T]\varepsilon_t - \Delta^t.\end{aligned} \tag{3.16}$$

Define $\Gamma_t = -P_t\phi_t\phi_t^T\varepsilon_t - \Delta^t$, then there is

$$\varepsilon_{t+i} = \varepsilon_t + \sum_{k=0}^{i}\Gamma_{t+k} \tag{3.17}$$

and

$$\phi_{t+i}^T[\varepsilon_t + \sum_{k=0}^{i-1}\Gamma_{t+k}] = \phi_{t+i}^T\varepsilon_{t+i}. \tag{3.18}$$

or

$$\phi_{t+i}^T\varepsilon_t = -\phi_{t+i}^T\sum_{k=0}^{i-1}\Gamma_{t+k} + \phi^T t + i\varepsilon_{t+i}. \tag{3.19}$$

Taking $\|\cdot\|_2^2$ to both sides of (3.19), there is

$$tr[\varepsilon_t^T\phi\phi_{t+i}^T\varepsilon_t] = \| -\phi_{t+i}^T\sum_{k=0}^{i-1}\Gamma_{t+k} + \phi^T t + i\varepsilon_{t+i}\|_2^2. \tag{3.20}$$

Make sum of (3.20) from $i = 0$ to $i = N-1$, it is obtained that

$$tr\{\varepsilon_t^T\sum_{i=0}^{i=N-1}[\phi_{t+i}\phi_{t+i}^T]\varepsilon_t\} = \sum_{i=0}^{i=N-1}[\| -\phi_{t+i}^T\sum_{k=0}^{i-1}\Gamma_{t+k} + \phi^T t + i\varepsilon_{t+i}\|_2^2]. \tag{3.21}$$

Apply the condition (1),

$$\begin{aligned}N\alpha\|\varepsilon_t^T\|_2^2 &\le 2\{\sum_{i=0}^{i=N-1}[\|\phi_{t+i}^T\sum_{k=0}^{i-1}\Gamma_{t+k}\|_2^2 + \|\phi^T t + i\varepsilon_{t+i}\|_2^2]\} \\ &\le 2\{\sum_{i=0}^{i=N-1}[N\beta\|\sum_{k=0}^{i-1}\Gamma_{t+k}\|_2^2 + \|\phi^T t + i\varepsilon_{t+i}\|_2^2]\}.\end{aligned} \tag{3.22}$$

Then

$$\begin{aligned}
\| \varepsilon_t^T \|_2^2 &\leq \frac{2}{N\alpha}\{\sum_{i=0}^{i=N-1}[N^2\beta \| \sum_{k=0}^{i-1}\Gamma_{t+k} \|_2^2 + \| \phi^T t + i\varepsilon_{t+i} \|_2^2]\} \\
&= \frac{2N\beta}{\alpha}\sum_{i=0}^{i=N-1}\sum_{k=0}^{i-1} \| \Gamma_{t+k} \|_2^2 + \frac{2}{N\alpha}\sum_{i=0}^{i=N-1} \| \phi^T t + i\varepsilon_{t+i} \|_2^2 .
\end{aligned} \tag{3.23}$$

Taking limit to the (3.23), there is

$$\begin{aligned}
&\lim_{t\to\infty} \| \varepsilon_t^T \|_2^2 \\
&\leq \limsup_{t\to\infty}\{\frac{2N\beta}{\alpha}\sum_{i=0}^{i=N-1}\sum_{k=0}^{i-1} \| \Gamma_{t+k} \|_2^2 + \frac{2}{N\alpha}\sum_{i=0}^{i=N-1} \| \phi^T t + i\varepsilon_{t+i} \|_2^2\} \\
&\leq \limsup_{t\to\infty}\{\frac{2N^3\beta}{\alpha} \| \Gamma_t \|_2^2 + \frac{2}{\alpha} \| \phi^T t\varepsilon_t \|_2^2\} \\
&\leq \limsup_{t\to\infty}\{\frac{2N^3\beta}{\alpha} \| -P_t\phi_t\phi_t^T\varepsilon_t - \Delta^t \|_2^2 + \frac{2}{\alpha} \| \phi^T t\varepsilon_t \|_2^2\} \\
&\leq \limsup_{t\to\infty}\{\frac{2N^3\beta}{\alpha}[\phi_t^T P_t^2\phi_t \| \phi_t^T\varepsilon_t \|_2^2 + \| \Delta^t \|_2^2] + \frac{2}{\alpha} \| \phi^T t\varepsilon_t \|_2^2\} \\
&\leq \limsup_{t\to\infty}\{\frac{2N^3\beta}{\alpha}[\frac{1}{(q-N+1)\alpha} \| \phi_t^T\varepsilon_t \|_2^2 + \| \Delta^t \|_2^2] + \frac{2}{\alpha} \| \phi^T t\varepsilon_t \|_2^2\} \\
&\leq \limsup_{t\to\infty}\{(\frac{2N^3\beta}{\alpha^2} + \frac{2}{\alpha}) \| \phi_t^T\varepsilon_t \|_2^2 + \frac{2N^3\beta}{\alpha} \| \Delta^t \|_2^2\} \\
&\leq \limsup_{t\to\infty}\{(\frac{2N^3\beta}{\alpha} + \frac{2}{\alpha})\frac{2(N+1)(q+N)\beta^2}{\alpha} \| \Delta^t \|_2^2 + \frac{2N^3\beta}{\alpha} \| \Delta^t \|_2^2\} \\
&= \limsup_{t\to\infty}\{(\frac{4N^3(N+1)(q+N)\beta^3}{\alpha^3} + \frac{4(N+1)(q+N)\beta^2}{\alpha^2} + \frac{2N^3\beta}{\alpha}) \| \Delta^t \|_2^2\}.
\end{aligned} \tag{3.24}$$

From (3.24), the estimation error is a strictly monotone increasing function of q and from the last theorem q should not be less than N, then the optimal simple number is the smallest one in the possible set, that is N. ♯

The Definition 3.3.1 need to look for a fixed number N to make the identification of the system, and Theorem 3.3.4 shows that N can be really found out for the model (3.13). But sometimes it is difficult to find this kind of number, especially for some time varying systems and nonlinear systems, there is not a fixed number of sample points to make the optimal estimation. Another definition of the identifiability which is more general to the Definition 3.3.1 is given in the following.

Definition 3.3.5 *Suppose the nonlinear model $\mathcal{M}$ with discrete measurement, Θ is the parameters in the model, consider the identification method $\mathcal{I}$, the identification results $\hat{\Theta}_t$ using $\mathcal{I}$ is unique, then the identification method $\mathcal{I}$ is said to be globally identifiable for model $\mathcal{M}$.*

Based on Definition 3.3.5, we will consider a time varying system with time delay:

Theorem 3.3.6 *Consider the model,*

$$y(t) = \phi^T(u(t), t, d^t)\Theta^t, \tag{3.25}$$

with d^t is unknown time delay in the system, other variables are the same to the Theorem 3.3.3,

1. *$u(t)$ is persistently excited, i.e., there exists α, β satisfying $0 < \alpha \leq \beta < \infty$ and a positive integer N, such that for the successive N samples, the (3.9) is satisfied.*
2. *$d^t \in D$, D is a finite countable set*
3. *parameter changing rate is bounded, i.e., there exist two positive values M_1, such that $\Delta^t = \mid \Theta^t - \Theta^{t-1} \mid \leq M_1$.*

then the Least Square estimator is globally identifiable for model (3.25).

In order to prove this theorem, the lemma should be given beforehand.

Lemma 3.3.7 *Suppose two different system which can start at any given time point, can be described as:*

$$y_i(t) = \phi^T(t)\Theta_i^t, \qquad i = 1, 2. \tag{3.26}$$

here $\Theta_i^t, i = 1, 2$ are different time varying parameters in these two systems respectively. If

- *the system matrix $\phi^T(t) \not\equiv 0$;*
- *the parameters vector $\Theta_1^t \not\equiv \Theta_2^t$,*

then for any given time variable $t_0 > 0$, there exists $t \geq t_0$ such that $y_1(t) \neq y_2(t)$.

This lemma can be proved by method of contrapositive.

Proof: Suppose the conclusion is not right, then there exists one $t_1 > 0$ such that for any $t \geq t_1$, $y_1(t) = y_2(t)$. Considering the difference between (3.26), then it is obtained that for any $t \geq t_1$, $\phi^T(t)(\Theta_1^t - \Theta_2^t) = 0$. Since $\phi^T(t) \not\equiv 0$, $\Theta_1^t = \Theta_2^t$ for $t \geq t_1$. If we assume that the system start at time t_1, then the two different systems (3.26) are the same system. It is contradictive to the precondition. It is proved the conclusion is correct. $\sharp$

Now we turn to prove the theorem 3.3.6.

Proof: In the system (3.25), $d^t \in D$, D is a finite countable set, the d^t is bounded. Suppose $D = \{d_i^t\}_{i=1,\cdots,M}$, for each fixed d_i^t, according to the Theorem 3.3.4 and 3.3.3, there exists N_i such that the Least Square estimator is (3.25)-N_i globally identifiable.

Let $(\hat{\Theta}_i^t)$ is the identification results based on time delay d_i^t, the estimated output is noted as $\hat{y}_i(t)$, and the error of the identification is defined respectively as

$$e_i(t) = \sum_{k=t_0}^{t} \| \hat{y}_i(t) - y(t) \|_2^2 \,. \tag{3.27}$$

If there exists only one $c \in \{1, \cdots, M\}$, such that

$$e_c(t_{N_{\max}}) = \min_{i=\{1,\cdots,M\}} e_i(t), \tag{3.28}$$

where $N_{\max}$ is defined as $\max\limits_{i=\{1,\cdots,M\}} \{N_i\}$. Then the couple $(\hat{\Theta}_c^t, d_c^t)$ is the optimal result of identification and it is unique.

If the considered c is not unique, suppose there are two index $c1$ and $c2$ such that

$$e_{c1}(t_{N_{\max}}) = e_{c2}(t_{N_{\max}}) = \min_{i=\{1,\cdots,M\}} e_i(t). \tag{3.29}$$

According to the Lemma 3.3.7, for the estimated output $\hat{y}_{c1}(t)$ and $\hat{y}_{c2}(t)$, there exists $t_1 \geq t_0$, such that $\hat{y}_{c1}(t_1) \neq \hat{y}_{c2}(t_2)$. Suppose $e_{c1}(t_1) > e_{c2}(t_1)$, then the couple $(\hat{\Theta}_{c2}^t, d_{c2}^t)$ can be seen as the optimal solution of the identification and it is the unique solution of the identification.

In the whole, the Least Square estimator is (3.25) globally identifiable. $\sharp$

Here the Theorem 3.3.6 only points that the LS estimator is globally identifiable for the time varying linear model with time delay. But since its time varying property and the unknown parameter is include the time delay, it is hardly to determine a fixed optimal sample number to make the estimation.

3.3.4 Iterative LS Method

Now, turn to the system identification for model (3.6). Assume that the conditions of Theorem 3.3.6 are satisfied, then LS estimator could be applied to make system identification. Moreover, the thesis extends this method to an iterative one, called as iterative LS method. This iterative LS method to make the identification of the model (3.6) is summarized as follows.

Consider N is the number of latest samples of the output and input which is the length of the moving windows for each estimation step, and define $\Upsilon^t \triangleq [\alpha^t \;\; \beta^t]^T$. From (3.6), based on N couples of input and output signals, a mixed integer optimization problem can be defined to solve the problem of system identification as:

$$\min_{\substack{l:\ positive\ integer \\ \Upsilon \in \Omega}} \| B_N - A_N(l)\Upsilon \|_2^2, \tag{3.30}$$

where B_N is a stack of the measured outputs

$$B_N \triangleq [y(k)\; y(k+1)\; \cdots\; y(k+N-1)]^T. \tag{3.31}$$

$A_N(l)$ is a stack of measured inputs and outputs, depending on the delay parameter l,

$$A_N(l) \triangleq \begin{pmatrix} y(k-1) & u(k-l-1) \\ y(k-2) & u(k-l-2) \\ \vdots & \vdots \\ y(k-N) & u(k-l-N) \end{pmatrix}. \tag{3.32}$$

Ω represents the possible range of Υ, which is determined by the system gain K^t and time constant T_p^t in the (3.4).

The optimization (3.30) sometimes leads to a non-convex nonlinear programming problem. However, the Branch-Bound (BB) method combined with Least Square method could still work out the solution in a reasonable efficient way, with respect to some potential pre-knowledge of the system, such as the possible range of the time delay. Thereby, a procedure using BB and LS method is applied here if we can have some way to determine $l_{\min} \leq l \leq l_{\max}$.

Then, the Iterative identification based on the LS method can be performed if N has been already chosen.

- In the each step, first construct a loop starting from $l_{\min}$ and ending at $l_{\max}$ by taking the increment of l as 1 at each step.

- For each iteration of l, based on the latest N couples of input and output obtain the LS solution $\Upsilon(l)$ to optimization (3.30) by using the specific number of l, and record the corresponding prediction error, where

$$\Upsilon(l) = (A_N^T(l)A_N(l))^{-1}A_N^T(l)B_N. \tag{3.33}$$

- The pair of (Υ^{t*}, l^{t*}) which leads to the minimal prediction error among all iterations in all steps with regards to l moving from $l_{\min}$ to $l_{\max}$, is the optimal candidate for (3.30) based on the corresponding N sample couples.

- The parameters in the original system (3.3) with (3.4), T_p^t and K^t, can be obtained from $\Upsilon^{t*} = [\alpha^{t*} \quad \beta^{t*}]^T$ as $T_p^{t*} = -\frac{T_s}{\ln \alpha^{t*}}$ and $K^{t*} = \frac{\beta^{t*}}{1-\alpha^{t*}}$. Time delay T_d^t is estimated as $l^{t*}T_s$.

- Repeat the former four steps when a new couple data of input and output is obtained.

One couple of input and output will lead to one parameters identification. Then, the on-line system identification can be performed according to the previously proposed scheme.

In order to release the computation load, the former procedure can be improved from iterative LS method to recursive LS method in order to solve the optimization (3.30). Following the same procedure, only (3.33) changes to the recursive format at the kth step and fixed l. Define $\varphi_k = [y(k)u(k-l)]^T$, $\phi_k = [\varphi(k)\varphi(k-N)]$ then the recursive LS solution could be:

$$\begin{aligned} \Upsilon_k(l) &= \Upsilon_{k-1}(l) + P_k(l)\varphi_k[y(k) - \varphi_k^T\Upsilon_{k-1}(l)] \\ P_k(l) &= P_{k-1}(l) - P_{k-1}(l)\phi_k[I + \phi_k^T P_{k-1}(l)\phi_k]^{-1}\phi_k^T P_{k-1}(l). \end{aligned} \tag{3.34}$$

3.4 Simulation

In the following simulation tests, two different scenarios are conducted: Time Invariant System and Time Varying System. Noted that in order to show the merits of this method, the method defined in Matlab system identification toolbox is applied to compare with our method. The matlab toolbox (MT) method to estimate the process model (3.3) with (3.4) is based on the prediction error method (PEM), see (100; 101) for details. In the thesis, the on-line system identification can be performed using matlab function 'pem' based on the recent N samples data.

3.4.1 Case A: Time Invariant System Test

In this part, Time Invariant System is considered. The system considered is described as:

$$y(s) = G(s)u(s), \tag{3.35}$$

with transfer function

$$G(s) = \frac{K}{T_p s + 1} \exp^{-T_d s}. \tag{3.36}$$

Here $G(s)$, K, T_p, T_d, which do not have the superscript t, mean that they do not change with time in the system running.

In the test, parameters of the system are set as $T_d = 2.05$, $T_p = 2$ and $K = 4$. Fig. 3.1 displays the input and output signals. In order to show the performance of the iterative LS method and MT method, two different sample numbers N are adopted. Note that the iterative LS method need to wait until more than $N + l_{\max}$ samples obtained to collect the enough data. It means the identification procedure will be started after $T_s(N + l_{\max})$. Here the time delay is refined in time interval from 0 to 5 seconds.

Firstly, the sample interval is chosen as $T_s = 0.1 second$.

- **50 Samples Estimation:** A moving window with 50 samples is used to make the estimation. The results could be seen in the following figures, Fig. 3.2 and Fig. 3.3 for LS method, Fig. 3.4 and Fig. 3.5 for MT method.

- **100 Samples Estimation:** In the second test, the length of sample window is changed to 100. The results could be seen in the following figures, Fig. 3.6 and

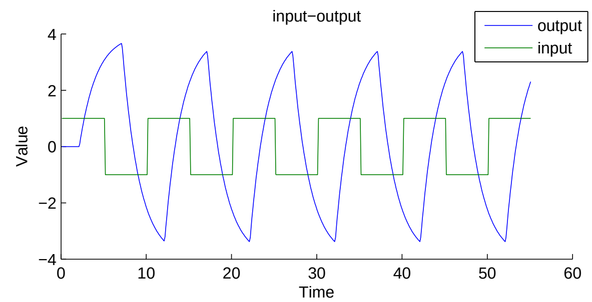

Figure 3.1: The input and output data for case A

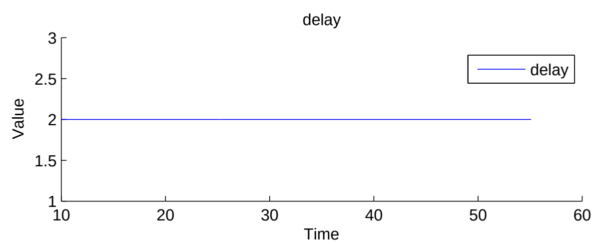

Figure 3.2: The delay estimation using LS for 50 samples in case A

Fig. 3.7 for LS method. For MT method, the results are quite good. The time delay T_d is estimated as 1.954, while K is 4.0 and T_p is 2. The value has little change in the whole procedure.

Secondly, in order to show the influence of the sample interval to the parameters identification, under the same condition of the above 100 samples estimation, only T_s is changed to 0.25 to make another test. The results could be seen in Fig. 3.8, Fig. 3.9 for LS method. The MT method approximated that T_d is 1.7727, K is 4.0 and T_p is 2.0.

The computation times of the numerical tests for the two methods are listed in Table 3.1. The whole procedure for the system identification is running in the simulation under the same computation condition.

From the tests, the following discussion could be made:

- **Time delay:** Only regarding time delay (with enough samples and small sample interval), both small delay and large delay are checked in the test, but large delay case is not listed here. For the small time delay, if the time delay is near the sample time, the result of iterative LS method is more accurate than MT method. Otherwise, MT method is better. It is because in LS method the time

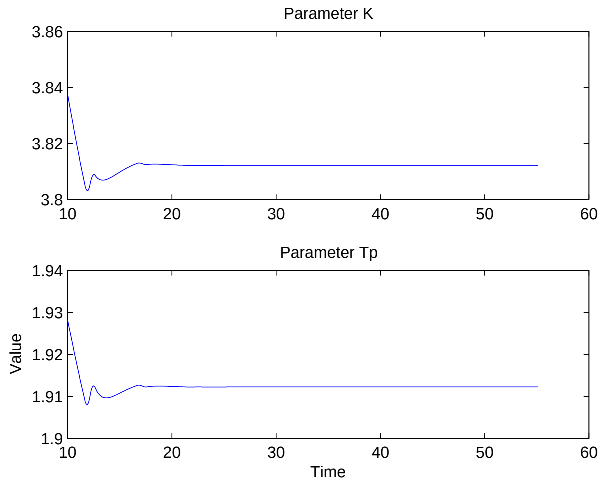

Figure 3.3: The parameters identification using LS for K and T_p for 50 samples in case A

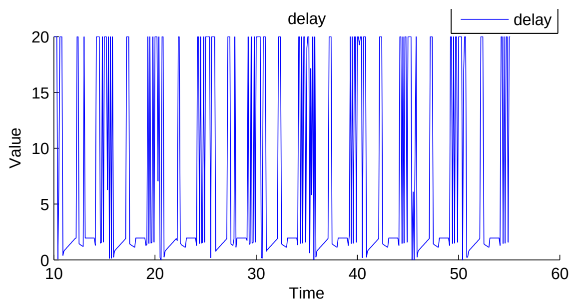

Figure 3.4: The delay estimation using MT for 50 samples in case A

Table 3.1: The computation times for the simulation (second)

	LS Method	**MT Method**
Condition	CUP–T2300, RAM–1GB, software–matlab 7.6.0	
50 Samples	0.531928 seconds	1600.054329 seconds
100 Samples	0.548145 seconds	1800.222629 seconds

delay is estimated by the sampled input delay in the discretization of the system, while the MT method just applies moving covariance to estimate the time delay directly. For the relatively larger time delay, LS method is much better than MT

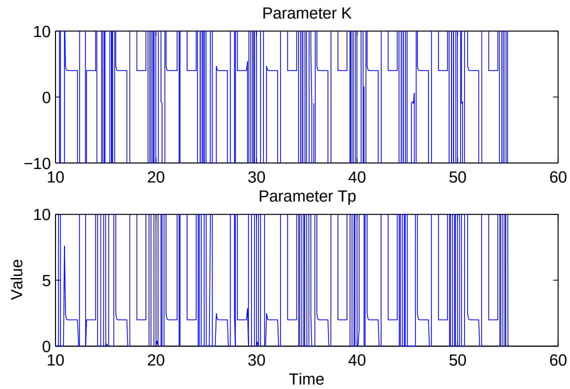

Figure 3.5: The parameters identification using MT for K and T_p for 50 samples case A

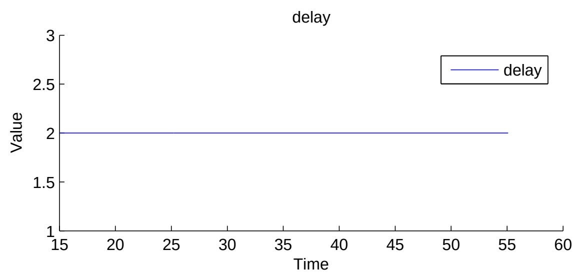

Figure 3.6: The delay estimation using LS for 100 samples in case A

method. The numerical simulation shows that if the time delay exceeds the 40 samples, MT method will return a warning and the result will be worse. But LS method does not have this problem. It could deal with all time delay estimation, only the performances could have a little difference.

- **Parameters identification:** Under the good condition–not too large time delay, enough samples and small sample interval, both two methods could make the parameters identification and show good performance. LS method has some fluctuations at first, then tends to a fixed value that only has a small deviation to the true value. The error is below 5%. MT method is much better than LS method. The estimated value using MT method only remain one fixed value that is quite close to the real value. This is because LS method first estimates the parameters of the discrete version of the system and then converts to the real parameters. No doubt it will decrease the accuracy of the original parameters

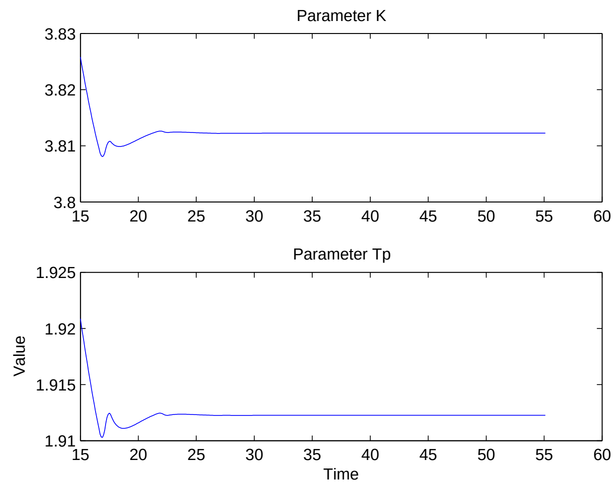

Figure 3.7: The parameters identification using LS for K and T_p for 100 samples in case A

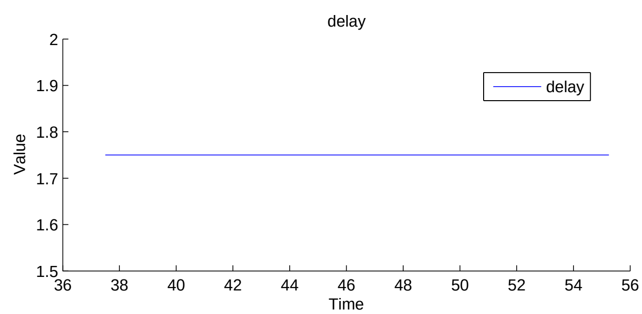

Figure 3.8: The delay estimation using LS for 100 samples with $T_s = 0.25$ in case A

estimation.

- **Sample interval:** The numerical simulations apply different sample interval times in order to indicate the sample interval could affect the performance of the estimations. From the simulation, estimation using small sample interval will be more accurate than using large sample interval. From the above tests, using large sample interval could lead to a relatively larger deviation to the real time delay for MT method. Even for some other larger sample intervals, results of the parameters estimation using MT method are not so good. But the choice of sample interval has less influence to the iterative LS method.

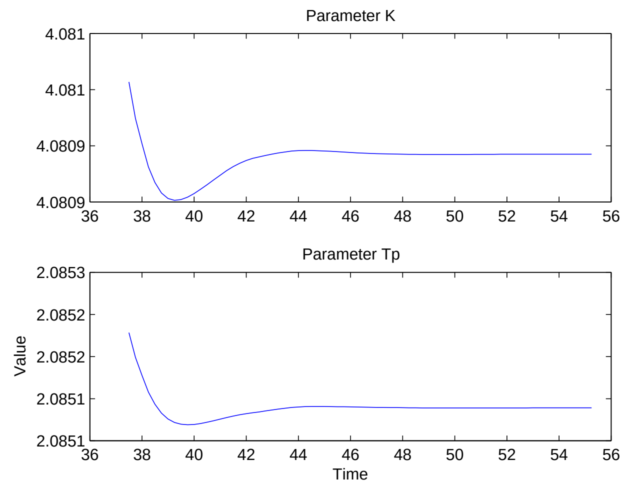

Figure 3.9: The parameters identification using LS for K and T_p for 100 samples with $T_s = 0.25$ in case A

- **Moving window:** The length of data window used to make the estimation could affect the results. The more data used, the more accurate the estimation is. But for MT method, too few data could lead to an unsatisfied estimation more obviously. It need much more data than the LS method to complete the estimation.

- **Computation load:** From Table 3.1 which shows the time two methods used, it can be observed that LS method need much less computation time than MT method.

From the test in case A, the iterative LS method has a good robustness to different conditions of system identification, such as sample interval, length of data and so on, which sometimes could affect the performance of MT method greatly. But, we also observed that MT method gives us a better estimation of the parameters under the good condition at the cost of a little more computation load.

3.4.2 Case B: Time Varying System Test

In this part, the time varying system is considered in the test. The system is considered as (3.3) with (3.4) as well, where the parameters of the system are as follows:

$$K^t = 3, T_p^t = 1, T_d^t = 3.05, \qquad \textit{when running time } t < 30; \tag{3.37}$$

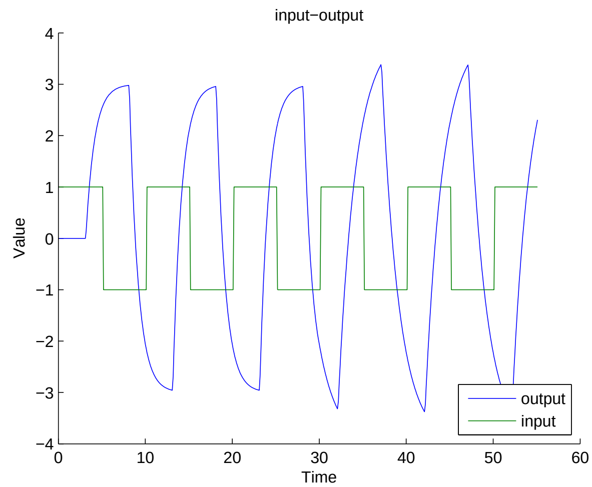

Figure 3.10: The input and output data for case B

$$K^t = 4, T_p^t = 2, T_d^t = 2.05, \qquad \textit{when running time } t \geq 30. \tag{3.38}$$

Actually, this system is a kind of switching system. The iterative LS method is still used to make the identification. The data comes from the simulation using the simulink of such a system. The input and output could be seen in Fig. 3.10. In this test, two different sample numbers are considered.

The sample interval is chosen as $T_s = 0.1$. The time delay is assumed in the range of 5 seconds.

- **50 Samples Estimation:** A moving window with 50 samples is used to make the estimation. The procedure begins at 10th second. The results could be seen in the following figures, Fig. 3.11 and Fig. 3.12 for LS method. The result of the MT method is not listed since 50 samples are not enough to make the estimation, as a result, the performance of the estimatioin is quite bad.

- **100 Samples Estimation:** In the second test, the length of sample window is changed to 100. The procedure begins at 15th second. The results could be seen in the following figures, Fig. 3.13 and Fig. 3.14 for LS method. For MT method, see Fig. 3.15 and Fig. 3.16.

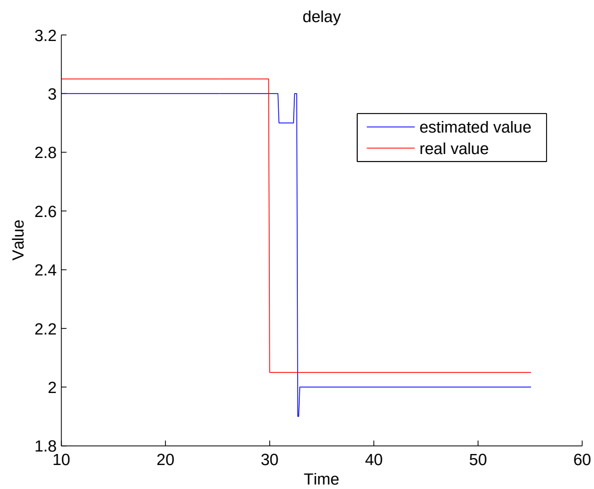

Figure 3.11: The delay estimation using LS for 50 samples with in case B

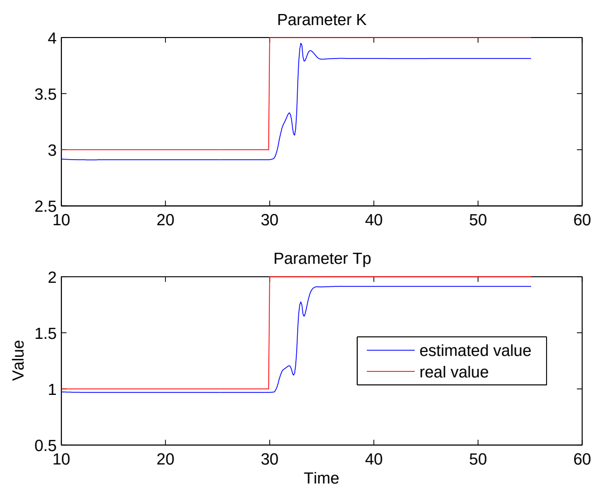

Figure 3.12: The parameters identification using LS for K and T_p for 50 samples in case B

According to the tests, the LS method showed good performance regarding to the precision for both of different sample points. In the different tests, the estimated value is stable when the procedure begins. When the system has a switching, the LS method need some delay to react to this switching. This delay is less than half length of windows. Then the estimated value will bias from the original stable value and some fluctuation emerge. In a while less than one length of windows time, the estimated

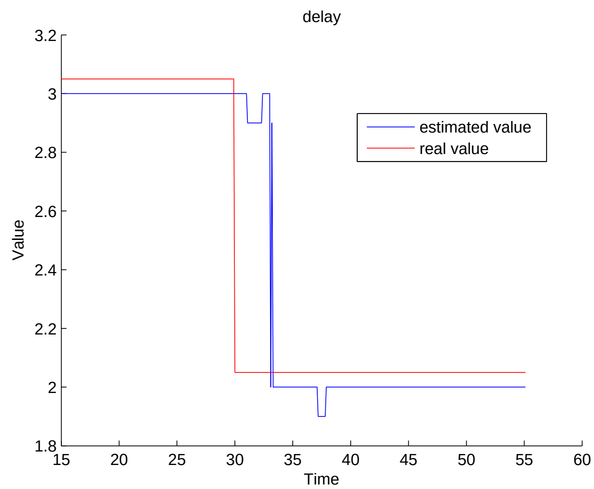

Figure 3.13: The delay estimation using LS for 100 samples with in case B

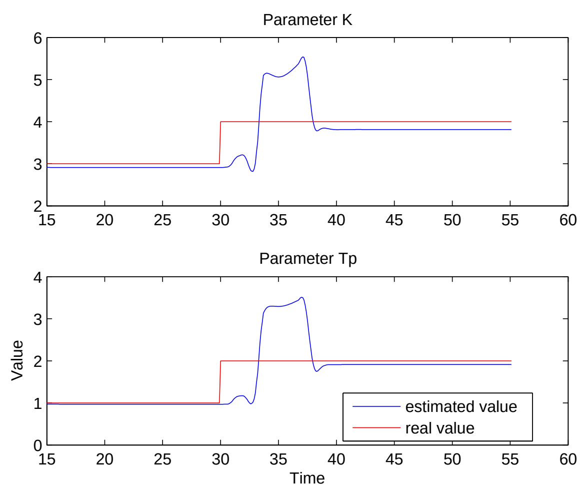

Figure 3.14: The parameters identification using LS for K and T_p for 100 samples in case B

value will recover to another stable value. It is obvious that the more sample points used, the more delay is. And since the inevitable error due to the discretization of the model, the estimated value has a little difference to the true value of the parameters, about one time interval for the dead time estimation and corresponding error for the other parameters identification. This can be reduced by decreasing the sampling interval. It can be seen for the two tests with different sampling points, 50 and 100 samples,

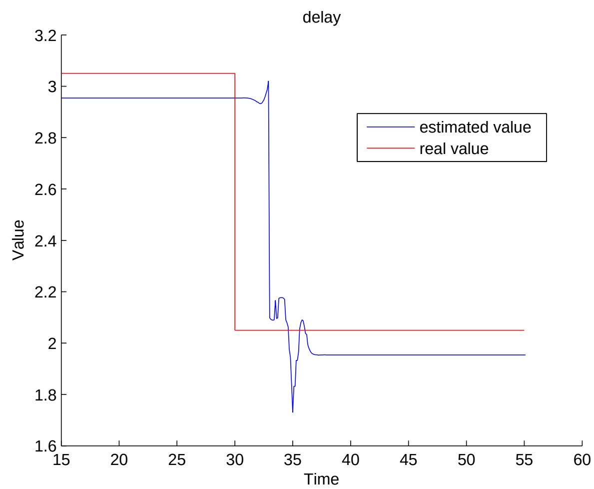

Figure 3.15: The delay estimation using MT for 100 samples with in case B

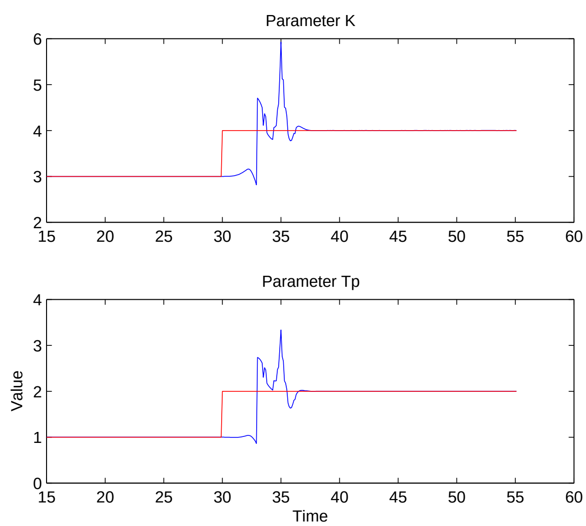

Figure 3.16: The parameters identification using MT for K and T_p for 100 samples in case B

the accuracy is nearly the same. But for MT method, it can not make the parameter identification using 50 points, the matlab returned not enough data alarm. Compared it with LS method for the 100 samples estimation, the MT method has the error for the time delay estimation as well. But the results for the other parameters are really better than LS method. And during the period when the system has a switching, the

MT method also has the fluctuation in the estimation and it is really severe than the LS method. Moreover, during a number of simulations, it has been observed that if the system time delay is more than 40 samples time, the MT method often returned a warning and the estimation result often really bad. But the LS method does not have this kind of problem. In general, it can be concluded that the proposed LS method is quite promising for TV-FOPDT model identification in terms of accuracy and flexibility.

3.5 System Identification for TV-FOPDT model with Input Depended Dead Time

In this part, the TV-FOPDT model adopted is the one with input depended dead time. Furthermore, the measurement of the system output is added with a Gaussian noise.

3.5.1 TV-FOPDT Model with Input Depended Dead Time

The system described in a Time-Varying FOPDT (TV-FOPDT) model with input depended dead time is defined in the following.

$$Y(s) = G^{u(t),t}(s)U(s), \tag{3.39}$$

with transfer function

$$G^{u(t),t}(s) = \frac{K^t}{T_p^t s + 1} \exp^{-T_d^{u(t),t} s}. \tag{3.40}$$

And the measurement is

$$x(s) = y(s) + \omega(s). \tag{3.41}$$

where $Y(s)/U(s)$ is the Laplace-transform of the system output/input $y(t)/u(t)$. K^t, T_p^t and $T_d^{u(t),t}$ are the system gain, time constant, and time delay (dead-time), respectively. Different with the standard FOPDT model, all these system parameters can be time-dependent, especially the time delay can also depend on the input signal. This dependence feature is represented by the corresponding subscript. $x(s)$ is the measured output of the system and $\omega(s)$ is the noise in the output measurement, which is assumed as a Gaussion process with 0 mean and variance Q.

The same problem of the system identification problem is considered as well. The only difference is the output $x(s)$, which is based on the measurement model (3.41).

Following the similar procedure as in the last section, the model (3.39) can be approximated by its discrete-time equivalence, i.e.,

$$Y(z) = G^{u(k),k}(z)U(z), \tag{3.42}$$

with

$$G^{u(k),k}(z) = \frac{K^k(1-\alpha^k)}{z^{l^{u(k),k}}(z-\alpha^k)}.$$

Here $\alpha^k \triangleq \exp^{-\frac{T_s}{T_p^k}}$, and T_s is the sample interval. It should be noticed that K^k and T_p^k are not the same as K^t and T_p^t in (3.40). The latter two are piecewise-constant (constant during every sampling period) timed functions, while the former two in (3.42) are sampled sequences. The relationship of these two description is that K^k is equal to K^t, T_p^k is equal to T_p^t at each sampling time, i.e., $K^k = K^t$ and $T_p^k = T_p^t$ when $t = kT_s$ for any k. Thereby, K^k, T_p^t are called as the kth sampled (time-varying) system gain, the kth sampled (time-varying) time constant (**?**). Here $l^{u(k),k}$ is the discrete approximation of the kth sampled system delay $T_d^{u(k),k}$ (the kth sampled (time-varying) time delay, $T_d^{u(k),k} = T_d^{u(t),t}$ when $t = kT_s$ for any k), and it is defined as an integer with the property:

$$l^{u(k),k}T_s \leq T_d^{u(k),k} \leq (l^{u(k),k}+1)T_s \tag{3.43}$$

Define $\beta^k \triangleq K^k(1-\alpha^k)$, then TV-FOPDT model with input depended dead time (3.6) can be transferred into a difference equation model described as

$$y(k) = \alpha^k y(k-1) + \beta^k u(k - l^{u(k),k} - 1), \tag{3.44}$$

for $k = l^{u(k),k}+1, l^{u(k),k}+2, \cdots \infty$.

The output measurement is not changed, but the measured output signal can only be obtained at each sampled time:

$$x(k) = y(k) + \omega(k). \tag{3.45}$$

Then, the original continuous-time model identification problem of (3.39) with parameters K^t, T_p^t and $T_d^{u(t),t}$ is converted to estimate parameter sequences of α^k, β^k and $l^{u(k),k}$ for a stochastic discrete-time system (3.44) based on a number of sampled input and measured output obtained by (3.45). This random discrete-time system identification problem is called the discreteized approximation of the original continuous-time identification problem.

3.5.2 Iterative LMSP method

Since the measured output is added with noise, the previous LS estimator need to be extended in probability meaning. Furthermore, in order to make the estimation more accurate at each iterative step, a forgetting factor is added in the proposed algorithm.

The method proposed in the following, is named as Least Mean Square Prediction (LMSP) identification method, in order to handle this system identification problem for the TV-FOPDT model with input depended dead time.

Suppose that the considered system (3.39) is running at kth sampling step and take N as the number of latest samples of the measured output and input into consideration, where N is the length of the moving data window used in each estimation step. Define $\theta^k = [\alpha^k \;\; \beta^k]^T$, then the parameters identification of the system (3.44) at the kth sampling step can be formulated as a Stochastic Mixed Integer Non-Linear Programming (SMINLP) problem, which is defined as:

$$\min_{\substack{l^{u(k),k}:\ positive\ integer \\ \theta^k \in \Theta^k}} \mathbf{E}\{\| B_N^k - A_N^k(l^{u(k),k})\theta^k \|_2^2\}, \tag{3.46}$$

where B_N^k is a stack of N latest measured outputs with forgetting factor at the current kth sampling step, i.e.,

$$B_N^k \triangleq [x(k) \; \rho x(k-1) \; \cdots \; \rho^{N-2}x(k-N+2) \; \rho^{N-1}x(k-N+1)]^T. \tag{3.47}$$

$A_N^k(l^{u(k),k})$ is a stack of N inputs and measured output with forgetting factor at the current kth sampling step which can generate B_N^k, depending on the delay parameter $l^{u(k),k}$, i.e.,

$$A_N^k(l^{u(k),k}) \triangleq \begin{pmatrix} x(k-1) & u(k-l^{u(k),k}-1) \\ \rho x(k-2) & \rho u(k-l^{u(k),k}-2) \\ \vdots & \vdots \\ \rho^{N-2}x(k-N+1) & \rho^{N-2}u(k-l^{u(k),k}-N+1) \\ \rho^{N-1}x(k-N) & \rho^{N-1}u(k-l^{u(k),k}-N) \end{pmatrix}. \tag{3.48}$$

Θ^k represents the possible range of θ^k, which is determined by the limits of the original system gain K^t and time constant T_p^t in (3.40) at the current sampling time kT_s. Here ρ is called as forgetting factor, which is used in order to decrease the effect of old data to the estimation at the current sampling time.

If the system runs with no time delay, or the time delay is the prior knowledge, the optimization problem (3.46) can be simplified to a problem of the minimization of the Mean Squared Error (MSE). In general, this SMINLP problem (3.46) may lead to some non-convex issue due to the unknown time delay $l^{u(k),k}$. But if some pre-knowledge about time delay in system can be obtained, such as the potential upper and lower limits of the time delay(s) for the entire system or each sampling step, an iterative numerical algorithm can be performed by combining the BB method, which is one typical method for MINLP problem, and the LMS technique for efficiently solving this SMINLP problem. The algorithm is called as an iterative LMS algorithm, which is summarized in the following:

- *Preparation:* The upper and lower limits for system time delay(s) in terms of some integer number multiplying with sampling period need to be given. Without losing generality, we assume that $l_{\min}^{u(k),k} \leq l^{u(k),k} \leq l_{\max}^{u(k),k}$ and $l_{\min}^{u(k),k}$, $l_{\max}^{u(k),k}$ are known before the procedure. The sampling rate T_s, sliding window length N and forgetting factor ρ need to be decided before the procedure.

- *Data collection period:* In the beginning, the algorithm only collects the sampled data until the process reaches a specific sampling step, denoted this step as k_{ini}, where $N + l_{\max}^{u(k_{ini}),k_{ini}} = k_{ini}$. It is to guarantee that there is enough data to construct matrix (3.47) and (3.48).

- *Iteration period:* The iterative identification starts from the k_{ini} step in an on-line manner. k is denoted as the sampling step and there is $k \geq k_{ini}$, a computing loop is constructed with regard to $l^{u(k),k}$ starting from $l_{\min}^{u(k),k}$ and ending at $l_{\max}^{u(k),k}$ by taking the unit increment.

  - For each iteration (k) of $l^{u(k),k}$ ($l_{\min}^{u(k),k} \leq l^{u(k),k} \leq l_{\max}^{u(k),k}$), solve the LMS problem (3.46) and record the corresponding prediction error. The LMS method with forgetting factor is adopted in this paper. The analytical solution has the format as:

$$\begin{aligned} \hat{\theta}^k(l^{u(k),k}) = ((A_N^k(l^{u(k),k}))^T Q^{-1} A_N^k(l^{u(k),k}))^{-1} \\ (A_N^k)^T(l^{u(k),k}) Q^{-1} B_N^k, \\ Cov(\hat{\theta}^k) = ((A_N^k(l^{u(k),k}))^T Q^{-1} A_N^k(l^{u(k),k}))^{-1}, \end{aligned} \tag{3.49}$$

    where $Cov(\hat{\theta}^k)$ means the covariance of $\hat{\theta}^k$.

- $(\hat{\theta}^k(\hat{l}^{u(k),k}), \hat{l}^{u(k),k})$ which leads to the minimal prediction error among all iterations moving from $l_{\min}^{u(k),k}$ to $l_{\min}^{u(k),k}$, denoted as $(\theta^{k*}(l^{u(k*),k*}), l^{u(k*),k*})$, is chosen as the optimal solution for (3.46) at the current step.
- The estimation of the kth sampled system parameters of (3.6) for the current sample, T_p^k and K^k, can be obtained from $\theta^{k*}(l^{u(k*),k*}) = [\alpha^{k*} \;\; \beta^{k*}]^T$ by

$$T_p^k = -\frac{T_s}{\ln \alpha^{k*}} \;\; and \;\; K^k = \frac{\beta^{k*}}{1-\alpha^{k*}}, \tag{3.50}$$

  and the sampled time delay $T_d^{u(k),k}$ is estimated as $l^{u(k*),k*}T_s$.
- Repeat the above steps when a (couple) new data of input and measured output is obtained.

According to the above procedure, the system identification for TV-FOPDT model with input depended dead time can be executed in an on-line manner. Note that in this system identification, only the result of the parameters estimation is focused on, so the covariance of estimated parameter which can be calculated by the second part of (3.49) is not recorded.

The previous method applies LS to make the system identification. The requirement of the LS is that the measurement noise should be uncorrelated with the system variable. Under this condition, the LS estimator is unbiased and consistent (151). However, in many cases, the measurement noise and some system variable are unmeasured, causal variables collapsed into the noise term are correlated, then the LS estimator is generally biased and inconsistent (151).

For this reason, the LS estimator in the algorithm need to be revised as Instrumental Variable (IV) methods, which is the generalization of the LS estimate. The main idea of the IV method is to modify the LS method so that it can be one consistent estimator for an arbitrary noises. Accordingly, the IV method modifies the former (3.49) as

$$\begin{aligned} \hat{\theta}^k(l^{u(k),k}) &= (Z_k^T Q^{-1} Z_k)^{-1} Z_k^T Q^{-1} B_N^k, \\ Cov(\hat{\theta}^k) &= (Z_k^T Q^{-1} Z_k)^{-1}, \end{aligned} \tag{3.51}$$

where Z_k is the chosen instrumental variable, which is correlated with the system variables and uncorrelated with noises. The major problem with the IV approach is the generation of the instrumental variables. The basic idea is that by pre-filtering the deterministic input, and it is possible to generate an IV vector Z_k, which is highly correlated with the noise-free process vector. In addition, it will be uncorrelated with any other noise in the system provided the input command is noise-free (151).

This IV method shows it is a consistent and unbias estimator for the system with arbitrary disturbance or noises. But in order to get more efficiency, this IV method should be adopted its recursive manner. Define $\varphi_k = [y(k)u(k-l)]^T$, $\phi_k = [\varphi(k)\varphi(k-N)]$, the recursive IV procedure can be summarized as following:

$$\begin{aligned} \hat{\theta}^k(l^{u(k),k}) &= \hat{\theta}^{k-1}(l^{u(k),k}) + P_k(l)Z_k[y(k) - \varphi_k^T \hat{\theta}^{k-1}(l^{u(k),k})] \\ P_k(l) &= P_{k-1}(l) - P_{k-1}(l)Z_k[I + \phi_k^T P_{k-1}(l)Z_k]^{-1}\phi_k^T P_{k-1}(l). \end{aligned} \tag{3.52}$$

3.5.3 Numerical Examples

A number of numerical simulations are applied to make the test of the proposed system identification method for TV-FOPDT model with input depended time delay.

The system considered is a switching FOPDT model with input depended time delay. The time delay of the system is dependent on the input signal $u(t)$ in the manner that $T_d^{u(t),t} = 0.5u(t)$. Other parameters are set as:

$$\begin{cases} T_p^t = 1,\ K^t = 3, & when\ \ t < 30\ seconds; \\ T_p^t = 2,\ K^t = 4, & when\ \ t \geq 30\ seconds. \end{cases}$$

Here t is the system running time. It means that the system has a switching at 30th second. The noise in the measurement of the output follows the distribution $\mathcal{N}(0,\ 0.001)$.

The test condition in the first case is set as $T_s = 0.1$ second, the sample number for the estimation $N = 40$, forgetting factor $\rho = 0.95$ and the pre-knowledge of the sampled time delay is assumed as $l_{\max}^{u(k),k} = 30$, $l_{\min}^{u(k),k} = 0$. According to the proposed method, it need to wait more than 7 seconds (the data collection period is $(40+30)\times 0.1 = 7$ seconds) to start the identification procedure in the beginning. In the test, the identification begins at 100th sampling time. The system is simulated in the simulink with the step input signal. Fig. 3.17 shows the input signal and measured output signal.

Fig. 3.18 and Fig. 3.19 display the results of the system identification. And in order to investigate the relation between the input and time delay, the rate between the time

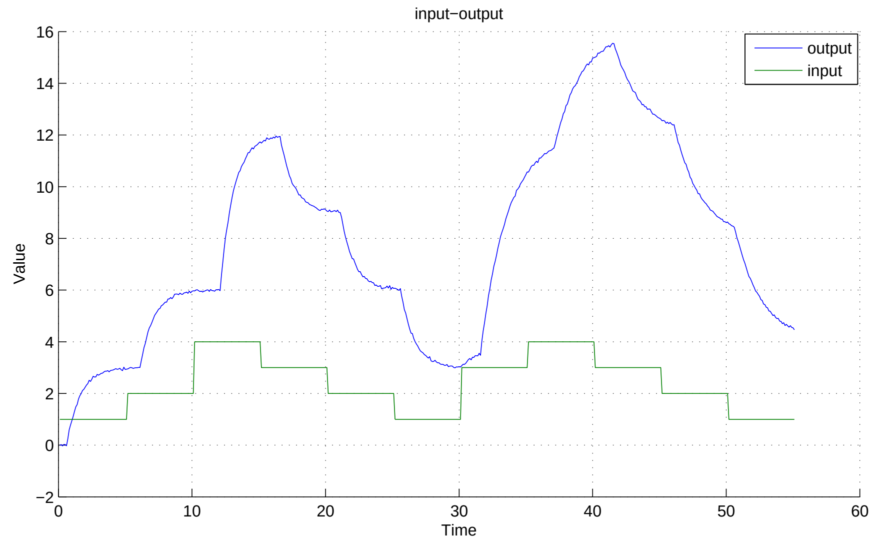

Figure 3.17: The input and output data for the first test

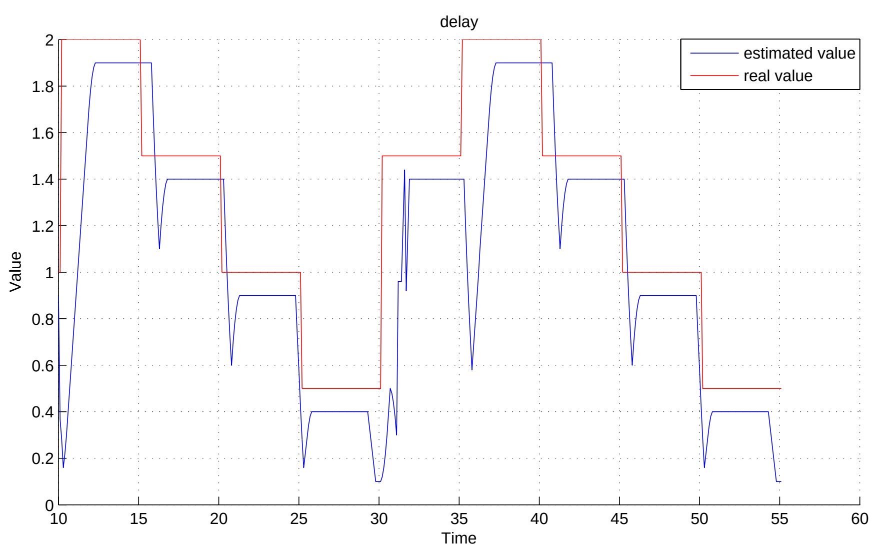

Figure 3.18: The time delay estimation for the first test

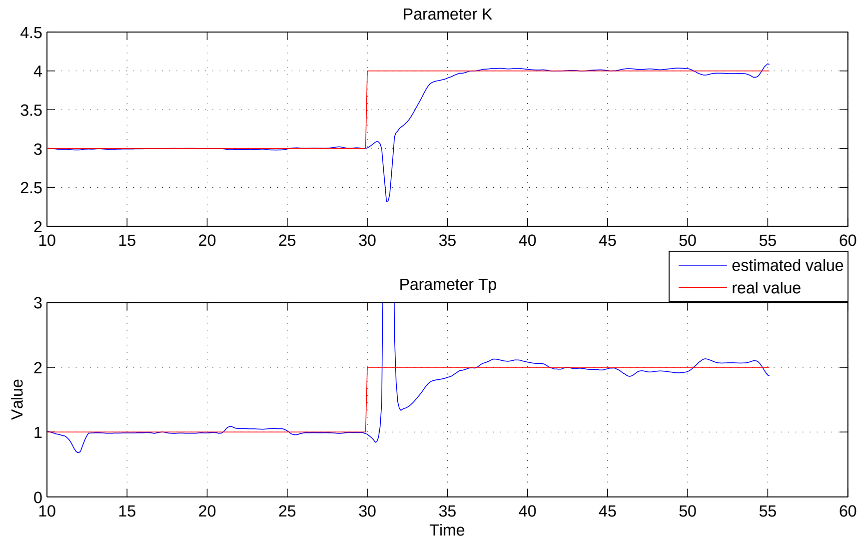

Figure 3.19: The parameters identification for the first test

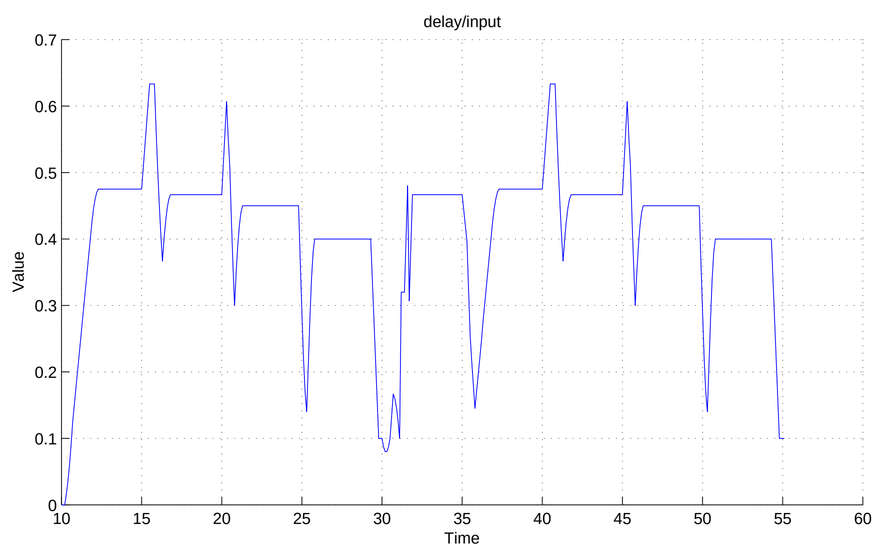

Figure 3.20: The estimated delay to input relationship for the first test

delay and input is calculated at each estimation step. From the simulation results, the following discussion could be made:

- **Time delay estimation:** The result of the estimation for time delay is showed in Fig. 3.18. Before the 10th second, it is the data collection period. From 10th second, the estimation procedure begins. It need some time to attain the steady time delay estimation. This time is about 2 seconds, which is less than the time of one moving window length ($40 \times 0.1 = 4$ seconds). When the time delay changed in the system, the estimation need a little time (less than 1 second) to react to this change. Then the steady estimation was disturbed and a peak appeared. After a short time, the estimation value will stabilized to a new value that is quite close to the true time delay again. In each time period when the time delay changes, the same phenomenon will emerge to the estimated value. But at 30th second, the system switched to a completely different system, in which not only time delay changed but also the other parameters changed. Unlike the other time when the time delay changed, before the changing time, the estimated value has been already different with the former steps. And in a period of about 2 seconds, it arises more than 2 peaks before it is back to the steady estimated value. Two different factors, both time delay change and system switching, work together to lead to this estimated value fluctuating more than before. It is observed that except for the time period when the system has a change, only small steady estimation errors (about 0.1 second) to the real time delay can be observed. This small estimation error is due to the fact that this identification solution is determined by T_s (see (3.30)).

- **Parameters identification:** The result of identification to the other two parameters expect time delay could be seen in Fig. 3.19. Regarding the estimation of system gain K^t, unlike the time delay estimation, it does not need time to attain the steady estimated value. From the beginning at 10th second, it showed a quite good performance to this parameter identification. The estimated value is nearly the same to the true value. When K^t changed at 30th second, the estimated value would be away from the original steady estimated value. The estimated value had a large peak before it returns to another new steady value. But the time it needed (about 5 seconds) is much more than the time used in the time delay estimation. At all sample time where the estimation value is steady, the error for the estimation of K^t is below 1%. For another parameter T_p^t, in the beginning, it has a small deviation and quickly returns back to the steady estimation value which is quite close to the real value 1. At the time the system changed, it can be observed a rather large peak appeared (in order to show most estimation result in detail, the value is omitted in the figure). In more than 5 seconds, it recovered to the steady value. But it seems not so steadier than the estimation for the parameter K^t. It is believed that the unavoidable error of the time delay estimation affects much more on this time constant than the system gain.

- **Time delay and input:** In order to show the relation between time delay and input signal, the rate between estimated time delay and measured input signal at each sampling time is calculated, which can be seen in Fig. 3.20. It can be observed that except for the time period the time delay changed, the rate is in the range of 0.4-0.5 at each sample points showed steady estimation. The result approximately shows that how the time delay depended on the input.

In order to show the moving widow length can affect the estimation result, a number of other tests are conducted. In each test, only N is changed with the first test to make the estimation. The results could be seen from Fig. 3.21 to Fig. 3.24.

From these tests, it can be observed:

- **Sample number:** From Fig. 3.21 to Fig. 3.24, different sample numbers are adopted. According to the results, estimation using 50 samples could provide the smoothest results at the cost of the delay to detect the parameters change in the system. But the result using 30 samples not only is less smooth, but also had an obvious decreasing in the accuracy.

Conclusion from above simulation tests:

From a number of the simulations, for a TV-FOPDT modeled system, in which the time delay depended on the input and other parameters of the system have a sudden change at some time, the proposed method, using SMINLP programming based on the BB and LMS method, can provide a reasonably accurate and prompt estimation for the time delay and parameters. The choosing of length of moving widow could affect

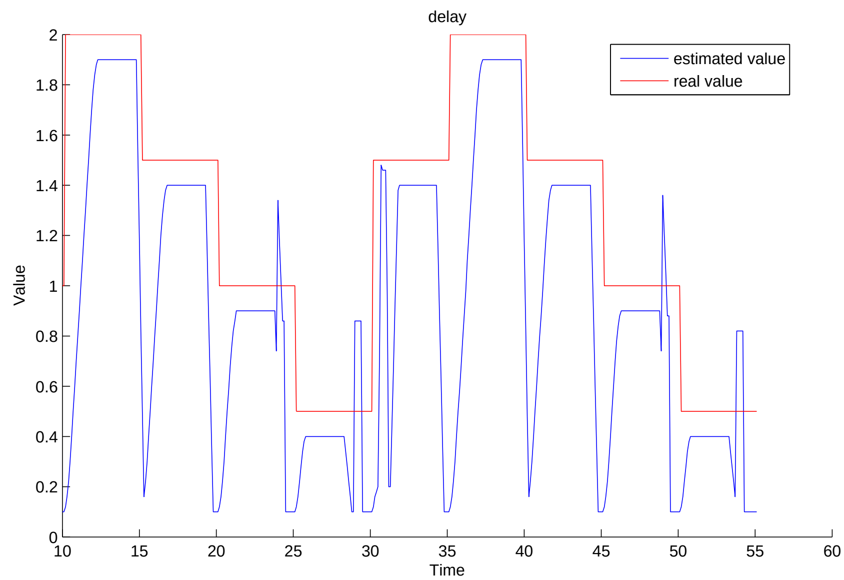

Figure 3.21: The time delay estimation based on 30 samples

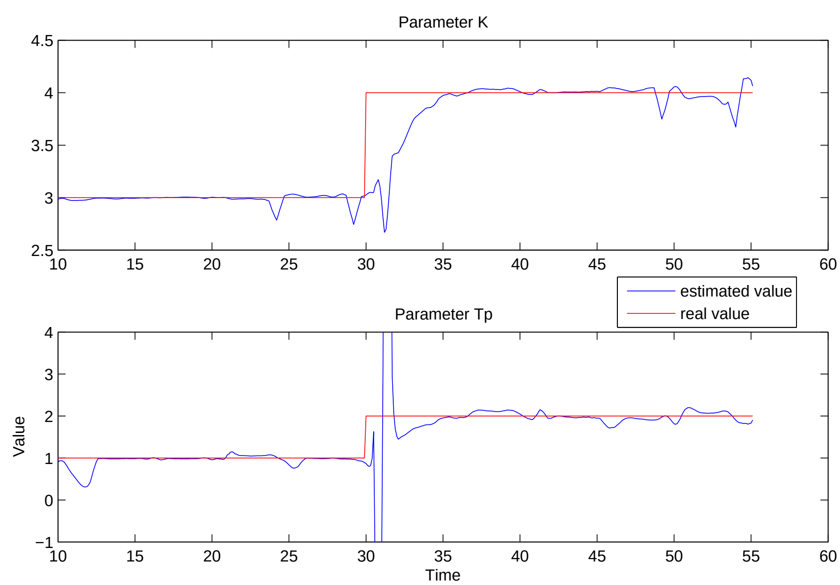

Figure 3.22: The parameters identification based on 30 samples

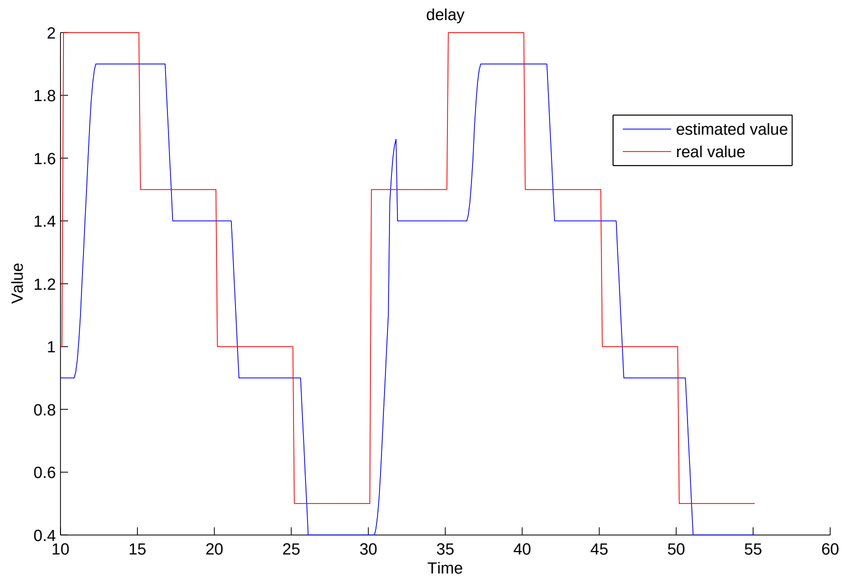

Figure 3.23: The time delay estimation based on 50 samples

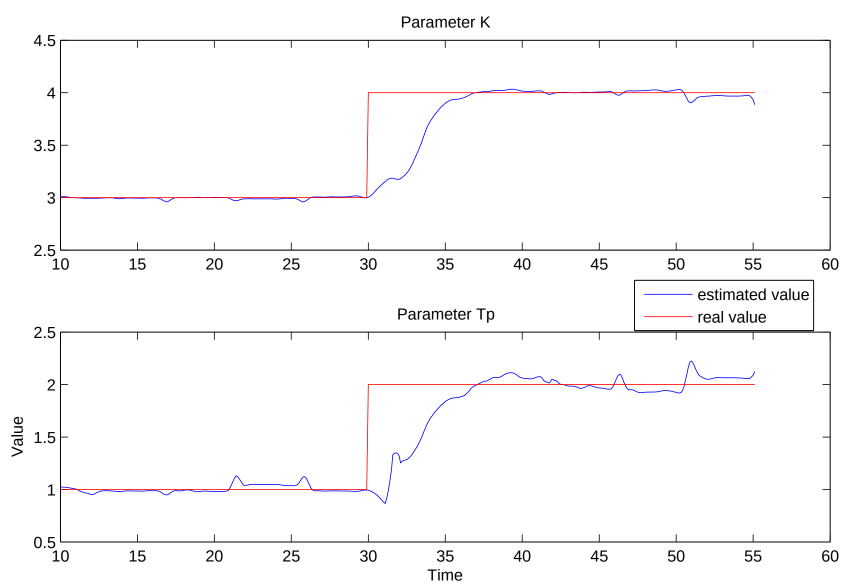

Figure 3.24: The parameters identification based on 50 samples

the performance of the estimation. Less data leads that the estimation can not capture the property of the system and results in an unsatisfied performance. In general, the more samples are used for estimation in each step, the much smoother estimation result could be obtained. But too many sample points can decrease the estimation to reflect time varying property of the parameters. It need to look for a balanced sample number in the estimation.

3.6 Multi-Input FOPDT Identification

In the previous sections, a Time-Varying FOPDT (TV-FOPDT) model even with input dependent dead time, is proposed. All of the above mentioned models and the corresponding identification methods are only suitable for SISO system situation.

But from the application point of view, many systems may be affected by more than one issue besides the known input variable, such as some disturbance from the physical mechanics, unknown noise and so on. Bearing it in the mind, the thesis extends the proposed TV-FOPDT methods into MISO case.

3.6.1 MISO TV-FOPDT Model Formulation

An MISO TV-FOPDT model considered here can be defined in the following manner:

$$Y(s) = G_1^t(s)U_1(s) + G_2^t(s)U_2(s), \tag{3.53}$$

with transfer functions

$$G_1^t(s) = \frac{K_1^t}{T_p^t s + 1} e^{-T_d^t s}. \tag{3.54}$$

and

$$G_2^t(s) = \frac{K_2^t}{T_p^t s + 1}. \tag{3.55}$$

The measurement is

$$x(t) = y(t) + \omega(t). \tag{3.56}$$

Here $u_1(t)$ is a known part of input. $u_2(t)$ is an unknown part of input, which is defined as the system's *disturbance*. $y(t)$ is the ”noise free” output, and $X(s)/U_i(s),\ i = 1,2$ is Laplace-transform of the system output/input. $x(t)$ is the noisy system output, and the noise $\omega(t)$ is zero-mean white Gaussian noise with variance Q. T_d^t is the time delay happened in the input, T_p^t is the system time constant, and K_1^t, K_2^t are system gains for different parts of inputs $u_1(t)$ and $u_2(t)$, respectively. The superscript t of variables also mean the time varying feature of the corresponding variables.

It is assumed that the time constants in $G_1^t(s)$ and $G_2^t(s)$ are the same in this consideration, which indicts that both of the two part of input affect the output in the same dynamic manner. It is also assumed that the time delay only affect to the known part of input $u_1(t)$, and here K_2^t is supposed to be known beforehand. All unknown factors relevant to $G_2^t(s)$ was modeled into the unknown part of input $u_2(t)$.

The considered MISO TV-FOPDT identification problem can be also formulated as: to identify system parameters K_1^t, T_p^t and T_d^t, as well as to simultaneously estimate the unknown input $u_2(t)$ based on the sampled data of control input $u_1(t)$ and output $y(t)$, in an on-line optimal manner (153).

3.6.2 Iterative LMS Method

In order to apply the same idea to make the system identification of the MISO TV-FOPDT model, the method proposed in the previous sections need to be extended to the multi-input cases.

As the same procedure, the continuous-time system (3.53) with (3.54) and (3.55) is approximated by its discrete-time equivalence followed by the same technique in the last section, i.e.,

$$Y(z) = G_1^k(z)U_1(z) + G_2^k(z)U_2(z), \tag{3.57}$$

with

$$G_1^k(z) = \frac{K_1^k(1-\alpha^k)}{z^{l^k}(z-\alpha^k)}, \tag{3.58}$$

and

$$G_2^k(z) = \frac{K_2^k(1-\alpha^k)}{z-\alpha^k}, \tag{3.59}$$

where $\alpha^k \triangleq \exp^{-\frac{T_s}{T_p^k}}$, and T_s is the sampled interval. As stated in last section, $\{K_i^k\}_{i=1,2}$ and T_p^k are not the same to $\{K_i^t\}_{i=1,2}$ and T_p^t in (3.54) and (3.55): The former ones are piecewise-constant (constant in each sampling interval) functions, while the latter ones are real timed functions. Their relationships can be described as $\{K_i^k\}_{i=1,2}$ are equal to $\{K_i^t\}_{i=1,2}$ and T_p^k is equal to T_p^t at each sampling time, i.e., $K_i^k = K_i^t,\ i = 1,2$ and $T_p^k = T_p^t$ when $t = kT_s$ for any nonnegative integer k. Hence, we call $\{K_i^k\}_{i=1,2}$, T_p^k as the *kth sampled (time-varying) system gains, the kth sampled (time-varying) time constant*, respectively. The l^k in (3.58) is the discrete approximation of the *kth sampled system delay* T_d^k ($T_d^k = T_d^t$ when $t = kT_s$ for any nonnegative integer k), with the property $T_d^k \approx l^k T_s$ (153).

Define $\beta^k \triangleq K_1^k(1-\alpha^k)$, $\gamma^k \triangleq u_2(k)(1-\alpha^k)$, then model (3.53) with (3.54) and (3.55) can be converted to

$$\begin{cases} y(k) = y_1(k) + y_2(k) \\ y_1(k) = \alpha^k y_1(k-1) + \beta^k u_1(k-l^k-1) \\ y_2(k) = \alpha^k y_2(k-1) + \gamma^k K_2^k \end{cases} \tag{3.60}$$

Make a sum of the last two equations in (3.60)) and use the first equation, the following model can be obtained:

$$y(k) = \alpha^k y(k-1) + \beta^k u_1(k-l^k-1) + \gamma^k K_2^k, \tag{3.61}$$

for $k = l^k+1, l^k+2, \cdots \infty$.

The measured output signal is collected at each sampled time:

$$x(k) = y(k) + \omega(k). \tag{3.62}$$

Take (3.61) and (3.62) together, then there exists

$$x(k) = \alpha^k x(k-1) + \beta^k u_1(k-l^k-1) + \gamma^k K_2^k + \omega^{'}(k), \tag{3.63}$$

for $k = l^k+1, l^k+2, \cdots \infty$. Here $\omega^{'}(k)$ is a new Gaussian noise that $\omega^{'}(k) \triangleq (1-\alpha^k)\omega(k)$.

Then, the original parameter identification problem of continuous-time model is converted to identify the parameters α^k, β^k, γ^k and l^k for a stochastic discrete-time system (3.63) based on a number of sampled input signals and measured outputs.

The considered system identification problem of (3.63) can be formulated as a Stochastic Mixed Integer Nonlinear Programming (SMINP) problem according to the same procedure in previous sections. Then choosing the Bound and Branch strategy (51) to handle the corresponding mixed integer optimization, the LMS method can be applied to cope with each optimal parameter identification under the assumption of boundness of time delays as well.

Assume that system (3.53) is running at kth sampling step and let N be the number of latest sample pairs of the measured output and input used to make the estimation at kth step. This N is also the length of the sliding window used in every estimation step. Define $\theta^k \triangleq [\alpha^k \;\; \beta^k \;\; \gamma^k]^T$, then the identification/esitmation problem at the kth sampling step can be formulated as:

$$\min_{\substack{l^k \in L \\ \theta^k \in \Theta^k}} \quad \mathbf{E}\{\| B_N^k - A_N^k(l^k)\theta^k \|_2^2\}, \tag{3.64}$$

where B_N^k is a vector variable consisting of N latest measured outputs with forgetting factor at the current kth sampling step, i.e.,

$$B_N^k \triangleq [x(k) \;\; \rho x(k-1) \;\; \cdots \;\; \rho^{N-2}x(k-N+2) \;\; \rho^{N-1}x(k-N+1)]^T. \tag{3.65}$$

$A_N^k(l^k)$ is a system matrix which depends on time delay parameter, and is generated using N pairs of input and measured output with a forgetting factor, it can be constructed by:

$$A_N^k(l^k) \triangleq \begin{pmatrix} x(k-1) & u_1(k-l^k-1) & K_2^{k-1} \\ \rho x(k-2) & \rho u_1(k-l^k-2) & \rho K_2^{k-2} \\ \vdots & \vdots & \vdots \\ \rho^{N-2}x(k-N+1) & \rho^{N-2}u_1(k-l^k-N+1) & \rho^{N-2}K_2^{k-N+} \\ \rho^{N-1}x(k-N) & \rho^{N-1}u_1(k-l^k-N) & \rho^{N-1}K_2^{k-N} \end{pmatrix}. \tag{3.66}$$

Θ^k stands for the possible range of θ^k, L means the boundaries of time delay, ρ is a so-called forgetting factor, which is used to decrease the effect of the old data to the new estimation at the current sampling time. It is much useful especially for the cases that some of system characteristics may be time varying (40). In the following part, the forgetting factor is selected in the interval $[0.95, \; 1]$.

Note that if there is no time delay in the system model, or the time delay is known beforehand, the optimization problem of the system (3.64) can be degenerated to a standard LMS problem. Moreover, if some pre-knowledge of time delay in the system can be known or obtained, such as the upper boundary and lower boundary at each sampling step, an iterative algorithm can be performed by searching the optimal solution in the entire possible region of time delay in an one-by-one manner as previous Iterative LS methods. For each iteration, LMS problem can be solved by applying some standard techniques referred in (100). In general, the LMS based method require to enumerate all the possible situation with regard to time delay. But this method can guarantee that the solution is globally optimal in most cases.

The same as the previous assumption, the boundary of time delay is described by some integer numbers multiplying with sampling period, i.e., $l^k_{\min} \leq l^k \leq l^k_{\max}$ and $l^k_{\min}$, $l^k_{\max}$, which are known beforehand. Moreover, in order to make parameter identification, the sampling interval T_s, the sliding window length N and forgetting factor ρ need to be decided as well before the procedure.

In the start, the algorithm need to wait to collect the enough sampled data to construct matrices (3.41) and (3.42) until a specific sampling step. Suppose this initial step as k_{ini}, where the condition $N + l^{k_{ini}}_{\max} \leq k_{ini}$ should be satisfied. Then, the main identification procedure can start from the k_{ini} step. Let sampling step $k \geq k_{ini}$, the whole scheme is in the following:

- A computing loop is constructed with regard to l^k starting from $l^k_{\min}$ and ending at $l^k_{\max}$ by taking the unit increment to l^k. For each iteration (k) of l^k ($l^k_{\min} \leq l^k \leq l^k_{\max}$), solve the LMS problem (3.40) and record the corresponding prediction error. The LMS problem has an analytical solution as:

$$\begin{aligned} \hat{\theta}^k(l^k) &= ((A^k_N(l^k))^T A^k_N(l^k))^{-1}(A^k_N(l^k))^T B^k_N, \\ Cov(\hat{\theta}^k) &= ((A^k_N(l^k))^T \hat{Q}(k)^{-1} A^k_N(l^k))^{-1}, \end{aligned} \tag{3.67}$$

  where $\hat{\theta}^k(l^k)$ stands for the estimation of θ^k at current iteration with discrete time delay l^k, $Cov(\hat{\theta}^k)$ means the covariance of $\hat{\theta}^k$, and $Q(k) \hat{=} (1-\hat{\alpha}^k)^2 Q$ is the covariance of $\omega^{'}(k)$.

- $(\hat{\theta}^k(\hat{l}^k), \hat{l}^k)$ which leads to the minimal prediction error among all iterations with regards to l^k moving from $l^k_{\min}$ to $l^k_{\min}$, denoted as $(\theta^{k*}(l^{k*}), l^{k*})$, is chosen as the optimal solution for (3.40) at the current step.

- The estimation of the kth sampled system parameters of (3.58) and (3.59) for the current sample, i.e., $\hat{T}^k_p$, $\hat{K}^k_1$ and $\hat{u}_2(k)$, can be obtained from $\theta^{k*}(l^{k*}) =$

$[\alpha^{k*} \ \ \beta^{k*} \ \ \gamma^{k*}]^T$ using the following relation:

$$\hat{T}_p^k = -\frac{T_s}{\ln \alpha^{k*}}, \tag{3.68}$$

$$\hat{K}_1^k = \frac{\beta_1^{k*}}{1-\alpha^{k*}}, \tag{3.69}$$

$$\hat{u}_2(k) = \gamma^{k*}/(1-\alpha^{k*}), \tag{3.70}$$

and the sampled time delay l^k is estimated as l^{k*}. As a result, $\hat{K}_1^k$, $\hat{T}_p^k$, $\hat{u}_2(k)$ and $l^{k*}T_s$ are set as the approximations of the original parameters and unknown input for the continuous system (3.3) at the current sampled step.

When a new (couple) data of input and measured output is obtained, the above procedure will be repeated. Thereby the system idenficiation/estimation for the model (3.53) can be executed in an on-line iterative manner. It can be noticed that the original method proposed in previous sections can become a special case of the considered problem here, i.e., corresponding to $u_2(t) \equiv 0$.

3.6.3 Numerical Examples

In the following, the proposed method in section 3.6 and the method used in section 3.5 are both applied and compared. For simplicity, the proposed method is noted as *new method*, while the latter one is noted as *old method*.

Case A-I: Data generated from a system with unknown input

Consider a switching TV-FOPDT system, where system parameters are set as: when $t < 30$ seconds, there are

$$T_p^t = 2,\ K_1^t = 3,\ K_2^t = 3,\ T_d^t = 3.05;$$

when $t \geq 30$ seconds, the parameters change to

$$T_p^t = 3,\ K_1^t = 4,\ K_2^t = 4,\ T_d^t = 2.05.$$

The noise in the measurement of the output follows the distribution $\mathcal{N}(0,\ 0.001)$. The sampling period is set as $T_s = 0.1$ second. The length of sliding window is selected as $N = 50$, and the forgetting factor $\rho = 0.95$. Assume we have the pre-knowledge of the sampled time delay like $l^k_{\text{max}} = 40$ and $l^k_{\text{min}} = 0$. According to the proposed method, it need to wait more than 8 seconds (the data collection period is $(50+40) \times 0.1 = 9$ seconds) to start the identification procedure in the beginning. In the test, the identification begins after 100th sampling time, i.e., after 10 seconds. The data is collected by simulated the considered system with a sweep signal as the control input, and a multi-step signal as the unknown input with the property

$$u_2(t) = \begin{cases} 1, & t < 40 \\ 1.2, & 40 \leq t < 60 \\ 2, & t \geq 60. \end{cases}$$

The known input signal and measured output obtained from this simulation are illustrated in Fig. 3.25.

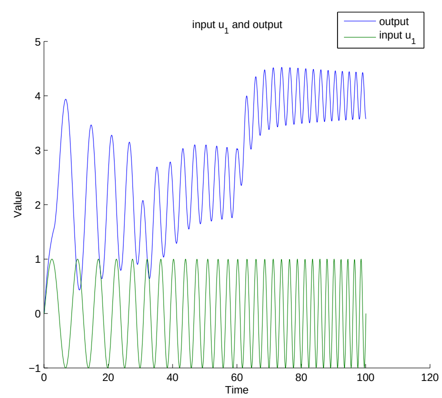

Figure 3.25: The known input and output data for Case A-I

Fig. 3.26, Fig. 3.27, Fig. 3.28 and Fig. 3.29 display the results of the system estimation for time delay, system gain of known input, time constant and unknown input, respectively. Here the red line plots the real value, the blue one shows the estimated value using the proposed method and the green one is the result using the original method proposed in the prevous section. From the simulation results, the following observation could be made. All of the results, including parameter identification, time delay estimation and unknown input estimation, show nearly the same characteristics.

It is obvious that the proposed method exhibited much better results than the old one did, which is supposed to be used only for SISO TV-FOPDT case. For the proposed method, the identification algorithm starts at 10th second. Since the system is

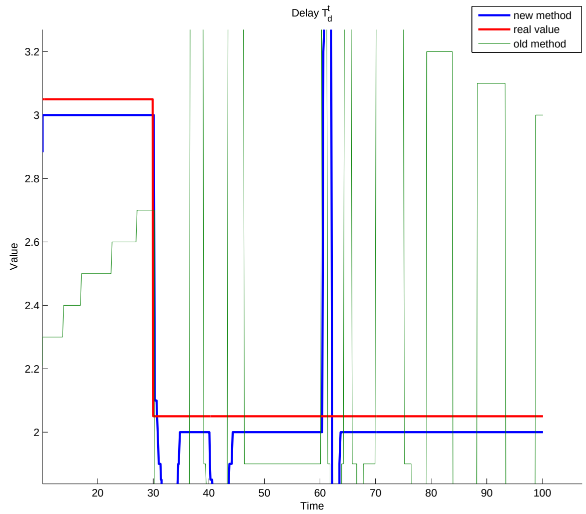

Figure 3.26: The time delay estimation for Case A-I

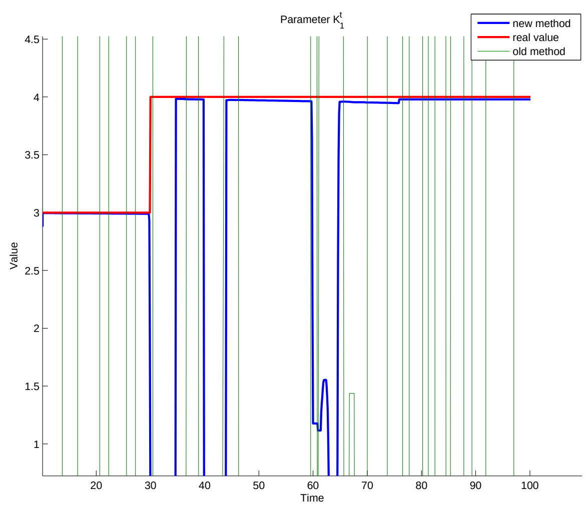

Figure 3.27: The identification result of K_1^t for Case A-I

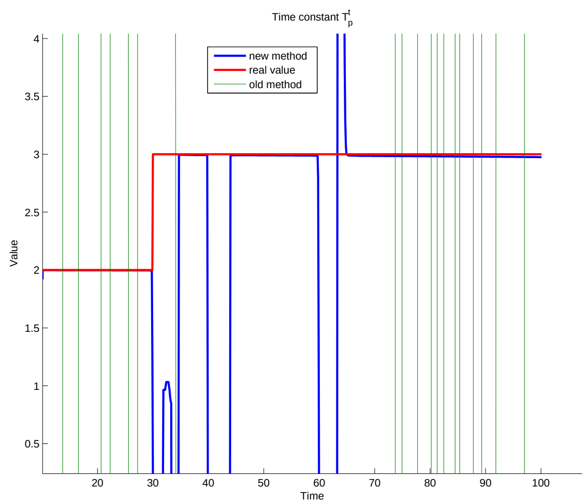

Figure 3.28: The estimated time constant for Case A-I

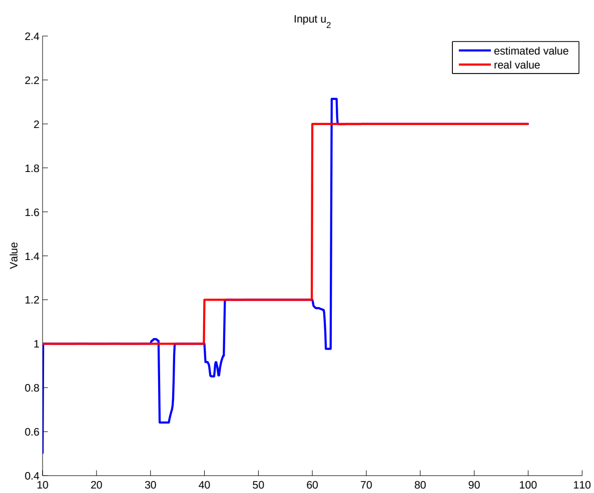

Figure 3.29: The estimated unknown input for Case A-I using proposed method

already at a steady situation, the estimations showed reasonably good approximations and precisions. This stable estimation lasted for about 20 seconds until the system had a switching at 30th second. Some deviations are clearly observed during a short period after the switch of system parameters (30 sec.). The fluctuation period is approximately equal to one window length (50∗0.1=5 seconds) before the estimated parameters started to converge to new steady-state values. The same phenomenon happened when the unknown input has jumps at 40th second and 60th second, respectively.

Regarding the accuracy, the time delay estimation showed some small steady-state estimated error and they are below 2% in most cases. These offsets are mainly due to the discretization of the system model, thereby it can be reduced by increasing the sampling frequency. All results of the other three estimation showed the steady-state error are less than 1% to the real values in most steady state cases.

Case A-II: Data generated from a system without unknown input

In this test, the data used for estimation is generated by applying the same input as used in Case A-I except that there is no unknown input, which means $u_2 \equiv 0$. Both identification methods are tested and compared in the following.

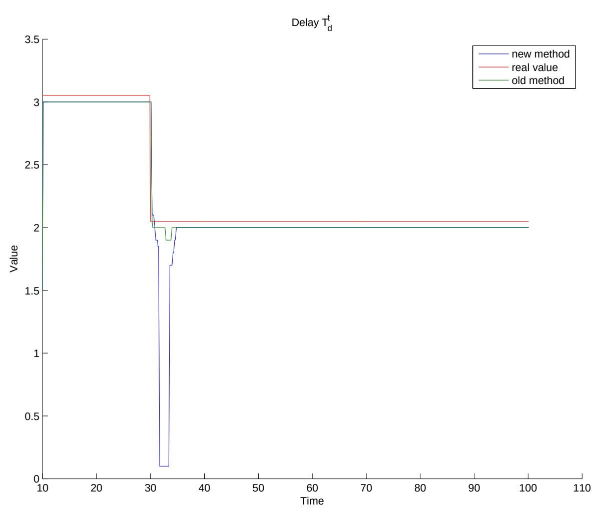

Figure 3.30: The time delay estimation for Case A-II

The results turned out that both methods showed almost same performances except different amplitudes of fluctuations after the switching point (30 sec), where the system

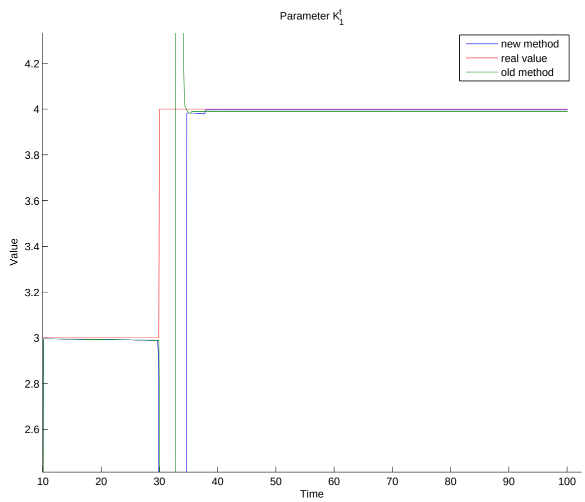

Figure 3.31: The identification result of K_1^t for Case A-II

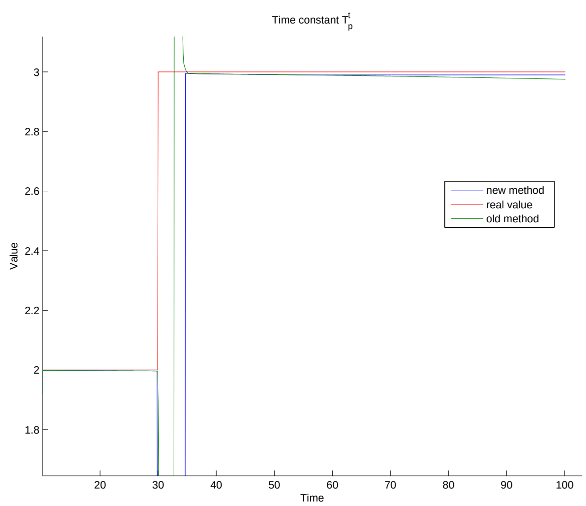

Figure 3.32: The estimated time constant for Case A-II

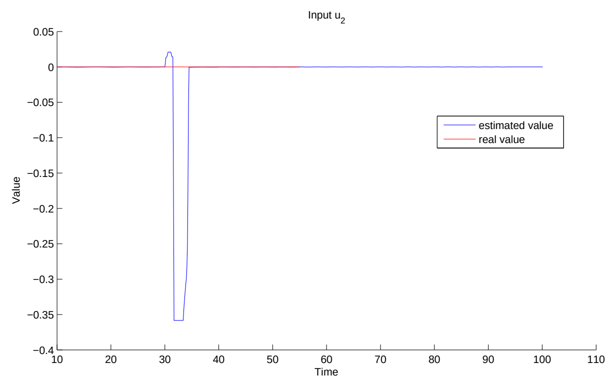

Figure 3.33: The estimated unknown input for Case A-II using the proposed method

parameters abruptly changed. The *new method* led to larger fluctuations than the *old method*. This is because that the *new method* got a wrong estimation (non-zero) of the unknown input for a short while, as shown in Fig. 3.33, which caused further deviations to all parameter estimations. Otherwise, we can conclude that both methods can provide almost same estimation performances.

3.7 Application for Superheat Modeling

3.7.1 Refrigeration and Superheat System

Figure 3.34: Refrigeration system

One of typical refrigeration systems follows the principle with vapor compression by using some types of refrigerant as the heat transfer medium. Generally, one refrigeration system composes of four basic components, that are expansion valve, evaporator, compressor and condenser. A vital variable that can greatly affect the efficiency of this kind of system is the filling refrigerant in the evaporator. The important factor to evaluate this refrigerant filling is the superheat, which can be defined as the difference between the outlet temperature of the gas and the inner temperature of the evaporator. This kind of superheat can be controlled by adjusting the degree to open the expansion valve. In order to maximally utilize the potential of the evaporator, the superheat needs to be maintained as low as possible.

Most existing commercial refrigeration systems use either a thermostatic expansion valve or a kind of on-off control of the expansion valve. These types of control are easy and simple for design and implementation, however they often do not lead to (smooth) comfort and energy-efficient performance. Some advanced feedback control methods are expected for this type of system. Nevertheless, no matter what kind of methods were used, generally, one mathematical model of the considered superheat dynamic is often required in order to have a automatical control designing and tuning process. The dynamics of superheat in a refrigeration system must be very complicated, which can consist of high nonlinearities and time varying properties. The detailed model of the evaoprator/superheat can be set up according to the conservation of mass, momentum and energy on the refrigerant, air and tube wall etc. However, this category of detailed model often causes some difficulties during the control design stage because of the complexity of the system model. For these reasons, in order to get a simple model of the system, Li (92) proposed an empirical model to decouple the superheat and capacity control, where the superheat system was modeled by so called First-Order Plus Dead-Time (FOPDT) model. However, a FOPDT model can only make sense for some local operating points. Later, Russmus and Lars (135) proposed another kind of nonlinear First-Order (FO) model in 2009, based on the first modeling principle. Their considered nonlinear FO model can be seen as an extension of the standard FO model by means that both of the system gain and time constant of the model were taken as functions of the inputs and disturbances, and hence an adaptive control of superheat was developed based on back-stepping method. However, the acquisition of this nonlinear FO model need many assumptions to be founded due to the physical modeling principle, and many of these assumptions are either impossible or difficult to be examined in the reality. Moreover, the time-delay feature of the superheat dynamic is not explicitly expressed in this model either.

But the proposed model TV-FOPDT, including some inputs dependent dead time, can describe the former status and solve the problems. In the thesis, Time-Varying FOPDT (TV-FOPDT) model is applied to model the superheat dynamic in a supermarket refrigeration system.

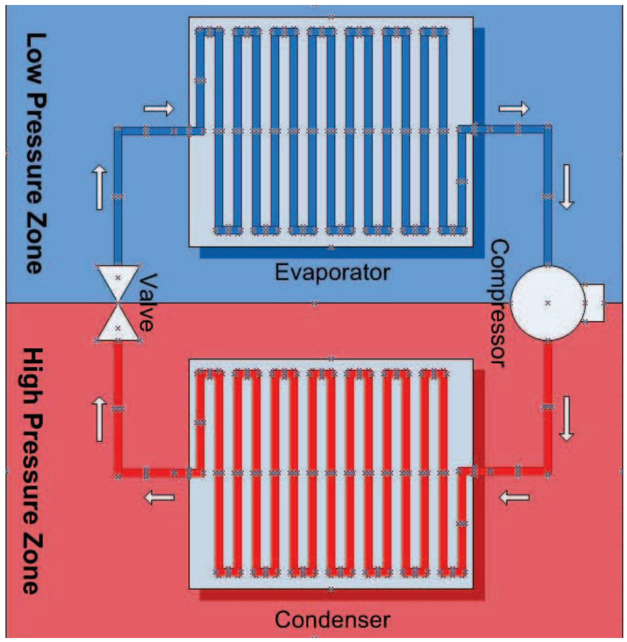

Figure 3.35: A superheat model

A popular superheat dynamic can be seen in the Fig. 3.35.

3.7.2 Superheat Dynamic Identification

Two different systems are considered in the following system identification. The considered refrigeration system is a supermarket display case cooler as shown in Fig. 3.34. Compared with a freezer, the display case cooler has a less efficient (adjustable) air curtain. Two sensors are installed to gain the superheat measurement. One pressure sensor is placed close to the inlet tube of the evaporator. Then the evaporation temperature is estimated based on this pressure measurement and the knowledge of refrigerant type. A thermostat transducer is placed at the evaporator outlet to measure the gaseous refrigerant temperature (172).

Case S1: TV-FOPDT with noise First, the system model is chosen as with only one transfer function (3.39) and the measurement model is described as (3.41). It is obvious there is noise in the measurement of the system.

The experimental data is collected from a real system installed at Danfoss A/S (172). The sampling period T_s is selected as 2 seconds. Moreover, it has been noticed from the experience that time delay of the real system is no more than 300 seconds, i.e., the upper limit of the time delay can be set up as 150 samples. The input data is the measurement of the percentage in the openness of the expansion valve, and the output data is the calculated temperature of superheat based on two sensor (both inner and outer) measurements. In order to significantly excite the considered system, the designed input signal is composed of a number of asymmetrical relay cycles. One set of input and output data is illustrated in Fig. 3.36.

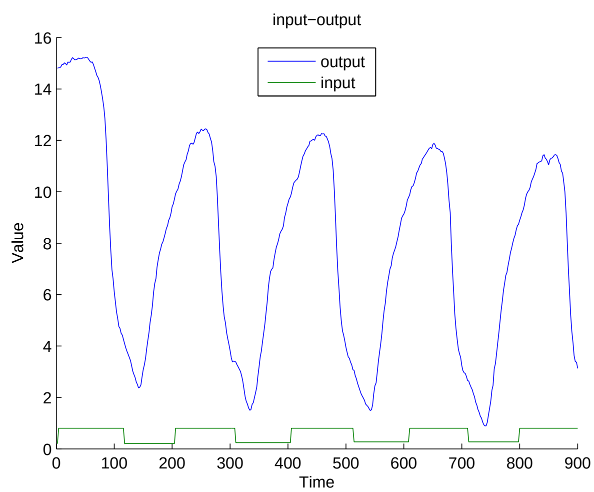

Figure 3.36: Input/Output Data for Case S1

A rectangular window with a length of 200 samples is used. Thereby the first estimation result comes at the next step after the 350th sampling step, i.e., 200 (window length) + 150 (maximal delay) =350. The estimated system time delay is indicated in Fig. 3.37. It can be noticed that during the period from the beginning to the 744th sampling step, the estimation stayed at a value of 32 sec. From the 746th sampling step, the estimated value stabilized around a value of 234 sec. The estimated system delay only significantly changed twice regarding to this tested experiment. The identification results of (sampled) system gain and time constant can be seen in Fig. 3.38. The time varying feature of these two parameters is quite obvious. In general, the estimated system gain has a trend to slightly increase until reaching some steady-state while the estimated system time constant has a trend to slightly decrease until reaching some steady-state. This test also showed that the superheat gradually converge to its expected working point (10 degree for this case). It has been found in [10] that the system parameters of a nonlinear FO model of the superheat dynamic are relevant to system input, output and disturbance as well. The coupling between the superheat dynamic and the compressor behavior is also studied in [7]. Thereby, a 3-D plot of the estimated time delay w.r.t. the input and output signals is shown in Fig. 3.39. From this observation, it seems that the system time delay mainly depends on the output value.

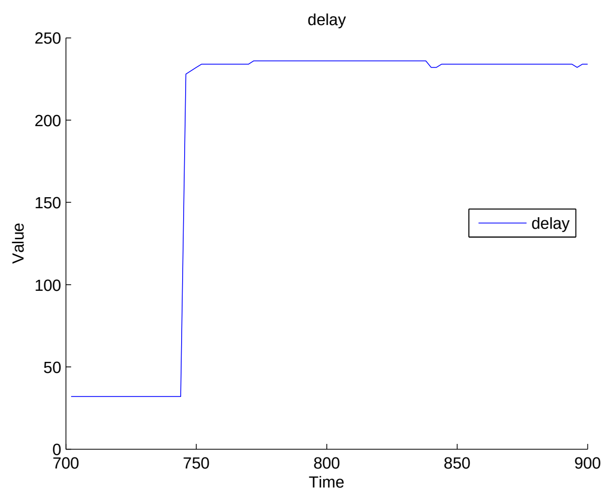

Figure 3.37: Delay Estimation for Case S1

Case S2: MI TV-FOPDT model

In this part, the data generated from a real refrigeration system, which is another set different to the case S1, is used to estimate a Multi-Input TV-FOPDT model of the superheat dynamic in the considered system. Some conditions are the same to the former case, i.e., the sampling period T_s is still selected as 2 seconds. Moreover, it has been noticed that the system time delay is no more than 400 sec., i.e., we can set

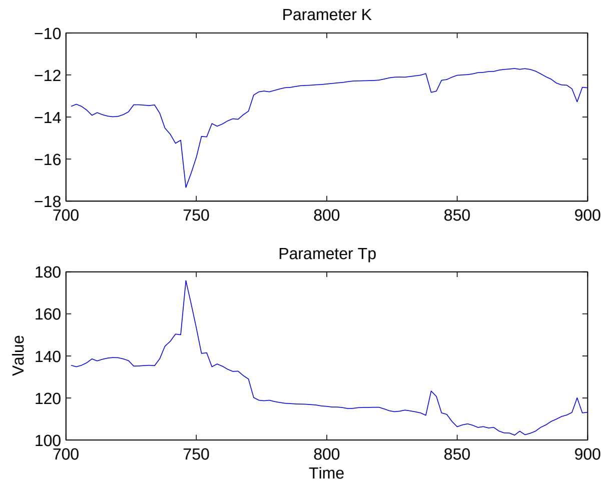

Figure 3.38: Parameter Estimation for Case S1

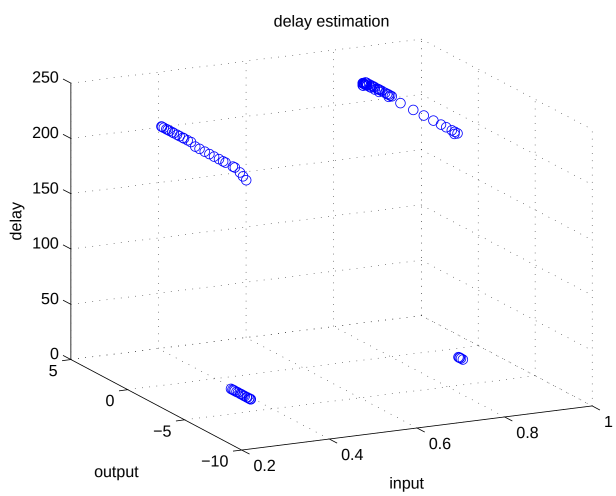

Figure 3.39: 3-D Plot for Case S1

up the upper limit of the time delay as 200 samples. The known part of input data is the measurement of the openness percentage of the expansion valve, and the output data is the calculated superheat (temperature) based on two sensor measurements. The designed input signal consists of a number of asymmetrical relay cycles. One set of input and output data is illustrated in Fig. 3.40.

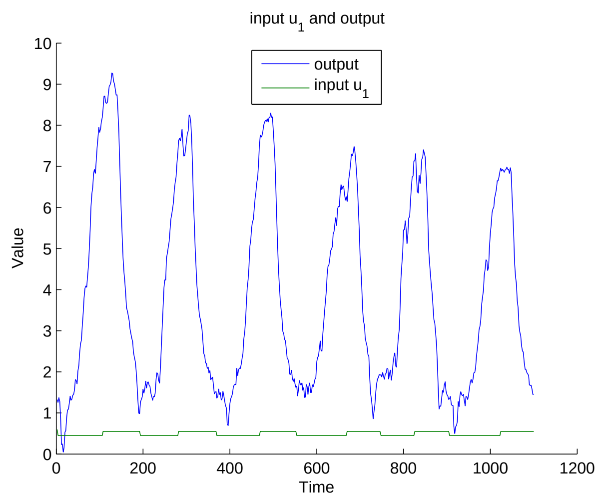

Figure 3.40: The input and output data from a real system

Under the assumption that the superheat (temperature) dynamic can be approximated by system model (3.53), we define parameter K_2^t as 2. It should be noticed that the value of K_2^t does not critically affect the estimation results, even though it could influence the estimated aptitude of the unknown input. Theoretically, it can be set as any value. A sliding window with a length of 200 samples is used. Thereby the first estimation result comes at 400 sampling step, i.e., 200 (window length) + 200 (maximal delay) =400, this means that the first estimation should start at 800 second. In this test, both the *new method* and *old method* are employed as well. The estimated system parameters are illustrated in Fig. 3.41, Fig. 3.42 and Fig. 3.43, respectively. At this moment, we are not always sure that the proposed *new method* works better than the *old method* did. From the so-far observed results, we can conclude that superheat model in this refrigeration system should take the disturbances into consideration, which could be due to the influences of compressor and/or the ambient thermal environment. Furthermore, since we expect a model which is suitable for modeling superheat dynamic in large operating region, there is no doubt that TV-FOPDT model should be one of the good candidates.

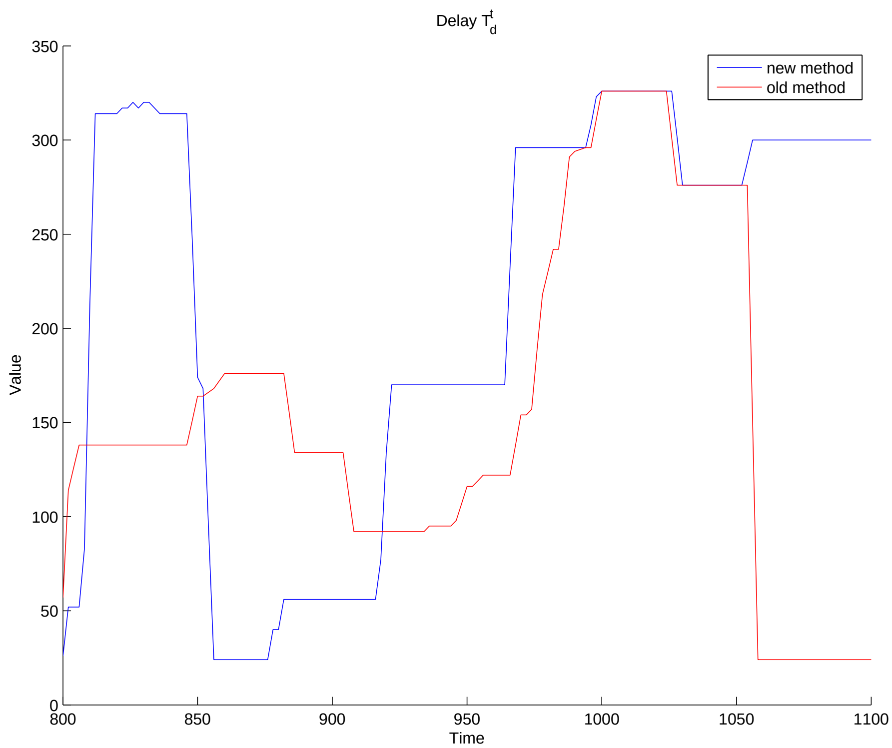

Figure 3.41: Time delay estimation for the real system

3.8 Conclusion

This Chapter considered a TV-FOPDT system identification problem. The models consist of three different kinds, simple TV-FOPDT, TV-FOPDT with input dependent dead time and Multi-Input FOPDT. The first two models can together called as SISO TV-FOPDT model compared with MI TV-FOPDT model. From the model studying, MI TV-FOPDT model can describe the disturbance input much better.

Correspondingly, a number of identification algorithms to estimate the time dependent parameters, as well as the unknown input for the MI TV-FOPDT model, are proposed. By regarding all unknown parameters as the ones need to be identified including

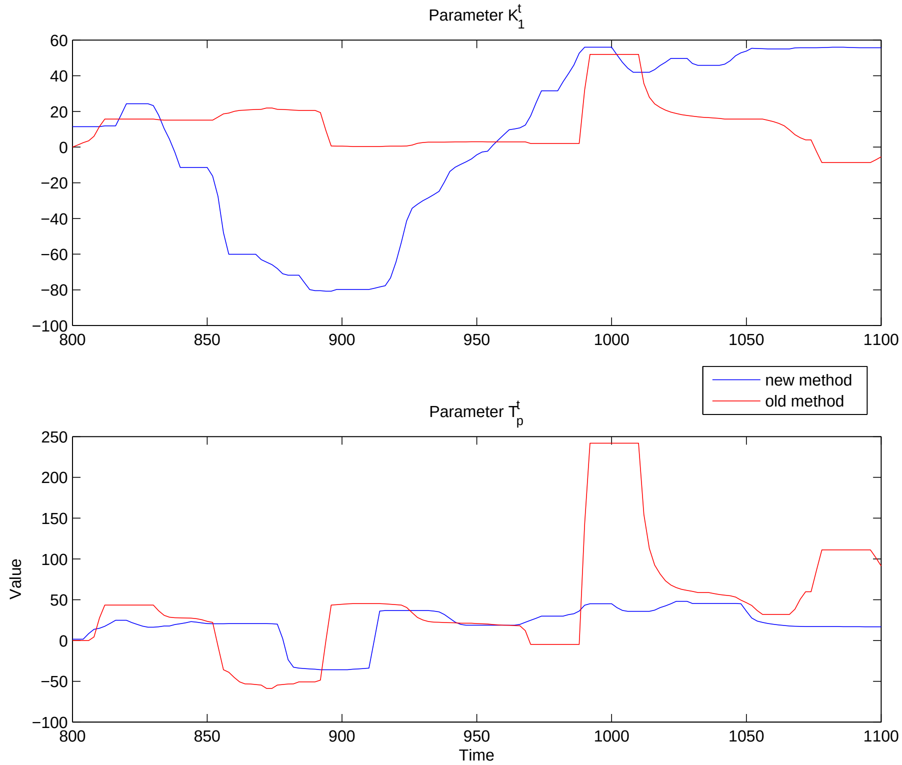

Figure 3.42: System gain and time constant estimation for the real system

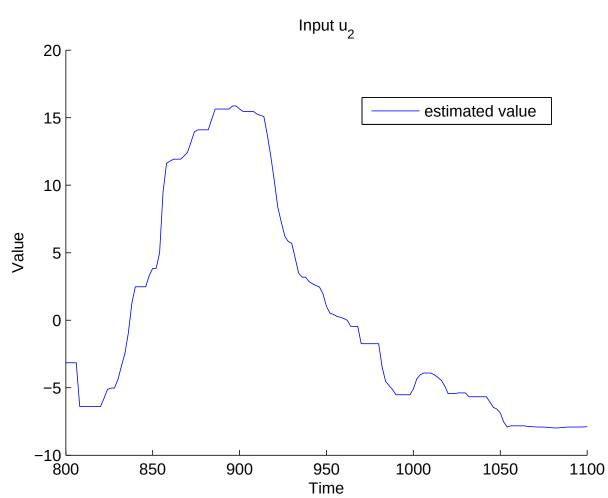

Figure 3.43: Unknown input estimation for the real system

the unknown input, the considered problem can be formulated as a Stochastic Mixed Integer Nonlinear Programming (SMINP) problem. Then bound-branch method for handling the mixed integer programming, the Least Mean Square (LMS) for handling the optimal parameter identification, together with the sliding window with forgetting factor for data selection, are adopted and combined to handle the formulated problem. The method can make the system identification in an on-line manner.

The proposed approaches are tested on a number of numerical examples and compared with the relevant methods. For the application, it is applied to model the superheat dynamic in a supermarket refrigeration system. There is no doubt that the MI TV-FOPDT provides more flexibility to model complex systems in a more realistic manner.

Chapter 4

Conclusions and Future Work

4.1 Thesis Conclusions

The thesis considered the techniques of parameter identification for two different kinds of nonlinear models, i.e., nonlinear ISDE model and nonlinear FOPDT model. The approaches to make the corresponding parameter identification are proposed, which are called as UKF plus ML method and Mixed Integer Nonlinear Programming (MINP) based method. The thesis make contribution to several points to the development of system identification. Firstly, the thesis suggests to apply some new nonlinear models to describe the systems more accurately. Secondly, some new methods are proposed to make parameter identifications of the corresponding models and these methods have their own merits. Thirdly, some theorems proved in the thesis can provide some theoretical support to the new models and parameter identification methods, such as the identifiability and convergence issues.

Nonlinear ISDE Model

- The merits of using Itô SDE model lies in that it can describe the structure of the random feature of system in a more accurate way and the mature ISDE theory can provide a theoretical support to this model.

- A nonlinear system identification approach was proposed to make the estimation of the system modeled by ISDE. The approach combined the Unscented Kalman Filter and Maximum Likelihood to make the parameter identification.

The characteristics and advantages of the proposed method is its relatively good precision, accuracy and computation load regarding to the parameter identification.

- The consistency and normality of the proposed method, UKF plus ML method, were proved under conditions of boundness for system functions including their derivatives and parameter possible ranges.

- A number of numerical tests were formulated to make the evaluation of the new scheme. The results showed it can provide a good performance in terms of the accuracy, convergence at a small extra cost of the computation load.

All in all, the ISDE model has its own unique merits to describe the random systems. Since it can describe the system in which the random part can be related with the state variable, the ISDE model can model the system with fault that may depend on the state variable. Moreover, the mature theory on ISDE can provide a useful support to the system analysis. For example, the Itô formula can simplify some system with state related random features to ones without state related random features. Then the technique of system identification can be applied simply to the system. The proposed approach of nonlinear system identification, UKF plus ML method, is proved to be consistency and normality under the corresponding conditions. It can guarantee the estimation using UKF plus ML method is correct for some kinds of systems. Furthermore, the normality property can show the confidential level of the estimation. Moreover, from a number of tests, it showed better performance in accuracy and convergence than direct Kalman Filter technique and EKF plus ML methods at cost of computation load.

Nonlinear FOPDT Model

- The identifiability of the time varying models are particularly defined based on the model structure, identification method and sampling points. Under the new definition, the condition that can guarantee the identifiability of nonlinear FOPDT is derived.

- The Time Varying FOPDT model, even with the input dependent dead time, was proposed. A method based on the Stochastic Mixed Integer Nonlinear Programming (SMINP) was developed to make the estimation of the parameters with time delay for the system. The approach applied Branch-Bound method and Least Mean Square method to solve the concerned problem.

- A number of numerical tests were formulated to make the evaluation of the algorithm. The results showed it can provide a good performance in terms of the accuracy and speed. The method was applied to make the estimation of a real system which models superheat in a refrigeration system.

All in all, the proposed nonlinear FOPDT has much more flexibility in modeling some complex processes much better than the traditional FOPDT model. The identifiability analysis showed that under some conditions the nonlinear FOPDT identification can be guaranteed using LS based method. The simulation results showed it is a fast and flexible method. But the sampling number and input signals can affect the performance of the accuracy.

4.2 Future Work

Nonlinear ISDE Model

Firstly, the UKF plus ML method can identify some models with state depended random features which can be simplified to ones without state depended random features using Itô formula. To find the approach to identify other models with state depended random features can be part of future works for the system identification of ISDE model.

Secondly, how to extend the parameter identification method proposed here to be a recursive version to make the on-line identification in order for the computation efficiency as well as the FDD purpose, is still open. The most difficulty lies in how to handle the time varying delay estimation recursively.

Nonlinear FOPDT Model

Firstly, whether the identifiability analysis of the nonlinear FOPDT can be extended to other nonlinear models or not need to be further investigated and studied in the future.

Secondly, it is undoubtable that nonlinear FOPDT model can not be used to describe all the system. For this reason, to find out what kind of system can be described using nonlinear FOPDT model is part of the future work. And how can we make the controller design or system reconfiguration based on the estimation of nonlinear FOPDT can also be the following work. Moreover, the correlation between the convergence rate of the selected identification algorithm and the time varying features of unknown parameters need to be further deeply investigated.

Bibliography

[1] S. Ahmed, B. Huang and S. L. Shah, Parameter and delay estimation of continuous-time models using a linear filter, *Journal of Process Control*, Vol. 16, pp. 323-331, 2006.

[2] J. Aldrich, R. A. Fisher and the Making of Maximum Likelihood 1912-1922, *Statistical Science*, 1997, Vol. 12, No. 3, pp. 162-176.

[3] G. H. M. De Arruda and P. R. Barros, Estimation of FOPDT models from open and closed loop relay experiments, *Brazilian Congress of Automatic*, 2004, pp. 1-6.

[4] K. J. Åström. Introduction to stochastic control theory, *Academic Press*, New York, USA, 1970. 10, 17, 20

[5] K. J. Åström and T. Söderström, Uniqueness of the Maximum Likelihood Estimates of the Parameters of an ARMA Model, *IEEE Transactions on Automatica Control*, Vol. AC-19, No. 6, December 1974. 27, 37

[6] K. J. Åström and T. Hägglund, PID Controllers: Theory, Design and Tuning, *Instrument Society of America*, 1995. 3

[7] K. J. Åström, H. Pnagopoulos and T. Hägglund, Design of PI controllers based on non-convex optimization, *Automatica*, Vol. 34, No. 5, 1998, pp. 585-601.

[8] A. Beghi, L. Cecchinato and M. Rampazzo, On-line, auto-tuning control of Electronic Expansion Valves, *International Journal of Refrigeration*, Vol. 34, 2011, pp. 1151-1161.

[9] S. Björklund, A Survey and Comparison of Time-Delay Estimation Methods in Linear Systems, Thesis No. 1061, Linköping Studies in Science and Technology.

[10] S. Björklund and L. Ljung, A Review of Time-Delay Estimation Techniques, *Technical reports from the Control & Communication group in Linköping*, 30th December 2003. 6

[11] Q. Bi, W. J. Cai, E. L. Lee, Q. G. Wang, C. C. Hang and Y. Zhang, Robust identification of first-order plus dead-time model from step response, *Control Enineering Practice*, Vol.7, 1999, pp.71-77.

[12] J. P. N. Bishwal, Parameter Estimation in Stochastic Differential Equations, *Springer-Verlag Berlin Heidelberg*, 2008. 10, 20

[13] T. Bohlin and S. F. Graebe, Issues in nonlinear stochastic grey-box identification, *Intrernational Journal of Adaptive Control and Signal Processing*, No. 9, 1995, pp. 465-490.

[14] S. Boyd and L. Vandenberghe, Convex Optimization, *Cambridge University Press*, 2009. 34, 35

[15] K. Brazauskas and D. Levisauskas, Adaptive transfer function-based control of nonlinear process. Case study: Control of temperature in industrial methane tank, *Int. J. Adapt. Control Signal Process.*, Vol.21, 2007, pp. 871-884. 12

[16] L. Y. Cao and H. M. Schwartz, Exponential convergence of the Kalman filter based parameter estimation algorithm, *International Journal of Adaptive Control and Signal Processing*, Vol. 17, 2003, pp. 763-783.

[17] A. B. Carripio, Tuning of Industrial Control Systems, *Instrument Society of America*, 1990, pp. 43-44. 111

[18] C. K. Charles and G. R. Chen, Kalman Filtering with Real-Time Applications, *Springer Series in Information Sciences*, 17 (4th ed.), Springer, New York, 2009. 29, 70, 72

[19] C. H. Chen and S. D. Fassois, Maximum Likelihood Identification Of Stochastic Wiener-Hammerstein-Type Non-Linear Systems, *Mechanical Systems and Signal Processing 6(2)*, 1992, pp. 135-153.

[20] J. Chen and R. J. Patton, Robust model-based fault diagnosis for dynamic systems, Kluwer, Boston, 1999. 8, 9

[21] C. T. Chou and M. Verhaegen, Subspace algorithms for the identification of multivariable dynamic errors-in-variables models, *Automatica*, Vol. 33, 10, 1997, pp.1857-1869. 5

[22] M. J. Crowder, Maximum Likelihood Estimation for Dependent Observations, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 38, No. 1(1976), pp. 45-53. 37, 41

[23] Danfoss, Thermostatic Expansion Valves, *Fitters Notes*, *Refrigeration and Air Conditioning Dept.*, Danfoss, 2005.

[24] R. David, Statistics and Data Analysis for Financial Engineering. *Springer, ISBN 978-1-4419-7786-1*, 2010. 36

[25] T. Deck and T. G. Theting, Robust Parameter Estimation for Stochastic Differential Equations, *Acta Applicandae Mathematicae*, Vol. 84, 2004, pp. 279-314.

[26] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1), pp. 1-38, 1977. 34

[27] T. Delsole and X. S. Yang, State and Parameter Estimation in Stochastic Dynamical Models, *ECMWF Workshop on Model Uncertainty*, 20-24 June, 2011.

[28] S. X. Ding, Model-based Fault Diagnosis Techniques Design Schemes, Algorithms, and Tools, *Springer*, 2008.

[29] F. Ding and T. Chen, Combined parameter and output estimation of dual-rate system using an auxiliary model, *Automatica*, Vol. 40, No. 10, October 2004, pp. 1739 C 1748.

[30] J. Ding, F. Ding and S. Zhang, Parameter identification of multi-input, single-output systems based on FIR models and least squares principle, *Applied Mathematics and Computation*, Vol. 197, 2008, pp. 297-305.

[31] A. J. Dorogovcev, The consistency of an estimate of a parameter of a stochastic differential equation, *Theory Probab. Math. Statist. 10*, pp. 73-82, 1976. 10

[32] D. Dougherty and D. Cooper, A practical multiple model adaptive strategy for single-loop mpc, *Control Eng. Pract.* Vol. 11, 2003, pp. 141-159.

[33] S. V. Drakunov, W. Perruquetti, J. P. Richard and L. Belkoura, Delay identification in time-delay systems using variable structure observers, *Annual Reviews in Control*, Vol. 30, 2006, pp. 143-158.

[34] Guang-ren Duan and Ai-guo Wu, Robust fault detection in linear systems based on PI observers, *International Journal of Systems Science*, Vol. 37, NO. 12, 10 October 2006, pp. 809-816.

[35] W. Eric and V. D. M. Rudolph, The unscented kalman filter for nonlinear estimation, *Proc. of IEEE Symposium 2000 (AS-SPCC)*, Lake Louise, Alberta, Canada, October 2000. 30, 31, 32

[36] G. Fedele, A new method to estimate a first-order plus time delay model from step response, *Journal of The Franklin Institute*, Vol. 346, 2009, pp. 1-9.

[37] J. C. Feng, H. J. Fan and C. K. Tse, Convergence Analysis of the Unscented Kalman Filter for Filtering Noisy Chaotic Signals, *1-4244-0921-7/07, IEEE*, 2007.

[38] A. R. Ferreira and J. M. Fernandes, A survey on time delay system estimation, *Proc. ECC'97*, Brussels, Belgium, 1-4 July 1997.

[39] R. A. Fisher, On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character,* Vol. 222, (1922), pp. 309-368. 9

[40] T. R. Fortescue, L. S. Kershenbaum and B. E. Ydstie, Implementation of self-tuning regulators with variable forgetting factors, *Automatica*, Vol. 17, No. 6, 1981, pp. 831-835. 149

[41] P. M. Frank, Analytical and qualitative model-based fault diagnosis: A survey and some new results, *European Journal of control,* Vol. 2, pp 6-28, 1996. 12

[42] H. Garnier and L. P. Wang, Identification of Continuous-time Models from Sampled Data, *Springer, London*, 2008.

[43] V. Genon-Catalot and J. Jacod, Estimation of the Diffusion Coefficient for Diffusion Processes: Random Sampling, *Board of the Foundation of the Scandinavian Journal of Statistics*, Vol. 21, 1994, pp. 193-221. 25

[44] J. Gertler, Fault Detection and Diagnosis in Engineering Systems, New York, Marcel Dekker, 1998. 12

[45] M. Gevers, L. Minskovic, D. Bonvin and A. Karimi, Identification Of A Two-Input System: Variance Analysis, *16th IFAC World Congress*, Prague, July 4-8, 2005.

[46] D. Ghosh, Maximum Likelihood Estimation of the Dynamic Shock-Error Model, *Journal of Econometrics*, 41, 1989, pp. 121-143. 27, 38

[47] S. Gibson and B. Ninness, Robust maximum-likelihood estimation of multivariable dynamic systems, *Automatica*, 41(10), pp. 1667-1682, October 2005. 35

[48] G. C. Goodwin, J. A. Carlos and R. E. Skelton, Conditions for Local Convergence of Maximum Likelihood Estimation for ARMAX Models. *13th IFAC Symposium on System Identification*, Rotterdam, The Netherland, 2003. 37

[49] E. W. Gralda and J. W. MacArthurb, A moving-boundary formulation for modeling time-dependent two-phase flows, *International Journal of Heat and Fluid Flow*, Vol.13, No.3, 1992, pp. 266-272.

[50] M. S. Grewal and A. P. Andrews, Kalman Filtering, Theory and Practice Using MATLAB, *John Wiley* & *Sons, Inc.*, 2008.

[51] I. E. Grossman and N. V. Sahinidis (eds), Special issue on mixed integer programming and its application to engineering, Part I, *Optim. Eng.*, Vol. 3, No. 4, Kluwer Academic Publishers, 2002. 148

[52] P. Hall and C. C. Heyde, Rates of Convergence in the Martingale Central Limit Theorem, *Ann. Probab. Vol. 9*, No. 3, pp. 395-404, 1981. 26

[53] L. Han, J. Sheng, F. Ding and Y. Shi, Recursive least squares identification for multirate multi-input single-output systems, *American Control Conference*, Hyatt Regency Riverfront, St. Louis, MO, USA June 10-12, 2009.

[54] T. Hachino, ZJ. Yang and T. Tsuji, On-Line Identification of Continuous Time-Delay Systems Using the Genetic Algorithm, *Electrical Engineering in Japan*, Vol. 116, No. 6, 1996, pp. 115-126.

[55] L. P. Hansen, Large sample properties of generalized method of moments, *Econometrica*, 50, pp. 1029-1054, 1982. 27

[56] A. Harvey, S. J. Koopman and N. Shephard, State Space And Unobserved Component Models, Theory And Applications, *Cambridge University Press*, 2004.

[57] S. Haykin, Neural networks: a comprehensive fundation. (2nd ed.) Upper Saddle Rever, New Jersey, Prentice Hall, 1999. 3

[58] B. Huang and Q. G. Wang, Overview of Emerging Bayesian Approach To Nonlinear System Identification, *SICOP Round Tables on Nonlinear Model Identification, International Workshop on Solving Industrial Control and Optimization Problems*, Cramado, Brazil, April 6-7, 2006.

[59] M. A. Hopkins and H. F. VanLandingham, Pseudo-Linear Identification: Optimal Joint Parameter and State Estimation of Linear Stochastic MIMO Systems, *American Control Conference*, pp.1301 - 1306, 1988. 70

[60] A. J. Isaksson, A. Horch and G. A. Dumont, Event-triggered deadtime estimation from closed-loop data, *Proceedings'of the American Control Conference*, Arlington, VA June 25-27, 2001, pp. 3280-3285.

[61] Rolf Isermann, Process Fault Detection Based on Modeling and Estimation Methods–A Survey, *Automatica*, Vol. 20, No. 4, pp. 387-404, 1984. 8, 9, 12

[62] Rolf Isermann, Fault Diagnosis of Machines via Parameter Estimation and Knowledge Processing–Tutorial Paper, *Automatica*, Vol. 29, No. 4, pp. 815-835, 1994. 8, 12

[63] Rolf Isermann, Supervision, Fault Detection and Fault Diagnosis Methods–An Introduction, *Control Eng. Practice*, Vol. 5, No. 5, pp. 639-652, 1997. 8, 9

[64] Rolf Isermann, Model-Based Fault Detection and Diagnosis–Status and Applications, *Annual Reviews in Control*, Vol. 29, No. 1, 2005, Pages 71-85. 70

[65] G. Jacovitti and G. Scarano, Discrete Time Techniques for Time Delay Estimation, *IEEE Transactions on Signal Processing*, Vol. 41, No. 2, 1993, pp. 525-533. 7

[66] A. H. Jazwinski, Stochastic Processes and Filtering Theory, *Academic Press, New York and London*, 1970. 10

[67] J. C. Hull, Options, Futures, and Other Derivatives, *Prentice Hall*, 1997. 18, 19, 23

[68] J. Jeisman, Estimation of the Parameters of Stochastic Differential Equations, *Doctoral Thesis in Queensland University of Technology,* December, 2005.

[69] Z. Ji and M. Brown, Joint State and Parameter Estimation For Biochemical Dynamic Pathways With Iterative Extended Kalman Filter: Comparison With Dual State and Parameter Estimation, *The Open Automation and Control Systems Journal*, 2009, Vol. 2, pp. 69-77.

[70] D. H. Johnson and D. E. Dudgeon, Array Signal Processing, *Concepts and Techniques*, Prentice Hall, 1993. 7

[71] S. Julier, J. Uhlmann and H. D. Whyte, A new approach for filering nonlinear systems, *Proceedings of the 1995 American Control Conference*, Vol. 3, IEEE Press, June 1995, pp. 1628-1632.

[72] J. Julier and J. K. Uhlmann, A new extension of the Kalman Filter to nonlinear systems, *Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Control*,1997. 32

[73] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators, *IEEE Transactions onn Automatic Control*, Vol. 45, No. 3, March 2000, pp. 477-482. 68

[74] V. Kadirkamanathan, P. Li, M. H. Jaward and S. G. Fabri, Particle filtering-based fault detection in non-linear stochastic systems, *International Journal of Systems Science*, 2002, Vol. 33, No. 4, pp. 259-265.

[75] J. Kano, O. Tonomura, M. Kano and S. Hasebe, State and Parameter Estimation for Tubular Microreactors Using Particle Filter, *ICROS-SICE International Joint Conference*, August 18-21, 2009, pp. 3278-3282.

[76] R. Kalman and R. Bucy, New results in linear prediction and filtering theory, *Journal of Basic Engineering (ASME)*, 83D, 1961. 11, 28

[77] T. Katayama, Subspace Methods for System Identification, *Springer,* 2005.

[78] P. Kazimierczyk, Maximum-Likelihood Approach in Parametric Identification of Stochastic Differential Models of Engineering Systems, *Structural Safety*, Vol. 8, 1990, pp. 29-44. 27

[80] T. Kealy and A. O. Dwyer, Closed Loop Identification of A First Order Plus Dead Time Process Model Under PI Control, *Proceedings of the Irish Signals and Systems Conference*, University College, Cork, June 2002, pp. 9-14.

[80] T. Kealy and A. O. Dwyer, Comparison of Open- and Closed-Loop Process Identification Techniques In The Time Domain, *Proceedings of the 3rd Wismarer Automatisierungssymposium, Wismar, Germany*, September 2002, Paper 1.3-4.

[81] P. E. Kloeden and E. Platen, Numerical solution of stochastic differential equations, Springer, Berlin, Germany, 1992.

[82] R. Kohn, Local and Global Identification and Strong Consistency in Time Series Models, *Journal of Econometrics*, Vol. 8, 1978, pp. 269-293.

[83] N. R. Kristensen, H. Madsen and S. B. Jørgensen, Parameter estimation in stochastic grey-box models, *Automatica*, 40, 2004, pp 225-237. 11, 28, 33, 61, 70

[84] P. R. Kumar and P. Varaiya, Stochastic Systems: Estimation, Identification, and Adaptive Control, *Prentice Hall*, Inc., Englewood Cliffs, New Jersey, 1986.

[85] K. E. Kwok, S. L. Shah, B. A. Finegan and G. K. Kwong, An observational trial of a computerized drug delivery system on two patients, *IEEE Trans. Cont. Sys. Tech.*, Vol.5, No.4, 1997, pp. 385-393. 11, 12, 110

[86] L. F. S. Larsen, Model Based Control of Refrigeration Systems, PhD thesis in Aalborg University, Denmark, 2005.

[87] L. F. S. Larsen, R. Izdi-Zamanabadi and R. Wisniewski, Supermarket Refrigertion System-Benchmark for Hybrid System Control, *Proc. of ECC07*, Kos, Greece, 2007. 110

[88] A. Le Breton, On continuous and discrete sampling for parameter estimation in diffution processes, *Math. Program. Stud. 5*, pp. 124-144, 1976. 10, 26

[89] J. Lee, W. Cho and T. F. Edgar, Control system design based on a nonlinear first-order plus time delay model, *J. Proc. Cont.*, Vol. 7, No. 1, pp. 65-73, 1997. 11, 12, 110

[90] J. Lee and J. Lyou, Fault Diagnosis and Accommodation of Linear Stochastic Systems with Unknown Disturbances, *The Institute of Control, Automation and Systems Engineers,* Korea, Vol. 4, No. 4, December 2002, pp. 270-276.

[91] E. L. Lehman and G. Casella, Theory of Point Estimation (2nd edition). *New York, Springer*, 1998. 36

[92] H. Li, S. Jeong, J. Yoon and S. You, An empirical model for independent control of variable speed refrigeration system, *Applied Thermal Engineering*, Vol.28, 2008, pp. 1918-1924. 158

[93] P. Li and V. Kadirkamanathan, Particle Filtering Based Likelihood Ratio Approach to Fault Diagnosis in Nonlinear Stochastic Systems, *IEEE Transactions on Systems, Man, and Cybernetics–Part C: Applications and Reviews*, Vol. 31, No. 3, August 2001.

[94] L. Ljung, Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *Automatic Control, IEEE Transactions*, Vol. 24, 1979, pp 36-50. 10, 28

[95] L. Ljung, Extended least squares method, Systems and Control Encyclopedia (M.G. Singh Ed), *Pergamon Press,* pp 1581, 1987. 10

[96] L. Ljung, Approaches to Identification of Nonlinear Systems, *29th Chinese Control Conference (CCC),* 29-31 July 2010. 3, 28

[97] L. Ljung, Perspectives on System Identification, *Annual Reviews in Control*, Vol. 34, No. 1, April 2010, pp. 1 C 12.

[98] L. Ljung, Development of System Identification, *Proceedings of the 13th World Congress of IFAC*, July 1996, San Francisco, California, vol G, pp. 141-146.

[99] L. Ljung and T. Glad, On Global Identifiability for Arbitrary Model Parameterizations, *Automatica*, Vol. 30, No. 2, 1994, pp. 265-276. 114

[100] L. Ljung, System Identification–Theory for the User, second edition, *Prentice Hall*, 1999. 2, 3, 9, 113, 114, 115, 124, 150

[101] L. Ljung, Prediction Error Estimation Methods, *Circuits Systems Signal Processing,* Vol. 21, No. 1, 2002, pp. 11-21. 5, 124

[102] L. Ljung, Identification of Nonlinear Systems, *plenary paper, Proc. 9th International Conference on Control, Automation, Robotics and Vision*, ICARCV 2006, Singapore, December 2006.

[103] B. P. Lo, A. J. Haslam and C. S. Adjiman, Parameter estimation for stochastic differential equations: algorithm and application to polymer melt rheology, *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, 2006.

[104] J. R. Magnus and H. Neudecker, Matrix Differential Calculus. Wiley Series in Probability and Statistics, *Wiley*, 1999, pp. 171-173. 42

[105] R. Mamat and P. J. Fleming, Method for on-line identification of a first order plus dead-time process model, *Electronics Letters*, 20th July 1995, Vol. 31, No. 15, pp. 1297-1298.

[106] G. J. McLachlan and T. Krishnan, The EM Algorithm and Extensions (2nd edn), *John Wiley Sons, Inc.*, Hoboken, NJ, 2008. 35

[107] H. Madsen, Statistically determined dynamical models for climate processes, PhD thesis, IMSOR, DTH, Lyngby, 1985. 5, 11, 20, 28

[108] H. Melgaard and H. Madsen, CTLSM version 2.6-a program for parameter estimation in stochastic differential equations, Technical report, IMSOR, No. 1, 1993. 11, 28

[109] R. V. D. Merwe and E. A. Wan, Sigma-Point Kalman Filters for Nonlinear Estimation and Sensor-Fusion-Applications to Integrated Navigation, *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2004, pp. 16-19. 32, 64

[110] A. G. Michelsen, J. Stoustrup and R. Izadi-Zamanabadi, Model Based Plug and Play Process Control of a Supermaket Refrigeration System: A Heauristic Approach, *Proceedings of the European Control Conference*, Budapest, Hungary, August 23-26, 2009, pp. 4090-4095.

[111] S. Mondal, G. Chakraborty and K. Bhattacharyya, Unknown Input High Gain Observer for Parametric Fault Detection and Isolation of Dynamical Systems, *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009*, Vol. 2.

[112] Ulrich Munz and Pedro J. Zufiria, Diagnosis of unknown parametric faults in non-linear stochastic dynamical systems, *International Journal of Control*, Vol. 82, No. 4, April 2009, pp. 603-619.

[113] A. K. Nandi, Subsample time delay estimation of narrowband ultrasonic echoes, *IEEE Ultrason, Ferroelect, Freq. Contr.*, vol. 42, 1995, pp. 993-1001. 7

[114] B. Ni, S. L. Shah and D. Xiao, A Time Delay Estimation Method for MIMO Systems, *Proceedings of th 17th World Congress*, The International Federation of Automatic Control, Seoul, Korea, July 6-11, 2008, pp. 5539-5545.

[115] J. N. Nielsen, H. Madsen and H. Melgaard, Estimating Parameters in Discretely, Partially Observed Stochastic Differential Equations, *Technical report*, IMM, Department of Mathematical Modelling, Denmark, May 10, 2000.

[116] J. N. Nielsen, H. Madsen and Peter C. Young, Parameter estimation in stochastic differential equations: an overview, *Annual Reviews in Control*, Vol. 28, pp 83-94, 2000. 26

[117] J. N. Nielsen and M. Vestergaard, Estimation In Continuous-Time Stochastic Volatility Models Using Nonlinear Filters, *International Journal of Theoretical and Applied Finance*, Vol. 3, No. 2, 2000, pp. 279-308. 6

[118] J. N. Nielsen and H. Madsen, Applying the EKF to stochastic differential equations with level effects, *Automatica*, Vol. 37, 2001, pp. 107-112.

[119] H. Niemann and J. Stoustrup, Detection of Parametric Faults, *the 15th IFAC World Congress*, Barcelona, 21-26 July 2002.

[120] J. Nilsson, I. Skog and P. Händel, Joint State and Measurement Time-Delay Estimation of Nonlinear State Space Systems, *Invited Paper in 10th International Conference on Information Science, Signal Processing and their Applications*, 2010.

[121] R. D. Nowak, Nonlinear System Identification, *Circuits Systems Signal Processing*, Vol. 21, No. 1, 2002, pp. 109-122.

[122] F. Orderud, Comparison of Kalman Filter Estimation Approaches for State Space Models with Nonlinear Measurements, Sem Saelands vei 7-9, No. 7491 Trondheim.

[123] H. Ouarti and T. F. Edgar, The Use of Approximate Models and Exact Linearization for Control of Nonlinear processes, *Proc. Amer. Cont. Conf.*, 1993, pp. 2268. 11, 110

[124] L. Overbeck and T. Ryden, Estimation in the Cox-Ingersoll-Ross Model, *Econometric Theory,* Vol. 13, 1997, pp. 430-461. 60

[125] A. Pagan, Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients, *Journal of Econometrics 13*, 1980, pp. 341-363. 37, 38

[126] J. H. Park, H. Park and I. B. Lee, Closed-loop on-line process identification using a proportional controller, *Chemical Engineering Science*, Vol. 53, No. 9, pp. 1713-1724, 1998.

[127] H. G. Park and M. Zak, Gray-Box Approach for Fault Detection of Dynamical Systems, *Journal of Dynamic Systems, Measurement, and Control*, September 2003, Vol. 125, pp. 451-454.

[128] J. Parkum and C. Wagner, Identification and control of a dry-expansion evaporator, *10th IFAC symposium on system identification*, 01 January, 1994, pp. 539-544.

[129] R. J. Patton, Fault-Tolerant Control Systems: The 1997 Situation, *Proceedings Safeprocess 1997*, Hull, UK, 1997, pp. 1033-1054. 1997 8, 12

[130] R. J. Patton, P. M. Frank and P. N. Clark, Issues of fault diagnosis for dynamic systems, *Springer*, Berlin, 2000. 8

[131] B. L. Pence, H. K. Fathy and J. L. Stein, Recursive Maximum Likelihood Parameter Estimation for State Space Systems using Polynomial Chaos Theory, *Automatica*, Vol. 47, No. 11, November 2011, pp. 2420 C 2424.

[132] S. Qin and T. Badgwell, A survey of industrial model predictive control technology, *control Eng. Pract. 11*, 2003, pp. 733-764. 112

[133] J. R. Raol, G. Girija and J. Singh, Modelling and Parameter Estimation of Dynamic Systems, *The Institution of Electrical Engineers*, 2004.

[134] H. Rasmussen, Adaptive Superheat Control of a Refrigeration Plant using Backstepping, *International Conference on Control, Automation and Systems*, COEX, Seoul, Korea, Oct. 14-17, 2008, pp. 653-658.

[135] H. Rasmussen and L. F. S. Larsen, Nonlinear superheat and capacity control of a refrigeration plant, *17th Mediterranean Conference on Control and Automation*, Thessaloniki, Greece, Jun. 24-26, 2009, pp.1072-1077. 110, 158

[136] J. R. Richard, Time-delay systems: an overview of some recent advances and open problems, *Automatica*, Vol. 39, pp. 1667-1694, 2003.

[137] I. Rivals and L. Personnaz, Black-box modeling with state-space neural networks, *Neural Adaptive Control Technology*, World Scientific, 1996, pp. 237-264. 4

[138] M. Saccoman, S. Audoly, G. Bellu and L. D. Angio, Parameter Identifiability of Nonlinear Biological Systems, *Positive Systems, LNCIS 294, pp. 87-93, 2003. Springer-Verlag Berlin Heidelberg 2003*.

[139] T. B. Schön, A. Wills and B. Ninness, System identification of nonlinear state space models, *Automatica 47*, 2011, pp. 39-49. 35

[140] D. J. Scott (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Adv. Appl. Prob. 5* (119–137). 51

[141] D. E. Seborg, T. E. Edgar and D. A. Mellichamp, Process Dynamics and Control - 2nd edition, Wiley, USA, 2004. 112

[142] J. Shalchian, A. Khaki-Sedigh and A. Fatehi, A subspace based method for time delay estimation, *Proceeding of the 4th international symposium on communications*, Limassol, Cyprus, 3-5 March 2010, pp. 1-4.

[143] G. W. Shin, Y. J. Song, T. B. Lee and H. K. Choi, Genetic Algorithm for Identification of Time Delay Systems from Step Responses, *International Journal of Control, Automation, and Systems*, vol. 5, no. 1, February 2007, pp. 79-85.

[144] H. Singer, Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood versus Extended Kalman Filter and Itô-Taylor Expansion, *Journal of Computational and Graphical Statistic*, Vol. 11, No. 4, pp. 972-995, 2002. 33

[145] A. Sitz, U. Schwarz, J. Kurths and H. U. Voss, Estimation of parameters and unobserved components for nonlinear systems from noisy time series, *Physical Review E*, Vol. 66, 2002, pp. 1-9.

[146] T. Söderström and P. Stoica, Instrumental Variable Methods For System Identification, *Circuits Systems Signal Processing*, VOL. 21, No. 1, 2002, pp. 1-9.

[147] R. M. Staley and P. C. Yue, On System Parameter Identiflability, *Information Sciences 2*, 1970, pp. 127-138.

[148] D. S. Stoffer and K. D. Wall, Bootstrapping State Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter, *Journal of the American Statistical Association*, Vol. 86, No. 416, Theory and Methods, December 1991. 33

[149] J. Stoustrup and H. Niemann, Fault Detection and Isolation in Systems with Parametric Faults, *Invited paper, Proceedings of the IFAC World Congress*, Vol. P, Beijing, China, July 1999, pp. 139-144.

[150] J. Stoustrup and H. Niemann, Fault Estimation-A standard Problem Approach, *International Journal of Robust and Nonlinear Control*, Vol. 12, 2002, pp. 649-673. 12

[151] T. Söderström and P. Stoica, Instrumental Variable Methods for System Identification, *Circuits Systems Signal Processing*, VOL. 21, No. 1, 2002, pp. 1-9. 138, 139

[152] Z. Sun and Z. Yang, Study of Nonlinear Parameter Identification Using UKF and Maximum Likelihood Method, *Proc. of 2010 IEEE Conference on Control Applications (CCA)*, YOKOHOMA, Japan, September 8-10, 2010, pp. 671-676. 70

[153] Z. Sun and Z. Yang, Time-Varying FOPDT System Identification with Unknown Disturbance Input, *Proceedings of the 2012 IEEE International Conference on Control Applications (CCA)*, IEEE Press, 2012, pp. 364-369. 147, 148

[154] S. W. Sung, I. B. Lee and B. K. Lee, Sung, On-line process identification and automatic tuning method for PID controllers, *Chemical Engineering Science 53*, 1998, pp. 1847-1859. 112

[155] K. R. Sundaresan and P. R. Krishnaswamy, Estimation of time delay, time constant parameters in time, frequency and Laplace domains, *Canadian Journal of Chemical Engineering* 56, 1978, pp. 257 C 262. 112

[156] Y. Tan, Time-Varying Time-Delay Estimation For Nonlinear Systems Using Neural Networks, *Int. J. Appl. Math. Comput. Sci.*, 2004, Vol. 14, No. 1, pp. 63-68.

[157] Y. Z. Tsypkin, E. D. Avedyan and O. V. Gulinskiy, On Convergence of the Recursive Identification Algorithms, *IEEE Transactions on Automatic Control*, Vol. AC-26, No. 5, OCTOBER 1981.

[158] M. C. VanDyke, J. L. Schwartz and C. D. Hall, Unscented Kalman Filtering for Spacecraft Attitude State And Parameter Estimation, *American Astronautical Society*, 2004, No. AAS-04-115.

[159] M. S. Varziri, Parameter Estimation in Nonlinear Continuoustime Dynamic Models With Modelling Errors and Process Disturbances, *Doctoral Thesis, Queen's University*, 2008.

[160] M. S. Varziri, K. B. McAuley and P. J. McLellan, Parameter and state estimation in nonlinear stochastic continuous-time dynamic models with unknown disturbance intensity, *The Canadian Journal of Chemical Engineering*, Vol. 86, 2008, pp. 827-837.

[161] V. Verdult, Nonlinear System Identification: A State-Space Approach, PhD thesis, University of Twente, 2002.

[162] M. Verhaegen and P. Dewilde, Subspace model identification, part i: The output-error state-space model identification class of algorithms, *International Journal of Control*, 56, 1992, pp. 1187-1210.

[163] M. Verhaegen and D. Westwick, Identifying MIMO Hammerstein systems in the context of subspace model identification methods, *International Journal of Control* 63 (2), 1996, pp. 331-349. 5

[164] A. Visioli, Practical PID control, Springer 2006. 6, 12, 111, 112

[165] Z. Xiang, R. Wang and Q. Chen, Fault Tolerant Control of Switched Nonlinear Systems With Time Delay Under Asynchronous Switching, *Int. J. Appl. Math. Comput. Sci.*, 2010, Vol. 20, No. 3, pp. 497-506.

[166] C. Yahya, On-line Fault Detection and Isolation for State Space Models Using a Combination of Extended Kalman Filters and the Standardized Innovation Approach, *Chemical Product and Process Modeling*, Vol. 2, Iss. 3, Article 27, 2007.

[167] Z. Yang and M. Blanke, the Robust Control Mixer Module Method for Control Reconfiguration, *Proceedings of the American Control Conference,* Chicago, Lllinois, June 2000, Vol. 5, 3407-3411.

[168] Z. Yang and D. L. Hicks, Multiple Objective Robust Control mixer Method for Synthesis of Reconfigurable Control, *Proceedings of the American Control Conference*, Vol. 6, Anchorage, AK, May 8-10, 2002, pp. 5092-5097.

[169] Z. Yang, M. Blanke and M. Verhaegen, Robust control mixer method for reconfigurable control design using model matching, *IET Control Theory Appl., Vol. 1, No. 1*, January 2007, pp. 349-357. x, 71, 73, 74, 75

[170] Z. Yang and D. M. A. Hussain, Mode Identification for Multi-Mode Switching Systems Based on Multi-Sampled Data, *Proceedings of 2007 International Symposium on Intelligent Signal Processing and Communication Systems,* Nov.28-Dec.1, 2007 Xiamen, China, pp. 176-179.

[171] Z. Yang, C. Andersen and R. Izadi-Zamanabadi, On-line Auto-Tuning of PI control of the superheat for a supermarket refrigeration system, *IEEE Multi-Conference on Systems and Control*, Denver, USA, Sept. 28-30, 2011, pp. 792-797. 110

[172] Z. Yang, Z. Sun and C. Anderson, Nonlinear FOPDT Model Identification for the Superheat Dynamic in a Refrigeration System, Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society, IECON11, IEEE Press, 2011. pp. 634-629. 160

[173] J. Yin, V. L. Syrmos and D. Y. Y. Yun, System Identification using Nonlinear Filtering Met hods with Applications to Medical Imaging, *Proceedings of the 39th IEEE Conference on Decision and Control Sydney, Australia December,* Vol. 4, 2000, pp. 3313-3318.

[174] P. C. Young, An Instrumental Variable Method for Real-time Identification of a Noisy Process, *Automatica*, Vol. 6, Pergamon Press, 1970, pp. 271-287.

[175] P. Young, Parameter estimation for continuous-time models-a survey, *Automatica*, 17, pp. 23-29, 1981.

[176] P. C. Young, Optimal IV Identification and Estimation of Continuous-Time TF Models, *15th Triennial World Congress, Barcelona, Spain*, 2002, pp. 1002.

[177] Y. Zhang, G. Dai, H. Zhang and Q. Li, A SVD-Based Extended Kalman Filter and Applications to Aircraft Flight State and Parameter Estimation, *Proceedings of the American Control Conference*, Vol. 2, Baltimore, Maryland, June 1994, pp. 1809-1813.

[178] Y. M. Zhang and J. Jiang, Bibliographical Review on Reconfigurable Fault-tolerant Control Systems, *IFAC Annual Reviews in Control*, 32(2), pp. 229-252, Dec. 2008. 71

[179] Z. Sun and Z. Yang, System Identification for Nonlinear FOPDT Model with Input Dependent Dead-Time, *Proceedings of the 15th International Conference on System Theory, Control and Computing*, ICSTCC 2011, IEEE Press, 2011. 12, 113

[180] W. X. Zheng, Identification of Noisy Input-Output Models Using the Least-Squares Based Methods, *Proceedings of the 47th IEEE Conference on Decision and Control Cancun, Mexico, Dec. 9-11*, 2008, pp. 3269-3274.

[181] Y. Q. Zou and W. P. Heath, Global convergence conditions in maximum likelihood estimation, *International Journal of Control*, Vol. 85, No. 5, May 2012, pp. 475-490.

[182] B. Øksendal, Stochastic differential equations, 5ed, *Springer*, 2000. 10, 20, 21, 22, 23, 24, 30

Appendix A

Personal CV

Name: Zhen Sun
Gender: Male
E-mail: zhensunsdu@gmail.com
Birthday: 30-December-1980
Education Background:

01.2011–06.2013:

- PhD student at Department of Energy Technology, Aalborg University Esbjerg, Denmark

12.2008–12.2010

- PhD student at Automation and Control, Department of Electronic Systems, Aalborg University Esbjerg, Denmark

09.2004–06.2007

- M.S. of Probability and Statistics, School of Mathematics and System Sciences, Shandong University, China

10.1999–07.2003

- B.S. of Information and Computation, School of Mathematics, Ocean University of China, China
- B.S. of Finance, School of Economy, Ocean University of China, China

Appendix B

Publication List

- Z. Sun and Z. Yang, Study of Nonlinear Parameter Identification Using UKF and Maximum Likelihood Method, *IEEE Conference on Control Applications Proceedings*, 10.2010, pp. 671-676.

- Z. Yang, Z. Sun and C. Anderson, Nonlinear FOPDT Model Identification for the Superheat Dynamic in a Refrigeration System, *Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society*, IECON11, IEEE Press, 2011. pp. 634-629.

- Z. Sun and Z. Yang, System Identification for Nonlinear FOPDT Model with Input Dependent Dead-Time, *Proceedings of the 15th International Conference on System Theory, Control and Computing*, ICSTCC 2011, IEEE Press, 2011, pp. 1-6.

- Z. Sun and Z. Yang, Application of Joint Parameter Identification and State Estimation to a Fault-Tolerant Robot System, *Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society*, IECON11, IEEE Press, 2011, pp. 498-503.

- Z. Sun and Z. Yang, Time-Varying FOPDT System Identification with Unknown Disturbance Input, *Proceedings of the 2012 IEEE International Conference on Control Applications (CCA)*, IEEE Press, 2012, pp. 364-369.

- Z. Yang and Z. Sun, Time-Varying FOPDT Modeling and On-line Parameter Identification, will appear in *Proc. of 13th IFAC Symposium on Large Scale Complex Systems: Theory and Applications*, LSS2013, 2013.
- Z. Sun and Z. Yang, Consistency Conditions of a Class of Nonlinear System Identification Using Combined UKF and ML Methods, *submitted to IEEE Conference on Decision and Control*, 2013.

This present report combined with the above listed scientific papers has been submitted for assessment in partial fulfilment of the PhD degree. The scientific papers are not included in this version due to copyright issues. Detailed publication information is provided above and the interested reader is referred to the original published papers. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty of Engineering and Science, Aalborg University.