UNIVERSITY OF BELGRADE

SCHOOL OF ELECTRICAL ENGINEERING

Nasar Aldian Ambark Shashoa

FAULT DETECTION AND ISOLATION IN THERMAL POWER PLANT STEAM SEPARATOR

doctoral dissertation

Belgrade,2013

UNIVERSITY OF BELGRADE

SCHOOL OF ELECTRICAL ENGINEERING

Nasar Aldian Ambark Shashoa

FAULT DETECTION AND ISOLATION IN THERMAL POWER PLANT STEAM SEPARATOR

doctoral dissertation

Belgrade,2013

UNIVERZITET U BEOGRADU

ELEKTROTEHNIČKI FAKULTET

Nasar Aldian Ambark Shashoa

DETEKCIJA I IZOLACIJA OTKAZA U SEPARATORU PARE TERMO-ENERGETSKIH POSTROJENJA

doktorska disertacija

Beograd, 2013

Mentor:

Prof. dr Željko Đurović
Univerzitet u Beogradu, Elektrotehnički fakultet

Članovi komisije:

Prof. dr Branko Kovačević
Univerzitet u Beogradu, Elektrotehnički fakultet

Prof. dr Vladimir Stevanović.
Univerzitet u Beogradu, Mašinski fakultet

Doc. dr Goran Kvaščev
Univerzitet u Beogradu, Elektrotehnički fakultet

Doc. dr Veljko Papić
Univerzitet u Beogradu, Elektrotehnički fakultet

Datum:_____

Acknowledgement

Naslov:

Detekcija i izolacija otkaza u separatoru pare termo-energetskih postrojenja

Sažetak:

Nedozvoljena devijacija najmanje jednog karakterističnog parametra ili osobine nekog sistema od njegovog uobičajnog ponašanja se može proglasiti greškom, odnosno otkazom u sistemu. Otkazi smanjuju efikasnost sistema, kvalitet proizvoda i ponekad mogu dovesti do potpunog zaustavljanja procesa, odnosno pada sistema. Ovakve pojave ne uzrokuju samo ekonomske gubitke već u nekim slučajevima mogu dovesti i do ljudskih žrtava. Rana detekcija otkaza može biti način sprečavanja ili smanjenja ranije pomenutih gubitaka. Stoga su monitoring sistema i detekcija otkaza postali esencijalni deo modernih sistema upravljanja. Kontrolni uređaji koji se u novije vreme eksploatišu da bi se unapredile performanse industrijskih procesa uključuju sofisticirane tehnike projektovanja digitalnih sistema i kompleksni hardver (ulazno-izlazni senzori, aktuatori, komponente i procesorske jedinice).

Da bi se smanjila verovatnoća pojavljivanja otkaza na ovoj opremi potrebno je projektovati sistem za automatsko nadgledanje procesa koji bi se koristio za što je moguće raniju detekciju i izolaciju otkaza. U poslednje tri decenije, problem detekcije i izolacije otkaza u dinamičkim procesima je privukao veliku pažnju stručne javnosti i razvijen je čitav spektar pristupa koji se baziraju na modelima sistema. Za projektovanje sistema za robusnu detekciju i dijagnozu otkaza je značajno poraslo interesovanje jer se korišćenjem istih mogu smanjiti gubici u proizvodnji, sprečiti kvarovi na opremi i povećati sigurnost osoblja. Pouzdan sistem je sistem koji ima sposobnost za:

1. Što raniju detekciju otkaza

2. Tačnu dijagnozu istih

3. Što brži povratak sistema u nominalni režim rada

U ovoj disertaciji je predložen novi pristup detekciji i identifikaciji otkaza u generatorima pare u okviru termo-energetskih postrojenja. Realni sistem, na kome su vršena merenja i predložen sistem za detekciju i izolaciju otkaza, je TEKO B1 blok

termoelektrane Kostolac u Srbiji, čija je nominalna snaga 330 MW. Za takvu prirodu procesa i dostupne podatke, implementirani sistem za detekciju i izolaciju otkaza predstavlja kombinaciju dva poznata pristupa, jedan koji se bazira na modelima procesa i drugi koji se bazira na merenjima. Analizirana su tri moguća tipa otkaza, u senzorima nivoa vode, protoka vode i protoka pare. Prvi korak u predloženom algoritmu je identifikacija procesa. Merenje nivoa vode u separatoru su izvršena pod ekstremno visokim pritiscima pare uz nestacionarni dotok vode i ispuštanje pare. Stoga su dostupni podaci o nivou vode veoma nepouzdani, a prisutan je i sporadični, ali veoma jak, šum merenja.

Kao posledica toga, standardne procedure za identifikaciju sistema nisu mogle da daju zadovoljavajuće rezultate i u ovoj tezi je predložena alternativa koja predstavlja robusnu estimaciju parametara sistema. U sledećem koraku je izvršeno statističko testiranje hipoteza pri čemu se ne koriste vektori merenja već se vektori parametara sistema izvode na osnovu procedure robusne identifikacije. Rezultati pokazuju izuzetnu efikasnost u detekciji i izolaciji otkaza jednog od moguća tri tipa otkaza koji su analizirani.

Pristup detekciji otkaza koji je predložen u ovoj tezi, u poređenju sa rešenjima dostupnim u literaturi, je specifičan u sledećem:

1. Zasnovan je na linearnom modelu procesa sa odgovarajućim brojem ulaza i izlaza koji se identifikuje u zatvorenoj sprezi.

2. Predložena tehnika identifikacije je robusna po svojoj prirodi što je veoma važno u slučajevima kada je prisutan sporadičan jako izražen šum merenja.

3. Otkazi nisu detektovani na osnovu reziduala u merenjima, što je uobičajeni pristup u literaturi, već na osnovu parametara identifikovanog modela.

4. Predloženi metod za detekciju i izolaciju otkaza je kombinacija tehnika na bazi modela i na bazi merenja tako da je moguće napraviti kompromis između verovatnoće pojavljivanja lažnog alarma i vremenskog kašnjenja u detekciji otkaza.

5. Predloženi metod je primenjen na realnom sistemu separatora u termoelektrani i pokazao se kao veoma efikasan.

Ključne reči:

Detekcija i izolacija otkaza, separator pare, identifikacija sistema, Neyman-Pearson test

Naučna oblast:

Tehničke nauke, Elektrotehnika

Uža naučna oblast:

Automatika

UDK broj:

621.3

Title:

Fault detection and isolation in thermal power plant steam separator

Abstract:

An un-permitted deviation of at least one characteristic property or parameter of a system from standard condition is referred as a fault. Faults result in reduced efficiency of the system, reduced quality of the product, and sometimes complete breakdown of the process. This not only causes economic losses but may also result in fatalities. An early detection of faults can assist to avert these losses. Therefore, fault detection and process monitoring is becoming an essential part of modern control systems. The control devices which are nowadays exploited to improve the overall performance of industrial processes involve both sophisticated digital system design techniques and complex hardware (input-output sensors, actuators, components and processing units).

In such a way, the probability of failure occurrence on such equipment may result significant and an automatic supervision control should be used to detect and isolate anomalous working conditions as early as possible. Since the last three decades, the problem of fault detection and isolation in dynamic processes has received great attention and a wide variety of model-based approaches has been proposed and developed. A robust fault detection and diagnosis (FDD) system design has attained increased attention for reducing production loss, avoiding equipment damage, and increasing human safety. A more dependable system is a system that has the ability to:

    1**.** Detect faults as fast as possible.

    2**.** Diagnose them accurately.

    3**.** Recover the system to the nominal performance as much as possible.

This thesis presents a fault detection and identification approach for steam generators at thermal power plants. The real system in which the measurements were made and the FDI algorithm implemented is located at the TEKO B1 Unit of the Kostolac Thermal Power Plant in Serbia, whose nominal power output is 330MW. Given the nature of the process and available data, the implemented FDI algorithm is a trade-off of sorts between the model-based and the data-driven approach. Three possible types of faults

are analyzed, of water level, water flow and steam flow sensors. The first step of the proposed algorithm is to identify the process. water level measurements in a separator are conducted under extremely high steam pressures, accompanied by constant unsteady water inflow and steam drain. therefore, available water level data are high unreliable and there is a sporadic of high-intensity measurement noise.

As a consequence, standard process identification procedures have been shown not to yield satisfactory results and the thesis proposes a robust alternative to parameter estimation. The next step included statistical testing of the hypotheses, not using the measured data vector but the parameter vector derived from a robust identification procedure. The results demonstrated exceptional detection and isolation efficiency of one of the three possible and most frequent faults which were analyzed.

The approach to sensor fault detection proposed in this work, compared to the solutions reported in the literature, is specific in the following respects:

1**.** It is based on a linear model of the process, with the corresponding number of inputs and outputs, identified in a closed loop.

2**.** The proposed identification technique is robust by its very nature, which is very important in the case of systems where sporadic high-intensity measurement noise is present.

3**.** Faults are not detected based on measurement residuals, which is the usual approach in the literature, but based on the parameter vectors of the identified model.

4**.** The proposed fault detection and isolation method is a combination of model-based techniques and the data-driven approach, such that a simple trade-off is possible between the probability of false alarm and the fault detection time delay;

5**.** The proposed method was applied in a real steam separator at a thermal power plant and demonstrated as highly efficient.

Keywords:

Fault detection and isolation algorithms, steam separators, systems identification, Neyman-Pearson test

Scientific Area:

Electrical Engineering

Specific Scientific Area:

Control Engineering

UDK number:

621.3

Table of Contents

Chapter Three

Pattern Recognition Methods

Chapter Four

Description of Thermal Power Plants

Chapter Five

Sensor Fault Detection and Isolation in Thermal Power Plants Steam Separator

List of Figures

CHAPTER ONE

INTRODUCTION

1.1 Background

Ever since the first stone tool was invented man has always been concerned about the condition of the machines he uses. For the major part of the human history the only way to learn about the malfunctions and their locations was by the five human senses for example touching to feel heat or vibration, smelling for fumes from overeating etc. With the passage of time the machines have been developing and modern control systems becoming more and more complex, usually including large number of components such as (sensors, actuators, computers, etc). However, the importance of product quality, safety and reliability is increasing in the industrial processes. In any industrial process, it is essential that maintenance is provided to ensure that the equipment runs safely, reliably and normally.

Properly maintained industrial plants have significant benefits, such as higher productivity, equipment which has a longer lifespan and, as a consequence, lower production costs. An effective and efficient maintenance plan requires that information concerning the condition of the equipment can be accessed on a timely basis. In the early 19th century, maintenance was only carried out following a failure as there was a lack of means to understand the status of machinery. Since that time, routine maintenance has been performed in order to find and fix problems before a fault occurs. However, time period based maintenance inspection is still not sufficient, particularly for incipient faults. With the development of electronic technology, a low-cost, on-line condition monitoring system has become realistic for industrial applications. Predictive maintenance is, therefore, achievable via deliberated fault detection and diagnosis (FDD) algorithms.

Fault detection and diagnostic (FDD) systems have a variety of potential uses, including monitoring the performance of the equipment over time, alerting users of the presence of a fault, improving the quality of service by identifying the cause of faulty operation, or even changing the operation of the equipment to reduce the effect of the fault. The functionality of these FDD systems can thereby greatly improve the performance of industrial process and assist the service technician in expeditiously fixing and maintaining equipment. So for any application, the benefits of FDD can be divided into two general categories: improved safety and reduced cost to operate.

Interest in the theoretical and experimental development of FDD methods in general is growing, due to the dependence of society on systems of increasing complexity. These systems can be found nearly everywhere in the more recently, from power plants to airplanes to consumer electronics.

The traditional attitude of the designer, that the system will function reliably for a long period of time if the design is "good enough", is slowly giving into the realization that many systems will be operated in conditions unforeseen by the designer, and that the cost of faults in such systems can sometimes be greater than the system itself. The implementation of FDD and condition monitoring systems is vital in such circumstances, to ensure that the system will continue to function well in the face of considerable operative uncertainty.

1.2 Objectives

The idea of parameter estimation approach fault detection is to generate estimations of parameters within a dynamic system, and compared with the parameters of the reference model to generate a symptom, called the residual, which carries the information of faults. In ideal situations, when there are no modeling uncertainties, disturbances, and noise, the estimations will completely match with the measurements in fault-free case and the residual will be zero. Any deviation of residual from zero will give an indication of faults. However, the presence of modeling uncertainties, disturbances, and noise is inevitable.

Therefore, the aim is to design adaptive recursive M-robust FDD such that the affect of modeling uncertainties, disturbances, and noise on the residual is reduced while the affect of faults is considerably increased. Now, instead of setting deviation of residual from zero as indicator of faults, a threshold which cares for the effect of modeling uncertainties, disturbances, and noise should be selected and if the residual exceeds the selected threshold, it gives an indication of the presence of faults. The selection of threshold plays a very important role in the performance of a fault detection system, if it is selected too low, some of modeling uncertainties, disturbances, and noise will cause the residual to cross the threshold and appear as faults, this is definitely not desired. Conversely, if the threshold is selected too high, some of the faults will not enable the residual to cross the threshold, and hence will remain undetected.

To design fault detection and isolation approach for separators in thermal power plant, the objectives with this thesis work are

- Design a robust version of recursive identification of linear dynamic discrete-time systems in the presence of non-Gaussian impulsive noise within a measurement sequence, shown to be rather insensitive to outliers.

- Classify the parameter vectors of the identified model. To do this step, dimension reduction of initial parameter vectors into two-dimensional space Based on Scattering Matrices is performed and a suitable choice of classifier selection is chosen and designed (the Neyman-Pearson type).

- The Choice of the Neyman-Pearson type gives chance to plot the relation between time delay of the transition of the reduced vector from the nominal mode class to the steam fault class with μ. This figure showing that a trade-off between the probability of false alarm ε_0 and the time delay between fault occurrence and detection.

1.3 Thesis structure

The work, completed to achieve the above objectives, is presented in this thesis, and the structure is outlined as follows:

Chapter 1: Introduction to the study. The background and objectives of the project are presented.

Chapter 2: Overview of Fault Detection and Diagnosis Techniques

This chapter introduces the terminology used in the field of fault detection and diagnosis. Types of faults are described in detail. An overview is provided of fault detection and diagnosis methodology. Model-free fault detection methods are presented; physical redundancy, special sensors, limit checking, spectrum analysis and logical reasoning are reviewed. Model-based fault detection methods are also presented. The three main methods of model based residual generation are: parameter estimation, parity relation and observer-based approaches are discussed. Fault diagnosis methods are explained briefly. Finally, robustness in fault detection system is illustrated.

Chapter 3: Pattern Recognition Methods and Their Application in FDD.

The definition of pattern recognition is reviewed. Bayesian classification methods are introduced; the bayes decision rule for minimum error, the bayes decision rule for minimum cost, The neyman - pearson test, minimax criterion and Sequential hypothesis testing are analyzed, and simulation program for every method illustrate application of these methods fault detection and diagnosis.

Parametric classifier methods are presented; Firstly, linear discriminant function methods, two methods of linear discriminant function are analyzed, Optimum design procedures and we describe two procedures (Fisher's linear discriminant and Optimum design for normal distributions) and other desired outputs and search techniques. Next Quadratic classifier design is analyzed. Also simulation program for every method illustrate application of these methods fault detection and diagnosis.

Non Parametric Density Estimation is illustrated; two methods of Non Parametric Density Estimation are also analyzed, histogram method and kernel density estimation method and simulation program for every method illustrate estimation of density function. Feature selection methods are discussed, the discrete karhunen-Loeve expansion and scatter Matrices and separability criteria for feature selection are described. Simulation program for every method show the three dimensions reduce to two dimensions.

Chapter 4: Description of Thermal Power Plants

Fuel and furnace section is discussed; coal conveyor, stoker and pulverizer are explained. Boiler section is described, classification of boilers, superheater, reheater, economizer, condenser and air path are presented. Steam turbine section is reviewed, classification of steam turbine, casings, nozzles, rotors, bearings, shaft packing glands, seam turbine performance and steam turbine generators are illustrated. stack gas path and cleanup is described, air preheater, ash disposal and smoke and dust removal are explained. Cooling towers are presented.

Chapter 5: One Approach to Steam-separator Fault Detection and Isolation in Thermal Power Plants

Some basic concepts is presented, system identification, An outlier, maximum likelihood estimation, adaptive control, robustness in fault detection system and Quantile-Quantile (QQ)-plot technique are overviewed.

Robust adaptive parameter identification is proposed, formulation of the problem, recursive M-robust parameter estimation, review of the QQ-plot technique, new adaptive M-robust recursive algorithm, algorithm description and details of the algorithm steps are explored and simulated.

Description of the Case Study (Steam Separator) and Identification Procedure is shown, overview, system description and data acquisition and process modeling and model validation are described.

Fault detection based on parametric process identification parameters is performed. Dimension reduction based on the discrete karhunen-loeve expansion and scattering matrices and separability criteria are analyzed. Also, Classifier selection and design is performed. Finally, Statistical change detection and isolation is demonstrated.

Chapter 6: Conclusions of this study and discussion of further work.

CHAPTER TWO

OVERVIEW OF FAULT DETECTION AND DIAGNOSIS TECHNIQUES

2.1 Introduction

For the improvement of reliability, safety and efficiency advanced methods of supervision, fault-detection and fault diagnosis become increasingly important for many technical processes. This holds especially for safety related to power plants and chemical plants. FDD plays a vital role to provide information on faults/failures in the system and to enable appropriate reconfiguration to take place. Therefore the main function of FDD is to detect a fault or failure and to find its location so that corrective action can be made to eliminate or minimize the effect on the overall system performance. Initial FDD applications in chemical and industrial plants used threshold testing to check system data. Using this method, a fault can be detected when a measured value crosses a given threshold. This classical limit-value-based method is simple and reliable; however, it only responds to a relatively large change to a feature, therefore a detailed fault diagnosis becomes impossible [1].

The classical approaches are limit or trend checking of some measurable output variables based on hardware or physical redundancy methods which use multiple sensors, actuators, components to measure and control a particular variable. Typically, a voting technique is applied to the hardware redundant system to decide if a fault has occurred and its location among all the redundant system components. The major problems encountered with hardware redundancy are the extra equipment and maintenance cost, as well as the additional space required to accommodate the equipment. With increasing system complexity and requirements for reliability, a quantitative model of a practical system was required and many investigations were therefore made using analytical approaches or functional redundancy.

In the analytical redundancy scheme, the resulting difference generated from the comparison of different variables is called a residual or symptom signal. In brief the residual should be zero when the system is in normal operation and should be different from zero when a fault has occurred. This property of the residual is used to determine whether or not faults have occurred. The idea was to generate signals that represent inconsistencies between normal and faulty system operation. Based on analytical model, the algorithms, such as observers, parity equations and parameter estimation, were designed for inconsistency signal generation (also known as residuals generation).

These model-based FDD methods have been widely implemented in many industrial fields, such as nuclear power plants, railway vehicles, jet engines, power plants and chemical plants [2].

2.2 Overview

2.2.1 System

System is a collection of objects arranged in an ordered form to serve some purpose. Everything not belonging to the system is part of the environment.

One may characterize the system by input-output (cause and effect) relations. What constitutes a system depends on the point of view of the observer. The system may be, for example, an amplifier consisting of electronic components, or a control loop including that amplifier as one of its parts, or a chemical processing unit having many such loops, or a plant consisting of a number of units or a number of plants operating together as a system in the environment of a global economy [3].

2.2.2 Process

A process is a processing plant that serves to manufacture homogeneous material or energy products. Industries that use such processing plants are called process industries. The common process industries are oil, chemicals, electrical power, glass, mining, metals, drugs, and food. From a control point of view, different kinds of variables in a process interact and produce observable variables.

The observable variables of interest to us are usually called outputs. The process is also affected by external variables. External variables that can be manipulated by us are inputs of the process. Other external variables are called disturbances. Then, a process is said to be dynamic when the current output value depends not only on the current external stimuli but also on their earlier values [3].

2.2.3 Model

A model is a representation of the essential aspects of a system (process) which presents knowledge of that system (process) in a usable form.

There are many types of models. People are most familiar with mental models which do not involve any mathematical formalization. To drive a car, for example, the driver has

a mental model about the relationship between the turning of the steering wheel and the turning of the car direction, and between the accelerator and the acceleration of the car.

For manual control of an industrial process, the process operator needs the knowledge about how the process outputs will respond to various control actions. Sometimes it is appropriate to describe the properties of a system by means of tables or plots. Such descriptions are called graphical models. Bode plots, step responses and impulse responses of linear systems are of this type.

For the application of modern systems and control theory, it is necessary to use mathematical models that describe the relationships among the system variables in terms of difference or differential equations. In fact, the use of mathematical models is not limited to the control community; a major part of the engineering field deals with the use of mathematical models for simulations and designs [3].

2.3 Terminology

Before moving further it is advisable to exactly define the terms related to fault detection which will be used again and again in this work.

The definitions of terms commonly used in the fault detection and diagnosis field.

These definitions are based on information obtained from the SAFEPROCESS Technical Committee is considered "on-going "in the sense that new definitions and updates are being made.

2.3.1 States and Signals

- **Fault:** A non-permitted deviation of at least one characteristic property or parameter of the system from acceptable/usual/standard condition.[4]

- **Failure:** a permanent interruption of a system's ability to perform a required function under specified operating conditions.[5]

- **False Alarm:** is an indication of a fault, when in actuality a fault has not occurred [6].

- **Malfunction:** An intermittent irregularity in the fulfillment of a system's desired function.[3]

- **Error:** A deviation between a measured or computed value of an output variable and its true or theoretically correct one.[3]

- **Disturbance:** An unknown and uncontrolled input acting on a system.

- **Missed Detection:** when there is not indication of a fault, through a fault has occurred.[6]

- **Residual:** A fault indicator, based on a deviation between measurements and model based computations.[6]

2.3.2 Functions

- **Fault Detection:** determination of faults present in a system and time of detection.

- **Fault Isolation:** determination of kind, location and time of detection of a fault. It follows fault detection.

- **Fault Identification:** determination of size and time-variant behavior of a fault. It follows fault isolation.

- **Fault Diagnosis:** determination of kind, size, location and time of a fault. It follows fault detection and includes fault isolation and identification.

- **Monitoring:** A continuous real-time task of determining the conditions of a physical system, by recording information, recognizing and indication anomalies in the behavior.

- **Supervision:** Monitoring physical and taking appropriate actions to maintain the operation in the case of fault [5].

- **Residual Computation:** residual value is computed from the known variable.

- **Residual Evaluation:** the residual is evaluated in order to detect, isolate and identify faults.

- **Protection:** Means by which a potentially dangerous behavior of the system is suppressed if possible, or means by which the consequences of a dangerous behavior are avoided.

11

2.3.3 System Properties

- **Reliability:** ability of a system to perform a required function under stated conditions, within a given scope, during a given period of time.

- **Safety:** ability of a system to not cause danger to persons or equipment or the environment.

- **Availability:** probability that a system or equipment will operate satisfactorily and effectively at any point in time [4].

- **Dependability:** A system that has a high reliability in terms of high availability and where the consequences of a fault are limited to the system it self, i.e. Local faults do not developed into failure at plant level [7].

2.3.4 Types Of Faults

The types of faults depend basically on their location within the system, the number of components that can be affected and their temporal evolution. Taking into account the effects of the faults, so the classification of faults are based on

2.3.4.1 Location In The Physical System

Depending on whether the fault is located in the sensor, actuator or in one of the components, we have the sensor fault, actuator fault or the component fault respectively.

- **Sensor Faults**: in closed loop systems, the measurements obtained by sensors are used to generate the control inputs and any fault in sensors can cause operating points that are far from the optimal ones. This results in degradation in the performance of the system. It is therefore, very important to detect these faults. Typical examples of sensor faults are: bias, drift, performance degradation (or loss of accuracy) and calibration error as illustrated in figure 2.1. Solid lines show the actual values whereas the dotted lines show the measured values [7].
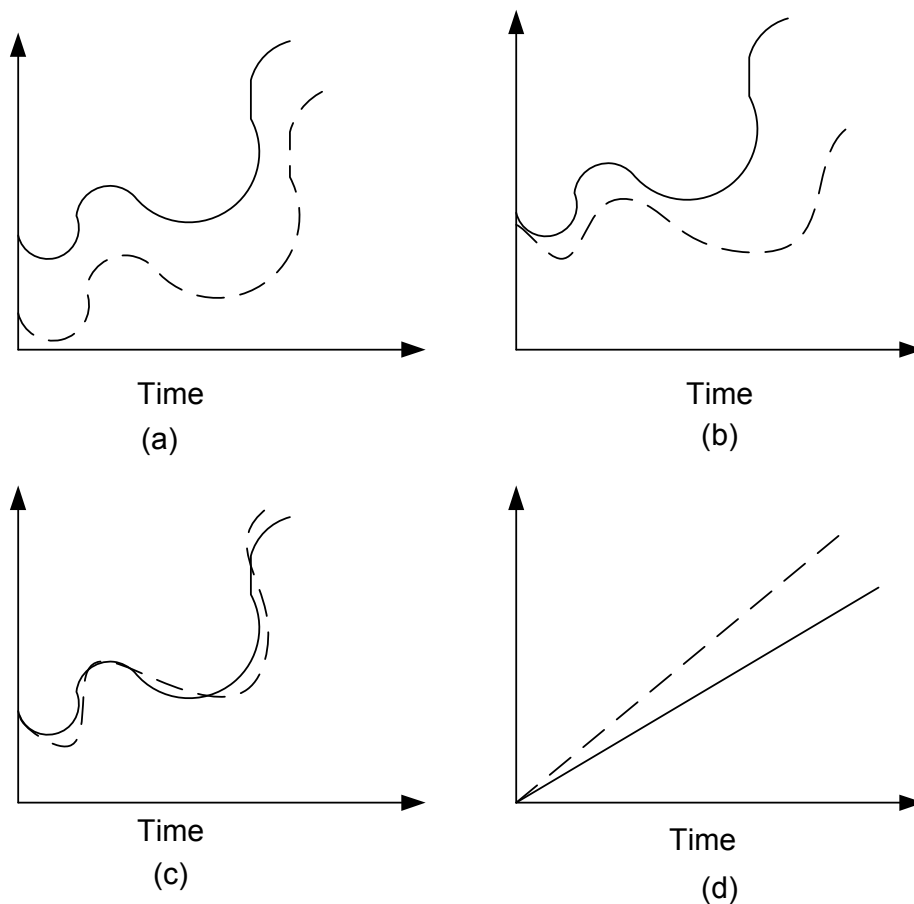
Figure 2.1: Graphical depiction of different kinds of sensor faults.

(a) Bias, (b) Drift, (c) Loss of accuracy, (d) Calibration error

- **Actuator Faults**: actuators are needed to transform control signals into proper actuation signals such as torques and forces to drive the system.· Actuator Faults represent partial or complete loss of control action. Total actuator fault can occur, for instance, as a result of a breakage, cut or burned wiring. This is an example of a completely lost actuator (stuck actuator). Partially failed actuator produces only a part of the normal (i.e., under nominal operating condition) actuation. It can result from, e.g., hydraulic or pneumatic leakage, increased resistance or fall in the supply voltage [9].

- **Component Faults:** these are the faults which appear in the components of plant. Components faults alter the physical parameters of the plant which, in

turn, results in change of its dynamical properties. The common reason for these faults is usually wear and tear, aging of components etc. some examples of components faults are leakages in tanks, breakages or cracks in gearbox system, change in friction due to lubricant deterioration etc. components faults may result instability of the process, therefore, it is extremely important to detect these faults [8].

2.3.4.2 Mathematical Properties

Depending on whether the faults are additive faults (those which correspond to sensor and actuator faults) or multiplicative faults (those which correspond to components faults):

- **Additive Process Faults**. These are unknown inputs acting on the plant, which are normally zero and which, when present, cause change in the plant outputs independent of the known inputs, such faults best describe plant leaks, loads, etc

- **Multiplicative Process Faults**. These are changes (abrupt or gradual) in some plant parameters. They cause changes in the plant outputs which depend also on the magnitude of the known inputs. Such faults best describe the deterioration of plant equipment, such as surface contamination, clogging, or the partial or total loss of power [10].

2.3.4.3 The Time Dependency of Faults

Faults can also be categorized according to whether these have developed slowly in the system (incipient faults), arisen suddenly like a step change (abrupt faults) or occurred in discrete intervals (intermittent faults)

- **Abrupt Faults:** These are faults that appear abruptly in a time instant. For example in a power supply break down [6]. Abrupt faults have more severe effects and may result in damage of equipments.

- **Incipient Faults:** these are faults that increase gradually changes from the nominal values to the faulty values [6]. Incipient faults grow slowly and result in degradation of equipments.

- **Intermittent Faults:** If the faults term change from the nominal value to the faulty value and returns to the nominal value after a short period of time then it is called intermittent fault [11].

Abrupt, Incipient and Intermittent faults are shown in figure 2.2.
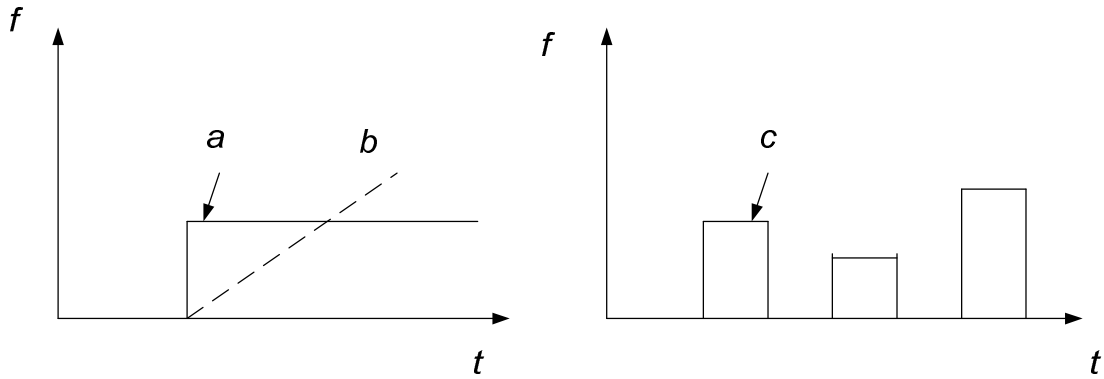
Figure 2.2: Time- dependency of faults: (a) abrupt; (b) incipient; (c) intermittent

The detection performance of the diagnostic technique is characterized by a number of important and quantifiable benchmarks namely

- **Fault Sensitivity:** The ability to detect faults of reasonably small size.

- **Reaction Speed:** The ability of the technique to detect faults with reasonably small delay after their arrival.

- **Robustness:** The ability of the technique to operate in the presence of noise, disturbances and modeling errors, with few false alarms.

Isolation performance is the ability of the diagnostic system to distinguish faults depends on the physical properties of the plant, on the size of faults, noise disturbances and model errors, and on the design of the algorithm. The tasks to be performed in the fault detection and diagnosis can be shown by the following diagram figure 2.3 [11].
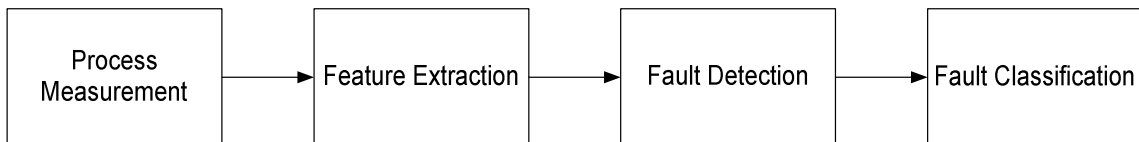
Figure 2.3: The fault and Isolation Task

2.3.5 Approaches to Fault Detection and Diagnosis

The methodology used in fault detection and diagnosis is clearly dependent on the process and the kind of available information [6]. The methods of fault detection and diagnosis may be classified into two major groups: first group which do not utilize the mathematical model of the plant (the model-free techniques) and the next which do (the model-based methods). They are described in the following sections.

2.4 Model-Free Fault Detection Methods

This fault detection and isolation method does not use the mathematical model of the plant range from physical redundancy and special sensors through limit-checking and spectrum analysis to logical reasoning. Some of the prominent model-free methods are as follows:

2.4.1 Physical Redundancy

In this approach, multiple sensors are installed to measure the same physical quantity. Any serious discrepancy between the measurements indicates a sensor fault. One of the drawbacks of the physical redundancy method is that it leads to extra hardware costs and extra weights.

2.4.2 Special Sensors

Sometimes special sensors may be installed explicitly for detection and diagnosis. These may be limit sensors (measuring e.g., temperature or pressure), other special sensors may measure some fault-indicating physical quantity, such as sound, vibration, elongation, etc [10].

2.4.3 Limit Checking

The most simple and frequently used method for fault detection is the limit checking of a directly measured variable $Y(t)$. The measured variables of a process are monitored and checked if the measured variable exceeds the limit of threshold, it gives indication of fault. Limit checking of absolute values of the measurements and the limit checking of derivative (trend) of the measurements are the two most simple and widely used approaches for fault detection [5].

2.4.3.1 Limit Checking of Absolute Values

Generally, two limit values called thresholds are present a maximal value Y_{max} and minimal value Y_{min}. A normal state is when

$$Y_{min} < Y(t) < Y_{max} \qquad \Rightarrow No \quad Fault \qquad (2.1)$$

This means that the process is in normal situation if the monitored variable stays within a certain tolerance zone exceeding of one of the thresholds then indicates a fault somewhere in the process as shown in figure 2.4.·

$$Y(t) < Y_{min} \quad or \quad y(t) > Y_{max} \qquad \Rightarrow Faulty \qquad (2.2)$$
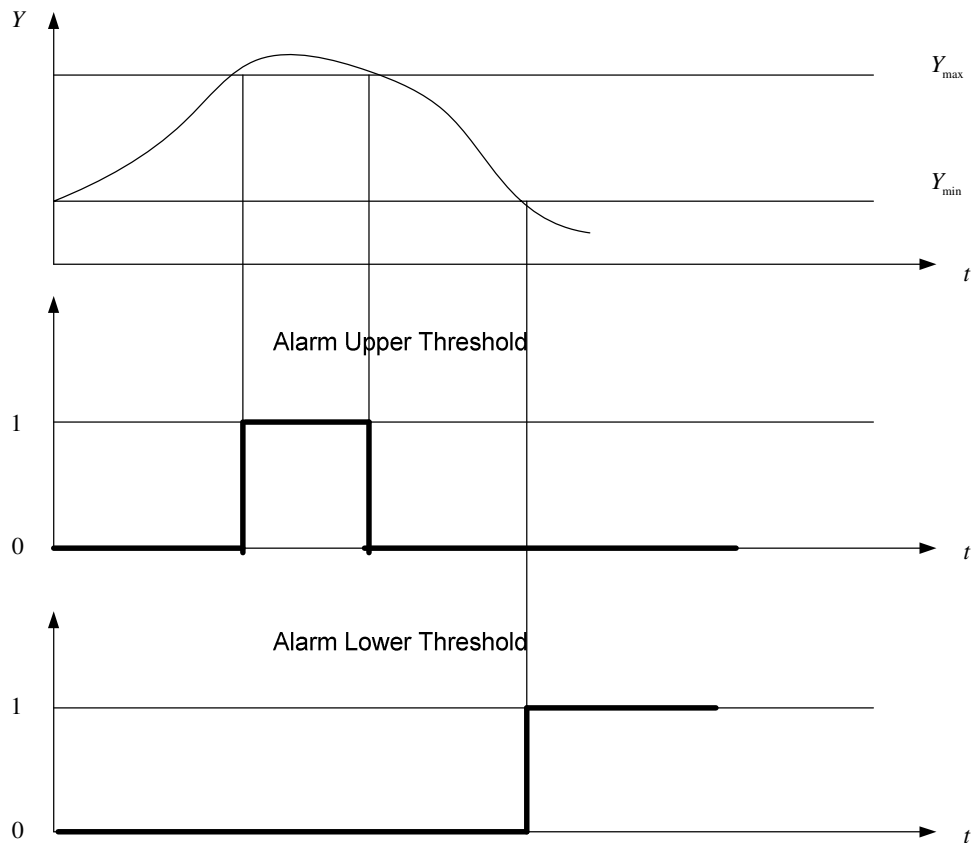
Figure 2.4: Limit Checking of Absolute Values

The thresholds are mostly selected based on experience and represent a compromise.

If selected too narrow some fluctuations and disturbances will cause an alarm of fault and if selected too wide, some of the small magnitude faults may not be detected.

Therefore a trade-off between too narrow and too wide threshold exists [5].

2.4.3.2 Limit Checking of Trend

A further simple possibility is to calculate the first derivative $\dot{Y} = {dY(t)}/{dt}$, the trend of the monitored variable and to check if

$$\dot{Y}_{min} < \dot{Y}(t) < \dot{Y}_{max} \tag{2.3}$$

$$\dot{Y}(t) < \dot{Y}_{min} \quad or \quad \dot{Y}(t) > \dot{Y}_{max} \qquad \Rightarrow Faulty \tag{2.4}$$

If relatively small thresholds are selected, an alarm can be obtained earlier than for limit checking of the absolute value as shown in figure 2.5 [12].
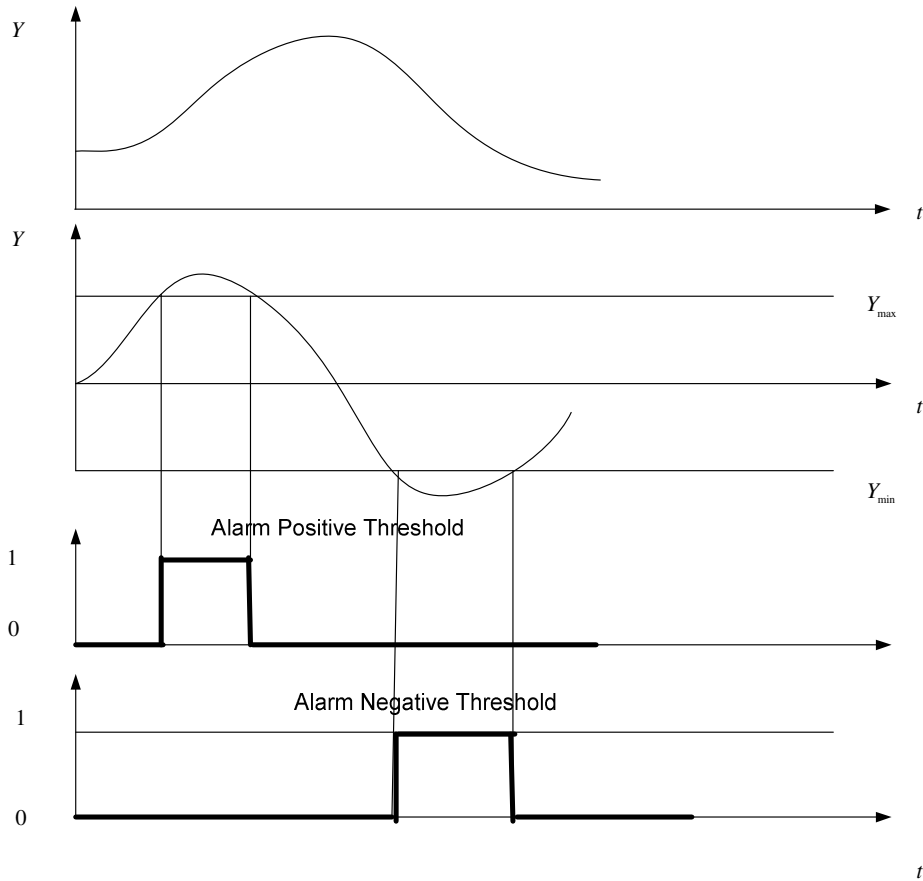
Figure 2.5: Limit checking of trend

The advantage of Limit checking approach is simple, reliable and can be easily implemented for steady-state situations. However, the drawback is that faults can be

detected only when these grow large enough to cross the limit. This may cause more damage to the process as compared to that if it was detected earlier and suitable remedies had been taken [7].

2.4.4 Spectrum Analysis

Analysis of the spectrum of the measured plant variables may also be used for detection and isolation. Most plant variables also exhibit a typical frequency spectrum under normal operating conditions. Any deviation from this is an indication of the abnormality. Some type of faults may also have their own characteristic signature in the spectrum, facilitating fault isolation [10].

2.4.5 Logic Reasoning

Logical reasoning techniques form a broad class which is complementary to the methods outlined above, in that they are aimed at evaluating the symptoms obtained by the detection hardware or software. The system may process the information presented by the detection hardware/software or may interact with a human operator inquiring from him about the particular symptoms and guiding him through the entire logical process [10].

2.5 Model-Based Fault Detection Methods

Model-based approaches to fault detection in dynamic systems have been received much attention over the last decades, both in research context and in the domain of application studies on real plants. The aim of model-based fault diagnosis is to generate information about faults which have occurred in target systems using actual measurements. The model-based method is referred to as an analytical redundancy, which is low-cost compared to hardware redundancy in some safety-critical applications, provided that a model can precisely simulate the behavior of a real system. Typically, the target system is considered as a continuous-variable dynamic system, which has an input U and an output Y, the detection methods generate residuals r, parameter estimates $\hat{\Theta}$, which are called features.

Consistency checking in analytical redundancy is normally achieved through a comparison between measured signals with estimated values. The estimation is generated by a mathematical model of the considered plant. The comparison is done

using the residual quantities which are computed as differences between the measured signals and the corresponding signals generated by the mathematical model. In theory, the residuals must be either zero in a fault free case, to indicate that no fault occurs, or non-zero in the case of a fault. However, in practice, deviations normally exist with different magnitudes. A threshold is, therefore, required for sensitivity adjustment. The value at which a threshold is set determines whether the FDD system has enough sensitivity to detect a fault or not. The balance is a trade-off between detection accuracy and false alarm.
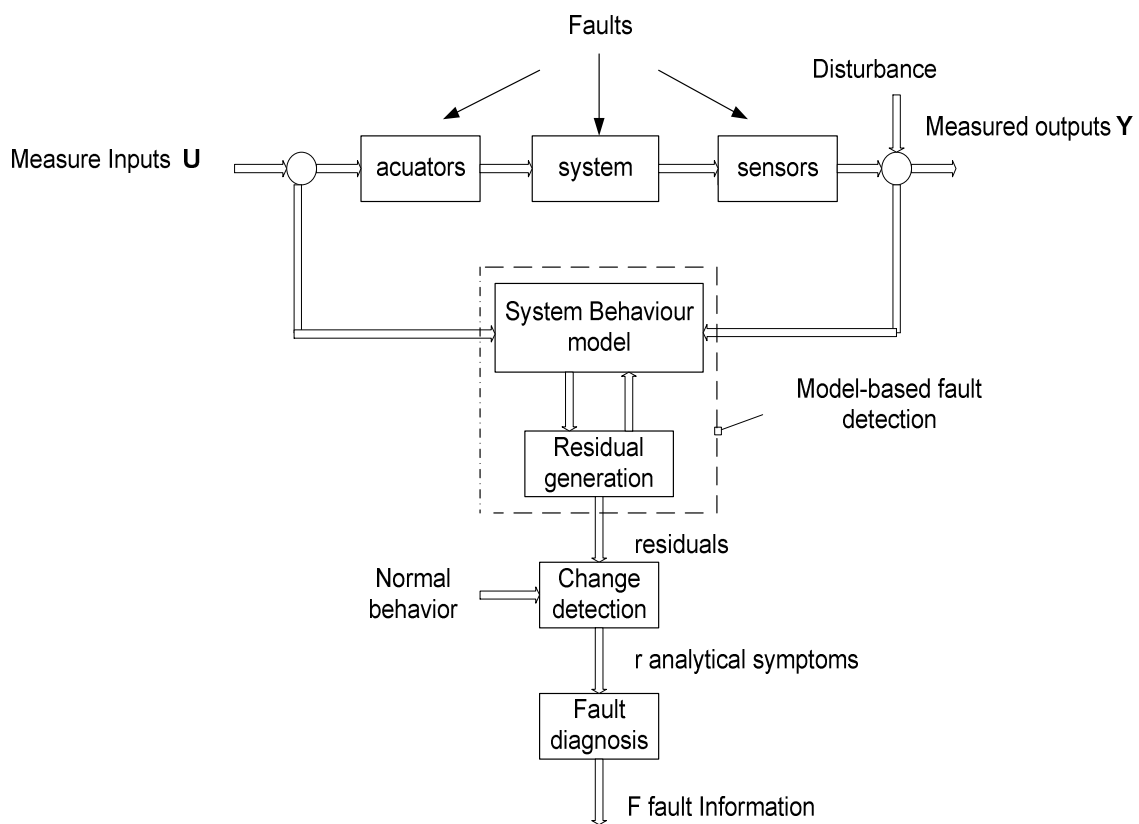
Figure 2.6: General scheme of process modal-based fault-detection and diagnosis

Figure 2.6 shows a general scheme of model-based fault detection and diagnosis. In this figure, the whole system consists of actuators, the target system and sensors, where the faults can be grouped as actuator faults, system faults or sensor faults. Disturbance (noise) is added on the sensor output. Both input and output are physical measurements, which are compared with the prediction from the system behavior model for residual generation. The residual generator aims to produce a set of inconsistencies to indicate

whether a fault is present. Normal behavior information is used as an input to the system behavior model to detect any change in system features and to produce symptoms to aid further diagnosis [1].

Analytical redundancy makes use of a mathematical model of the system under investigation and it is therefore often referred to as the model based approach to fault diagnosis. Basically, an intact FDD system includes three stages (procedures) with different functions: system modeling, residual generation and residual evaluation.

- **System Modeling**

    A precise mathematical model is required to predict system performance. For most systems, precise mathematical models are often very difficult to obtain. The robustness of the FDD scheme is often achieved by designing algorithms where the effects of model uncertainties and unmodeled dynamic disturbances on residuals are minimized and sensitivity to faults is maximized.

- **Residual Generation**

    First this block generates residual signals using available inputs and outputs from the monitored system. Its output should be normally zero or close to zero under no fault condition. The procedure used to compute residuals is called residual generation. Such a procedure is used to extract fault symptoms from the system, with the fault symptom represented by the residual signal. Most of the contribution in the field of model based FDD focuses on the residual generation problem, since the residual evaluation (decision making) becomes relatively easy if residuals are well designed [9].

- **Residual Evaluation (Decision Making)**

    This block examines symptoms for the likelihood of faults and a decision rule is then applied to determine if any faults have occurred. The residual evaluation block may perform a simple threshold test (geometrical methods) on the instantaneous values of the residuals. On the other hand, it may consist of statistical methods e.g. generalized likelihood ratio testing or sequential probability ratio testing.

The advantage of model-based fault detection and diagnosis are

- Higher FDD performance can be obtained, for example, more types of faults can be detected and the detection time is shorter.

- FDD can be performed over a large operating range.

- FDI can be performed passively without disturbing the operation of the process.

- High diagnosis performance can be obtained in spite of presence of disturbances.

- Reliance on hardware redundancy can be reduced, which means that the cost and weight can be reduced

The disadvantage of model-based FDD is, quite naturally, the need for a reliable model and possibly a more complex design procedure. So the accuracy of the model is usually the major limiting factor of the performance of a model based FDD system [7]. The generation of residual (i.e. symptoms) is the main issue in model-based fault diagnosis. A variety of methods are available in the literature for residual generation and the following sections present some of the most common popular analytical redundancy residual generation techniques:

- Parameter estimation

- Parity relation

- Observer-Based Approaches [8].

2.5.1 Parameter Estimation

In most practical cases, the process parameters are not known at all, or they are not known exactly enough. Therefore, they can be determined by means of parameter estimation methods, by measuring input and output signals, $u(z)$ and $y(z)$, if the basic structure of the model is known. This approach is based on the assumption that the faults are reflected in the physical system parameters and the basic idea is that the parameters of the actual process are estimated on-line using well-known parameter estimations methods [13].

The results are thus compared with the parameters of the reference model, obtained initially under fault free assumptions. Any discrepancy can indicate that a fault may have occurred. Now we explain two methods of least square (LS) parameter

Estimation, The nonrecursive (Off-Line) Parameter Estimation Equation Methods and Recursive (On-Line) Parameter Estimation Equation Methods [7].

2.5.1.1 The Nonrecursive (Off-Line) Parameter Estimation

It is assumed that the process can be described by the linear, dynamic, time-invariant, discrete-time system, which can be represented by a difference equation with constant parameters:

$$y(k) + a_1 y(k-1) + ... + a_n y(k-n) = b_1 u(k-1) + ... - b_m u(k-m) + e(k)$$

$$y(k) = -\sum_{i=1}^{n} a_i y(k-i) + \sum_{i=1}^{m} b_i u(k-i) + e(k) \tag{2.5}$$

Where $y(k)$ is the sequences of system output, $u(k)$ is measurable input and $e(i)$ is stochastic input, or noise, respectively, while the constants $a_i, i = 1,...,n$ and $b_j, j = 1,...m$ represent system parameters.

From figure 2.7, equation (2.5) can be rewritten in the following polynomial form:

$$\hat{A}(z^{-1}) y(z) - \hat{B}(z^{-1}) u(z) = e(z) \tag{2.6}$$

Where

$$A(z^{-1}) = 1 + \sum_{i=1}^{n} a_i z^{-i}, B(z^{-1}) = \sum_{i=1}^{m} b_i q^{-i} \tag{2.7}$$

Are the so-called characteristics and control polynomials, respectively. Equation (2.7) can also be rewritten in linear regression form as

$$y(z) = \psi^T (z)\Theta + e(z) \tag{2.8}$$

Where

$$\Theta^T = [a_1 ....a_n b_1 ....b_m] \tag{2.9}$$

Represents vector of constant system parameters and

$$\psi^T (k) = [-y(k-1),..........,-y(k-n), u(k-1),........., u(k-m)] \tag{2.10}$$

Represents a vector of input and output measurable samples and the equation error (residual) $e(k)$ is introduced as

$$e(k) = y(k) - \psi^T(k)\Theta \tag{2.11}$$

Eq. (2.11) can be written compactly as

$$E = Y - \Psi\Theta \tag{2.12}$$

Where

$$\Psi = \begin{bmatrix} \psi^T(1) \\ \psi^T(2) \\ . \\ . \\ \psi^T(N) \end{bmatrix} \tag{2.13}$$

And

$$E = [e(1) \quad e(2) \quad ....e(N)]^T \tag{2.14}$$
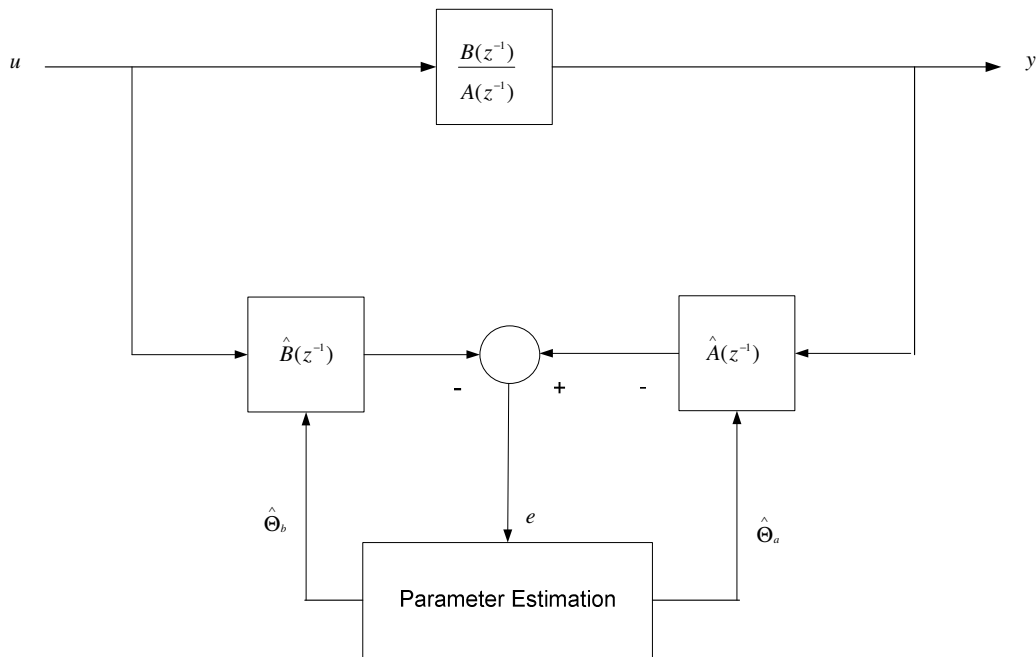
Figure 2.7: Modal Structure for Equation Error Methods

For the minimization of the error vector $e(i)$, the least-square method can be applied. To this end, define the following cost function [11].

$$J = E^T E = \sum_{i=1}^{N} e^2(k) \tag{2.15}$$

If Eq. (2.12) is substituted in equation (2.15), we obtain

$$J = E = (Y - \Psi\Theta)^T (Y - \Psi\Theta) \tag{2.16}$$

Hence

$$\frac{\partial J}{\partial \Theta} = -2\Psi^T (Y - \Psi\Theta)$$

Where the following formula was used

$$\frac{\partial}{\partial \Theta}[A\Theta] = \frac{\partial}{\partial \Theta}[\Theta^T A^T] = A^T \tag{2.17}$$

If we set $\frac{\partial J}{\partial \Theta}$ equal zero, we obtain

$$\Psi^T Y = \Psi^T \Psi\Theta \tag{2.18}$$

Relation (2.15) is known as the canonical equation and has a solution when the matrix $\Psi^T \Psi$ is invertible, in which case we have

$$\hat{\Theta} = [\Psi^T \Psi]^{-1} \Psi^T Y \tag{2.19}$$

Equation (2.19) represents nonrecursive (Off-Line) Parameter Estimation Equation of the Least Squares (LS) Method [11].

2.5.1.2 Recursive (On-Line) Parameter Estimation

In many practical cases, it is necessary that parameter estimation takes place concurrently with the system's operation. This parameter estimation problem is called on-line identification and its methodology usually leads to a recursive procedure for every new measurement (or data entry). For this reason, it is also called recursive least-squares estimate (RLS) or recursive identification. The proposed recursive algorithm is given by the following theorem.

Suppose that $\Theta(k)$ is the estimate of the parameters of the *nth* order system equation (2.5) for k data entries. Then, the estimate of the parameter vector $\Theta(k+1)$ for *k+1* data entries, with $(k = 1,2,3,....)$ is given by the expression

$$\hat{\Theta}(k+1) = \hat{\Theta}(k) + \gamma(k) \quad [y(k+1) - \psi^T(k+1)\hat{\Theta}(k)]$$

<div align="center">

new	old	correction	new	one−step−ahead
estimation	estimation	factor	measuremen	predictionof the new
				measuremen

</div>

(2.20)

The correcting vector is given by

$$\gamma(k) = P(k+1)\psi(k+1) = [\psi^T(k+1)P(k)\psi(k+1)+1]^{-1}\psi^T(k+1)P(k) \tag{2.21}$$

And the matrix $P(k+1)$ is calculated from the recursive formula

$$P(k+1) = [I - \gamma(k)\psi^T(k+1)]P(k) \tag{2.22}$$

With initial conditions

$$P(k) = \alpha I \quad and \quad \Theta(k) = 0 \tag{2.23}$$

With α large $(\alpha = 100,...,1000)$

2.5.2 Parity Relation

A straightforward way to detect process faults by parity equations is to compare the parity (consistency) of the analytical models with the actual outputs (measurements from sensors) of a real system. In theory, the result of parity equations (residuals) is zero under fault-free conditions, where an accurate and robust system model is a must [1]. The difference of signals between the actual output and the analytical model are expressed by residuals. Therefore residuals describe discrepancies between the process and the model and check for consistency, the design of the residuals can be made with transfer functions or in state-space formulation.

2.5.2.1 Parity Equations with Transfer Functions

In the form of a transfer function model, the process is described by transfer function of discrete system as

$$G_p(z) = \frac{y_p(z)}{u(z)} = \frac{B(z)}{A(z)} \tag{2.24}$$

And the process model described by

$$G_m(z) = \frac{y_m(z)}{u(z)} = \frac{\hat{B}(z)}{\hat{A}(z)} \qquad\qquad (2.25)$$

It worth noting that the model parameter and structure of the monitored process have to be known a priori

$$G_p(z) = G_m(z) + \Delta G_m(z) \qquad\qquad (2.26)$$

Where $\Delta G_m(z)$ describes model errors.

The residuals can be formulated by the output error or the polynomial error, similar to parameter estimation methods.

- For the output error residual $r(z)$ is

$$\begin{aligned}
r'(z) &= y_p(z) - y_m(z) = y_p(z) - G_m(z)u(z) \\
&= G_p(z)[u(z) + f_u(z)] + n(z) + f_y(z) - G_m(z)u(z) \qquad (2.27)\\
&= \Delta G_m(z)u(z) + G_p(z)f_u(z) + n(z) + f_y(z)
\end{aligned}$$

The methodology here described is depicted in Figure 2.8.
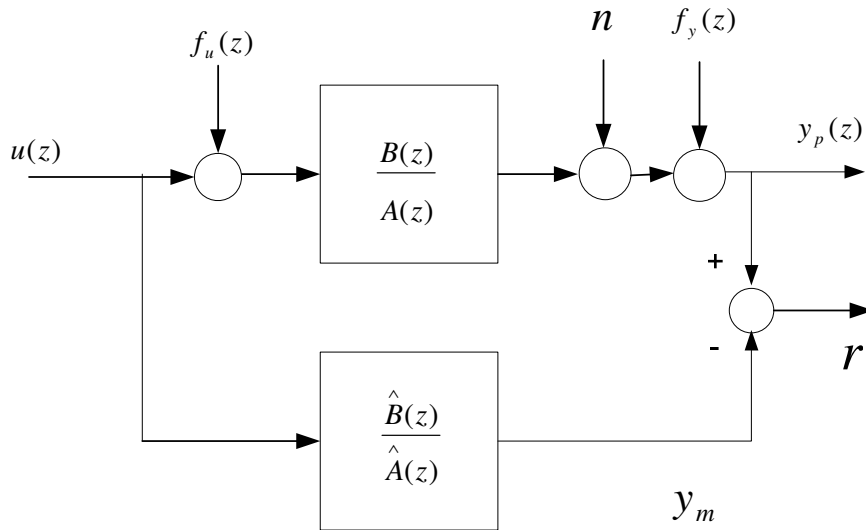
Figure 2.8: Scheme for Output Error via Parity Equation Method

The residual is zero for ideal matching of process and model, no additive faults $f_u(z)$, $f_y(z)$ and no noise. Usually, it shows deviations depending on the model error $\Delta G_m(z)$ and noise n and the input signal u. Another possibility to generate

- A polynomial error or equation error is as shown in Figure 2.9.
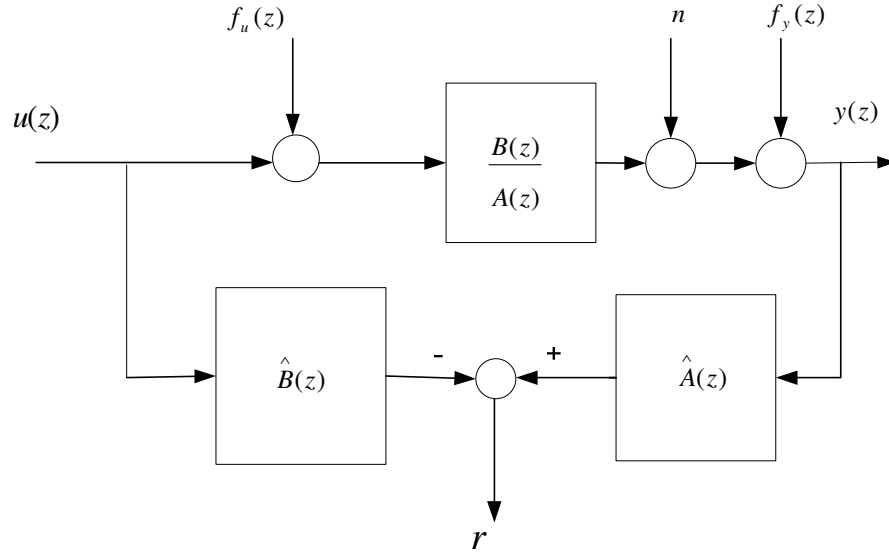
Figure 2.9: Scheme for Equation Error via Parity Equation Method

$$r(z) = \hat{A}(z)y_p(z) - \hat{B}(z)u(z)$$
$$= \hat{A}(z)[G_p(z)[u(z) + f_u(z)] + n(z) + f_y(z)] - \hat{B}(z)u(z)$$

(2.28)

Equation (2.27), (2.28) represent parity equations, $r^{'}$ and r called primary residuals.

2.5.2.2 Parity Equations with State Space Models

The parity equations with state space models for discrete time will be deriving in this part. According to Figure 2.10, the basic process equations for discrete time are

$$x(k+1) = Ax(k) + Bu(k) + Vv(k) + Lf(k)$$

(2.29)

$$y(k) = Cx(k) + Nn(k) + Mf(k)$$

(2.30)

Where $v(k)$ and $n(k)$ are measurable disturbance signals. $f(k)$ are additive faults which may be composed of additive faults $f_l(k)$ on the input and $f_m(k)$ on the output. To simplify the notations, the state-space model without faults and disturbances is used

$$x(k+1) = Ax(k) + Bu(k)$$

(2.31)

$$y(k) = Cx(k)$$

(2.32)

By introducing equation (2.31), equation (2.32) yields

$$y(k+1) = CAx(k) + CBu(k) \qquad (2.33)$$

And for the qth sample

$$y(k+q) = CA^q x(k) + CA^{q-1} BCBu(k) + \ldots\ldots\ldots + CBu(k+q-1) \qquad (2.34)$$

Here redundant equations are generated for different time instants. Now, the equations are summarized and lead to

$$Y(k+q) = Tx(k) + QU(k+q) \qquad (2.35)$$

Or time shifted by q backwards
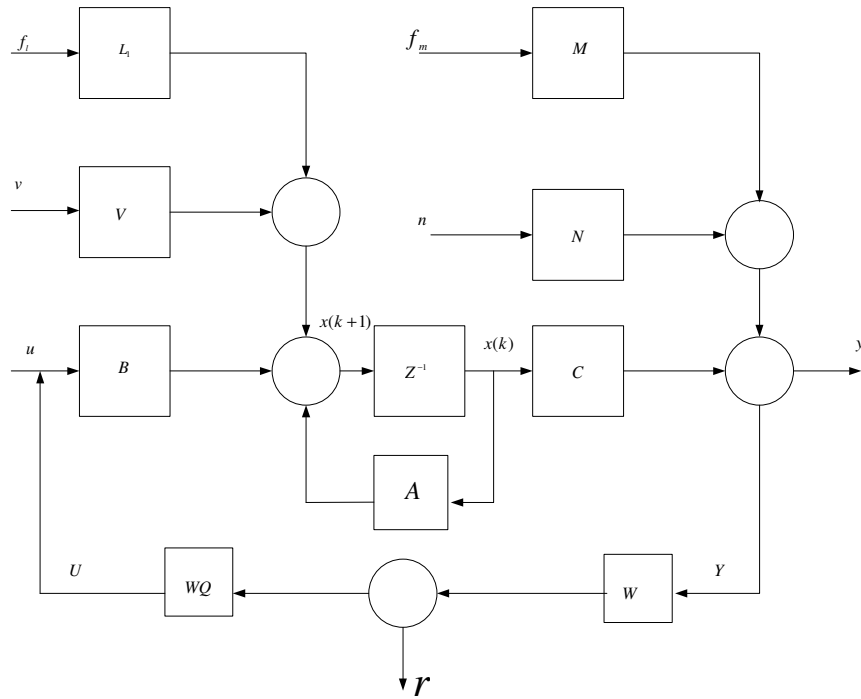
Figure2.10 Residual generation with parity equation in discrete time a state-space model

$$Y(k) = Tx(k-q) + QU(k) \qquad (2.36)$$

With the vectors

29

$$Y(k) = \begin{bmatrix} y(k-q) \\ y(k-q+1) \\ . \\ . \\ y(k) \end{bmatrix} , \quad U(k) = \begin{bmatrix} u(k-q) \\ u(k-q+1) \\ . \\ . \\ u(k) \end{bmatrix} \tag{2.37}$$

And the matrices

$$T = \begin{bmatrix} C \\ CA \\ CA^2 \\ . \\ . \\ CA^q \end{bmatrix} , \quad Q = \begin{bmatrix} 0 & 0 & . & . & . & . & 0 \\ CB & 0 & . & . & . & & \\ CAB & CB & .. & . & . & . & \\ . & & .. & . & . & . & .. \\ . & & .. & . & . & .. & . \\ CA^{q-1}B & CA^{q-2}B & . & . & .. & . & CB & 0 \end{bmatrix} \tag{2.38}$$

In order to remove the non-measurable states $x(k-q)$ and to obtain a parity vector useful for FDD, equation (2.36) is multiplied by W, such that

$$WT = 0 \tag{2.39}$$

This leads to residuals generated at time, k, can be written as

$$r(k) = WY(k) - WQU(k) \tag{2.40}$$

Where $Y(k)$ the measurement of system outputs and W is a vector for residual generating manipulation. A well designed residual generation vector is selected to achieve a specific structured residual response to faults and to decouple from unknown disturbances and model uncertainties.

2.5.3 Observer-Based Approaches

As state observers use output error between a measured process output and an adjustable model output, they are a further alternative for model-based fault detection. It is assumed, as in the case of parity equation approaches, that the structure and the parameters of the model are precisely known.

Several approaches have been proposed for fault detection which is based on the classical Luenberger state observer, kalman filter and so-called output observer. The output error is therefore used as a residual [12]. In order to obtain the structure of a

Luenberger state observer approach, the discrete time, time-invariant, linear dynamic model for the process under consideration in a state-space form is considered [15].

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k)$$

(2.41)

Where $u(k)$ is the input vector, $x(k)$ is the state vector and $y(k)$ is the output vector and assume that all matrices **A, B** and **C** of the system are known, a state observer is used to reconstruct the un measurable state variable based on measured inputs and outputs

$$\hat{x}(k+1) = A\,\hat{x}(k) + Bu(k) + He(k)$$
$$e(k) = r(k) = y(k) - C\,\hat{x}(k)$$

(2.42)

The observer scheme described by previous equation is depicted in Figure 2.11.
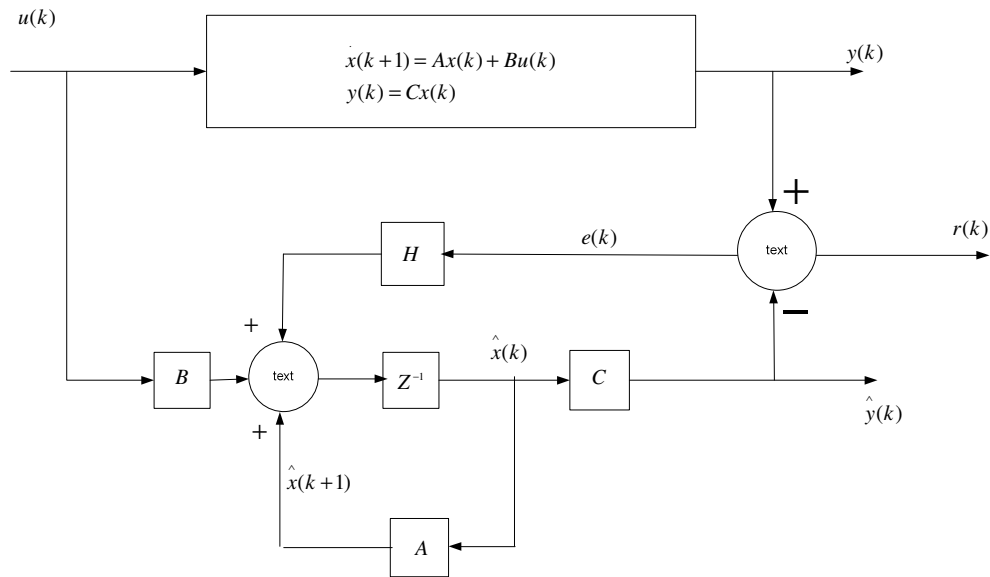
Figure 2.11: Process and state observer

For the state estimation error $e_x(k)$, it follows that

$$e_x(k) = x(k) - \hat{x}(k)$$
$$e_x(k+1) = x(k+1) - \hat{x}(k+1) = (A - HC)e_x(k)$$

(2.43)

The state error $e_x(k)$ vanishes asymptotically

$$\lim_{k \to \infty} e_x(k) = 0 \qquad (2.44)$$

If the observer is stable this can be achieved by proper design of the observer feedback matrix H. Let the process be influenced by disturbance and faults as depicted in Figure 2.12.
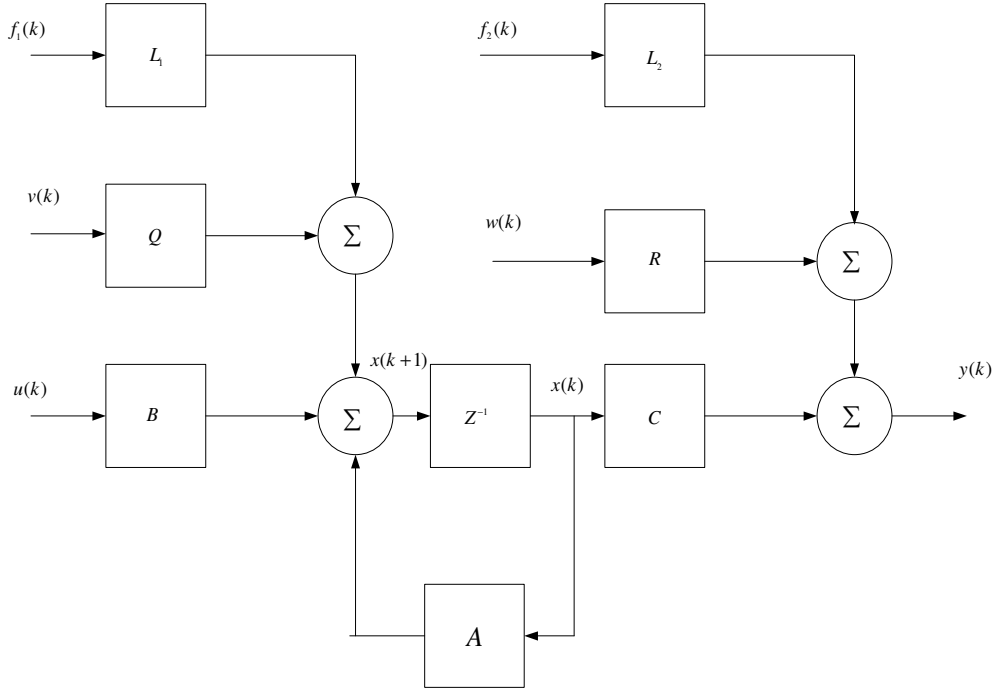
Figure 2.12: MIMO process with faults and noises

It is described by the following system

$$\begin{aligned}x(k+1) &= Ax(k) + Bu(k) + Qv(k) + L_1 f_l(k) \\ y(k) &= Cx(k) + Rw(k) + L_2 f_m(k)\end{aligned} \qquad (2.45)$$

Where $v(k)$ and $w(k)$ represent the non-measurable disturbance vector at the input and at the output respectively, $f_l(k)$ and $f_m(k)$ are input and output additive faults vector and L_1 and L_2, are fault entry matrices. For the state estimation error, under assumption that disturbance are neglected, the following equation hold

$$e_x(k+1) = (A - HC)e_x(k) + L_1 f_l(t) - HL_2 f_m(t) \qquad (2.46)$$

When the disturbance are neglected, the output becomes

$$y(k) = Cx(k) + L_2 f_m(k) \tag{2.47}$$

Also the output error (residual) becomes

$$\begin{aligned}
e(k) = r(k) &= y(k) - C\hat{x}(k) = Cx(k) + L_2 f_m(k) - C(x(k) - e_x(k)) \\
&= Ce_x(k) + L_2 f_m(k)
\end{aligned} \tag{2.48}$$

If faults appear as changes in the parameter matrices $\Delta A, \Delta B$ or ΔC, the process behavior becomes

$$\begin{aligned}
x(k+1) &= (A + \Delta A)x(k) + (B + \Delta B)u(k) \\
y(k) &= (C + \Delta C)x(k)
\end{aligned} \tag{2.49}$$

And the state $e_x(k)$ and the output error without disturbances

$$e_x(k+1) = (A - HC)e_x(k) + (\Delta A - H\Delta C)x(k) + \Delta Bu(k) \tag{2.50}$$

$$e(k) = r(k) = Ce_x(k) + \Delta Cx(k) \tag{2.51}$$

The changes $\Delta A, \Delta B$ and ΔC are multiplicative faults [12].

2.6. Fault Diagnosis Methods

The task of fault diagnosis consists of the determination of the type of fault with as many details as possible such as the fault size, location and time of detection. The diagnostic procedure is based on the observed analytical and heuristic symptoms and the heuristic knowledge of the process, as shown in figure 2.6. Symptoms are unusual changes of the features from its normal values. In fault free case the symptoms are zero.

The analytical symptoms in the model-based fault detection are the residuals.

The relationship between the symptom and the faults may be unknown. Therefore, classification and inference methods are used for fault diagnosis [7]. Next chapter will describe classification or pattern recognition methods for fault diagnosis application.

CHAPTER THREE

PATTERN RECOGNITION METHODS

3.1. Pattern Recognition Definition

Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified. We will refer to these objects using the generic term patterns.

Pattern recognition has a long history, but before the 1960s it was mostly the output of theoretical research in the area of statistics. As with everything else, the advent of computers increased the demand for practical applications of pattern recognition, which in turn set new demands for further theoretical developments. As our society evolves from the industrial to its postindustrial phase, automation in industrial production and the need for information handling and retrieval are becoming increasingly important. This trend has pushed pattern recognition to the high edge of today's engineering applications and research. Pattern recognition is an integral part in most machine intelligence systems built for decision making [16].

Applications of pattern recognition systems and techniques are numerous and cover a broad scope of activities. We enumerate only a few examples referring to several professional activities:

Agriculture:

Crop analysis and Soil evaluation

Astronomy:

Analysis of telescopic images and automated spectroscopy

Biology:

Automated cytology, Properties of chromosomes and Genetic studies

Civil administration:

Traffic analysis and control and Assessment of urban growth

Economy:

Stocks exchange forecast and Analysis of entrepreneurial performance

Engineering:

Fault detection in manufactured products, Character recognition, and Speech recognition, Automatic navigation system and Pollution analysis

Geology:

Classification of rocks, Estimation of mining resources, Analysis of geo-resources using satellite images and Seismic analysis

Medicine:

Analysis of electrocardiograms, Analysis of electroencephalograms and Analysis of medical images

Military:

Analysis of aerial photography and Detection and classification of radar and sonar signals, Automatic target recognition

Security:

Identification of fingerprints and Surveillance and alarm systems

As can be inferred from the above examples the patterns to be analyzed and recognized can be signals (e.g. ectrcocardiographic signals), images (e.g. aerial photos) or plain tables of values (e.g. stock exchange rates) [17].

3.2. Bayesian Classification

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the specific probability distributions of the patterns in each class are known [18].

3.2.1. The Bayes Decision Rule for Minimum Error

The pattern classification system assigns an observation to class i with the maximum discriminant function value

$$g_i(X) < g_j(X) \Rightarrow X \in \omega_i$$
$$g_i(X) > g_j(X) \Rightarrow X \in \omega_j$$

(3.1)

Where $g_j(X)$ is the discriminant function for class j given observation vector X, the statistics of the data in each class can provide analytical measures to determine the optimal discriminant functions in terms of minimizing the error rate, the average probability of error [19].

With ω_i the error rate can be minimized by using the discriminant function

$$g_i(X) = q_i(X)$$

(3.2)

Where $q_i(X)$ is a posteriori probability of ω_i given X

Equation (3.1) becomes

$$q_i(X) < q_j(X) \Rightarrow X \in \omega_i$$
$$q_i(X) > q_j(X) \Rightarrow X \in \omega_j$$

(3.3)

Equation (3.3) indicates that, if the probability of ω_i given X is larger than the probability of ω_j, X is classified to ω_i, and vice versa. In the sequel, we assume that the a priori probabilities P_i are known also the other statistical quantities the class-conditional probability density functions $p_i(X)$ are known [1].

Then the a posteriori probability calculated using bayes theorem, as

$$q_i(X) = \frac{P_i p_i(X)}{p(X)}$$

(3.4)

where $p(X)$ is the mixture density function.

If we will focus on the two-class case ω_1, ω_2 are the two classes in which our patterns belong and X be an observation vector.

The decision rule of equation (3.3) can be expressed

$$P_1 p_1(X) > P_2 p_2(X) \Rightarrow X \in \omega_1$$
$$P_1 p_1(X) < P_2 p_2(X) \Rightarrow X \in \omega_2$$

(3.5)

This is known as Bayes' rule for minimum error. [9]

$$l(X) = \frac{p_1(X)}{p_2(X)} > \frac{P_2}{P_1} \Rightarrow X \in \omega_1$$

$$l(X) = \frac{p_1(X)}{p_2(X)} < \frac{P_2}{P_1} \Rightarrow X \in \omega_2$$

(3.6)

The function $l(X)$ is called the likelihood ratio and we call $\frac{P_2}{P_1}$ the threshold value of

the likelihood ratio for the decision. Figure 3.1 and Figure 3.2 give a simple illustration

for a two-class discrimination problem. Class ω_1 is normally distributed with zero mean

and unit variance $p_1(X) = N(X|0,1)$. Class ω_2 is a normal mixture (a weighted sum of

normal densities) $p_2(X) = 0.6 * N(X|1,1) + 0.4 * N(X|-1,2)$.

Figure 3.1 plots $p_i(X) * Pi$ $i = 1,2,$ and the priors are taken to be $P_1 = 0.5, P_2 = 0.5$.

Figure 3.2 plots the likelihood ratio $l(X)$ and the threshold $\frac{P_2}{P_1}$. We see from this

figure that the decision rule equation (3.3) leads to a disjoint region for class ω_2 [20].
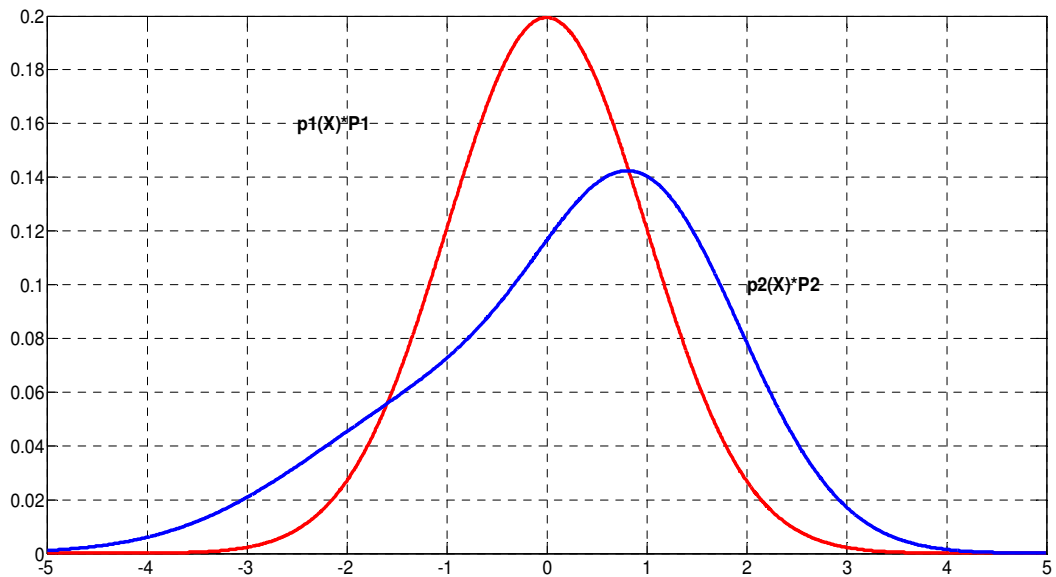
Figure 3.1: pi(X)*Pi, for classes 1, 2
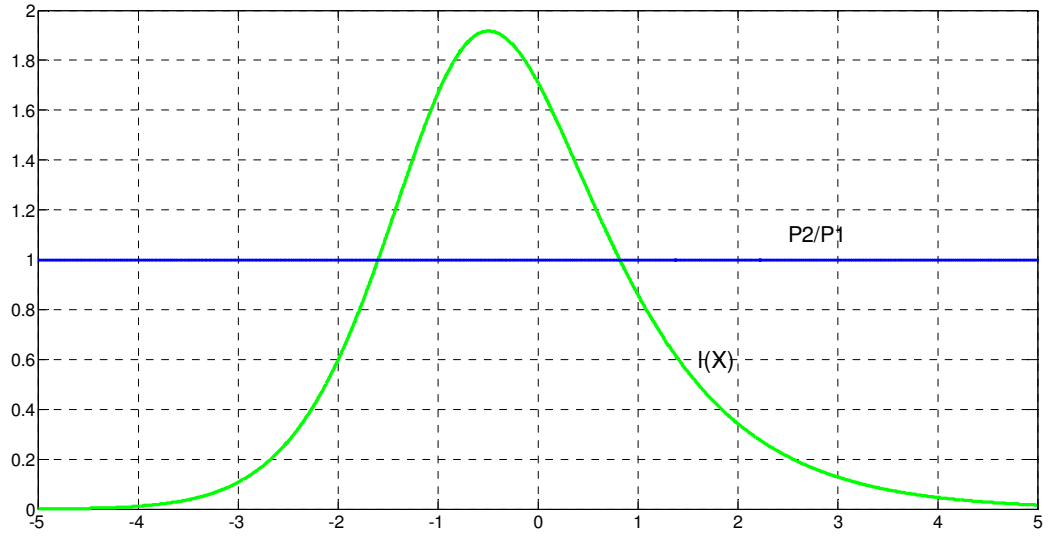
Figure 3.2: Likelihood function

When we write the minus-log likelihood ratio rather than writing the likelihood ratio itself, the decision rule of equation (3.6) becomes

$$h(X) = -\ln l(X) = -\ln p_1(X) + \ln p_2(X) > \ln \frac{P_1}{P_2} \Rightarrow X \in \omega_1$$

$$h(X) = -\ln l(X) = -\ln p_1(X) + \ln p_2(X) < \ln \frac{P_1}{P_2} \Rightarrow X \in \omega_2$$

(3.7)

Equations (3.3), (3.6), (3.7) are called the Bayes test for minimum error.

In special case when $p_i(X)$ are normal with expected vectors M_i and covariance matrices Σ. The decision rule of equation (3.7) becomes

$$h(X) = -\ln l(X) = \frac{1}{2}(X - M_1)^T \Sigma_1^{-1}(X - M_1) - \frac{1}{2}(X - M_2)^T \Sigma_2^{-1}(X - M_2) + \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} > \ln\frac{P_1}{P_2} \Rightarrow X \in \omega_1$$

$$h(X) = -\ln l(X) = \frac{1}{2}(X - M_1)^T \Sigma_1^{-1}(X - M_1) - \frac{1}{2}(X - M_2)^T \Sigma_2^{-1}(X - M_2) + \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} < \ln\frac{P_1}{P_2} \Rightarrow X \in \omega_2$$

(3.8)

In general, the decision rule of equation (3.7) or any other decision rule, does not lead to perfect classification. In order to evaluate the performance of a decision rule, we must calculate the probability of error $p(error) = \varepsilon$ that is the probability that a sample is assigned to the wrong class. The conditional error given $X, r(X)$ due to the decision rule of equation (3.3) is either $q_1(X)$ or $q_2(X)$ whichever smaller. That is,

$$r(X) = \min[q_1(X), q_2(X)] \tag{3.9}$$

The total error is computed by $E\{r(X)\}$.

$$\varepsilon = E\{r(X)\} = \int r(X)p(X)dX$$
$$= \int \min[P_1 p_1(X), P_2 p_2(X)]dX \tag{3.10}$$

$$\varepsilon = P_1 \int_{L2} p_1(X)dX + P_2 \int_{L1} p_2(X)dX \tag{3.11}$$

And expressed as

$$\varepsilon = P_1\varepsilon_1 + P_2\varepsilon_2 \tag{3.12}$$

Where ε_i, the probability of misclassifying patterns from class ω_i and this is given by

$$\varepsilon_1 = \int_{L2} p_1(X)dX \quad and \quad \varepsilon_2 = \int_{L1} p_2(X)dX \tag{3.13}$$

Then the total error which is called the bayes error, is

$$\varepsilon = P_1 \int_{L2} p_1(X)dX + P_2 \int_{L1} p_2(X)dX = P_1\varepsilon_1 + P_2\varepsilon_2 \tag{3.14}$$

In equation (3.14), we distinguish two types of errors: The first type error means decision that the sample is coming from the second class, and it actually belongs to the first one and analogously to that, the second type error means a decision that the sample belongs to the first class and it actually belongs to the second one. The total error is a weighted sum of these errors. The computation of the bayes error is a very complex problem except in some special cases. This is due to the fact that ε is obtained by integrating high-dimensional density functions in complex regions as seen in equation (3.12). Therefore, it is sometimes more convenient to integrate the density function of $h = h(X)$ of equation (3.5), which is one-dimensional

$$\varepsilon_1 = \int_{\ln(p_1/p_2)}^{+\infty} p_h(h|\omega_1)dh \quad and \quad \varepsilon_2 = \int_{-\infty}^{\ln(p_1/p_2)} p_h(h|\omega_2)dh \tag{3.15}$$

where $p_h(h|\omega_i)$ is the conditional density of h for ω_i.

To demonstrate the use of bayes decision rule for minimum error test in FDD, we suppose two-dimensional samples from two Gaussian classes are generated and the statistics of the classes are

$$M_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}; M_2 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}; \Sigma_2 = \begin{bmatrix} 2 & 0.4 \\ 0.4 & 2 \end{bmatrix} \tag{3.16}$$

Then, the decision boundary is given by a quadratic form equation (3.8).

The result is presented in Figure 3.3.It illustrates the decision boundary that separate two classes: the first class represents the regular regime of work; the second class represents a non-regular regime of work (which may mean the presence of failures or disturbances, which considerably change the work point of the entire system).
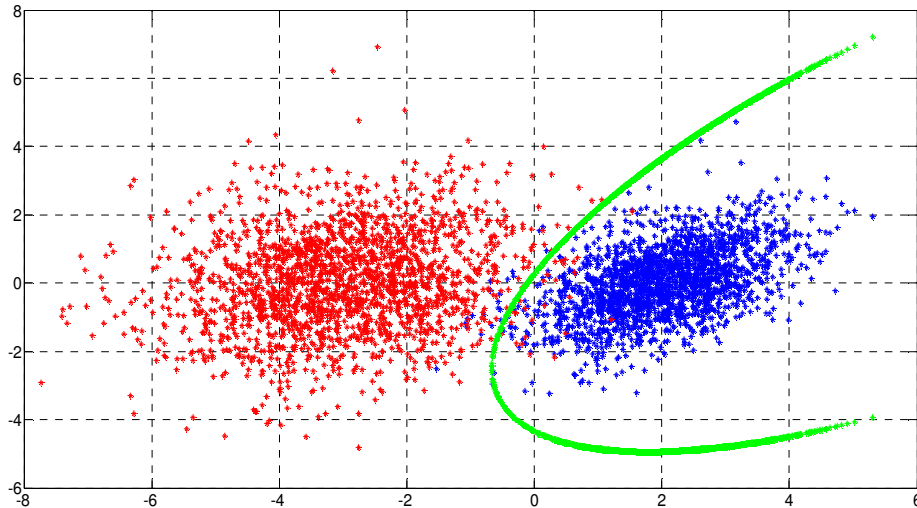
Figure 3.3: Decision boundaries for normal distribution

3.2.2. The Bayes Decision Rule for Minimum Cost

In the previous section, the decision rule selected the class for which the a posteriori probability $q_i(X)$ was the greatest. This minimized the probability of making an error. We now consider a somewhat different rule that minimizes an expected loss or cost. This is a very important concept since in many applications the costs associated with misclassification depend upon the true class of the pattern and the class to which it is assigned. For example in a medical diagnosis problem, the misclassification of a

cancer patient to normal may have a more damaging effect than the misclassification of a normal patient to cancer [20].

We make this concept more formal by introducing a loss that is a measure of the cost of making the decision C_{ij} that a pattern X belongs to class ω_i when the true class is ω_j.

$$C_{ij} = \cos t \quad of \quad deciding \quad a \quad pattern \quad X \in \omega_i \quad when \quad X \in \omega_j \tag{3.17}$$

In practice, it may be very difficult to assign or decide costs. In many situations, costs are a combination of several different factors measured in different units – money, time, quality of life. As a consequence, they may be the subjective opinion of an expert.

For the above two classes, the conditional cost or conditional risk of deciding a pattern X belong to class ω_i is defined as

$$r_1(X) = C_{11}q_1(X) + C_{12}q_2(X) = C_{11}P_1 p_1(X) + C_{12}P_2 p_2(X) \tag{3.18}$$

$$r_2(X) = C_{21}q_1(X) + C_{22}q_2(X) = C_{21}P_1 p_1(X) + C_{22}P_2 p_{22}(X) \tag{3.19}$$

The decision rule and the resulting conditional cost given $X, r(X)$ are

$$\begin{aligned} r_1(X) < r_2(X) &\Rightarrow X \in \omega_1 \\ r_1(X) > r_2(X) &\Rightarrow X \in \omega_2 \end{aligned} \tag{3.20}$$

And

$$r(X) = \min[r_1(X), r_2(X)] \tag{3.21}$$

The total cost of this decision is

$$\begin{aligned} r = E\{r(X)\} &= \int \min[r_1(X), r_2(X)] p(X) dX \\ &= \int \min[C_{11}P_1 p_1(X) + C_{12}P_2 p_2(X), C_{21}P_1 p_1(X) + C_{22}P_2 p_{22}(X)] dX \\ &= \int_{L1}[C_{11}P_1 p_1(X) + C_{12}P_2 p_2(X) dX + \int_{L2}[C_{21}P_1 p_1(X) + C_{22}P_2 p_2(X)] dX \end{aligned} \tag{3.22}$$

The boundary which minimizes r of equation (3.22) can be found as follows.

First, rewrite equation (3.22) as a function of L_1 alone. This is done by replacing $\int_{L1} p_i(X)dX$ with, $1 - \int_{L1} p_i(X)dX$ since L_1 and L_2 do not overlap and cover the entire domain. Thus,

$$r = (C_{21}P_1 + C_{22}P_2) + \int_{L_1} [(C_{11} - C_{21})P_1 p_1(X) + (C_{12} - C_{22})P_2 p_{22}(X)]dX \qquad (3.23)$$

Now our problem becomes one of choosing L_1 such that r is minimized. Suppose, for a given value of X, that the integrand of equation (3.23) is negative. Then we can decrease r by assigning X to L_1.

Thus the minimum cost decision rule is to assign to L_1 those $X\,s$ and only those $X\,s$ for which the integrand of equation (3.23) is negative. This decision rule can be stated by the following inequality:

$$(C_{12} - C_{22})P_2 p_2(X) < (C_{21} - C_{11})P_1 p_1(X) \Rightarrow X \in \omega_1$$
$$(C_{12} - C_{22})P_2 p_2(X) > (C_{21} - C_{11})P_1 p_1(X) \Rightarrow X \in \omega_2 \qquad (3.24)$$

By applying the bayes theorem equation (3.4) to the equation (3.24), the bayes decision rule for minimum cost is expressed as [20].

$$\frac{p_1(X)}{p_2(X)} > \frac{(C_{12} - C_{22})}{(C_{21} - C_{11})} \frac{P_2}{P_1} \Rightarrow X \in \omega_1$$
$$\frac{p_1(X)}{p_2(X)} < \frac{(C_{12} - C_{22})}{(C_{21} - C_{11})} \frac{P_2}{P_1} \Rightarrow X \in \omega_2 \qquad (3.25)$$

This decision rule is called the Bayes test for- minimum cost.

Equation (3.25) is equal to equation (3.6) for the special selection of the cost functions

$$C_{12} - C_{22} = C_{21} - C_{11} \qquad (3.26)$$

This is called a symmetrical cost function.

3.2.3. The Neyman - Pearson Test

An alternative to the Bayes decision rules for a two-class problem is the Neyman– Pearson test. In a two-class problem there are two possible types of error that may be made in the decision process. We may classify a pattern of class ω_1 as

belonging to class ω_2 or a pattern from class ω_2 as belonging to class ω_1. Let the probability of these two errors be ε_1 (error probability of class 1) and ε_2 (error probability of class 2) respectively as given in equation (3.13) [20].

The Neyman–Pearson decision rule is to minimize the error ε_1 subject to ε_2 being equal to a constant, say ε_0. If class ω_1 is termed the positive class and class ω_2 the negative class, then ε_1 is referred to as the false negative rate, the proportion of positive samples incorrectly assigned to the negative class; ε_2 is the false positive rate, the proportion of negative samples classed as positive. An example of the use of the Neyman–Pearson decision rule is in radar detection where the problem is to detect a signal in the presence of noise. There are two types of error that may occur; one is to mistake noise for a signal present. This is called a false alarm. The second type of error occurs when a signal is actually present but the decision is made that only noise is present. This is a missed detection. If ω_1 denotes the signal class and ω_2 denotes the noise then ε_2 is the probability of false alarm and ε_1 is the probability of missed detection. In many radar applications, a threshold is set to give a fixed probability of false alarm and therefore the Neyman–Pearson decision rule is the one usually used [21].

To determine this decision rule, we must find the minimum of

$$r = \varepsilon_1 + \mu(\varepsilon_2 - \varepsilon_0) \tag{3.27}$$

where μ is a Lagrange multiplier and ε_0 is the specified false alarm rate. Inserting ε_1 and ε_2 of equation (3.13) into equation (3.27)

$$
\begin{aligned}
r &= \int_{L_2} p_1(X)dX + \mu\{\int_{L_1} p_2(X)dX - \varepsilon_0\} \\
&= (1 - \mu\varepsilon_0) + \int_{L_1} \{\mu p_2(X) - p_1(X)\}dX
\end{aligned}
\tag{3.28}
$$

This will be minimized if we choose L_1 such that the integrand is negative, i.e.

$$if \quad \mu p_2(X) - p_1(X) < 0, \quad then \quad X \in \omega_1 \tag{3.29}$$

Or, in terms of the likelihood ratio,

$$\frac{p_1(X)}{p_2(X)} < \mu \Rightarrow X \in \omega_1$$

$$\frac{p_1(X)}{p_2(X)} > \mu \Rightarrow X \in \omega_2$$

(3.30)

Thus the decision rule depends only on the within-class distributions and ignores the a priori probabilities.

The threshold μ is chosen so that the following equation

$$\varepsilon_2 = \int_{L1} p_2(X)dX = \varepsilon_0$$

(3.31)

Or, use the density function of $h(X)$ of equation (3.15)

$$\varepsilon_2 = \int_{-\infty}^{-\ln(\mu)} p_h(h|\omega_2)dh = \varepsilon_0$$

(3.32)

However, in general μ cannot be determined analytically and requires numerical calculation. We conclude that Neuman-Pearson test proceeds from the assumption that the error of one type is incomparably more serious with significant consequences and therefore its probability is minimized ε_1, while the probability of the error of the second, less important type, is adopted as a constant ε_2. In order to analyze the applicability of The Neyman - Pearson Test, two-dimensional samples from two classes with the following probabilities functions are generated.

$$p_1(X) = \frac{1}{2\pi|\Sigma_1|^{1/2}} \exp(-\frac{1}{2}(X-M_1)^T \Sigma_1^{-1}(X-M_1))$$

(3.33)

$$p_2(X) = P_1 \frac{1}{2\pi|\Sigma_{21}|^{1/2}} \exp(-\frac{1}{2}(X-M_{21})^T \Sigma_{21}^{-1}(X-M_{21})) + P_2 \frac{1}{2\pi|\Sigma_{22}|^{1/2}} \exp(-\frac{1}{2}(X-M_{22})^T \Sigma_{22}^{-1}(X-M_{22}))$$ (3.34)

Then, the relation between μ and ε_2 is given by equation (3.31) and the statistics or parameters of the classes are

$$M_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix}; M_{21} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}; M_{22} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}; \Sigma_{21} = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}; \Sigma_{22} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

(3.35)

The result is presented in Figure 3.4. It illustrate the relation between different values of μ the corresponding a constant errors ε_2 equation (3.32).
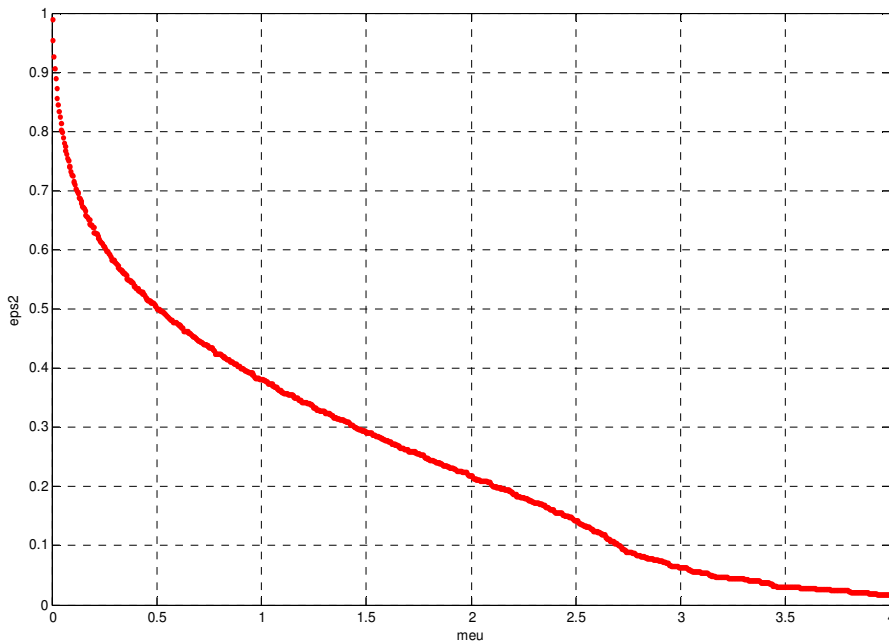
Figure 3.4: The Neyman-Pearson Test

3.2.4. Minimax criterion

The Bayes decision rules rely on knowledge of both the within-class distributions and the prior class probabilities. Therefore, in order to design a decision rule which minimizes the error, we need to know the values of P_i beforehand. After the design is completed, the decision rule stays optimum only if the $P_i^{'}S$ stay the same. Unfortunately in practice, $P_i^{'}S$ vary after the decision rule is fixed [21].

We want to design the classifier when we do not know the prior Probabilities. A reasonable approach is then to design our classifier so that the worst overall risk for any value of the priors is as small as possible. That is, minimize the maximum possible overall risk [18]. In order to understand this, we shall limit our discussion below to the two-class problem, and then write our overall risk equation (3.21) in terms of conditional risks:

$$r(X) = \min[r_1(X), r_2(X)]$$ \hfill (3.36)

First, let us express the cost of equation (3.23) in terms of P_1. Since $P_1 + P_2 = 1$, P_2 is uniquely determined by P_1. Inserting $P_2 = 1 - P_1$ into equation (3.23) and replacing $\int_{L1} p_1(X)dX$ by $1 - \int_{L2} p_1(X)dX$.

The total cost (the risk) of this decision is

$$r = C_{22} + (C_{12} - C_{22})\int_{L_1} p_2(X)dX + P_1[(C_{11} - C_{22}) + (C_{21} - C_{11})\int_{L_2} p_1(X)dX - (C_{12} - C_{22})\int_{L_1} p_2(X)dX] \quad (3.37)$$

To explain equation (3.36), we suppose two-dimensional samples from two classes are generated and the statistics of the classes are

$$M_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 8 \end{bmatrix}; M_{21} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}; M_{22} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}; \Sigma_{21} = \begin{bmatrix} 3 & 0.1 \\ 0.1 & 6 \end{bmatrix}; \Sigma_{22} = \begin{bmatrix} 2 & 0.4 \\ 0.4 & 9 \end{bmatrix} \quad (3.38)$$

In Figure 3.5, the curved line represents an example of the Bayes cost is plotted against P_1 where L_1 and L_2 are selected optimally for each P_1 (red curve). If L_1 and L_2 are fixed for $P_1 = 0.28$, for example, and if P_1 varies later unexpectedly, then r changes according to equation (3.37), as a linear function of P_1 which passing through the curved line at the point when $P_1 = 0.28$ (blue line), as shown in Figure 3.5 The maximum such error will occur at an extreme value of the prior, here at $P_1 = 1$ (blue line). In order to prevent this deterioration of performance, we choose L_1 and L_2 to make the coefficient of P_1 Zero in equation (3.37) regardless of the predicted value for P_1. Then, the straight line becomes the tangent at the point $C = 1.6722$ where the Bayes cost curve is maximum. Then, to minimize the maximum of such error, we should design our decision boundary for the maximum point of Bayes cost (here 1.6722), and thus the error will not change as a function of P_1 (green line). This decision rule is called the minimax rest.

Thus, in the minimax test, the boundary is designed to satisfy

$$(C_{11} - C_{22}) + (C_{21} - C_{11})\int_{L_2} p_1(X)dX - (C_{12} - C_{22})\int_{L_1} p_2(X)dX = 0 \quad (3.39)$$

And also the risk is independent of P_1 and from equation (3.37), the minimax risk is

$$r = C_{22} + (C_{12} - C_{22}) \int_{L_1} p_2(X) dX \qquad (3.40)$$
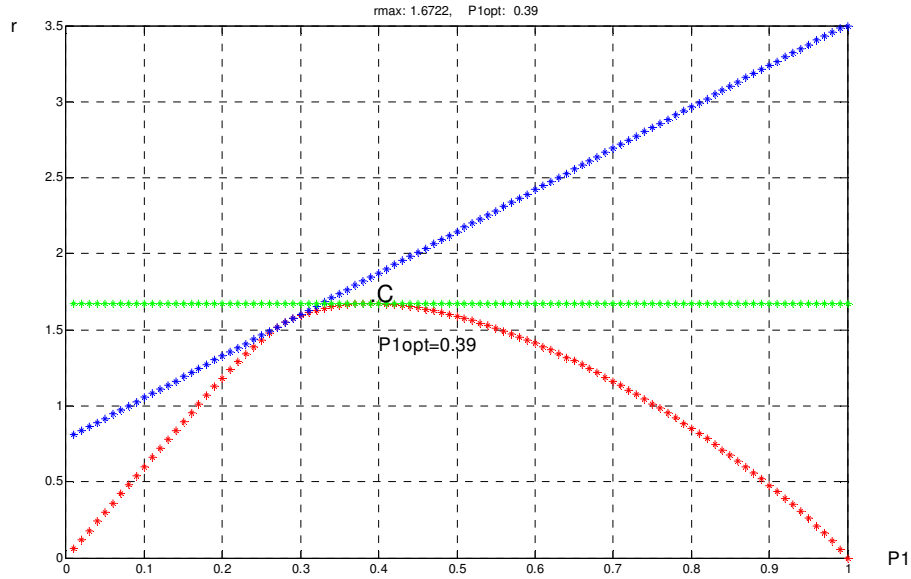
Figure 3.5: Bayes cost vs. P1

3.2.5. Sequential Hypothesis Testing

For the usual statistical tests the sample size is fixed before the data are taken, but for a sequential test the total sample size depends on the data and is thus a random variable. We are interested in sequential tests because they are economical in the sense that we may reach a decision earlier via a sequential test than via a fixed sample size test [22].

The sequential testing of hypotheses has the following mechanism of decision making. As the new sample comes, the algorithm adopts one of the following three decisions: the gathered samples either belong to the first class, or the gathered samples belong to the second class, or no decision can be made, and it is therefore necessary to wait for the next sample. This mechanism is executed so that the upper and lower thresholds of decision making are computed in advance, and the discrimination function is calculated with every new sample. If the function is under the lower threshold the first decision is made, if it is above the upper threshold, the second decision is made, and if neither is the case, the next sample is awaited [19].

The following subsection explain The Wald Sequential Test

- **The Wald Sequential Test**

Suppose $X_1, \ldots, X_n$ are independent and identically distributed random vectors observed sequentially using the joint density functions of these n vectors, $p_i(X_1, \ldots, X_n)$ $(i = 1,2)$ the minus-log likelihood ratio becomes

$$s = -\ln \frac{p_1(X_1, \ldots, X_n)}{p_2(X_1, \ldots, X_n)} = \sum_{i=1}^{n}[-\ln \frac{p_1(X_i)}{p_2(X_i)}] = \sum_{i=1}^{n} h(X_i) \tag{3.41}$$

Instead of fixing n, we may terminate the observations when s of equation (3.39) reaches a certain threshold value a and b. That is

$$\begin{aligned}
s_n &\le a \to X's \in \omega_1 \\
a &< s_n < b \to wait \quad for \quad the \quad sample \quad (n+1) \\
b &\le s_n \to X's \in \omega_2
\end{aligned} \tag{3.42}$$

Where s_n, is used instead of s to indicate the number of observations, a and b are thresholds to determine ω_1 and ω_2 respectively. This decision rule is called the Wald sequential test [18]. The parameters a and b are computed according to the desired classification errors of the first and second class:

$$a = -\ln \frac{1 - \varepsilon_1}{\varepsilon_2} \tag{3.43}$$

$$b = -\ln \frac{\varepsilon_1}{1 - \varepsilon_2} \tag{3.44}$$

Having in mind the structure of this classifier, the quantity s where $s = \sum_{i=1}^{m} h(X_i)$ should be either a or b with

$$\begin{aligned}
s &= a \quad (accept \quad \omega_1) \quad with \quad probability \quad 1 - \varepsilon_1 \quad when \quad X's \in \omega_1 \\
s &= a \quad (accept \quad \omega_1) \quad with \quad probability \quad \varepsilon_2 \quad when \quad X's \in \omega_2 \\
s &= b \quad (accept \quad \omega_2) \quad with \quad probability \quad \varepsilon_1 \quad when \quad X's \in \omega_1 \\
s &= b \quad (accept \quad \omega_2) \quad with \quad probability \quad 1 - \varepsilon_2 \quad when \quad X's \in \omega_2
\end{aligned} \tag{3.45}$$

Therefore,

$$E\{s|\omega_1\} = a(1 - \varepsilon_1) + b\varepsilon_1 \tag{3.46}$$

$$E\{s|\omega_2\} = a\varepsilon_2 + b(1 - \varepsilon_2) \qquad (3.47)$$

Introducing the mean values:

$$E\{s|\omega_i\} = \sum_{i=1}^{m} E\{s(X_i)|\omega_i\} = m\eta_i \qquad (3.48)$$

Where η_i is equal $E\{h(X_i)|\omega_i\} = E\{-\ln\dfrac{p_1(X)}{p_2(X)}|\omega_i\}$. Thus, the average number of observations needed to reach the decisions is

$$E\{m|\omega_1\} = \frac{a(1-\varepsilon_1) + b\varepsilon_1}{\eta_1} \qquad (3.49)$$

$$E\{m|\omega_2\} = \frac{a\varepsilon_2 + b(1-\varepsilon_2)}{\eta_2} \qquad (3.50)$$

In order to analyze the applicability of the Wald sequential test two-dimensional samples from two classes are generated. The statistics of the classes are the following:

$$M_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix}; M_{21} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}; M_{22} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}; \Sigma_{21} = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}; \Sigma_{22} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 2 \end{bmatrix} \qquad (3.51)$$

The following figure illustrate the efficiency of the Wald sequential test may be measured through the curve that shows the influence of desired classification errors to the needed number of observations.
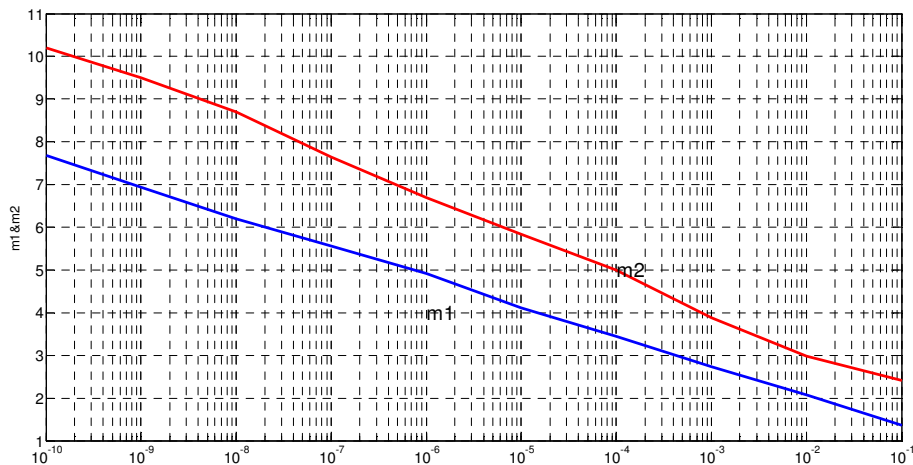
Figure 3.6: Required number of samples dependence on desired classification error

3.3. Parametric Classifier

Our major concern in Bayes likelihood ratio test was to design classifiers based on probability density or probability functions. In most applications, we must estimate these density functions using a finite number of sample observation vectors. However, they may be very complex or require a large number of samples to give accurate results [18].

Even if we can obtain the densities, the likelihood ratio test may be difficult to implement; time and storage requirements for the classification process may be excessive. Therefore, we are often led to consider a simpler procedure for designing a pattern classifier irrespective of the underlying distributions describing the training data. The most common choices are linear and quadratic and the major advantages of these procedures are their simplicity and computational attractiveness [16].

3.3.1. Linear discriminant function

Discrimination is a separation procedure that tries to find a discriminant function whose numerical values are such that the observations from several classes are separated as much as possible. An allocation procedure that uses a discrimination function as a well-defined rule in order to optimally assign a new observation to the labelled classes is called classification. It is evident that only good discrimination leads to good classification [23].

Let us once more focus on the two-class case and consider the family of discriminant functions that are linear combinations of the components of $x = (x_1, \ldots, x_n)^T$

$$
\begin{aligned}
h(X) = V^T X + v_0 < 0 \Rightarrow \omega_1 \\
h(X) = V^T X + v_0 > 0 \Rightarrow \omega_2
\end{aligned}
\tag{3.52}
$$

This is called a linear discriminant function, $V = [v_1, v_2, \ldots, v_n]^T$ is known as the weight vector and v_0 as the threshold. Our design work is to find the optimum coefficients of the weight vector and the threshold value for given distributions under various criteria.

3.3.1.1. Optimum Design Procedure

Equation (3.52) indicates that an n-dimensional vector X is projected onto a vector V, and that the variable, $y = V^T X$ in the projected one dimensional h-space is classified to either ω_1 or ω_2, depending on whether $y < -v_0$ or $y > -v_0$.

Figure 3.7 shows an example in which distributions are projected onto two vectors, V and V'. From Figure 3.7, we notice the error on V is smaller than that on V'. Therefore, the optimum design procedure for a linear classifier is to select V and v_0 which give the smallest error in the projected h-space.

When X is normally distributed, $h(X)$ of equation (3.52) is also normal. Therefore, the error in the h-space is determined by $\eta_i = E\{h(X)|\omega_i\}$ and $\sigma_i^2 = Var\{h(X)|\omega_i\}$ which are functions of V and v_0 [21]. Thus, The expected values and variances of $h(X)$ are

$$\eta_i = E\{h(X)|\omega_i\} = V^T E\{X|\omega_i\} + v_0 = V^T M_i + v_0 \tag{3.53}$$

$$\sigma_i^2 = Var\{h(X)|\omega_i\} = V^T E\{(X - M_i)(X - M_i)^T|\omega_i\}V = V^T \Sigma_i V \tag{3.54}$$
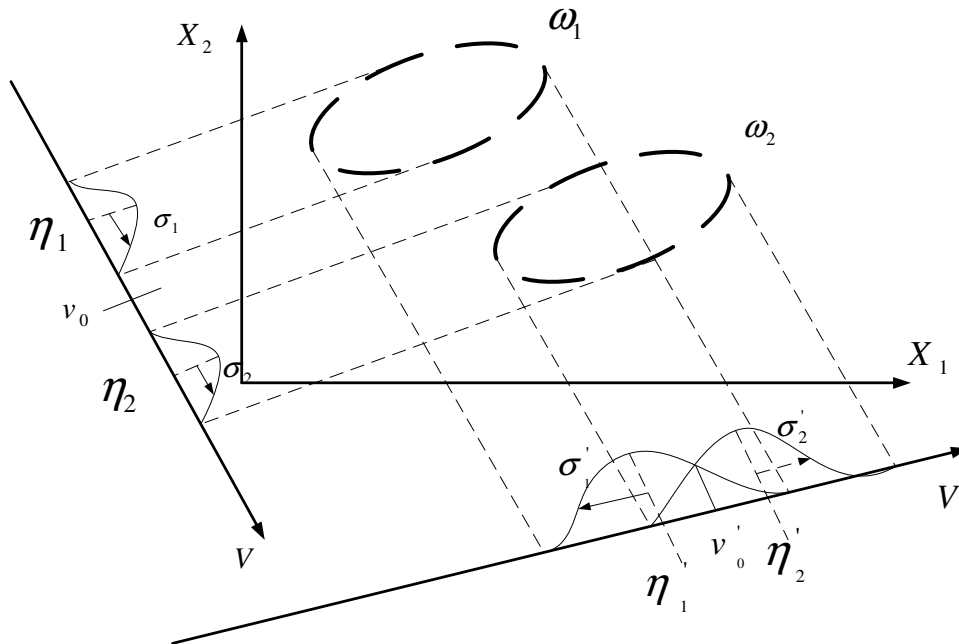
Figure 3.7: An example of linear mapping

Let $f(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$ be any criterion to be minimized or maximized for determining the optimum V and v_0. The derivative of f with respect to V and v_0 give two equations and their solution for V, gives the optimum V

$$V = [s\Sigma_1 + (1-s)\Sigma_2]^{-1}(M_2 - M_1) \qquad (3.55)$$

Where

$$s = \frac{\partial f / \partial \sigma_1^2}{\partial f / \partial \sigma_1^2 + \partial f / \partial \sigma_2^2} \qquad (3.56)$$

I. Fisher's linear discriminant

Fisher's linear discriminant is very popular among users of discriminant analysis. Some of the reasons for this are its simplicity and unnecessary of strict assumptions. For two classes of observations, the Fisher criterion is given by

$$f = \frac{(\eta_1 - \eta_2)^2}{(\sigma_1^2 + \sigma_2^2)} \qquad (3.57)$$

This criterion measures the difference of two means normalized by the averaged variance. The derivates of f with respect to σ_1^2 and σ_2^2 is substitute into equation (3.56) gives the value of $s = 0.5$, and from equation (3.53) the optimum V is

$$V = [0.5\Sigma_1 + 0.5\Sigma_2]^{-1}(M_2 - M_1) \qquad (3.58)$$

$h(X)$ With V of equation (4.58) is called the Fisher discriminant function and Fisher linear classifier, respectively. The Fisher criterion does not depend on v_0 because the subtraction of η_2 from η_1 eliminates v_0, from equation (4.53). Figure 3.8 shows the Fisher discriminant function that separate two classes and the parameters are

$$M_1 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 4 \end{bmatrix}; M_{21} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}; M_{22} = \begin{bmatrix} 2 \\ 9 \end{bmatrix}; \Sigma_{21} = \begin{bmatrix} 3 & 0.1 \\ 0.1 & 3 \end{bmatrix}; \Sigma_{22} = \begin{bmatrix} 2 & 0.4 \\ 0.4 & 6 \end{bmatrix} \qquad (3.59)$$
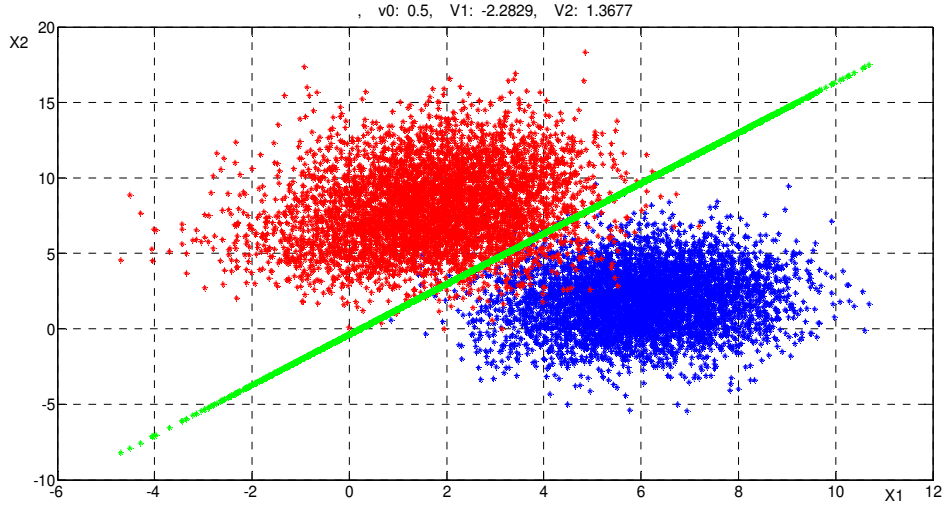
Figure 3.8: Fisher Discriminant function

II. Optimum Design for Normal Distributions

When the distributions of $h(X)$ are normal, we can find V and v_0 which minimize the Bayes error in the h-space. The Bayes error in the h-space is expressed as a function of η_i and σ_i^2 as

$$\varepsilon = P_1 \int_{-\eta_1/\sigma_1}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} d\xi + P_2 \int_{-\infty}^{-\eta_2/\sigma_2} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} d\xi \qquad (3.60)$$

The derivative of ε with respect to $\eta_1, \eta_2, \sigma_1^2$ and σ_2^2 give four equations and their solutions for v_0 must be selected to make the two density functions of $h(X)$ equal at $h(X) = 0$.

$$\frac{P_1}{\sqrt{2\pi}\sigma_1} e^{-(\eta_1/\sigma_1)^2/2} = \frac{P_2}{\sqrt{2\pi}\sigma_2} e^{-(\eta_2/\sigma_2)^2/2} \qquad (3.61)$$

Then the result is

$$s = \frac{-\eta_1/\sigma_1^2}{-\eta_1/\sigma_1^2 + \eta_2/\partial\sigma_2^2} \qquad (3.62)$$

$$[s\Sigma_1 + (1-s)\Sigma_2]V = (M_2 - M_1) \qquad (3.63)$$

Where s stays between 0 and 1 because $\eta_1 < 0$ and $\eta_2 > 0$. Thus, if we can find V and v_0, which satisfy equation (4.51).and equation (4.63), these V and v_0, minimize the error of equation (4.60). Unfortunately, since η_i and σ_i^2 are functions of V and v_0, the explicit solution of these equations has not been found. Thus, we must use an iterative procedure to find the solution. Before discussing the iterative process, we need to develop one more equation to compute v_0 from s and V. This is done by substituting η_1 and η_2 of equation (4.53).into equation (4.62).and by solving equation (4.62).for v_0. The result is

$$v_0 = -\frac{s\sigma_1^2 V^T M_2 + (1-s)\sigma_2^2 V^T M_1}{s\sigma_1^2 + (1-s)\sigma_2^2} \tag{3.64}$$

The iterative operation is carried out by changing the parameter s with an increment of Δs. First, calculate V for given s and using the V obtained to compute σ_i^2 by equation (4.54), v_0 by equation (4.64), and η_i by equation (4.53) in that sequence, after that calculate ε by equation (4.60). Then change s from 0 to 1. s Which minimizes ε can be found from ε vs. s plot. The advantage of this process is that we have only one parameter s to adjust. Figure 3.9 shows the minimum value of the error and the values of V and v_0 that verify it. The statistics of the classes are

$$M_1 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 6 \end{bmatrix}; M_{21} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}; M_{22} = \begin{bmatrix} 2 \\ 9 \end{bmatrix}; \Sigma_{21} = \begin{bmatrix} 3 & 0.1 \\ 0.1 & 8 \end{bmatrix}; \Sigma_{22} = \begin{bmatrix} 2 & 0.2 \\ 0.2 & 6 \end{bmatrix} \tag{3.65}$$


Figure 3.9: s vs Error

3.3.1.2. Other Desired Outputs and Search Techniques

It is recommended to design the classifier in pattern recognition is by using samples near the decision boundary because samples far from the decision boundary are less important to the design. However, if we fix the desired output $\gamma(X)$ and try to minimize the mean-square error between the actual outputs $h(X)$ and $\gamma(X)$, $h(X)$ contribute more to the mean-square error when it is large. This has long been recognized as a disadvantage of a mean-square error approach in pattern recognition. The following steps demonstrate the modification which reduces this effect. We can write the linear discriminant function equation (3.53) as

$$h(X) = -V^T X - v_0 > 0 \Rightarrow \omega_1$$
$$= [v_0 \quad v_1 \quad . \quad . \quad . \quad v_n] \begin{bmatrix} -1 \\ -x_1 \\ . \\ . \\ . \\ -x_n \end{bmatrix} \tag{3.66}$$

$$h(X) = V^T X + v_0 > 0 \Rightarrow \omega_2$$
$$= [v_0 \quad v_1 \quad . \quad . \quad . \quad v_n] \begin{bmatrix} 1 \\ x_1 \\ . \\ . \\ . \\ x_n \end{bmatrix} \tag{3.67}$$

We can be designated to

$$W^T = [v_0 \quad v_1 \quad . \quad . \quad . \quad v_n] \tag{3.68}$$

$$Z = [-1 \quad -x_1 \quad . \quad . \quad . \quad -x_n]^T \quad X \in \omega_1 \tag{3.69}$$

$$Z = [1 \quad x_1 \quad . \quad . \quad . \quad x_n]^T \quad X \in \omega_2 \tag{3.70}$$

So, the discriminant function is written as

$$h(X) = W^T Z = \sum_{i=0}^{n} w_i z_i > 0 \tag{3.71}$$

Where z_0 is either $+1$ or -1 depends on which class it is belong.

So, to design new classifier, we have to generate a new set of vectors $Z's$ and to calculate W^T that satisfy equation (3.71).

We concentrate our discussion on the two-class case ω_1, ω_2, if we have n samples for each class.

Then

$$\{X_1^1, X_2^1, \ldots, X_n^1\} \Rightarrow \{Z_1^1, Z_2^1, \ldots, Z_n^1\} \tag{3.72}$$

Where

$$Z_i^1 = \begin{bmatrix} -1 \\ -X_i \end{bmatrix} \quad X_i \in \omega_1, \quad i = 1, \ldots, n \tag{3.73}$$

And

$$\{X_1^2, X_2^2, \ldots, X_n^2\} \Rightarrow \{Z_1^2, Z_2^2, \ldots, Z_n^2\} \tag{3.74}$$

$$Z_i^2 = \begin{bmatrix} 1 \\ X_i \end{bmatrix} \quad X_i \in \omega_2, \quad i = 1, \ldots, n \tag{3.75}$$

Therefore

$$Z = [Z_1^1 \ Z_2^1 \ \ldots \ Z_n^1 \ Z_1^2 \ Z_2^2 \ \ldots \ Z_n^2] = [Z_1 \ Z_2 \ \ldots \ Z_{2n}] \tag{3.76}$$

First, we select the desired output $\gamma(X)$ and substituting it into equation (3.71) as

$$\left.\begin{array}{l} W^T Z_1 = \gamma_1 \\ W^T Z_2 = \gamma_2 \\ . \\ . \\ . \\ W^T Z_{2n} = \gamma_{2n} \end{array}\right\} = \left.\begin{array}{l} Z_1^T W = \gamma_1 \\ Z_2^T W = \gamma_2 \\ . \\ . \\ . \\ Z_{2n}^T W = \gamma_{2n} \end{array}\right\} \Rightarrow \begin{bmatrix} Z_1^T \\ Z_2^T \\ . \\ . \\ . \\ Z_{2n}^T \end{bmatrix} W = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ . \\ . \\ . \\ \gamma_n \end{bmatrix} \tag{3.77}$$

Next, we assume

$$U = \begin{bmatrix} Z_1^T \\ Z_2^T \\ . \\ . \\ . \\ Z_{2n}^T \end{bmatrix} \quad and \quad \Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ . \\ . \\ . \\ \gamma_n \end{bmatrix} \tag{3.78}$$

After that

$$UW = \Gamma \Rightarrow U^T U W = U^T \Gamma$$

The last step

$$W = (U^T U)^{-1} U^T \Gamma \tag{3.79}$$

The following figure is an example of other desired outputs and search techniques for two classes, the decision boundary that separate two classes and the parameters of the classes are

$$M_1 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 4 \end{bmatrix}; M_{21} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}; M_{22} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}; \Sigma_{21} = \begin{bmatrix} 3 & 0.1 \\ 0.1 & 3 \end{bmatrix}; \Sigma_{22} = \begin{bmatrix} 2 & 0.4 \\ 0.4 & 6 \end{bmatrix} \tag{3.80}$$
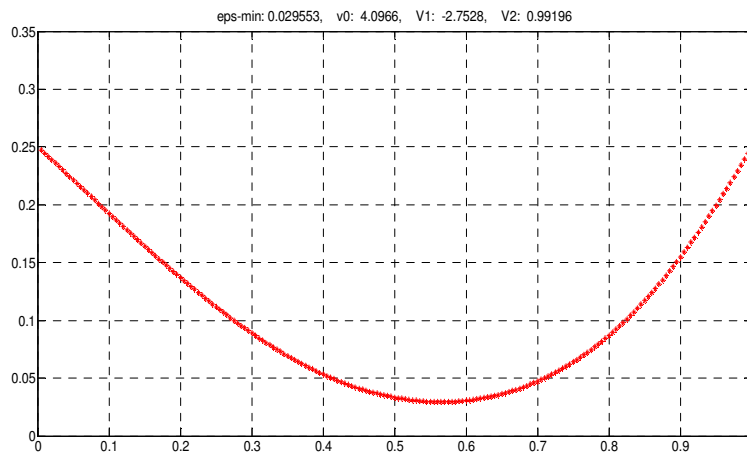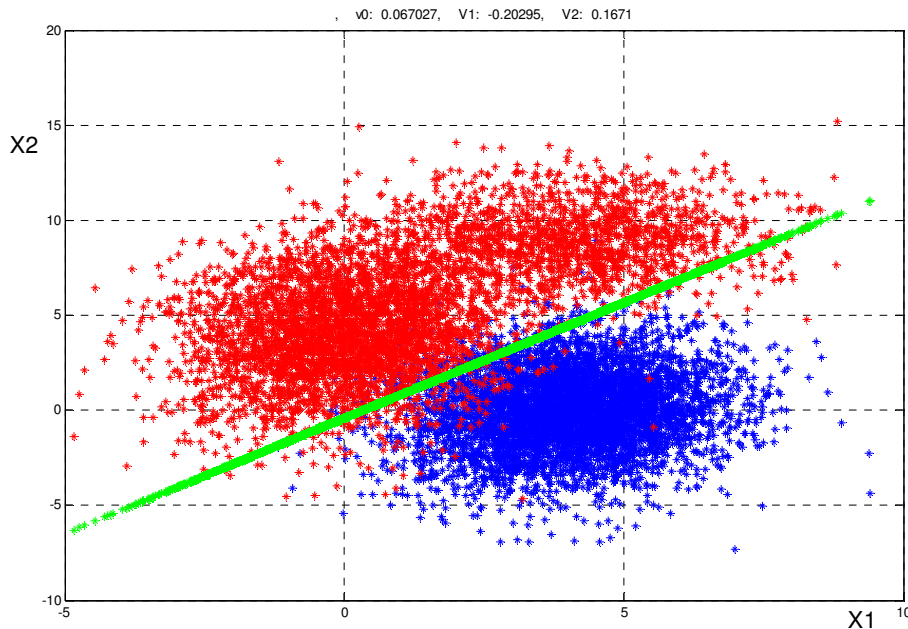

Figure 3.10: Other Desired Outputs and Search Techniques

3.3.2 Quadratic Classifier Design

When the distributions of X are normal for both ω_1, ω_2 .then bayes discriminant function becomes the quadratic equation as equation (3.8). However, to design a quadratic classifier, we must estimate M_i and Σ_i and inserting these estimates into equation (3.8). To avoid this disadvantage we use Quadratic Classifier Design [16].

The general quadratic classifier may be expressed as

$$h(X) = X^T Q X + V^T X + v_0 < 0 \Rightarrow \omega_1$$
$$h(X) = X^T Q X + V^T X + v_0 > 0 \Rightarrow \omega_2 \tag{3.81}$$

Where Q, V and v_0 are a matrix, vector, and scalar, respectively.

To interpret equation (3.81), we will limit our discussion on the two-class case ω_1, ω_2.

$$h(X) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v_0 < 0 \Rightarrow \omega_1 \tag{3.82}$$

$$h(X) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v_0 > 0 \Rightarrow \omega_2 \tag{3. 83}$$

When we deal with class one equation (3.82) becomes

$$h(X) = x_1^2 q_{11} + x_1 x_2 q_{21} + x_1 x_2 q_{12} + x_2^2 q_{22} + v_1 x_1 + v_2 x_2 + v_0 < 0$$
$$= \alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + v_1 x_1 + v_2 x_2 + v_0 < 0 \tag{3.84}$$

Where $\alpha_1 = q_{11}$, $y_1 = x_1^2$, $\alpha_2 = (q_{21} + q_{12})$, $y_2 = x_1 x_2$, $\alpha_3 = q_{22}$ and $y_3 = x_2^2$.

And

$$h(X) = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & v_1 & v_2 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 & x_1 & x_2 \end{bmatrix}^T + v_0 < 0 \tag{3.85}$$

Then

$$h(X) = W^T Z + v_0 < 0 \tag{3.86}$$

Where $W^T = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & v_1 & v_2 \end{bmatrix}$ and $Z = \begin{bmatrix} y_1 & y_2 & y_3 & x_1 & x_2 \end{bmatrix}^T$

To change $h(X) > 0$, equation (3.86) becomes

$$h(X) = -W^T Z - v_0 > 0$$

So

$$h(X) = \begin{bmatrix} v_0 & \alpha_1 & \alpha_2 & \alpha_3 & v_1 & v_2 \end{bmatrix} \begin{bmatrix} -1 & -y_1 & -y_2 & -y_3 & -x_1 & -x_2 \end{bmatrix}^T = K^T Q_1 > 0 \quad (3.87)$$

For class two is the same steps

$$h(X) = W^T Z + v_0 > 0 \Rightarrow \omega_2$$
$$h(X) = \begin{bmatrix} v_0 & \alpha_1 & \alpha_2 & \alpha_3 & v_1 & v_2 \end{bmatrix} \begin{bmatrix} 1 & y_1 & y_2 & y_3 & x_1 & x_2 \end{bmatrix}^T = K^T Q_2 > 0 \quad (3.88)$$

Then

$$Q = [Q_1 \quad Q_2] \quad (3.89)$$

To calculate K, first we select the desired output γ.

$$h(X) = K^T Q = Q^T K = \gamma \quad (3.90)$$

Next, we assume $Q^T = U$

$$UK = \gamma \Rightarrow (U^T U)K = U^T \gamma \quad (3.91)$$

Finally

$$K = inv(U^T U)U^T \gamma \quad (3.92)$$

Figure 3.11 an example of Quadratic Classifier Design for two classes
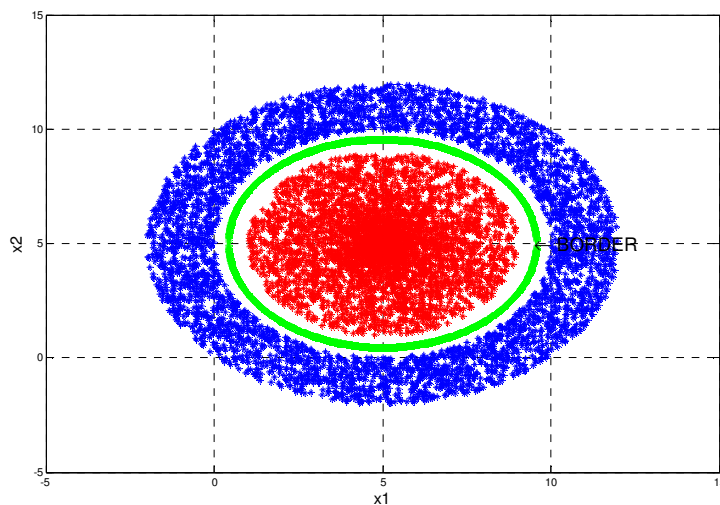
Figure 3.11: Quadratic Classifier Design

3.4. Non Parametric Density Estimation

Applying statistical pattern recognition techniques often requires an estimation of probability density functions of data samples. If the distribution of the data is known to follow a certain form with a few parameters, such as that of the Gaussian distribution, then the probability density can be easily evaluated using the estimated parameters of the distribution function. However, it is not always possible to assume a density function in a parametric form without causing significant error. In this case, a non-parametric approach must be taken by employing density estimation techniques [24].

There are many methods that have been used for statistical density estimation and the common methods are the histogram, the Kernel Density estimate and the k-nearest neighbor-density estimate.

3.4.1. Histogram Method

The histogram method is perhaps the oldest and the simplest method of density estimation. It is the classical method by which a probability density is constructed from a set of samples, also histogram is easy to create and are computationally feasible [25].

To construct a histogram one needs to select a starting point x_0 and the bin width b and we define the bins of the histogram to be the intervals

$$[x_0 + (i-1)h, \quad x_0 + ih], \quad i = 1, 2, ..., n .$$

The histogram is then defined by

$$f_{Hist}(x) = \frac{1}{n} \frac{Number \quad of \quad observation \quad in \quad the \quad same \quad bin \quad as \quad x}{b} \tag{3.93}$$

More generally one can also use bins of different widths, in which case

$$f_{Hist}(x) = \frac{1}{n} \frac{Number \quad of \quad observation \quad in \quad the \quad same \quad bin \quad as \quad x}{Width \quad of \quad bin \quad coniaining \quad x} \tag{3.94}$$

The following figure is an example of histogram.
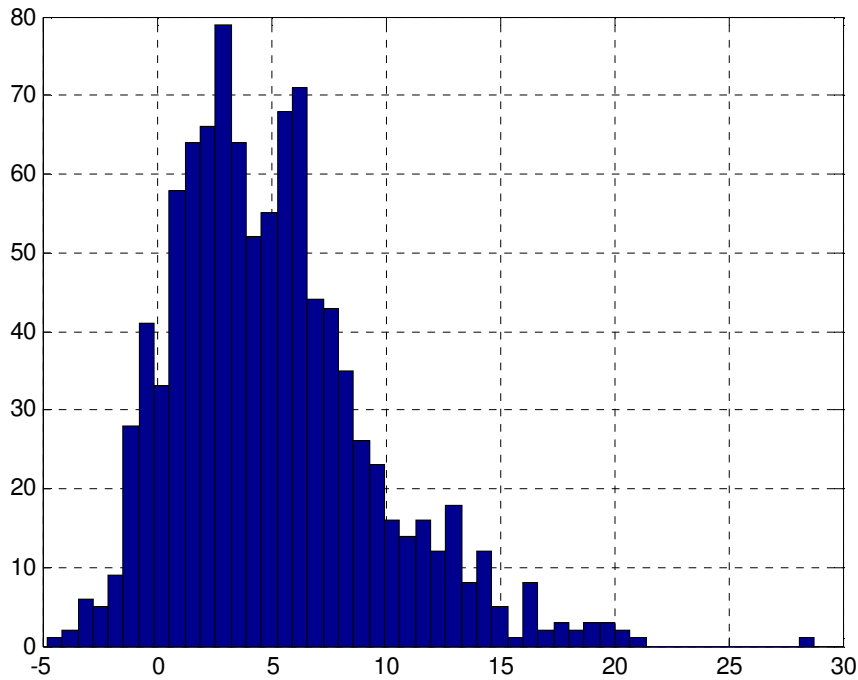
Figure 3.12: The Histogram Method

The discontinuity is the most drawback of histogram. It causes extreme difficulty if derivatives of the estimate are required, so the cause of using alternative methods is quite strong [23].

3.4.2. Kernel Density Estimation Method

The histogram is the picture of a density estimator that spreads the probability mass of each sample item uniformly throughout the interval (i.e. bin) it is observed in. Note that the observations are in no way expected to be uniformly spread out within any particular interval, so the mass is not spread equally around the observation unless it happens to fall exactly in the center of the interval [27].

In this section, we focus on the kernel density estimator that more fairly spreads out the probability mass of each observation, not arbitrarily in a fixed interval, but smoothly around the observation, typically in a symmetric way.

The kernel estimator is given by

$$\hat{f}_{Ker}(x) = \frac{1}{nh_n}\sum_{i=1}^{n}K(\frac{x-X_i}{h_n}) \tag{3.95}$$

Where the function $K(x)$ is called a kernel and smoothing function h_n is a positive sequence of bandwidths analogous to the bin width in a histogram.

The kernel function must satisfy the condition that

$$\int K(x)dx = 1 \tag{3.96}$$

Usually, the kernel is a symmetric probability density function, and the kernel function is often chosen in such a way that it has mathematically tractable properties, such as continuity or differentiability. Some examples include the Gaussian kernel function, the epanechnikov kernel function, or triweight kernel function [28].

The illustration of this procedure is shown in Figure 3.13
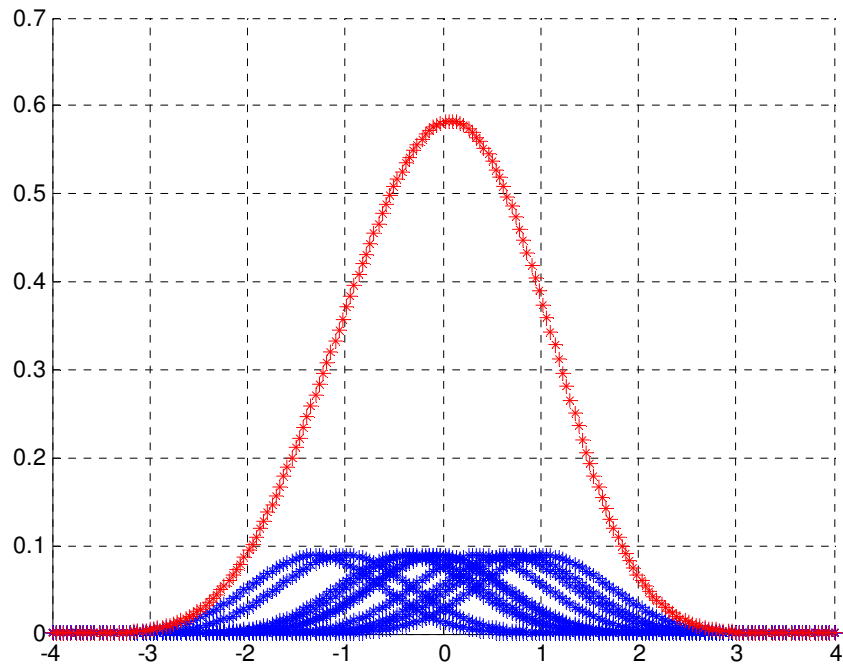
Figure 3.13: Kernel Density Estimation

CHAPTER FOUR

DESCRIPTION OF THERMAL POWER PLANTS

4.1. Introduction

The course of thermal power plants is concerned with the transformation of primary energy in fossil or fuel into electrical energy. The availability of electrical energy is a measure of national standard of living in any country. The electricity has been the most preferred form of energy in most countries [29]. A power plant may be defined as a machine or assembly of equipment that generates and delivers a flow of mechanical or electrical energy. The main equipment for the generation of electric power is generator. When coupling it to a prime mover runs the generator, the electricity is generated. The type of prime move determines the type of power plants.

The major power plants are Steam Power Plant, Diesel Power Plant, Gas Turbine Power Plant and Nuclear Power Plants. These power plants are called thermal power plant, because these convert heat into electric energy [30]. The thermal power plants are built usually in the vicinity of water sources needed to provide steam where the gas or crude oil or coal, are used as fuel for that. In thermal power plants, steam, internal combustion engines or gas turbines are the common prime movers of electric generators. Steam is an important medium of producing mechanical energy.

Steam has the advantage that it can be raised from water which is available in abundance. It does not react much with the materials of the equipment of power plant and is stable at the temperature required in the plant. Steam is used to drive steam engines, steam turbines etc. Steam power plant is most suitable where coal is available in abundance. Also it has the desirable characteristic such as higher efficiency, lower cost, ability to burn coal especially of high ash content, and inferior coals, reduced environmental impact in terms of air pollution, reduced water requirement and higher reliability and availability.

A steam power plant must have following equipments.

- Fuel and furnace section. To bring the fuel from the storage and burn it in furnace.

- Boiler and steam section. Heat generated in the furnace is utilized to convert water into steam.

- Turbine generator section to uses the heat energy of steam and performs work.

- Stack gas path and cleanup.

4.2. Fuel and Furnace Section

This is the part of the plant, where the fuel is arrived to boiler furnace. This component consists of Coal conveyor, Stoker and Pulverizer.

4.2.1. Coal Conveyor

With this coal is transported from coal storage place in power plant to the place near the boiler. There are many types of conveyors and the most commonly used are

- **Belt Conveyor**

Figure 4.1 shows a belt conveyor. It consists of an endless belt. Moving over a pair of end drums (rollers).At some distance a supporting roller is provided at the center. The belt is made, up of rubber or canvas. Belt conveyor is suitable for the transfer of coal over long distances. It is used in medium and large power plants. Belt conveyor has the advantage that it the initial cost of the system is not high and power consumption is also low. Average speed of belt conveyors varies between 200-300 r.p.m.

Figure 4.1: Belt conveyor

- **Screw Conveyor**

It consists of an endless helicoids screw fitted to a shaft as shown in Figure 4.2.The screw while rotating in a trough transfers the coal from feeding end to the discharge end. This system is suitable, where coal is to be transferred over shorter

distance and space limitations exist. The initial cost of the system is low. It suffers from the drawbacks that the power consumption is high. Rotation of screw varies between 75-125 r.p.m.

Figure 4.2: Screw Conveyor

- **Bucket Elevator**

It consists of buckets fixed to a chain as shown in Figure 4.3.The chain moves over two wheels. The coal is carried by the buckets from bottom and discharged at the top.
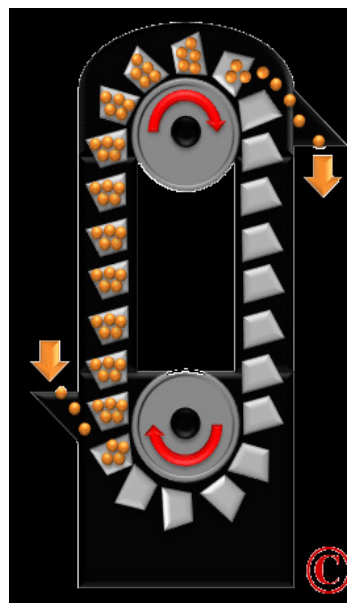
Figure 4.3: Bucket elevator

4.2.2 Stoker

The coal which is brought near by boiler has to put in boiler furnace for combustion. Stokers are mechanical devices located in the furnace and are designed to feed solid fuel onto a grate where the fuel burns as primary air is introduced, with over-fire air also being introduced for enhancing the process of combustion. Stokers are also designed to remove ash residues that remain after combustion.

They are used on large boilers, giving high heat release rates and employed for handling a variety of solid fuels such as coal, wood, bark, rice hulls and municipal waste [32].

Stokers essentially consist of:

a. Fuel feed system.

b. A moving or stationary grate assembly for supporting the burning fuel and admitting the majority of combustion air.

c. An over-fire air system for completing the combustion process and to reduce emissions such as NOx.

d. An ash-discharge system

Generally, two types of stokers systems are available

i. Overfeed stokers where the fuel is supplied from above the grate and air supply is done from below.

ii. Underfeed stokers where both the fuel and air supply are from under the grate.

i. Overfeed Stokers

Overfeed stokers are further classified into two types according to the method of feeding coal to the furnace and by the type of grate [33].

- Chain-grate or traveling-grate stoker

- Spreader stoker

- **Chain-Grate or Traveling-Grate Stoker**

Coal is fed onto one end of a moving steel chain grate. As grate moves along the length of the furnace, the coal burns before dropping off at the end as ash. The coal-feed hopper runs along the entire coal-feed end of the furnace as shown in Figure 4.4.

A coal grate is used to control the rate at which coal is fed into the furnace, and to control the thickness of the coal bed and speed of the grate. Coal must be uniform in size, as large lumps will not burn out completely by the time they reach the end of the grate. As the bed thickness decreases from coal feed end to rear end, different amounts of air are required- more quantity at coal-feed end and less at rear end [34].
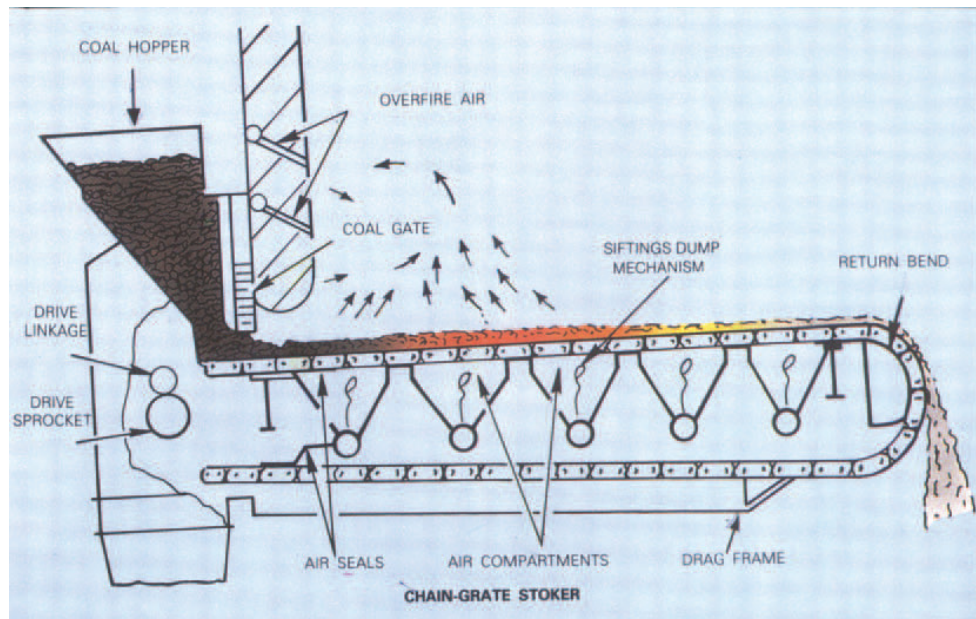
Figure 4.4: Chain -grate stoker

- **Spreader Stoker**

In this type, the fuel is spread uniformly over the grate as it is thrown into the furnace and combustion air enters from below. The fuel fines burn in suspension as they fall against the upward moving air flow. The heavier fuel gets burned on the grate and ash is removed from the discharge end.

Spreader stokers are the most common among the stokers in use presently and have the capacity to handle a wide variety of solid fuels. Spreader stokers consist of a variable feeding device a mechanism for throwing the fuel into the furnace and grates with suitable opening to admit air. Coal falls on the grate and combustion is completed as it slowly moves through the furnace.

The ash falls into the pit when the grates pass over the sprocket. Spreader stoker is shown in Figure 4.5 [32].
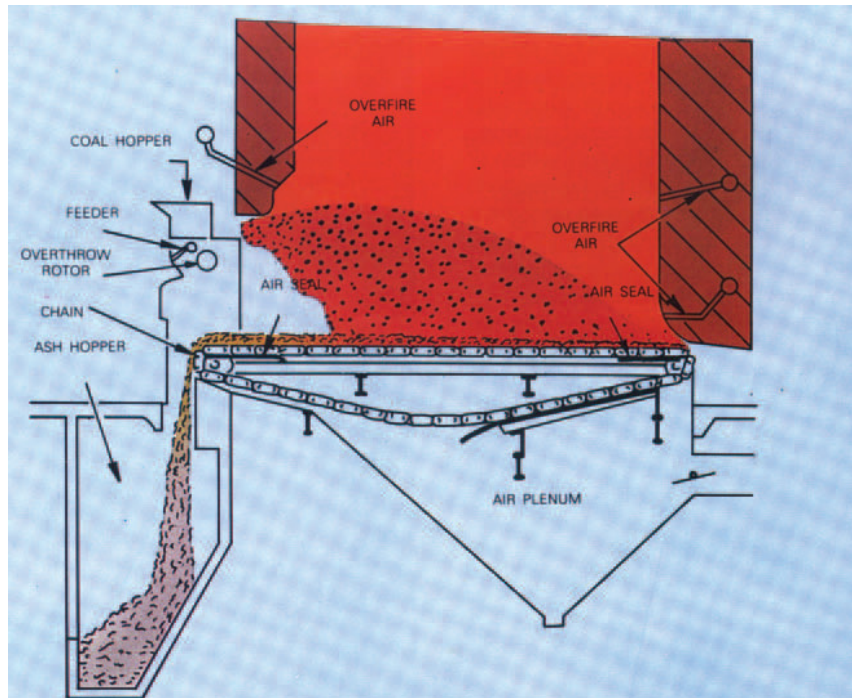
Figure 4.5: Spreader Stoker

ii. Underfeed Stokers

The underfeed stoker as shown in Figure 4.6 is generally used for house heating furnaces and boilers. This type of stoker is one in which the fuel is fed upward from underneath the furnace or boiler. The action of a screw or worm carries the fuel back through a retort from which it passes upward as the fuel above is being consumed.

The ash is generally deposited on dead plates on either side of the retort, from which it can be removed. Underfeed stokers can be designed for use with either anthracite or bituminous coal, but the individual pieces of coal should be uniform in size and no larger than 1 inch in diameter. It is desirable to treat the coal with oil in order to eliminate dust [35].
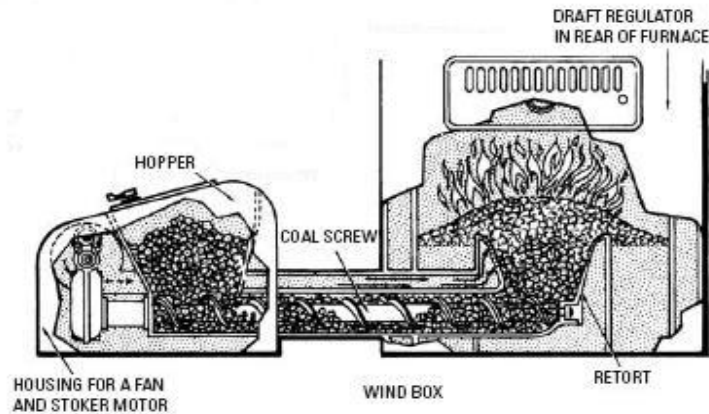
Figure 4.6: Underfeed Stoker

4.2.3. Pulverizer

Coal is pulverized (powdered) to increase its surface exposure thus permitting rapid combustion. Efficient use of coal depends greatly on the combustion process employed. For large scale generation of energy the efficient method of burning coal is confined still to pulverized coal combustion. The pulverized coal is obtained by grinding the raw coal in pulverising mills.

The essential functions of pulverising mills are as follows: Drying of the coal, grinding and separation of particles of the desired size. Proper drying of raw coal which may contain moisture is necessary for effective grinding. The various pulverising mills used are as follows [36]:

I.   Ball and Tube Mill

Ball tube mills are rotating horizontal cylinders containing steel or special alloy balls. Coal intermingled with the balls is crushed by the impact and grinding action of the balls as the mill rotates. Hot air flowing through the ball mill carries coal particles to classifiers situated at both ends of the mill as shown in Figure 4.7. Large particles rejected in the classifiers are returned to the grinding zone for further size reduction. Small particles passing through the classifier exit the mill through piping leading to burners in direct fired systems and to cyclone collectors in indirect fired (storage bin) systems.
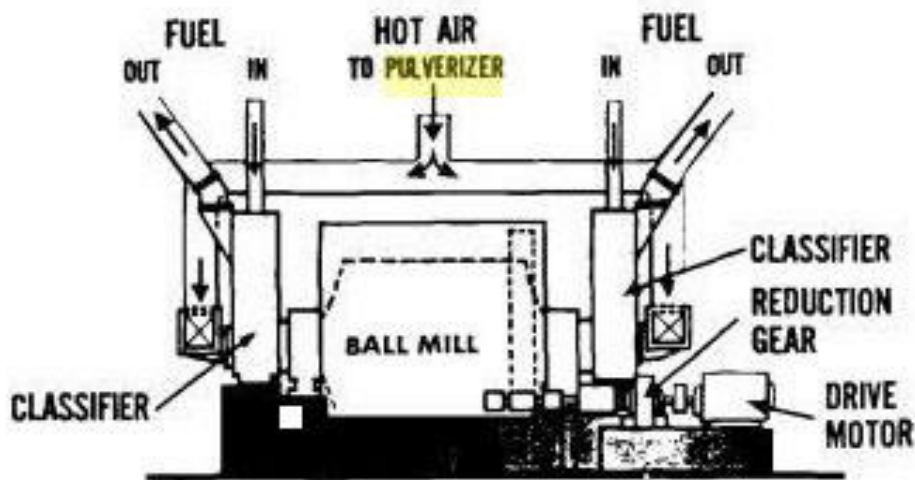
Figure 4.7: Ball and Tube Mill

II.   Ring and Ball Mill

Ball-ring mills are technical appliances designed for grinding (pulverising) raw materials, such as: hard coal, limestone, lime, etc. Ball-ring mills are applied in thermal power plants, energy industry, cement and lime processing, chemical works etc. This type consists of two rings separated by a series of large balls. The lower ring rotates, while the upper ring presses down on the balls via a set of spring and adjuster assemblies. Coal is introduced into the center or side of the pulverizer (depending on the design) and is ground as the lower ring rotates causing the balls to orbit between the upper and lower rings. The coal is carried out of the mill by the flow of air moving through it. The size of the coal particles released from the grinding section of the mill is determined by a classifier separator [33].

Ball-ring mills, thanks to their construction, are highly reliable machines with top operational qualities. Reduction of maintenance downtime makes it possible to reach an availability rate from 90 to 95%. The construction of ball-ring mills makes them resistant to hard inclusions that can occur in ground materials, so the machines are suitable for grinding hardest and most abrasive materials. Figure 4.8 illustrates Ring and Ball Mill [37].
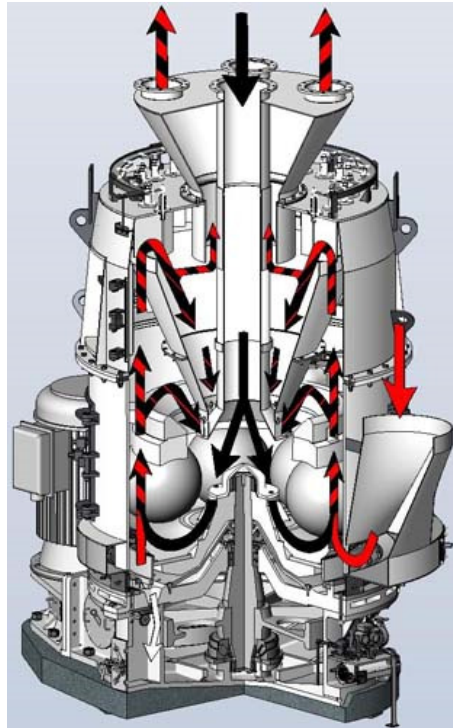
Figure 4.8: Ring and Ball Mill

4.3. Boiler and Steam Section

Boiler is an apparatus to produce steam. Thermal energy released by combustion of fuel is transferred to water, which vaporizes and gets converted into steam at the desired temperature and pressure. The steam produced is used for producing mechanical work in steam turbine.

A boiler should fulfill the following requirements

i. **Safety.** The boiler should be safe under operating conditions.

ii. **Accessibility.** The various parts of the boiler should be accessible for repair and maintenance.

iii. **Capacity.** The boiler should be capable of supplying steam according to the requirements.

iv. **Efficiency.** To permit efficient operation, the boiler should be able to absorb a maximum amount of heat produced due to burning of fuel in the furnace.

v. It should be simple in construction and its maintenance cost should be low.

vi. Its initial cost should be low.

vii. The boiler should have no joints exposed to flames.

The performance of a boiler may be measured in terms of its evaporative capacity also called power of a boiler. It is defined as the amount of water evaporated or steam produced in kg per hour [30].

4.3.1. Classification of Boilers

Most conventional steam boilers are classed as either fire-tube or water tube types.

I. Fire-Tube Boilers

Fire tube boilers are the most common heat generators used for production of hot water and steam. They have a round combustion chamber (also called the furnace) where combustion of fossil fuels takes place and tubes (also called fire- or smoke tubes). Behind the heat transfer wall is the boiler water that is heated up by energy released from combustion. In a steam boiler, the water turns to steam above the water line in the boiler and fills the steam space located at the top of the boiler in a cylindrically shaped drum, where it leaves through the steam valve.

A safety valve is set to allow escape of steam at pressures above normal operating pressure; this device is necessary on all boilers, because continued addition of heat to water in a closed vessel without means of steam escape result in a rise in pressure and ultimately in explosion of the boiler. Gases, products of combustion, leave the boiler after giving up heat to boiler internal walls from where the heat is transferred to the water. The furnace is filled with flame and gases, and tubes are filled with gases, giving the boiler the name - fire-tube. Fire-tube boilers have the advantage of being easy to install and operate. They are widely used in small installations to heat buildings and to provide power for factory processes. Figure 4.9 represent Fire-tube boilers.
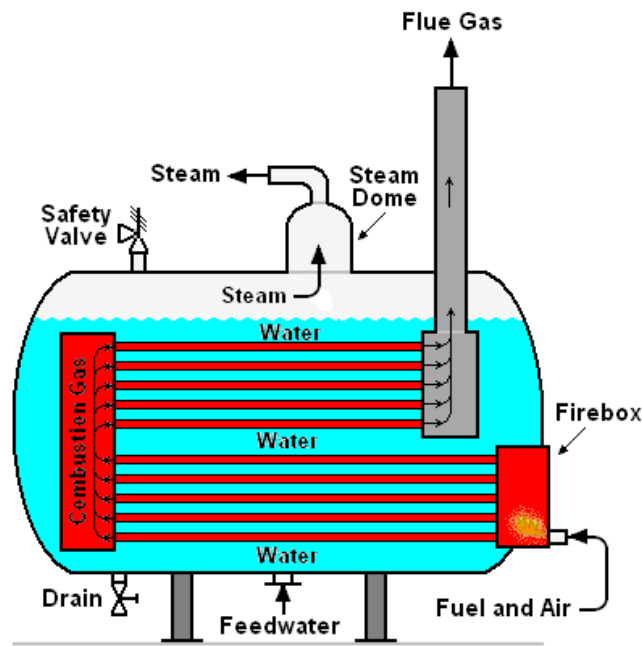
Figure 4.9: Fire-tube boiler

Various advantages of fire tube boilers are as follows.

o Low cost

▪ Fluctuations of steam demand can be met easily

• It is compact in size.

II.   Water-Tube Boilers

In these boilers water is inside the tubes and hot gases are outside the tubes. They consist of drums and tubes. They may contain any number of drums .The most common Water-tube generators are two-drum generators that feed water enters the boiler to one drum (here it is drum below the boiler).This water circulates through the tubes connected external to drums. A hot gas which surrounds these tubes will convert the water in tubes in to steam [38].

This steam is passed through tubes and collected at the top of the drum since it is of light weight. So the drum store steam and water (upper drum).The entire steam is collected in one drum and it is taken out from there. As the movement of water in the water tubes is high, so rate of heat transfer also becomes high resulting in greater efficiency. Most modern water boiler tube designs are within the capacity range

4,500–120,000 kg/hour of steam, at very high pressures [39]. Figure 4.10 is a diagram of Water Tube Boiler
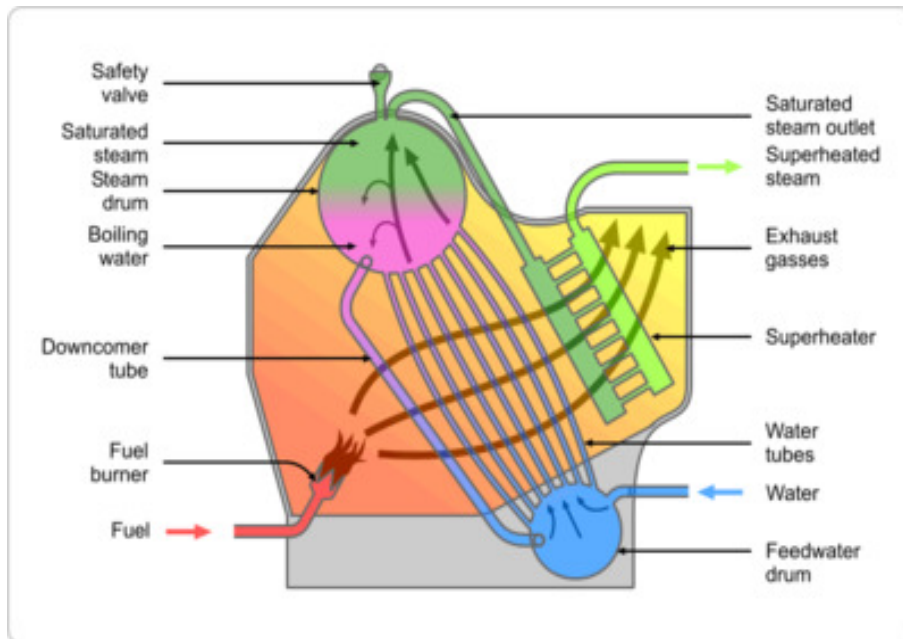
Figure 4.10: Water-tube boiler

Water tube boilers have various advantages as follows.

o Heating surface is large. Therefore steam can be generated easily.

▪ Large heating surface can be obtained by use of large number of tubes.

• Because of high movement of water in the tubes the rate of heat transfer becomes large resulting into a greater efficiency

• **Merits and Demerits of Water tube boilers Over Fire Tube boilers Merits**

i. Generation of steam is much quicker due to small ratio of water content to steam content. This also helps in reaching the steaming temperature in short time.

ii. Its evaporative capacity is considerably larger and the steam pressure range is also high-200 bar.

iii. Heating surfaces are more effective as the hot gases travel at right angles to the direction of water flow.

iv. The combustion efficiency is higher because complete combustion of fuel is possible as the combustion space is much larger.

v. The thermal stresses in the boiler parts are less as different parts of the boiler remain at uniform temperature due to quick circulation of water.

vi. The boiler can be easily transported and erected as its different parts can be separated.

vii. Damage due to the bursting of water tube is less serious. Therefore, water tube boilers are sometimes called safety boilers.

viii. All parts of the water tube boilers are easily accessible for cleaning, inspecting and repairing.

viiii. The water tube boiler's furnace area can be easily altered to meet the fuel requirements.

Demerits:

i. It is less suitable for impure and sedimentary water, as a small deposit of scale may cause the overheating and bursting of tube. Therefore, use of pure feed water is essential.

ii. They require careful attention. The maintenance costs are higher.

iii. Failure in feed water supply even for short period is liable to make the boiler over-heated [36].

- **A Good Boiler Must Possess The Following Qualities:**

i. The boiler should be capable to generate steam at the required pressure and quantity as quickly as possible with minimum fuel consumption.

ii. The initial cost, installation cost and the maintenance cost should be as low as possible.

iii. The boiler should be light in weight, and should occupy small floor area.

iv. All the parts of the boiler should be easily approachable for cleaning and inspection.

v. The boiler should have a minimum of joints to avoid leaks which may occur due to expansion and contraction.

vi. The water and flue gas velocities should be high for high heat transfer rates with minimum pressure drop through the system.

vii. The boiler should conform to the safety regulations as laid down in the Boiler Act.

4.3.2. Superheater

The superheater is situated at the hottest part of the boiler. It is employed to raise the steam temperature above the saturation temperature by absorbing heat from fuel gases as shown in Figure 4.11. The maximum temperature to which steam can be heated will depend upon the metallurgy and economy in initial cost and maintenance cost of the superheater. The superheating of steam makes it possible to recover more energy from steam which improves the cycle efficiency of the plant.

It also eliminates the formation of water vapor during conveying of steam in pipelines and during its early flow through the turbine blades. From the superheater, the steam is led to high-pressure turbine [40]. Non-superheated steam is called saturated steam or wet steam. Superheaters were applied to steam locomotives in quantity from the early 20th century, to most steam vehicles, and to stationary steam engines including power stations [41].
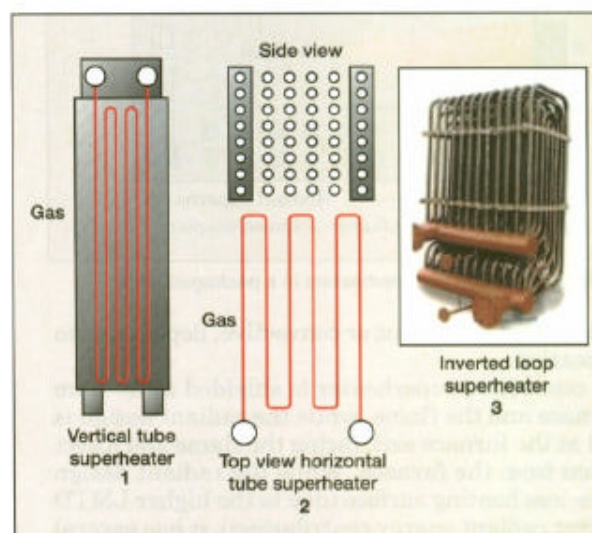
Figure 4.11: Super heater

4.3.3. Reheater

Power plant furnaces may have a reheater section containing tubes heated by hot flue gases outside the tubes. Exhaust steam from the high pressure turbine is rerouted

to go inside the reheater tubes to pickup more energy to go drive intermediate or lower pressure turbines [42].

4.3.4. Economizer

A boiler economizer is a heat exchanger device that captures the "lost or waste heat" from the boiler's hot stack gas. The economizer typically transfers this waste heat to the boiler's feed-water or return water circuit, but it can also be used to heat domestic water or other process fluids. Capturing this normally lost heat reduces the overall fuel requirements for the boiler. Less fuel equates to money saved as well as fewer emissions - since the boiler now operates at a higher efficiency. This is possible because the boiler feed-water or return water is pre-heated by the economizer therefore the boilers main heating circuit does not need to provide as much heat to produce a given output quantity of steam or hot water.

Again fuel savings are the result. Boiler economizers improve a boiler's efficiency by extracting heat from the flue gases discharged. Each economizer is specifically designed to match our client's boiler characteristics in order to maximize efficiency and the use of boiler room space. Because Systems Equipment Corporation Boiler Economizers are manufactured from stainless steel the usual corrosion problems encountered by our competitions designs are eliminated [43]. Figure 4.12 represent a boiler economizer
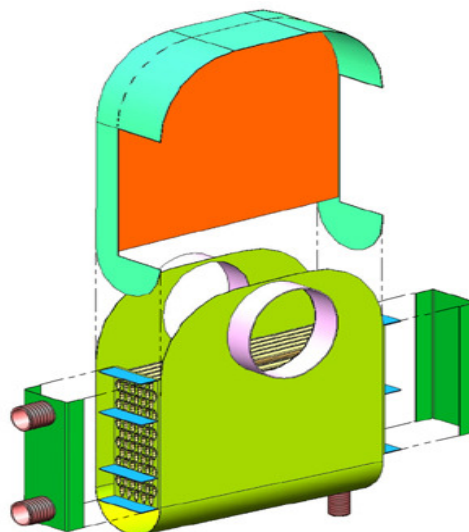
Figure 4.12: Economizer

4.3.5. Condenser

Steam after rotating steam turbine comes to condenser. Condenser refers here to the shell and tube heat exchanger (or surface condenser) installed at the outlet of every steam turbine in Thermal power stations of utility companies generally. These condensers are heat exchangers which convert steam from its gaseous to its liquid state, also known as phase transition.

The purpose is to condense the outlet (or exhaust) steam from steam turbine to obtain maximum efficiency and also to get the condensed steam in the form of pure water, otherwise known as condensate, back to steam generator or (boiler) as boiler feed water. Condensers are classified as Jet condensers or contact condensers and Surface condensers.

I.   Jet Condensers

In Jet condensers the steam to be condensed mixes with the cooling water and the temperature of the condensate and the cooling water is same when leaving the condenser. Types of jet condensers are

- **Parallel-Flow Type of Jet Condensers**

The exhaust steam and cooling water find their entry at the top of the condenser and then flow downwards and condensate and water are finally collected at the bottom, as shown in Figure 4.13 [36].
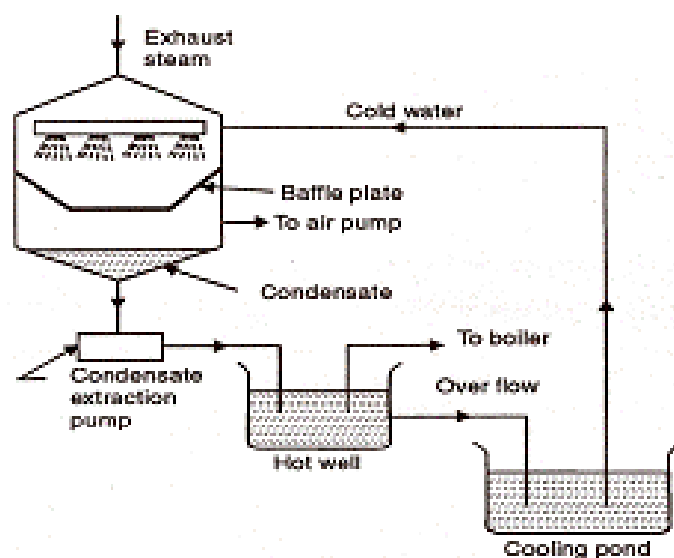
Figure 4.13: Parallel-Flow Type of Jet Condensers

- **High level or Barometric Condenser**

Figure 4.14 shows a high-level jet condenser. It is also called barometric condenser. In this type the condenser shell is placed at a height of 10.33 m (barometric height) above the hot well and thus the necessity of providing an extraction pump can be obviated [44].
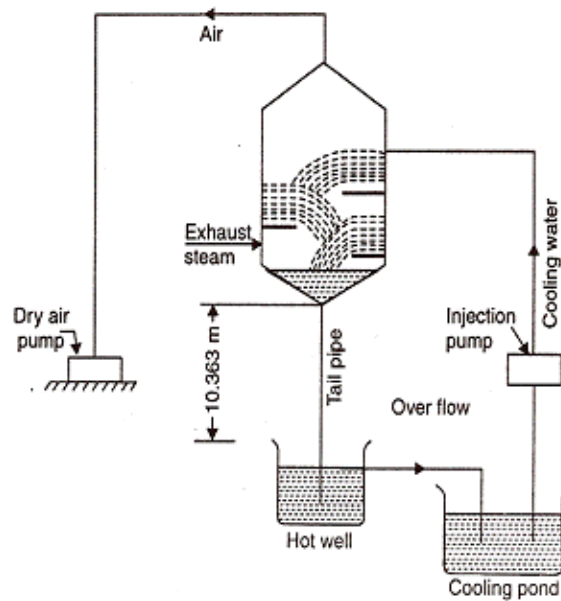
Figure 4.14: High level or Barometric Condenser

- **Ejector Condenser**

Figure 4.15 shows an ejector condenser. In this condenser cold water is discharged under a head of about 5 to 6 m through a series of hollow truncated cones. The steam and air enter the condenser through a non-return valve. Mixing with water condenses steam. Due to this decreased pressure exhaust steam along with associated air is drawn through the truncated cones and finally lead to diverging cone. Cooling water and air is discharged into the hot well [36].
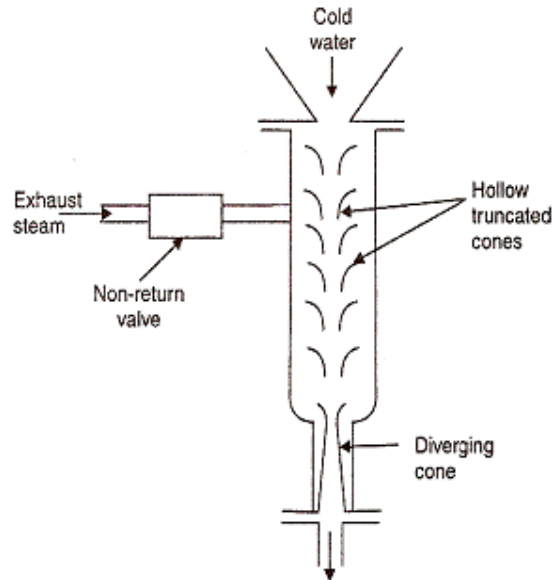
Figure 4.15: Ejector Condenser

II. Surface Condensers.

Here, there is no direct contact between the steam to be condensed and the circulating cooling water. There is a wall interposed between them through heat must be convectively transferred. The temperature of the condensate may be higher than the temperature of the cooling water at outlet and the condensate is recovered as feed water to the boiler. Both the cooling water and the condensate are separately with drawn. Because of this advantage surface condensers are used in thermal power plants. Final output of condenser is water at low temperature is passed to high pressure feed water heater; it is heated and again passed as feed water to the boiler. Since we are passing water at high temperature as feed water the temperature inside the boiler does not decrease and boiler efficiency also maintained [33].

The surface condensers may be classified as follows:

- **Down-Flow Type**

The cooling water enters the shell at the lower half section and after traveling through the upper half section comes out through the outlet. The exhaust steam entering shell from the top flows down over the tubes and gets condensed and is finally removed by an extraction pump as shown in Figure 4.16 [44].
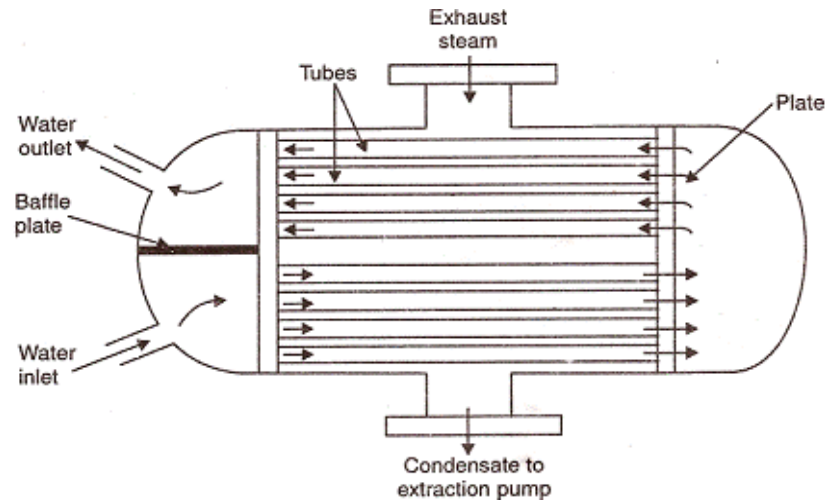
Figure 4.16: Down-Flow Type

- **Central Flow Condenser**

  In this type of condenser, the suction pipe of the air extraction pump is located in the centre of the tubes which results in radial flow of the steam. The better contact between the outer surface of the tubes and steam is ensured. Figure 4.17 illustrates central flow condenser
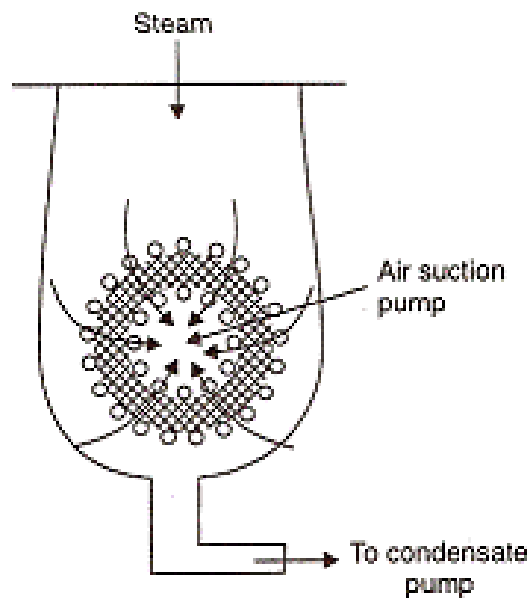
Figure 4.17: Central Flow Condenser

- **Evaporation Condenser**

  In this condenser, steam to be condensed is passed through a series of tubes and the cooling waterfalls over these tubes in the form of spray and it usually has one or

more fans. A steam of air flows over the tubes to increase evaporation of cooling water, which further increases the condensation of steam. Figure 4.18 represents Evaporation condenser [36].
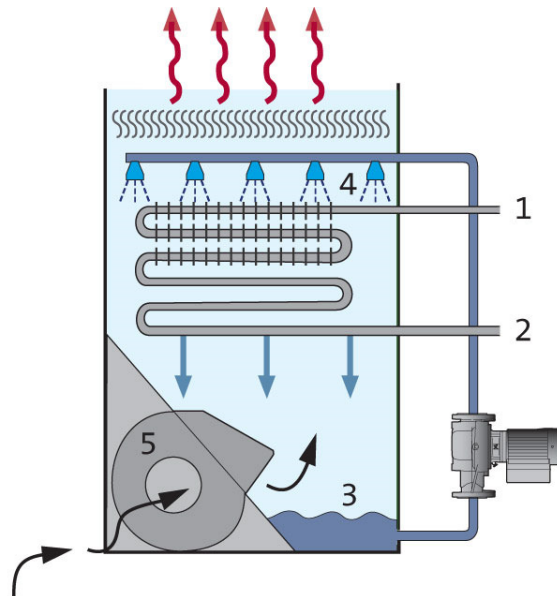
Figure 4.18: Evaporation condenser

4.3.6. Air Path

External fans are provided to give sufficient air for combustion. The forced draft fan takes air from the atmosphere and, first warming it in the air preheater for better combustion, injects it via the air nozzles on the furnace wall. The induced draft fan assists the FD fan by drawing out combustible gases from the furnace, maintaining a slightly negative pressure in the furnace to avoid backfiring through any opening [42].

4.4. Steam Turbine Section

A steam turbine is a mechanical device whose purpose is to convert thermal (or heat) energy into work. The steam turbine uses thermal energy from steam under pressure and converts it into rotary motion, or mechanical work.

The original version of the steam turbine was the steam engine. The steam turbine is highly thermodynamically efficient, meaning that it acts at a high level of performance without losing much energy in the process, and has a low power-to-weight ratio. This is a ratio of actual performance used to compare devices across their respective weights.

These qualities make the steam turbine ideal for power stations. Its thermo dynamical efficiency comes from the fact that it has multiple stages for steam expansion. Steam turbines come in a variety of sizes, from small 0.75 kW units that are used in pumps and compressors, to 1.5 million kW units that are used to generated electricity.

Energy from steam is converted into mechanical work via expansion. Expansion occurs through a series of nozzles (or fixed blades) and moving blades within the turbine. The nozzles are arranged with the circular turbine casing that is specifically designed to withstand steam pressure.

4.4.1. Classification of Steam Turbine

The design of turbine blades is based on two fundamental principles: impulse turbine and reaction turbine.

I.Impulse Turbine

If the flow of steam through the nozzles and moving blades of a turbine takes place in such a manner that the steam is expanded only in nozzles and pressure at the outlet sides of the blades is equal to that at inlet side; such a turbine is termed as impulse turbine because it works on the principle of impulse. In other words, in impulse turbine, the drop in pressure of steam takes place only in nozzles and not in moving blades. This is obtained by making the blade passage of constant cross-section area. As a result, the velocity of the steam decreases when it exits the blades, but the pressure remains constant. Therefore, energy is transferred by the change in velocity of the steam and not by pressure [36]. Figure 4.19 illustrates Impulse turbine
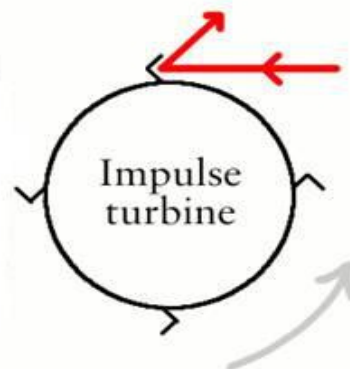
Figure 4.19: Impulse turbine

II.Reaction Turbine

The reaction turbine principle depends on the blade diverting the flow of steam and gaining kinetic energy in the process. Reaction turbines have a different blade cross section than impulse turbines. The reaction blades act like the wings of a plane, while the impulse blades act like an engine piston, just like the wing of an airplane, the kinetic energy is converted to power by decreasing the steam's velocity and lowering pressure. When the steam enters the blade and travels across it, there is a decrease in pressure on the upper surface and an increase in pressure on the lower surface. As a result, the force that drives the turbine is a reaction force [45]. Figure 4.20 shows Reaction turbine
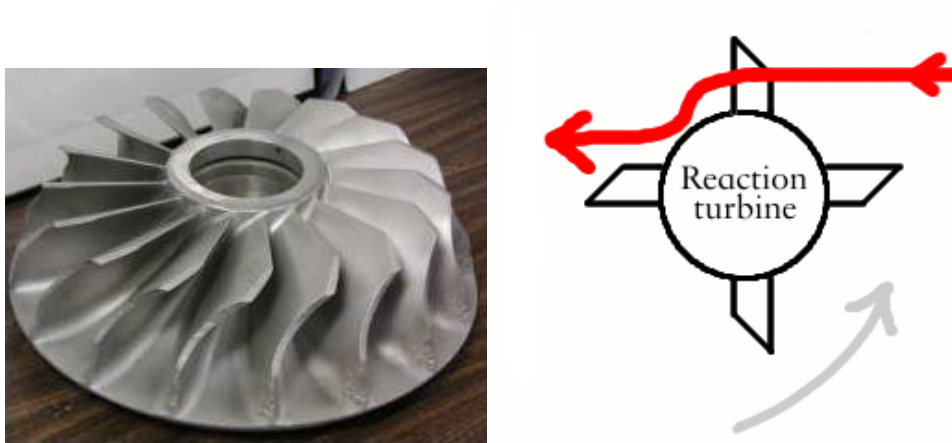
Figure 4.20: Reaction turbine

The next figure illustrate the difference between Impulse Turbine and Reaction Turbine with regard to steam flow
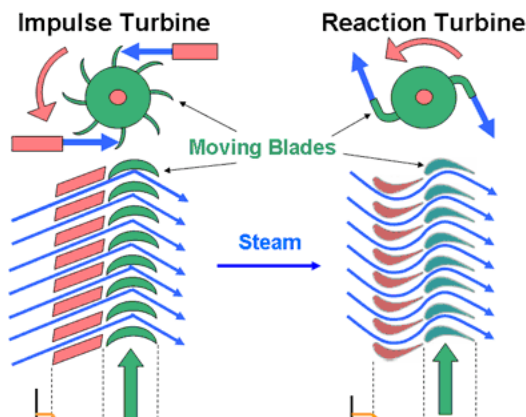
Figure 4.21: the difference between Impulse Turbine and Reaction Turbine

Other than the operating and controlling equipment, similarity exists in both the impulse and reaction turbines. These include casings, nozzles, rotors, bearings, and shaft glands.

4.4.2. Casings

The materials used to construct turbines will vary somewhat depending on the steam and power conditions for which the turbine is designed. Turbine casings shows in Figure 4.22 are made of cast carbon steel for non superheated steam applications. Superheated applications use casings made of carbon molybdenum steel [46].
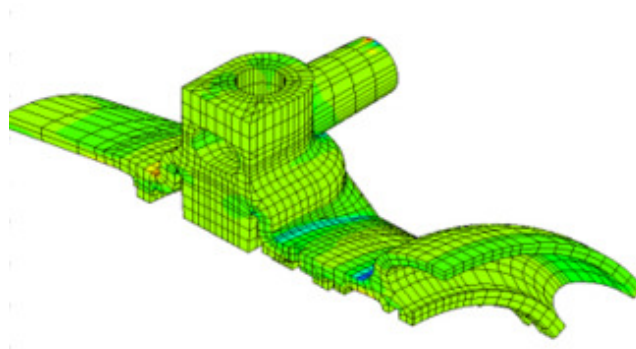
Figure 4.22: Casing

4.4.3. Nozzles

The primary function of the nozzles is to convert the thermal energy of steam into kinetic energy. The secondary function of the nozzles is to direct the steam against the blades [47]. Figure 4.23 illustrates Nozzles
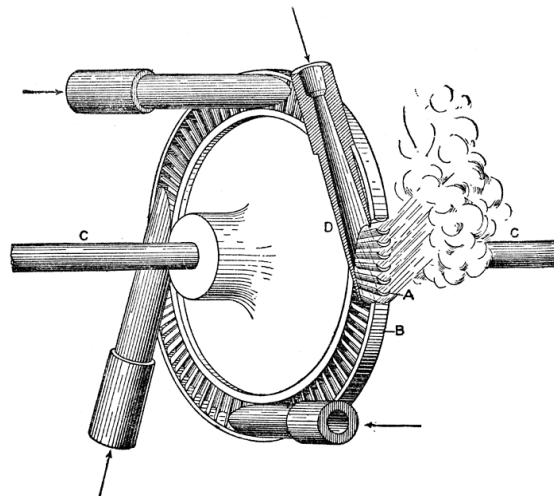
Figure 4.23: Nozzles

4.4.4. Rotors

Rotors as shown in Figure 4.24 (forged wheels and shaft) are manufactured from steel alloys. The primary purpose of a turbine rotor is to carry the moving blades that convert the steam's kinetic energy to rotating mechanical energy [46].

Figure 4.24: Rotors

4.4.5. Bearings

The rotor of every turbine must be positioned radially and axially by bearings. Radial bearings carry and support the weight of the rotor and maintain the correct radial clearance between the rotor and casing [46]. Figure 4.25 shows sliding surface bearing.
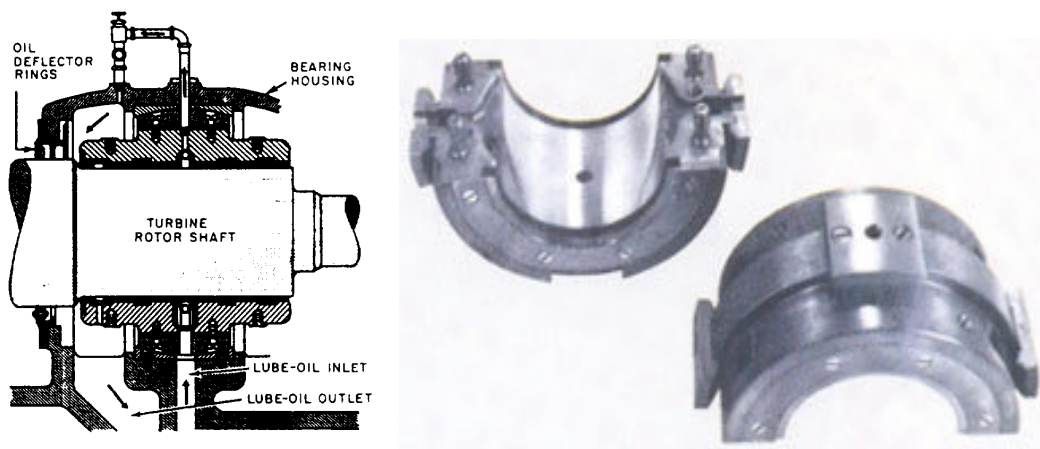
Figure 4.25: sliding surface bearing

4.4.6. Shaft Packing Glands

Shaft packing glands prevent the leaking of steam out of or air into the turbine casing where the turbine rotor shaft extends through the turbine casing. Labyrinth and carbon rings are two types of packing [47]. They are used either separately or in combination. Figure 4.26 illustrates Carbon packing gland.
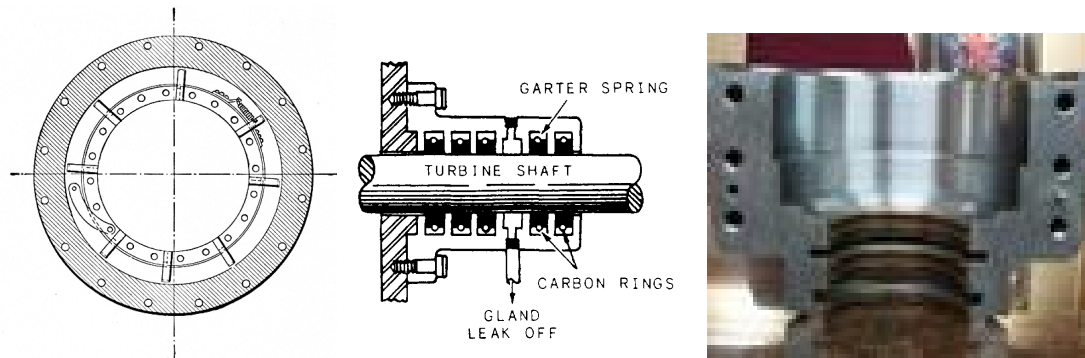
Figure 4.26: Carbon packing gland

Figure 4.27 illustrates overall construction of steam turbine
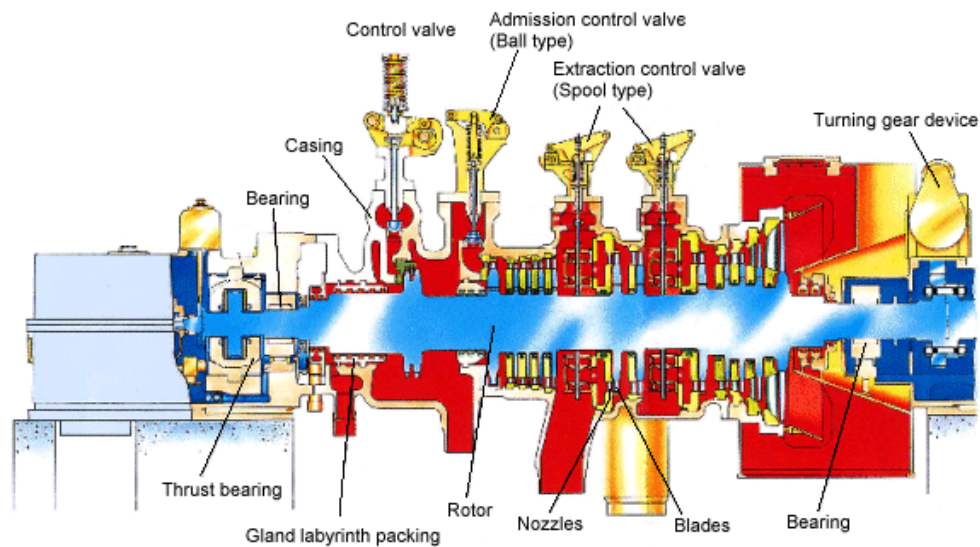
Figure 4.27: steam turbine

4.4.7. Steam Turbine Performance

Turbine performance can be expressed by the following factors:

i. The steam flow process through the unit-expansion line or condition curve.

ii. The steam flow rate through the unit.

iii. Thermal efficiency.

iv. Losses such as exhaust, mechanical, generator, radiation etc.

Mechanical losses include bearing losses, oil pump losses and generator bearing losses. Generator losses include will electrical and mechanical losses. Exhaust losses include the kinetic energy of the steam as it leaves the last stage and the pressure drop from the exit of last stage to the condenser stage. For successful operation of a steam turbine it is desirable to supply steam at constant pressure and temperature. Steam pressure can be easily regulated by means of safety valve fitted on the boiler. The steam temperature may try to fluctuate because of the following reasons:

i. Variation in heat produced due to varying amounts of fuel burnt according to changing loads.

ii. Fluctuation in quantity of excess air.

iii. Variation in moisture content and temperature of air entering the furnace.

iv. Variation in temperature of feed water.

v. The varying condition of cleanliness of heat absorbing surface [38].

4.4.8. Steam Turbine Generators

The Turbine is connected to a Generator and a generator is an alternator is an electromechanical device that converts the mechanical shaft energy it receives from the turbine into electrical energy. Steam turbine driven a.c. synchronous generators (alternators) are of two or four pole designs. These are three phase measuring machines offering economic, advantages in generation and transmission.

Generator losses appearing as heat must be constantly removed to avoid damaging the windings. Large generators have cylindrical rotors with minimum of heat dissipation surface and so they have forced ventilation to remove the heat. Large generators generally use an enclosed system with air or hydrogen coolant. Figure 4.28 represents turbine generator section [30].
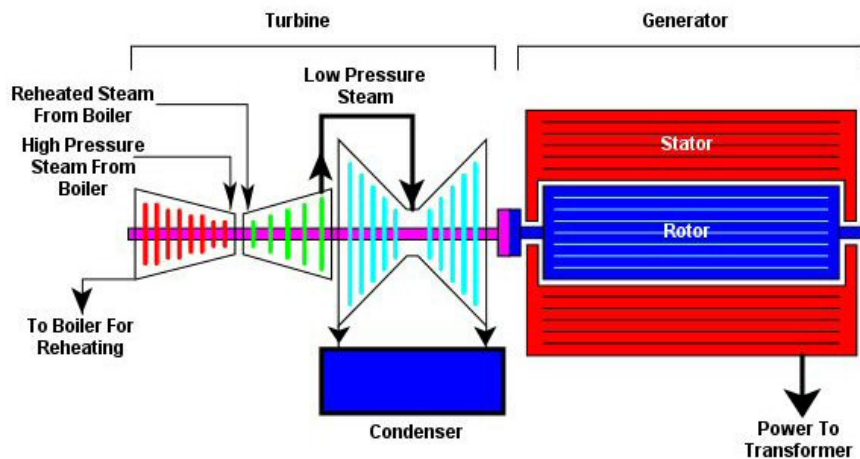
Figure 4.28: Turbine Generator

4.5. Stack Gas Path and Cleanup

4.5.1. Air Preheater

The remaining heat of flue gases is utilized by air preheater. It is a device used in steam boilers to transfer heat from the flue gases to the combustion air before the air enters the furnace and it is kept at a place near by where the air enters in to the boiler. The purpose of the air preheater is to recover the heat from the flue gas from the boiler to improve boiler efficiency by burning warm air which increases combustion efficiency, and reducing useful heat lost from the flue.

As a consequence, the gases are also sent to the chimney or stack at a lower temperature, allowing simplified design of the ducting and stack [33].

4.5.2. Ash Disposal

A large quantity of ash is, produced in steam power plants using coal. Ash produced in about 10 to 20% of the total coal burnt in the furnace. Handling of ash is a problem because ash coming out of the furnace is too hot, it is dusty and irritating to handle and is accompanied by some poisonous gases. It is desirable to quench the ash before handling due to following reasons:

1. Quenching reduces the temperature of ash.

2. It reduces the corrosive action of ash.

3. Quenching reduces the dust accompanying the ash.

Handling of ash includes its removal from the furnace, loading on the conveyors and delivered to the fill from where it can be disposed off.

The handling equipment should perform the following functions:

1. Capital investment, operating and maintenance charges of the equipment should be low.

2. It should be able to handle large quantities of ash.

3. Clinkers, soot, dust etc. create troubles; the equipment should be able to handle them smoothly.

4. The equipment used should remove the ash from the furnace, load it to the conveying system to deliver the ash to a dumping site or storage and finally it should have means to dispose of the stored ash.

5. The equipment should be corrosion and wear resistant [36].

4.5.3. Smoke and Dust Removal

In coal fed furnaces the products of combustion contain particles of solid matter floating in suspension. This may be smoke or dust. The production of smoke indicates that combustion conditions are faulty and amount of smoke produced can be reduced by improving the furnace design. The size of dust particles is designated in microns (1 μ = 0.001 mm). Dust particles are mainly ash particles called fly ash intermixed with some quantity of carbon ash material called cinders. Gas borne particles larger than 1μ in diameter are called dust and when such particles become greater in size than 100p they are called cinders. Smoke is produced due to the incomplete combustion of fuels, smoke particles are less than 10p in size.

The disposal smoke to the atmosphere is not desirable due to the following reasons:

1. A smoky atmosphere is less healthful than smoke free air.

2. Smoke is produced due to incomplete combustion of coal. This will create a big economic loss due to loss of heating value of coal.

3. In a smoky atmosphere lower standards of cleanliness are prevalent. Buildings, clothings, furniture etc. becomes dirty due to smoke. Smoke corrodes the metals and darkens the paints.

To avoid smoke nuisance the coal should be completely burnt in the furnace. The presence of dense smoke indicates poor furnace conditions and a loss in efficiency and capacity of a boiler plant. A small amount of smoke leaving chimney shows good furnace conditions whereas smokeless chimney does not necessarily mean a better efficiency in the boiler room. To avoid the atmospheric pollution the fly ash must be removed from the gaseous products of combustion before they leaves the chimney. The removal of dust and cinders from the flue gas is usually effected by commercial dust collectors which are installed between the boiler outlet and chimney usually in the chimney side of air preheater [38].

4.6. Cooling Towers

The condensate (water) formed in the condenser after condensation is initially at high temperature. This hot water is passed to cooling towers. A cooling tower is a heat removal system used to remove heat from the condensate (water). Cooling towers allow the water to be cooled and then returned for use in recirculation through the system. This saves enormous amounts of money, time and energy.

There are two ways in which cooling towers work to remove heat, evaporation or the use of air. Temperature measurements taken during each of these cooling processes are called the wet-bulb air temperature and the dry-bulb air temperature. The dry-bulb air temperature is used when heat is removed by exposing the water to air. The wet-bulb temperature is used when heat is removed by the process of evaporation.

I.   Cooling With Air

Water needing to be cooled is pumped to the top of the tower and then directed to flow down a designated path where the water forms into droplets. These droplets are met by a current of air that is blowing upward and past the water. The water is cooled by the air as it passes. It then collects at the bottom of the cooling tower structure where it is returned to the production process. Some air-cooled towers use large fans at the top of the structure to draw the air up [48].

II.   Cooling With Evaporation

It allows a small portion of the water being cooled to evaporate into a moving air stream to provide significant cooling to the rest of that water stream. The heat

from the water stream transferred to the air stream raises the air's temperature and its relative humidity to 100%, and this air is discharged to the atmosphere [45]. Figure 4.29 illustrates cooling tower
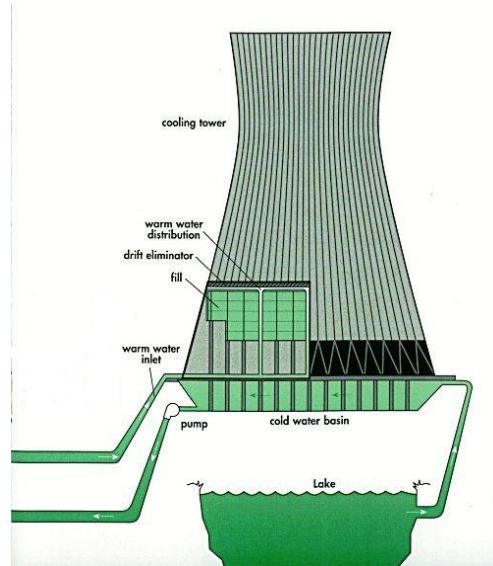
Figure 4.29: Cooling Towers

In the end, we can say in thermal power plants, mechanical power is produced by transforming thermal energy from combustion of a fuel, into rotational energy. The most common source of combustible fuel is fossil fuel: coil, oil and natural gases. The fuel is burned in boilers where the heat is used to convert water into steam. The steam drives turbines that drive power generators [49].A layout of a thermal power plant that uses coal as fuel can be seen in Figure 4.30
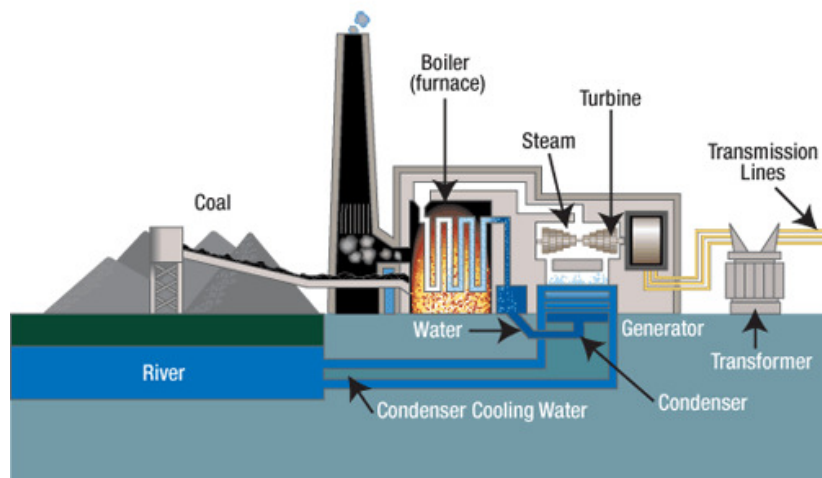
Figure 4.30: Thermal Power Plant That Uses Coal As Fuel

CHAPTER FIVE

Sensor Fault Detection and Isolation in Thermal Power Plants Steam Separator

5.1 Introduction

5.1.1 System Identification

A fundamental concept in science and technology is that of mathematical modeling. A mathematical model is a very useful of describing the knowledge we have about a process or system. The determination of a mathematical model of a process or system is known as system identification. In control systems, a mathematical model of a process or system is in most cases necessary for the design of the controller. The model is also necessary for the design of adaptive and robust control systems. A process or a system may be described by several models and one must be able to choose the suitable type of model for each specific application.

Mathematical models may be distinguished as parametric and nonparametric models. Parametric models obviously involve parameters: for example, the coefficients of differential or difference equations, of state equations, and of transfer functions. Nonparametric models do not involve parameters and are usually graphical representations, such as the Nyquist or Bode diagrams of a transfer function or impulse response function. If we deal with the problem of determining mathematical models of linear, discrete time-invariant, single-input–single-output (siso) dynamic systems, described by difference equations and An interesting feature is the determination of a recursive algorithm, which allows the estimation of the vector parameter θ for $N+1$ measurements, based on the following formula:

$$\theta(N+1) = \theta(N) + \Delta\theta = \theta(N) + \gamma(N)[y_{N+1} - \varphi^T(N+1)\theta(N)] \tag{5.1}$$

Where $\gamma(N)$ and $\varphi(N+1)$ are known vector quantities and y_{N+1} is the $N+1$ measurement of the output y of the system. [14].

5.1.2 An outlier

An outlier is defined as a point that lies very far from the mean of the corresponding random variable. This distance is measured with respect to a given threshold. Points with values very different from the mean value produce large errors during training and may have disastrous effects.

If the number of outliers is very small, they are usually discarded. However, if this is not the case and they are the result of a distribution with long tails, then the designer may have to adopt cost functions that are not very sensitive in the presence of outliers [35].

5.1.3 Maximum Likelihood Estimation

When the type of the distribution is known, the exact distribution (or density) function still depends on some parameters. The simplest examples of such parameters are the mean and the standard deviation of the distribution. Somewhat more complicated but basically the same is the situation when the random variable is the output of a dynamic system; now its distribution depends on the system parameters. Thus the probability density function may be written as $f_X(x|\theta)$, where θ, a scalar or a vector represents the parameter(s) of the distribution which are assumed to be known.

Often some of the parameters are not known; then observations of the random variable, together with the knowledge of the type of the distribution (and its other parameters) may be used to obtain an estimate for the unknown parameters [1].

Let $x_1,..........,x_n$ be random samples drawn from pdf $p(x|\theta)$. We form the joint pdf $p(X|\theta)$, where $X = \{x_1,..........,x_n\}$ is the set of the samples. Assuming statistical independence between the different samples we have

$$L(\theta) = p(X|\theta) \equiv p(x_1,..........,x_n|\theta) = \prod_{k=1}^{n} p(x_k|\theta) \tag{5.2}$$

This is a function of θ and it is also known as the likelihood function of θ with respect to X. For likelihood function $L(\theta)$, where θ is an unknown parameter [43]. Let θ_e be a value of the parameter such that $L(\theta_e) \geq L(\theta)$ for all possible values of θ.

Then, θ_e is called a maximum likelihood estimate (MLE) for θ, that means the parameter estimates are thus obtained by maximizing the likelihood function under the given observations that is [37].

$$\hat{\theta}(x) = \arg\max L(\theta) \tag{5.3}$$

Technically, the estimates are sought by computing the partial derivatives of the likelihood function and equating it with zero

$$\hat{\theta}(x) = \underset{\theta}{sol}[\partial L(\theta)/\partial\theta = 0] \tag{5.4}$$

With the additional requirement that the second derivative of the likelihood function, at the solution, has to be negative

$$[\partial^2 L(\theta)/\partial\theta^2\Big|_{\theta=\hat{\theta}} < 0 \tag{5.5}$$

5.1.4 Adaptive control

In common sense, "to adapt" means to change a behavior or characteristic to conform to new and unknown circumstances. In the sense of control theory and engineering, an adaptive controller is an "intelligent" controller that can modify its behavior in response to the variations in the dynamics of the process and the character of the disturbances.

As defined in and simply put, an adaptive control system is a system which adjusts automatically on-line the parameters of its controller, so as to maintain a satisfactory level of performance when the parameters of the system under control are unknown and/or time varying [35]. Adaptive control schemes have been applied in the paper industries, rolling mills, power plants, chemical reactors, cement mills, autopilots for aircrafts, and ships, etc.

The use of adaptive controllers may lead to improvement of product quality, increase in production rates, fault detection, and energy saving [38]. There are different types of adaptive control schemes namely least squares estimation, dynamic inversion with neural networks, and model reference adaptive control. In This chapter we will use recursive least square estimation in our simulation that we described this method in section (2.5).

5.1.5 Robustness in Fault Detection System

Usually, the parameters of the system vary with time, and the characteristics of the disturbances and noises are unknown so that they can not be modeled accurately. Since an accurate mathematical model of a physical process is not always available, there is often a mismatch between the actual process and its mathematical model, even if no

fault in the process occurs. This constitutes a source of false alarm, which can corrupt the performance of the fault detection and diagnosis system. The effect of modeling uncertainties, disturbances, and noise is therefore the most crucial point in the model based FDD concept, and the solution to these problems is the key for its practical applicability. To overcome these difficulties, FDD system has to be made robust to such modeling errors and disturbances.

In the context of automatic control, the term robustness is used to describe the insensitivity or invariance of the performance of control systems with respect to disturbances, model-plant mismatches or parameter variations, but in contrast to automatic control systems, they must not be robust to actual faults. On the contrary, while generating robustness to disturbances, the designer must maintain or even enhance the sensitivity of fault diagnosis schemes to faults [29]. An FDD system, which is designed to provide both sensitivity to faults and robustness to modeling errors and disturbances, is called a robust FDD scheme. During the last decades, much FDD research has focused on robust fault diagnosis of uncertain systems. Adaptive threshold can be used to increase the robustness to modeling uncertainties.

5.1.6 Quantile-Quantile (QQ)-plot technique

The quantile-quantile (Q-Q) plot is a graphical technique for examines whether or not a sample $x_1, ..., x_n$ has come from a distribution with a given distribution function $F(x)$ [10]. The following steps create a Quantile-Quantile-plot (Q-Q-plot)

1. Rank your data in ascending order: $x_{(1)}, ..., x_{(n)}$

2. for each data point $x_i; i = 1, 2, ..., n$ compute ρ_i. Where

$$\rho_i = (i-1)/(n-1) \approx r_i = (i-0.5)/n \tag{5.6}$$

3. For each ρ_i find the distribution quantile

$$y_i = F^{-1}(\rho_i) \approx F^{-1}(r_i) \tag{5.7}$$

$$y_i = F^{-1}(\rho_i) \approx F^{-1}(r_i); \ \rho_i = (i-1)/(n-1) \approx r_i = (i-0.5)/n; i = 1, 2, ..., n$$

4. Plot the distribution quantile y_i against the sample quantiles x_i (Q-Q plot) [56].

5. If $x_i = y_i$ for all i we get a straight line indicating that the observations have the same distribution function that supposes $F(\cdot)$.

5.2 Robust Adaptive Parameter Identification

5.2.1 Formulation of the Problem

Let us observe an abstract, linear, dynamic, time-invariant, discrete-time system, which can be represented by a difference equation with constant parameters:

$$y(i) = -\sum_{k=1}^{n} a_k y(i-k) + \sum_{k=1}^{m} b_k u(i-k) + \xi(i) \tag{5.8}$$

Where $y(i) \in R^1, u(i) \in R^1, \xi(i) \in R^1$ are the sequences of system output, measurable input and stochastic input, or noise, respectively, while the constants $a_i, i = 1,\ldots\ldots,n$ and $b_j, j = 1,\ldots\ldots,m$ represent system parameters. If a backward shift operator is introduced as $q^{-k} y(i) = y(i-k)$, equation (5.8) can be written in the following polynomial form:

$$A(q^{-1}) y(i) = B(q^{-1}) u(i) + \xi(i) \tag{5.9}$$

Where

$$A(q^{-1}) = 1 + \sum_{k=1}^{n} a_k q^{-k}, B(q^{-1}) = \sum_{k=1}^{m} b_k q^{-k} \tag{5.10}$$

Are the so-called characteristics and control polynomials, respectively. Equation (5.8) can also be written in a linear regression form:

$$y(i) = Z^T(i)\Theta + \xi(i) \tag{5.11}$$

Where the regression vector

$$Z^T(i) = [-y(i-1),...,-y(i-n),u(i-1),...,u(i-m)] \tag{5.12}$$

Represents a vector of input and output measurable samples, and $\Theta^T = \begin{bmatrix} a_1 \cdots a_n \, b_1 \cdots b_m \end{bmatrix}$ represents vector of constant system parameters. The problem of recursive system identification described by equation (5.9) is actually the problem of estimation of unknown parameters included in the vector Θ in real time, based on system input and

output signal measurements. Formulation of the identification problem then reduces to the minimization of empirical average losses [62]:

$$J_k(\Theta) = \frac{1}{k}\sum_{i=1}^{k} \rho(v(i,\Theta)) \tag{5.13}$$

Where $v(i,\hat{\Theta}) = y(i) - \hat{y}(i/\Theta)$ is the output prediction error or measurement residual, with $\hat{y}(i/\Theta)$ being a forecasting model, while $\rho(.)$ is a loss function.

Taking into account equation (5.8) or equation (5.11), the mean-square optimal forecasting model, minimizing the criterion $E\{v^2(i,\Theta)\}$, is given by [52]:

$$\hat{y}(i/\Theta) = [1 - A(q^{-1})]y(i) + B(q^{-1})u(i) = Z^T(i)\Theta \tag{5.14}$$

In the maximum likelihood scheme (ML), the loss function in equation (5.13) is chosen as

$$\rho(\cdot) = -\ln p(\cdot) \tag{5.15}$$

where $p(\cdot)$ is the noise probability density function (pdf) [63] .For the Gaussian noise, $p(\cdot)$ is the quadratic function, and the resulting algorithm minimizing equation (5.13) is the standard linear LS approach. However, the ML method is very sensitive to deviations of the real noise pdf from the assumed one, and in the presence of impulsive noise or outliers, it ceases to work [65]. Thus, one should modify the algorithm to make it more robust. Huber's ρ-function (ML loss function) uses for robust estimation

$$\rho(x) = -\ln(p(x)) = \begin{cases} \dfrac{x^2}{2\sigma^2}, if |x| \le k \\[2ex] \dfrac{k|x|}{\sigma^2} - \dfrac{k^2}{2\sigma^2}, if |x| > k \end{cases} \tag{5.16}$$

where σ^2 is the noise variance. More precisely, this pdf represents the worst-case pdf, minimizing the maximum asymptotic estimation error variance, within the class of ε-contaminated normal pdf's [55]:

$$P_s = \{p | p(\cdot) = (1-\varepsilon)N(\cdot | 0,\sigma^2) + \varepsilon h(\cdot), 0 \le \varepsilon \le 1\} \tag{5.17}$$

Where $N(\cdot|0,\sigma^2)$ denotes the zero-mean normal pdf with the variance σ^2, while $h(\cdot)$ is a zero-mean symmetric pdf with the variance $\sigma_0^2 >> \sigma^2$. Thus, the robust loss function in equation (5.196) is the ML function corresponding to the worst-case pdf within the pre-specified pdf's class equation (5.17), where the tuning constant k depends on the contamination index ε, i.e., $k = k(\varepsilon)$ [53] .This type of statistical robustness is known in the literature as min-max robustness over a family of distributions. Unfortunately, the highly-technical character of min-max robustness makes it relatively inaccessible to applied workers. Thus, in practice, one has to design an estimation procedure having a readily-apparent resistance property of insensitivity to outliers, along with desirable efficient robustness.

5.2.2 Recursive M-robust parameter estimation

Starting from the discussion in the preceding section, a particular M-estimator minimizes the sum of the weighted residuals equation (5.13), where the robust ρ-function is defined in equation (5.16). The tuning parameter k in equation (5.16) has to be chosen so as to provide the desired efficiency at the nominal Gaussian observation model. On the other hand, the derivative of the ρ-function in equation (5.16), the so-called influence function [66]

$$\Psi(x) = \rho^{'}(x) = \min\left(\max\left(\frac{x}{\sigma^2}, -\frac{k}{\sigma^2}\right), \frac{k}{\sigma^2}\right) \tag{5.18}$$

If recursive minimization of the criterion in equation (5.16) is performed, it is simple to arrive at a recursive identification procedure of the form [53].

$$\Theta(i) = \Theta(i-1) + \Gamma(i)Z(i)\Psi(v(i,\Theta(i-1))); \Theta(0) = \Theta_0 \tag{5.19}$$

$$v(i,\Theta) = y(i) - Z^T(i)\Theta \tag{5.20}$$

$$\Gamma(i) = \Gamma(i-1) - \frac{\Gamma(i-1)Z(i)Z^T(i)\Gamma(i-1)}{w^{-1} + Z^T(i)\Gamma(i-1)Z(i)}; \Gamma(0) = \gamma^2 I \tag{5.21}$$

Where γ^2 is a finite positive constant, and I is an identity matrix. Thus, the recursive M-robust algorithms are defined by equations (5.18), (5.19)-(5.21) where $\Theta(0)$ and $\Gamma(0)$

are the initial guesses. The LS procedure, minimizing the quadratic performance index is a suitable algorithm for generating the initial parameter estimates $\Theta(0)$.

In particular, for the zero-mean normal noise $\xi(i)$ with the variance σ^2, $p(\xi) \sim N(0, \sigma^2)$ the ML loss function $\rho(x) = x^2 / 2\sigma^2$ yielding the influence function $\Psi(x) = -[\ln p(x)]' = x/\sigma^2$ and $w = 1/\sigma^2$.

Application of the recursive M-robust scheme equations (5.18), (5.19)-(5.21) requires the noise variance σ^2 and the tuning constant k in equation (5.18) to be known.

As mentioned before, the parameter k represents a constant which should be selected with the aim of reaching the desired efficiency of the algorithm in the case of the nominal normal noise model in equation (5.17). In addition, since the value of noise variance σ^2 is generally unknown, it is essential to somehow estimate it.

There are many methods to estimate noise variance and find influence function. A popular, simple and frequently method used form of the robust variance estimate is the median of the absolute median deviations [57], [58],

$$d = median\{|y - median\{y\}|)/0.6745 \tag{5.22}$$

The divisor 0.6745 is used because then $d \approx \sigma$ if the sample size is large enough and if the sample actually arises from the normal pdf. The particular scheme of selecting d in equation (5.22) suggests appropriate values of the tuning constant k in equation (5.18). Namely, since $d \approx \sigma$, k is usually taken to be a value close to 1.5 and this procedure is known as the 1.5-Huber M-robust estimator. Another method is to estimate the unknown noise parameters in equation (5.17) using the measurement residuals and the QQ-plot method combined with data classification based on robustified residuals winsorization. The derived results are then used to design an adaptive M-robust influence function instead of equation (5.18).

Finally, we suggest the system parameter estimation problem to be solved robustly using the recursive M-robust algorithm equations (5.19)-(5.21). First, we provide a brief review of the QQ-plot method in Section (5.4), and then we apply this method in Section (5.5) to the posed adaptive M-robust parameter estimation problem.

5.2.3 Review of the QQ-plot Technique

Let us consider the case of random samples $\{z_i\}$ from a distribution $F(z)$ having the corresponding probability density function (pdf) $f(z)$. By ranking the samples $\{z_i\}$, $i = 1,...,n$, we obtain the non-decreasing sequence $\{z_i\}$, $i = 1,...,n$ such that $y_i \leq y_j$ for $i < j$. The probability that some observation y will have the rank i in the ordered sequence $\{y_i\}$ follows directly from the Bernoulli experiment [83], [85],

$$P(i/y) = \binom{n-1}{i-1} F^{i-1}(y)(1-F(y))^{n-i} \tag{5.23}$$

In this way, the conditional expectation $m_{i/y}$ and the conditional variance $\sigma_{i/y}^2$ of the random variable i (assuming the observation y), are given by [61].

$$m_{i/y} = E\{i/y\} = 1 + (n-1)F(y) \tag{5.24}$$

$$\sigma_{i/y}^2 = E\{i^2/y\} - E^2(i/y) = (n-1)F(y)(1-F(y)) \tag{5.25}$$

A plot of the ordered data y_i versus the quantity $F^{-1}(\rho_i) = F^{-1}((i-1)/(n-1))$ is called the QQ plot.

Also we can write

$$y_i = F^{-1}(\rho_i) \approx F^{-1}(r_i); \rho_i = (i-1)/(n-1) \approx r_i = (i-0.5)/n; i = 1,2,...,n \tag{5.26}$$

Where $F^{-1}(\cdot)$ is the inverse of the distribution function $F(\cdot)$, It should be noted that the values of rank scores r_i and ρ_i are close to each other. However, it is more convenient to use r_i, since it assigns the finite values $F^{-1}(r_i)$ to the first and the last sample, y_1 and y_n, respectively, in the case of a commonly used Gaussian, Cauchy or Laplace distribution $F(\cdot)$. Thus, if the QQ-plot in Equation (5.26) is fairly linear, then it indicates that the observations have the same distribution function, $F(\cdot)$ even in the tails. Moreover, if the observations y_i are in a strict sense white noise (i.e., they are independent and identically distributed (i.i.d.) with the distribution function, $F(\cdot)$ the relation equation (5.26) can be expressed in linear regression form

$$y_i = m + \sigma F_n^{-1}(r_i) = m + \sigma \tilde{r}_i \qquad (5.27)$$

Here, $F_n(\cdot)$ is the standard normal distribution generating centralized and normalized random variables $\tilde{r}_i = (y_i - m)/\sigma$ with a zero-mean and unit variance, while $m = E\{y\}$ and $\sigma^2 = E\{(y-m)^2\}$.

A typical noise record $\{y_i\}$, corresponding to the standard Gaussian distribution $F_n(\cdot)$, with a zero-mean and unit variance, is displayed in Figure (5.1(a)), while Figure (5.1(b)) shows the QQ-plot of these normal data.

a)   b)
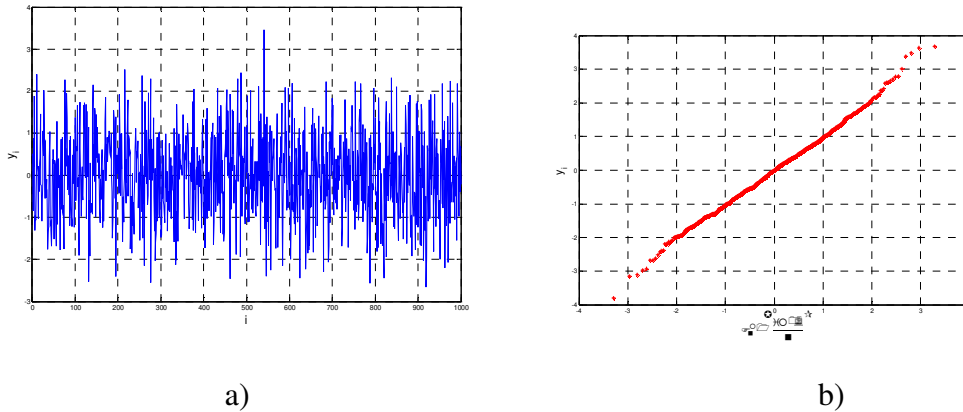
Figure 5.1: QQ-Plot of normal data: a) Normal data; b) Normal QQ-Plot

Starting from equation (5.27) one can estimate the unknown parameters m and σ, using LS algorithm [57-62].

$$\begin{bmatrix} \hat{m} \\ \hat{\sigma} \end{bmatrix} = (\Sigma^T \Sigma)^{-1} \Sigma^T Y \qquad (5.28)$$

Where the $2 \times n$ matrix Σ^T and $n \times 1$ vector Y are defined by

$$\Sigma^T = \begin{bmatrix} 1 & 1 & . & . & . & 1 \\ \tilde{r}_1 & \tilde{r}_2 & \cdot & \cdot & \cdot & \tilde{r}_n \end{bmatrix}; Y^T = [y_1 \, y_2 ... y_n] \qquad (5.29)$$

Furthermore, it can be shown that the mean-value estimator, $\hat{m}$ in equation (5.28), is consistent (i.e., that it is unbiased and improves as the number of observations increases) [72].

Unfortunately, the noise pdf in equation (5.17) deviates from the assumed Gaussian model, being characterized by heavier tails, due to the contamination pdf $h(\cdot)$. As mentioned before, the heavy-tailed aspect of the noise is associated with large glint spikes, called outliers. A typical heavy-tailed noise record is given in Figure 5.2(a), while Figure 5.2(b) shows the normal QQ-plot of this record, based on the standard normal distribution, $F_n(\cdot)$, in Equation (5.24).

a)   b)
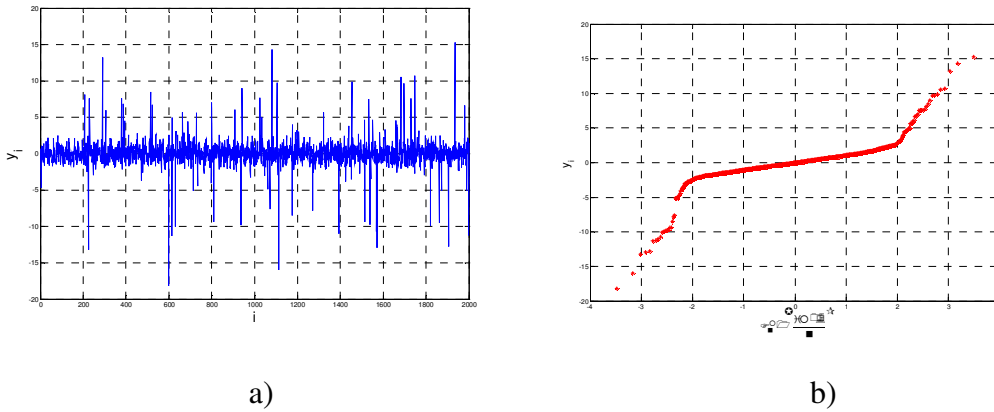
Figure 5.2: Normal QQ-Plot of contaminated normal data: a) contaminated Gaussian data; b) normal QQ-plot.

Figure 5.2 indicates that the data are non-Gaussian, while the normal QQ-plot shape shows that the real data distribution has Gaussian-like middles and heavier tails, as in equation (5.17). The outliers generated by the heavy tails of the underlying distribution have a considerable influence on the conventional linear LS estimates of the form equation (5.28), which are quite non-robust [63]. Thus, there is considerable hope that the performance of the estimates equations (5.28), (5.29) can be improved by robustifying the QQ-plot procedure in some way.

5.2.4 A New Adaptive M-robust Recursive Algorithm

Starting from the M-robust approach in section (5.3), we propose how to apply the classical QQ-plot technique from section (5.4), combined with data clustering based on a robustified winsorization approach, in order to estimate the unknown noise statistical parameters, ε, σ_0^2 and σ^2 in equation (5.17). Then we use these results to design an adaptive suboptimal M-robust influence function, instead of the min-max influence function in equation (5.17). Finally, the system parameter identification problem

equation (5.13) is solved robustly, using the recursive M-robust approach in equations (5.19)-(5.21). In order to clarify the presentation, we divided this section in the following three subsections.

● **Motivations**

As mentioned above, the ε- contaminated distribution family equation (5.17) is extremely important for the field of robust estimation, since it models a number of different applications in which there is the sporadic phenomenon of high-intensity irregular measurements called outliers [68-71]. Moreover, in this situation, it is appropriate for the contaminant $h(\cdot)$ in equation (5.17) to adopt the Gaussian pdf of zero mean and variance σ_0^2, which is considerably higher than the nominal one. Accordingly, the class of measurement noise pdf's in equation (5.17) reduces to

$$p = (1-\varepsilon)N(0,\sigma^2)+\varepsilon N(0,\sigma_0^2) \tag{5.30}$$

Which is described with three unknown parameters: contamination degree or intensity $\varepsilon \in [0,1]$ nominal variance of regular measurement noise, σ^2 and outlier variance, $\sigma_0^2 >> \sigma^2$. The optimal influence function $\Psi(\cdot)=[-\ln(p)]'$ is determined from the real noise pdf in equation (5.30) with known statistical parameters ε, σ^2 and σ_0^2.

Thus noise pdf is susceptible to the statistical parameters as well as the optimal influence function. In order to illustrate the importance of the statistical parameters ε, σ^2 and σ_0^2 in the noise model equation (5.30) to the selection of the influence function, let us look at Figure 5.3, representing the optimal influence function $\Psi(\cdot)=[-\ln(p)]'$ based on the ML approach.
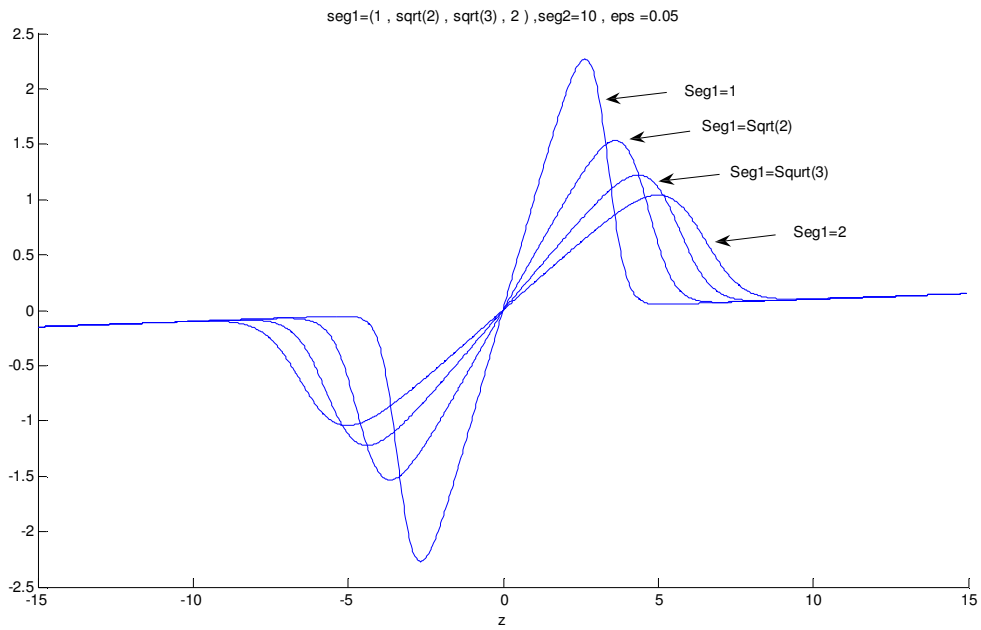
Figure5.3: ML influence functions for different statistical parameter values ε, σ^2 and σ_0^2 in the noise model Equation (5.30): a) for different nominal variance values σ^2,
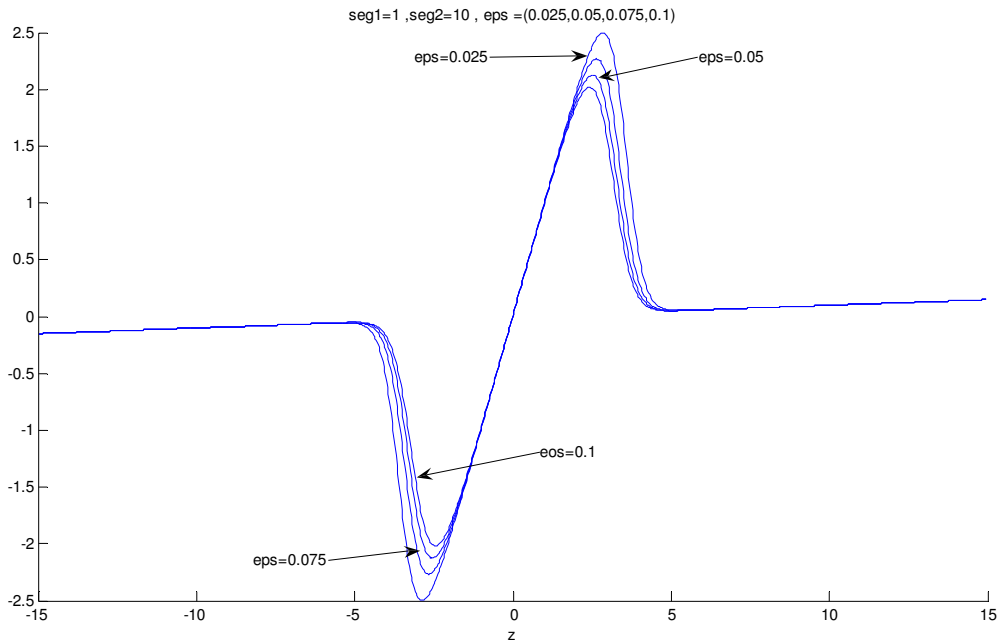
b) For different contamination intensity values ε
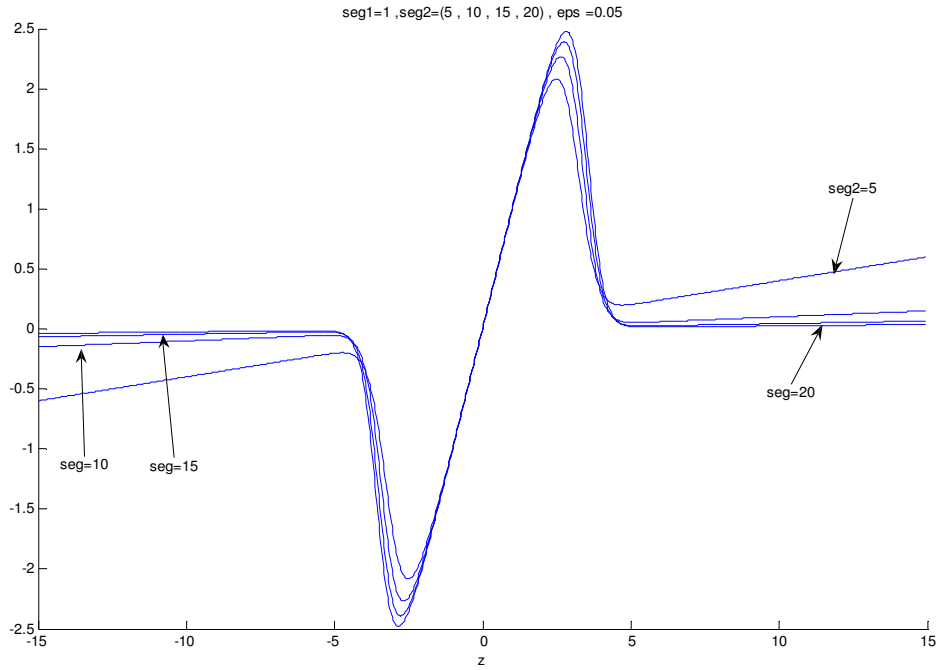
c) For different outlier variance values σ_0^2.

The following conclusions can be drawn on the basis of the obtained results. Each of the given ML functions has a shape with three distinct regions. The first region is the zone of low, positive or negative, values of arguments, in which the influence function is almost ideally linear. The variance of the nominal normal model σ^2 is exclusively responsible for the slope of this linear segment.

The third region is in charge of the presence of outliers in the structure equation (5.30) of measurement noise, and it is also linear, but within the range of intensive residuals of huge, positive or negative, values. Moreover, it is evident that the slope of the linear segment in this region is a consequence of the outlier variance σ_0^2. Finally, there is another region, the transitional zone between the first and the third region, which mostly depends on the intensity of contamination, ε and the variance of the nominal model, σ^2 while it is quite insensitive to the variance of outliers, σ_0^2. This is a domain in which both regular measurements and bad measurements may occur with almost equal probability.

Keeping this in mind, it becomes clear that some estimation of parameters ε, σ^2 and σ_0^2 can be valuable in the right choice of a sub-optimal ML-type influence function, which would ultimately yield a more efficient estimator equations (5.19)-(5.21) of system parameters. A suitable theoretical scope for the estimation of these parameters is the QQ-plot approach, as shown in the previous section.

• **Algorithm Description**

Starting from the previous discussion, a complete adaptive robustified parameter identification procedure consists of the following steps:

1. Data clustering using the QQ-plot together with the robust winsorization approach.

2. LS estimation of unknown statistical parameters of classified regular and irregular observations, generated by the Gaussian mixture pdf in equation (5.30).

3. Estimation of the contamination degree ε in equation (5.30).

4. Calculations of the suboptimal M-robust influence function $\Psi(\cdot)$ in the parameter estimation procedure equations (5.219)-(5.21).

5. System parameter estimation using the recursive M-robust approach.

Each of these steps will be described below in more details in the following subsection.

• **Details of the Algorithm Steps**

As denoted in the previous subsection, the proposed algorithm consists of the following five steps.

Step 1: Data Clustering Using a QQ-plot of Robustly Winsorized Data

The goal of this step is to classify the data within the sliding window (frame) of the n last samples within two classes corresponding to regular data and outliers, respectively.

Let the hypothetic curves $s_j = \alpha_j$ and $s_j = \beta_j$ divide the QQ-plot plain into two data subsets: the residuals s_j, that satisfy the regularity condition,

$$s_j < \alpha_j < \beta_j, j = 1,...,n \tag{5.31}$$

Belong to the class of regular data points, while the residuals, s_j satisfying the irregularity conditions,

$$s_j < \alpha_j \quad or \quad \beta_j < s_j, \, j = 1, 2, ..., n \qquad (5.32)$$

Belong to the class of bad data points or outliers. This situation is depicted in Figure 5.4. However, this procedure corresponds to the nonlinear transformation of ordered residuals

$$\Psi_i(s) = \begin{cases} \alpha_i; & s \le \alpha_i \\ s; & \alpha_i < s < \beta_i \\ \beta_i; & \beta_i \le s \end{cases} \qquad (5.33)$$
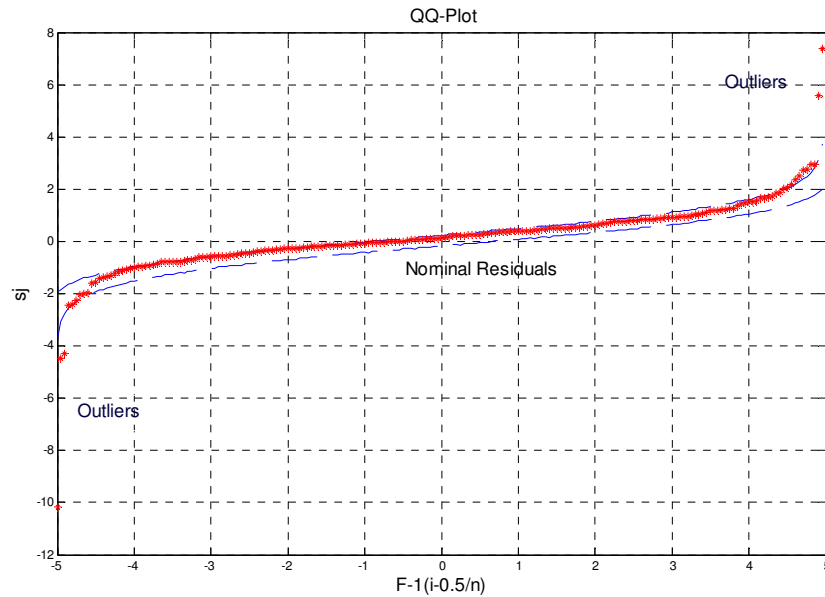
Figure 5.4: Data classification using a normal QQ-plot of the residual sequence generated by the ε-contaminated normal pdf in equation (5.30) with $\varepsilon = 0.05$.

The nonlinear transformation of data based on equation (5.36) with $s_j = -\beta_j$ is known in the literature as winsorization [76]. To choose parameters $\{\alpha_i, \beta_i\}, i = 1, 2, ..., n$, these parameters have to satisfy two requirements

1. To maximize the probability of a correct decision that the regular data will be classified as valid.

2. To minimize the probability of a wrong decision where bad data will be detected as valid.

In addition, the statistical parameters of regular and irregular samples are not known at this time. The final disadvantage can be overdetermined using residual centralization and normalization:

$$s_j \Leftarrow (s_j - E\{s_j\})/std(s_j) \tag{5.34}$$

Where $std(\cdot)$ is the standard deviation and $E\{\cdot\}$ denotes the mathematical expectation. Here α_i and β_i can be selected so that the probability mass in the regular region (α_i, β_i) is equal to a pre-specified limit p, which determines the desired efficiency under the nominal Gaussian model in equation (5.34). In this way, α_i and β_i have to satisfy the relation

$$\int_{\alpha_j}^{\beta_j} f\left(s_j / j\right) ds_j = F(\beta_j / j) - F(\alpha_j / j) = p \tag{5.35}$$

Where $f\left(s_j / j\right)$ is the conditional pdf of the residual s_j given its rank, j, in the ordered centralized and normalized sequence $\{s_j\}, j = 1, 2, ..., n$ equation (5.34).This pdf is expressed by the Bayesian formula [73],[75].

$$f(s_j / j) = P(j / s_j) f_n(s_j) / P(j) = n f_n(s_j) \binom{n-1}{j-1} F_n^{j-1}(s_j) \left[1 - F_n(s_j)\right]^{n-j} \tag{5.36}$$

Where $f_n(\cdot)$ and $F_n(\cdot)$ are zero-mean standard normal pdf and probability distribution functions, with unit variance, respectively, while $P(j) = 1/n$. Figure 5.5 depicts the a posteriori pdf's, $f(s / j), j = 1, ..., n$, in equation (5.38)for the window size $n = 25$, While Figure 5.6 shows the corresponding conditional expectations $m_{s/j}$ and the conditional variances $\sigma_{s/j}^2$ respectively.
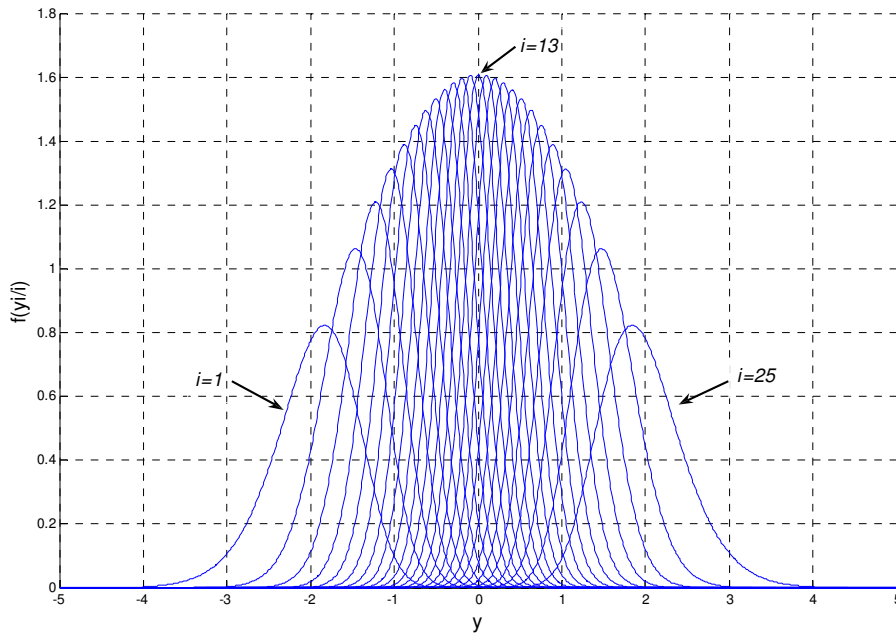
Figure 5.5: Conditional pdf $f(y/i)$ of the random observation i assuming its ith rank

in the ordered sequence $(i = 1, 2, ..., 25)$.

a)   b)
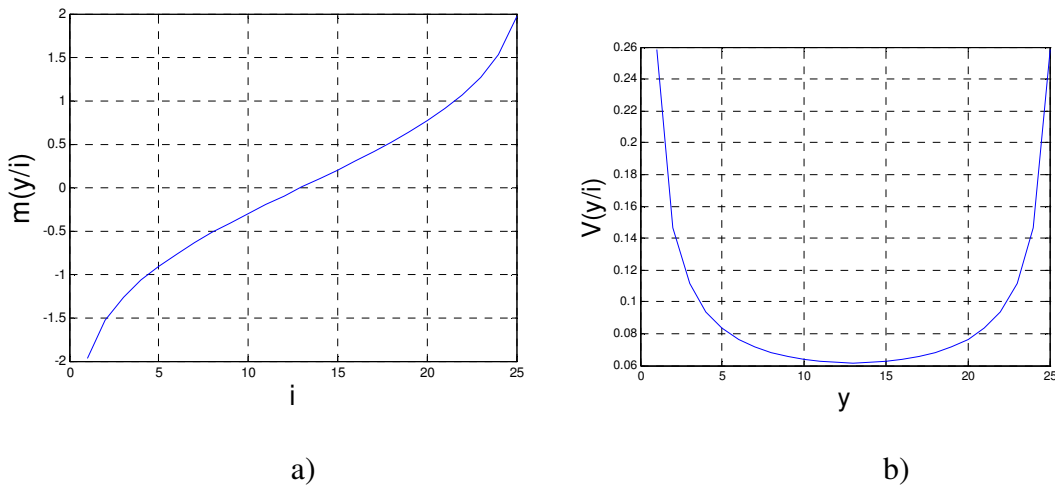
Figure 5.6: Conditional expectations: a) conditional mean; and b) conditional variance.

Obviously, although the *a priori* pdf, $f(s)$ is zero-mean and symmetric, the *a posteriori*

pdf's, $f(s/j)$, have no such properties, except for the rank $j = (n+1)/2$.

The choice of parameter p in equation (5.35) is very important, since too-large p

increases the probability of false alarm (i.e., the probability that the outlier will be

detected as a regular observation), while too-small p increases the probability of false detection (i.e., the probability that a valid observation will be detected as a bad one). Based on equation (5.35), one may conclude that the parameter p controls the trade-off between the degree of robustness and performance degradation under a nominal normal distribution. Regarding the structure of the noise distribution in equation (5.30), the value of p is related to the contamination degree, ε where the smaller the value of ε, the greater value of p, and vice versa. However, ε is not known in advance and one has to adopt p, a priori. A reasonable choice is to select p from the interval [0.9-0.99]. In general, equation (5.35) does not define two parameters uniquely, so we need an additional criterion. Similarly, as in detection theory, the natural requirement is to minimize the probability of false alarm (i.e., the probability of classifying the outliers as valid observations, if it has rank j in the ordered sequence, s_j, $j = 1,...,n$ in equation (5.34) which is given by:

$$P_{err}^j = \int_{\alpha_j}^{\beta_j} f_o\left(s_j \mid j\right) ds_j \tag{5.37}$$

Where $f_o\left(s_j \mid j\right)$ is the conditional pdf of outliers, with that in mind, and since the variance of the outlier is rather large, the integral in equation (5.37) can be approximated by the mean-value theorem from mathematical analysis [88], yielding:

$$P_{err}^j \approx \left(\beta_j - \alpha_j\right) \times const. \tag{5.38}$$

Thus, the condition of minimizing the probability of wrong classification reduces to the minimization of distance $D = D\left(\alpha_j, \beta_j\right) = \left|\beta_j - \alpha_j\right|$. In other words, we suggest that the solution to equation (5.35) be found under the constraint that the distance D is minimal. In this way, the most probable observations (i.e., the observations which are not outliers), will propagate untruncated through the nonlinearity equation (5.33). Thus taking into account equation (5.35) and definition of D, the posed problem reduces to the minimization task:

$$\beta_j = \arg\min_{\beta} J_j\left(\beta\right) ; \quad J_j\left(\beta\right) = \left|\beta - F^{-1}\left(-p + F\left(\beta / j\right) / j\right)\right| \tag{5.39}$$

$$\alpha_j = F^{-1}\left(-p + F\left(\beta_j / j\right) / j\right) ; \quad j = 1, 2, ..., n \tag{5.40}$$

Where $F(\cdot / j)$ and $F^{-1}(\cdot / j)$ are the conditional distribution and its inverse function, respectively. The minimization problem equation (5.39) is nonlinear and iterative methods are required to solve it. The outputs of the classification procedures equations (5.31), (5.32), (5.39) and (5.40) are subsets: the subset of regular observations $s_j^r, j = 1,...,n_r$ and the subset of bad data or outliers $s_j^o, \; j = 1,...n_o$.

Step 2: Estimation of Statistical Parameters of Regular and Bad Data Samples

Since the residual samples are classified into two classes, representing regular and bad data points, one can estimate the statistical parameters, mean value and variance, for each of these classes from observations, using the nonrecursive LS method equation (5.28), equation (5.29). This procedure is presented in Figure 5.7.
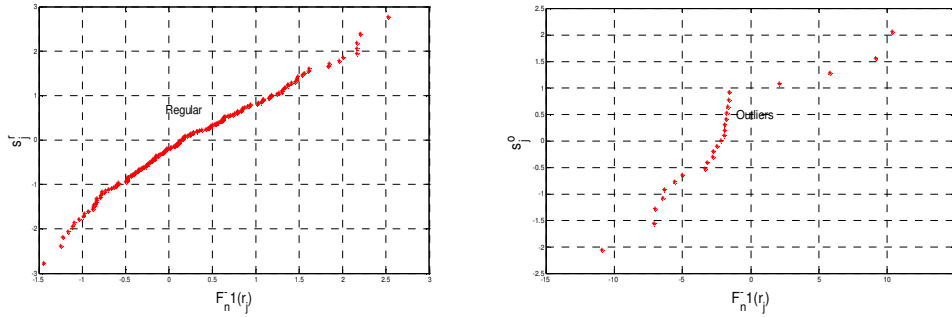
Figure 5.7: Data classification and QQ-plot representation of the classes

Where the resulting QQ-plot corresponding to mixed data in Figures 5.2 and Figures 5.4 is divided into two linear normal QQ-plots, representing the classes of regular and irregular data points respectively, it should be noted that this step needs the original non-normalized data to be used. Thus, by applying the LS algorithm equations (5.28), (5.29) on the observations within the classes, one obtains:

$$\begin{bmatrix} \hat{m}_r \\ \hat{\sigma}_r \end{bmatrix} = \left(\Sigma_r^T \Sigma_r \right)^{-1} \Sigma_r^T Y_r \; ; \Sigma_r^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \tilde{r}_1 & \tilde{r}_2 & \dots & \tilde{r}_{n_r} \end{bmatrix} \; ; Y_r^T = \begin{bmatrix} s_1^r & s_2^r & \dots s_{n_r}^r \end{bmatrix} \tag{5.41}$$

$$\begin{bmatrix} \hat{m}_o \\ \hat{\sigma}_o \end{bmatrix} = \left(\Sigma_o^T \Sigma_o \right)^{-1} \Sigma_o^T Y_o \; ; \; \Sigma_o^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \tilde{r}_1 & \tilde{r}_2 & \dots & \tilde{r}_{n_o} \end{bmatrix} \; ; Y_o^T = \begin{bmatrix} s_1^o & s_2^o & \dots s_{n_r}^o \end{bmatrix} \tag{5.42}$$

Here $\tilde{r}_i = F_n^{-1}(r_i)$, $r_i = (i - 0.5)/n \; i = 1, 2,...,n$ where i is the rank of the observation within the pre-specified data frame of size n, while $F_n(\cdot)$ is the standard normal

distribution. Moreover, s_i^r, $i = 1,...,n_r$ denotes the ordered observations within the first class of regular data, while s_i^o, $i = 1,...,n_o$ represents the ordered data within the second class of bad data or outliers.

Step 3: Estimation of Contamination Degree

The estimation of the contamination degree or intensity ε in equation (5.30) is given by [73]

$$\hat{\varepsilon} = \frac{n_o}{n}; \; n = n_o + n_r \tag{5.43}$$

Where n_r the number of regular observations within the first class and n_o is the number of outliers within the second class. However, the quality of this estimate may be poor, due to the residual samples near the limiting lines α_j and β_j that is, $s_j \approx \alpha_j$ or $s_j \approx \beta_j$, can be classified incorrectly with high probability. Therefore, we suggest the following iterative procedure for estimating the parameter ε, which originated from the ML-based classification [74].

$$\hat{\varepsilon}^{(l)} = \frac{1}{n} \sum_{j=1}^{n} q^{(l)}\left(s_j\right) \tag{5.44}$$

$$q^{(l+1)}\left(s_j\right) = \frac{\hat{\varepsilon}^{(l)} f_o\left(s_j\right)}{\left(1 - \hat{\varepsilon}^{(l)}\right) f_r\left(s_j\right) + \varepsilon^{(l)} f_o\left(s_j\right)} ; \; l = 0,1,2,...; j = 1,2,...,n \tag{5.45}$$

Where $f_r(\cdot)$ is the normal pdf with mean $\hat{m}_r$ and variance $\hat{\sigma}_r^2$, while $f_o(\cdot)$ is also a normal pdf but with the mean value $\hat{m}_o$ and variance $\hat{\sigma}_o^2$. The counter l represents the iteration. The procedure will be repeated until the relative difference between the new and the preceding estimate is relatively small (usually 10^{-3}). The required statistical parameters are generated in Step 2. The estimate equation (5.43) can be used as the initial guess $\hat{\varepsilon}^{(0)}$.

Step 4: Calculation of the Suboptimal M-robust Influence Function

The estimated noise parameters $\hat{\sigma}_o$, $\hat{\sigma}_r$ and $\hat{\varepsilon}$ allow for the estimation of the Gaussian mixture pdf in equation (5.30)

$$\hat{p}(\cdot) = (1 - \hat{\varepsilon}) N\left(\cdot \mid \hat{m}_r, \hat{\sigma}_r^2\right) + \hat{\varepsilon} N\left(\cdot \mid \hat{m}_o, \hat{\sigma}_o^2\right) \tag{5.46}$$

And consequently, the design of the suboptimal ML influence function becomes:

$$\hat{\Psi}(\cdot) = -\left[\ln\left(\hat{p}(\cdot)\right)\right]' \tag{5.47}$$

Step 5: Estimation of System Parameters

The suboptimal M-robust influence function from Step 4 is now applied to estimate system parameters robustly using the recursive M-robust algorithm equations (5.19)-(5.21).The structure of the complete algorithm is given by a flow chart Figure 5.8.
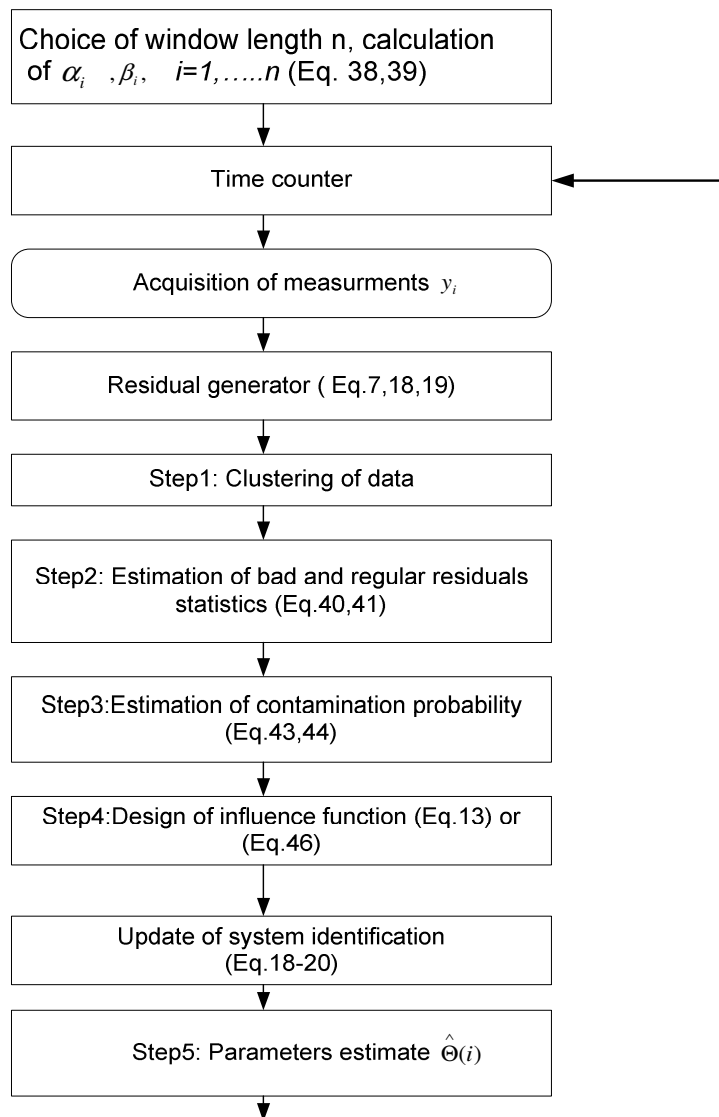

Figure 5.8: Flow chart of adaptive M-robust parameter identification scheme (ARA algorithm)

117

5.3 Description of the Case Study (Steam Separator) and Identification Procedure

5.3.1 Overview

This chapter presents a fault detection and identification approach for steam generators at thermal power plants. The real system in which the measurements were made and the FDI algorithm implemented is located at the TEKO B1 Unit of the Kostolac Thermal Power Plant in Serbia, whose nominal power output is 320 MW.

Given the nature of the process and available data, the implemented FDI algorithm is a trade-off of sorts between the model-based and the data-driven approach. The first step involved the identification of the process. In view of the nature of water level measurements in a separator and the fact that such measurements are conducted under extremely high steam pressures, accompanied by constant unsteady water inflow and steam drain, available water level data are highly unreliable and there is a sporadic presence of high-intensity measurement noise [75]. As a consequence, standard process identification procedures have been shown not to yield satisfactory results and this chapter therefore proposes a robust alternative to parameter estimation.

The next step included statistical testing of the hypotheses, not using the measured data vector but the parameter vector derived from a robust identification procedure. The results demonstrated exceptional detection and isolation efficiency of one of the three possible and most frequent faults which were analyzed.

5.3.2 System Description and Data Acquisition

Thermal power plants are the largest generators of electricity in Serbia, contributing more than 65% to the overall power supply. As such, their operational efficiency and stability needs to be maximized. Special emphasis is placed on reliable long-term operation in terms of negotiated delivery commitments, operation per design criteria for energy efficiency, and longevity of the facility.

It is, therefore, extremely important to monitor vital subsystems and their individual components, such that early detection of any change in characteristics, or faults, will prevent accidents, down time, and substantial financial loss. The chapter addresses steam drums in thermal power plant boilers [76]. A boiler is a unit in which the chemical energy of fossil fuel is converted into heat energy of steam as described in

chapter four. The steam separator is cylindrical, 24m high, with a diameter of 0:9 m. One of the requirements of the process is that the water level in the steam separator be maintained at a height of 8 m.

This requirement is fulfilled by the system operating in a closed loop, where a cascaded PID regulator, based on a fault signal (difference between the required and the measured water level) generates a control signal for feedwater pump discharge. Given the high rate of discharge of the feedwater pumps in the nominal mode (up to 350 kg=s) and the small cross-sectional area of the steam separator, the water level is a highly dynamic physical quantity such that operators do not allow even short-term open loop operation. In other words, all the signals were acquired and the system identified in a closed loop.

Figure 5.9 shows that regardless of control, the water level clearly fluctuates relative to the required reference level. These fluctuations are the consequence of multiple causes. The first cause is the variable heat load of the boiler, as a result of the varying caloric value of the coal. The second reason for the noticeable water level fluctuations relative to the reference level is the pronounced nonlinearity of the dead zone type, including hysteresis in the behavior of the feedwater pumps. Ultimately, large-variance measurement noise is apparent in Figure 5.9.

Noise originates from sudden evaporation and the appearance of large steam bubbles inside the vessel, which rapidly separate of the surface and create a pressure disturbance. As a result, the surface of the water is itself highly dynamic. However, regardless of the fact that these causes reduce separator water level control performance, the water level fluctuations inside the drum are a result of discrete actuator position changes. Thereby, they are the reason for variable water flow from the pumps, ensuring the persistent excitation needed for efficient process identification.
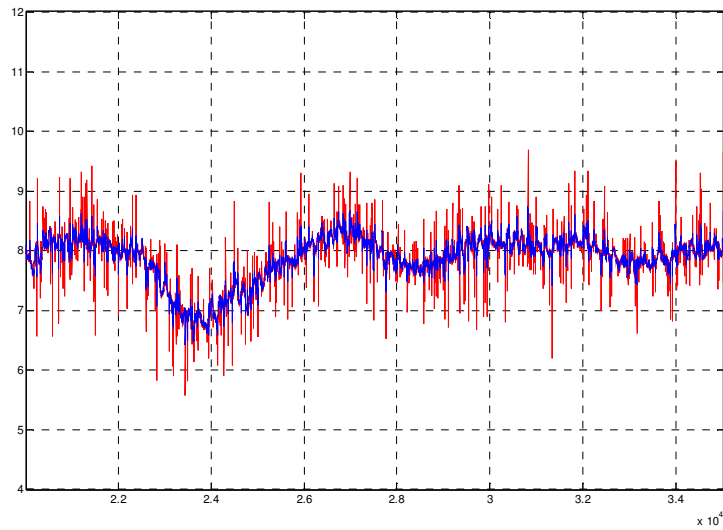
Figure 5.9: Measured (blue) and estimated (red) water levels, fault-free mode over a period of four hours.

In this subsystem, the most important measurements are those of feedwater flow and of steam flow at the steam drum outlet, as well as of water levels. These three physical quantities are measured indirectly, via corresponding differential pressures. Water flow and steam flow are measured by pressure transmitters (SIEMENS SITRANS PDSIII7MF4533) whose measurement range is $0 - 500$mbar, with an accuracy of $0:075\%$ and they are located on the apertures.

The water level in the steam separator is measured by differential pressure gauges of the SIEMENS SITRANS PDSIII7MF4433 type, whose measurement range is $0-1570$mbar and the accuracy is the same as above. Sensor failure due to ageing is rather frequent but this must not affect operational stability. Such situations can be prevented by timely replacement, provided that any change in characteristics is detected on time.

The most frequent faults of this subsystem are erroneous water level or flow measurements where the error is multiplicative (i.e., the measurement is scaled). Such faults are the hardest to detect and are virtually undetectable by simple methods which only check limits. These occurrences are often a result of sensor design, or of physical alterations of the Venturi tube apertures. The faults examined in this research are those of the water level sensor, steam flow sensor and feedwater flow sensor.

These faults were selected based on experience, because they are the most frequent faults in a steam separation system. At the thermal power plant where the tests were conducted, the steam separator water level was determined indirectly, by measuring the pressure at the bottom of the separator, while the fresh steam flow and feedwater flow sensors were based on the principle of determining the flow rate by measuring the pressure difference at the measuring aperture. In view of the conditions in which these sensors operate (e.g. high pressures, sedimentation of substances and erosion of material), they need to be calibrated relatively frequently.

 Periodic calibration is rather complex and extremely costly, particularly of the steam flow sensor, and it requires the production process to be shut down. Therefore, calibration is generally not undertakenas often as necessary. On the other hand, contrary to some other sensors which upon failure generate a frozen value or limit values from the range, these sensor faults are not readily apparent. Namely, these sensors errors of the multiplicative type occur as a result of measurement performance degradation. Such errors are the most difficult to discern because they are not characterized by abrupt changes in the measurement error and that is why the research results reported in this paper are deemed to be of significance for industrial applications.

5.3.3 Process Modeling and Model Validation

The analyzed industrial plant of steam separator is highly complex, the processes which take place are non-linear, the water steam boundary is not clear because the drum contains a mixture of water and steam in a diphasic state, and the steam quality and steam pressure affect the water level.However, given the ultimate goal (an efficient method for fault detection and isolation), we attempted to plausibly describe the process using a simple linear model with two inputs and one output, and to apply the model to detect and isolate sensor faults.

Given that the water level in the steam drum depends on the water flow to the drum and the steam flow from the drum and since an integrating effect is inherent in the process, the following discrete separator model in the form of discrete transfer functions is proposed

$$\Delta Y(z) = \frac{B_1(z)}{A(z)} f_{IN}(z) + \frac{B_2(z)}{A(z)} f_{OUT}(Z) \tag{5.48}$$

In equation (5.8) z is the complex variable of the Z-transformation, ΔY is the steam-drum water level increment, f_{IN} is the water flow to the steam drum, and f_{OUT} is the steam flow at the outlet, while the corresponding model polynomials are defined as

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_n z^{-n}$$
$$B_i(z) = b_{i1} z^{-1} + b_{i2} z^{-2} + ... + b_{in} z^{-n}$$

(5.49)

where n is the model order and $i = 1,2$.

Let us consider a linear, time-varible, discrete-time system, which can be represented by

$$y(i) = -\sum_{k=1}^{n} a_k y(i-k) + \sum_{k=1}^{n} b_{1k} u_1(i-k) + \sum_{k=1}^{n} b_{2k} u_2(i-k) + \xi(i)$$

(5.50)

Where $\{u_1(i), u_2(i)\}$ are the system inputs sequence, $y(i)$ is the system output sequence, $\xi(i)$ is the noise.

Equation (5.50) can be written as equation (5.11).

Where the regression vector is defined as

$$Z^T(i) = [-y(i-1)... - y(i-n)u_1(i-1)...u_1(i-n)u_2(i-1)...u_2(i-n)]$$

and the unknown parameters vector is $\Theta^T = [a_1...a_n b_{11}...b_{1n} b_{21}...b_{2n}]$.

To identify the process, the robust recursive identification method defined by relations equations (5.18)-(22) was used [79], [80]. Robustness is ensured by the introduction of an influence function of the Huber type Equation (5.18), and parameter k affects robust estimation efficiency and its value is generally taken to be $k = 1.42$, resulting in 95% efficiency. The measurement noise variance σ^2 is generally unknown, so that its *ad hoc* robust estimate, based on a series of most recent measurements y_i, is expressed in equation (5.22) [84-86].

Since the measurement noise variance estimation procedure defined by equation (5.22) is implemented periodically in a measurement window of suitable length, such an identification approach is adaptive and can successfully monitor the time-varying properties of the process. Identification takes place in real time, based on data from the closed-loop process. Special attention was devoted to the selection of the model order n

in equation (5.49) .The Akaike criterion [16], [21] was used to select the appropriate model order n:

$$AIC = \log\left(\frac{1}{N}\sum_{k=1}^{N} e^2(k)(1+\frac{2d}{N})\right) \tag{5.51}$$

Where N is the number of samples, and $d = 3n$ is the number of model parameters.

The measurement data used for this analysis were collected over a period of 24 hours during which there were no faults, and the entire identification procedure was repeated for different values of the model order $n = 1,...,9$. Figure 5.10 shows how the AIC criterion function varied for each of the chosen model orders n [87].

From figure 5.10, the minimal model order which adequately describes the system is $n = 3$. A higher model order n would result in higher model accuracy, but the number of model parameters d would increase.
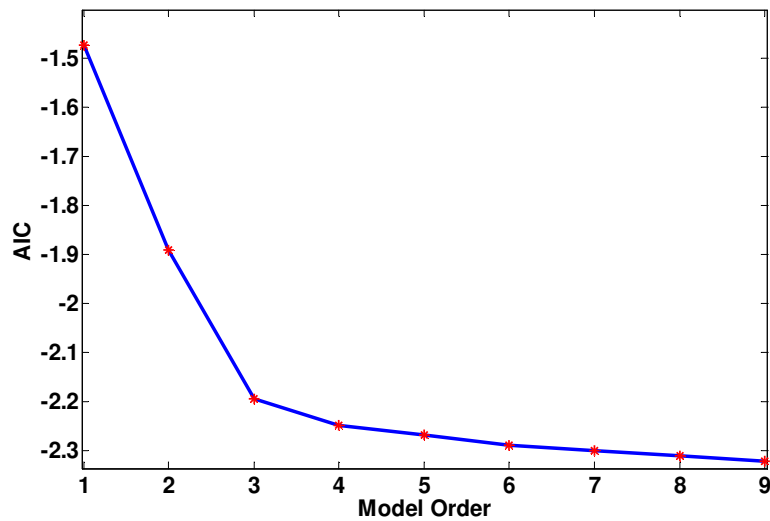
Figure5.10: Akaike criterion vs. model order n.

After adopting the adequate model order, the proposed robust procedure was applied and Figure5.9 shows estimated and measured water levels in a fault-free scenario, over a period of 4 hours. The signal sampling period is 1 s. The estimated steam-drum water level closely follows the measured water level and thus validates the proposed procedure and the adopted model. It follows that the parameters of the process were successfully identified, such that it was possible to describe the condition of the system

based on parameter movement. Figure5.11 shows a comparative analysis of the proposed adaptive robust algorithm (ARA) and the conventional recursive least-squares algorithm (rLS), using the movement of model (1) parameters over time. The contribution of robust estimation is apparent in the graphics, in that the occurrence of pulse noise does not have a significant disturbing effect like it does in the conventional recursive least-squares method. A common technique was used to validate the model; it is based on analysis of the autocorrelation function of the measurement residuals and the cross-correlation function between the input signals and residuals. Figure5.12 contains the respective plots which show that the normalized plot of the auto-correlation function is very similar to the Dirac function, with extremely low values (less than 0.07) for non-zero arguments. The normalized cross-correlation function (normalized with respect to the corresponding standard deviations) also features extremely low intensities for all time arguments Figure5.13.
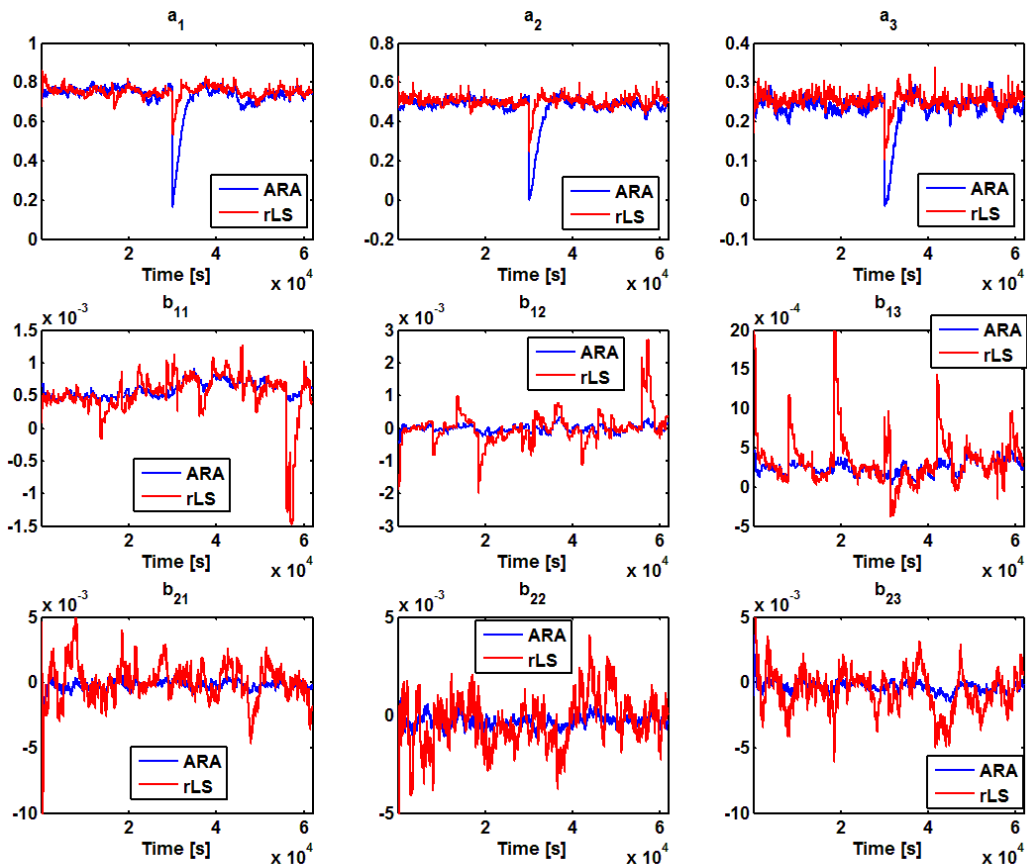
Figure5.11: Movement of estimated parameters in the nominal mode: proposed method (blue) and conventional recursive least-squares method (red).
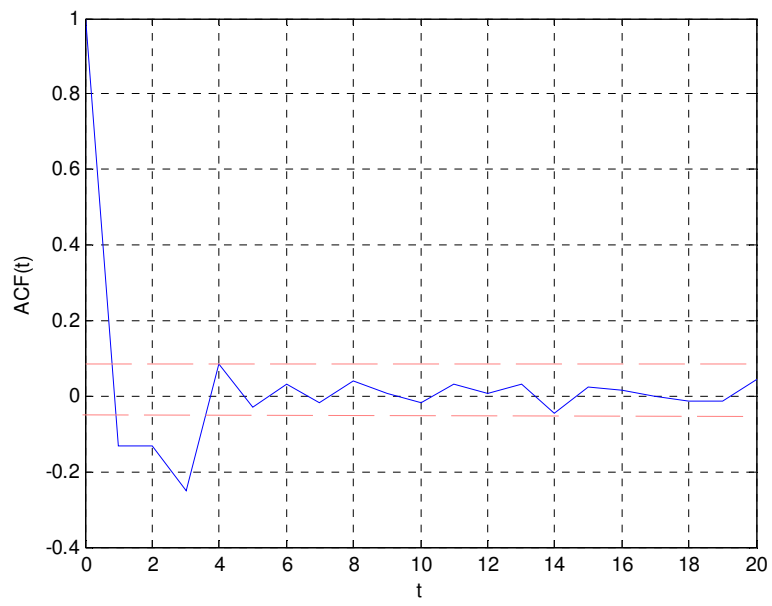
Figure 5.12: Autocorrelation function of the measurement residuals
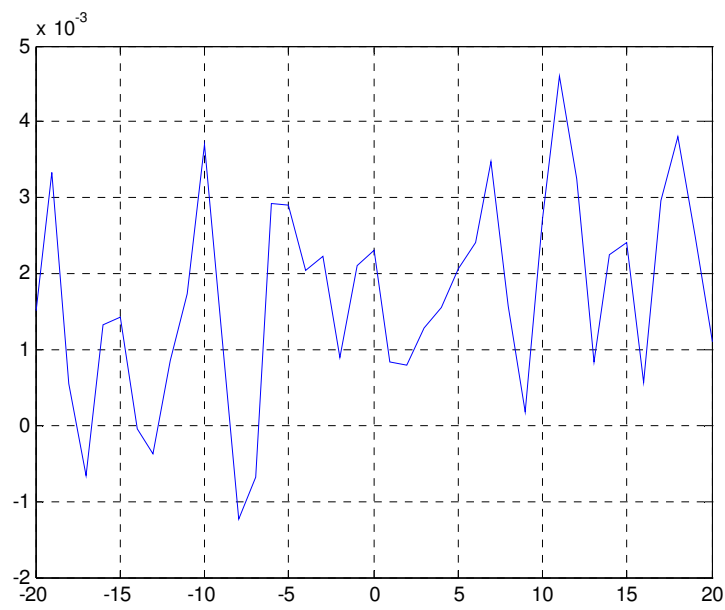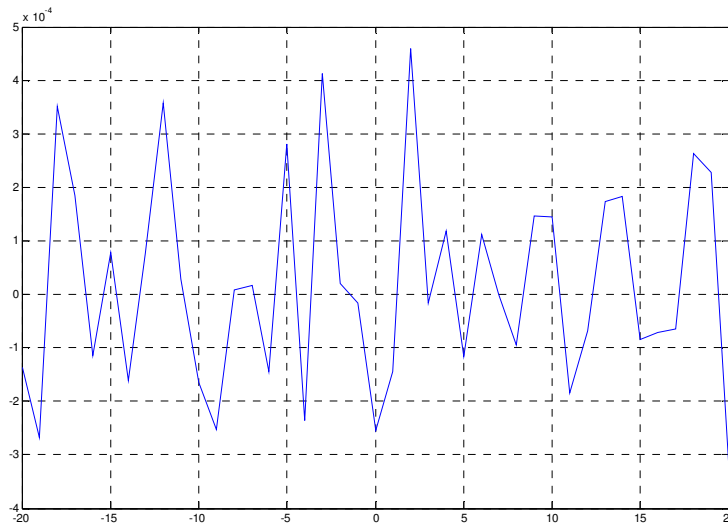
Figure 5.13: The normalized cross-correlation function of the input signals and measurement residuals

Figure 5.14 shows the movement of estimated parameters under the scenario where the steam-drum water level sensor is faulty and there is an error of the multiplicative type (i.e., the measurement shows only a certain percentage of the real value). The fault occurs in the middle of the interval shown. This type of fault is the hardest to detect given that measurements are within the permissible signal range [87].

The control system acts on the basis of a bad measurement and there is a higher probability that the actual water level might drop below the set limit. Based on the movement of the parameters of the identified system, of which there are nine, it is difficult to decide exactly when the fault occurred through analysis of a single parameter. Instead, a dimension reduction method needs to be applied, to ultimately detect the fault. The details of these procedures are discussed in the next section
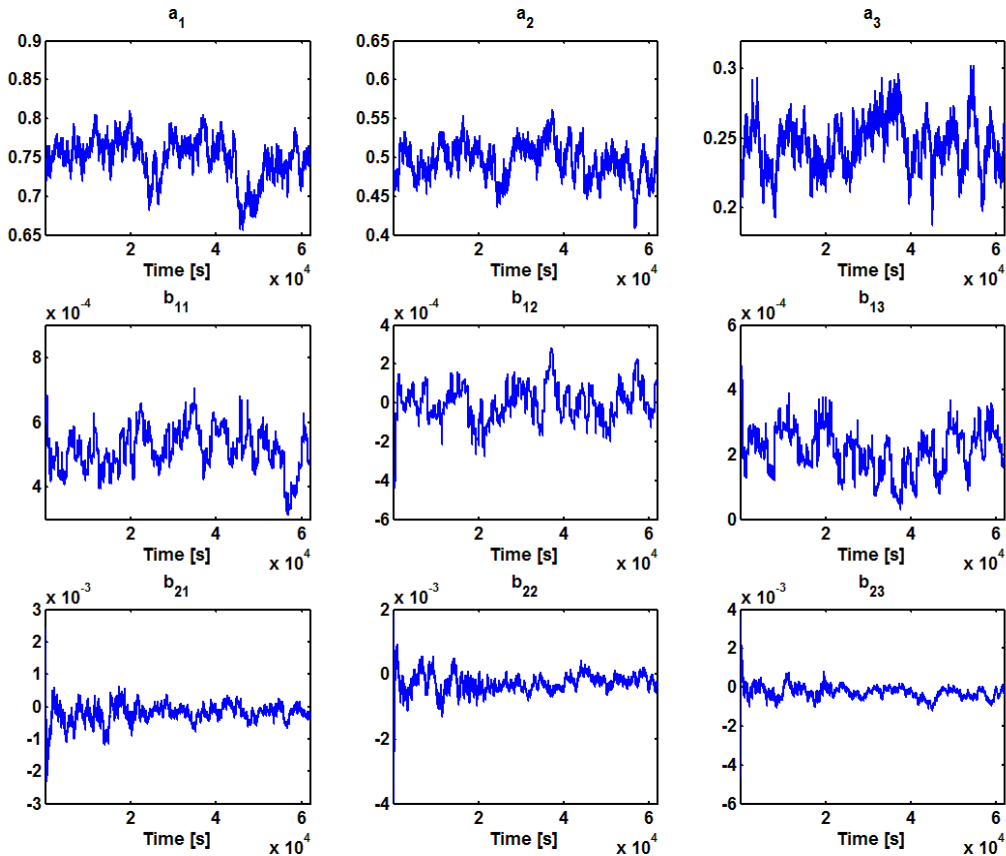
Figure 5.14: Movement of estimated system parameters when feedwater flow measurement fails at t=30000s

5.4 Fault Detection Based On Parametric Process Identification Parameters

The previous section contains a detailed description of the system and its most significant potential faults. First the proposed robust adaptive identification of the system was conducted for each of these faults, as well as for the nominal operating mode of the system, and then, using this identification, system model parameters were estimated for each of the scenarios. Since the adopted system model is of the third order, the number of derived parameters is nine. Given that a data classification method was used for fault detection, special attention needs to be paid to the parameters which are representative in terms of apparent post-fault behavioral changes. It is easy to tell that not all parameters are equally informative, so dimension reduction, from nine to two parameters (Y_1, Y_2) and classifier selection and design are proposed.

5.4.1 Dimension Reduction

In pattern classification, dimension reduction is often unavoidable. Namely, it is not uncommon for a process to be described by a large number of parameters, where not all parameters are of equal informative value, such that it is possible to describe the behavior of the process well enough using a smaller set of parameters. Numerous dimension reduction techniques have been developed, which largely seek out suitable transformation matrices $A_{n \times m}$, where n is the initial vector dimension, and m is the desired dimension $(m < n)$ that will allow for appropriate projection

$$Y_{m \times 1} = A_{n \times m}^T X_{n \times 1} \tag{5.52}$$

Of the initial measurement vectors X (in the present case this is the parameter vector of the identified model) onto reduced-dimension vectors Y, which need not have a physical meaning in the general case. Therefore, the major task now is summarized as follows. Given a number of features, how can one select the most important of them so as to reduce their number and at the same time retain as much as possible of their class discriminatory information? The procedure is known as feature selection or reduction [28]. There are many types dimension reduction methods, The Discrete Karhunen-Loeve Expansion and scattering matrices were selected and described and used the data to choose which the best for the work.

I. The Discrete Karhunen-Loeve Expansion

Let X be an n-dimensional random vector. Then, X can be represented without error by the summation of n linearly independent vectors as

$$X = \sum_{i=1}^{n} y_i \phi_i = \Phi Y \tag{4.53}$$

Where

$$\Phi = [\phi_1 \quad . \quad . \quad . \quad \phi_n] \tag{5.54}$$

And

$$Y = [y_1 \quad . \quad . \quad . \quad y_n] \tag{5.55}$$

The matrix Φ is deterministic and is made up of n linearly independent column vectors. Thus,

$$|\Phi| \neq 0$$

We may assume that the columns of Φ form an orthonormal set, that is,

$$\phi_i^T \phi_j = \begin{cases} 1 & for \quad i = j \\ 0 & for \quad i \neq j \end{cases} \tag{5.56}$$

We may call ϕ_i the ith feature or feature vector, and y_i the ith component of the sample in the feature (or mapped) space.

Suppose that we choose only $m (< n)$ of ϕ_i and that we still want, at least, to approximate X well. We can do this by replacing those components of Y, which we do not calculate, with preselected constants and form the following approximation

$$\hat{X}(m) = \sum_{i=1}^{m} y_i \phi_i + \sum_{i=m+1}^{n} b_i \phi_i \tag{5.57}$$

The resulting representation error is

$$\Delta X(m) = X - \hat{X}(m) = X - \sum_{i=1}^{m} y_i \phi_i - \sum_{i=m+1}^{n} b_i \phi_i = \sum_{i=m+1}^{n} (y_i - b_i) \phi_i \tag{5.58}$$

Note that both X and ΔX are random vectors. We will use the mean-square magnitude of ΔX as a criterion to measure the effectiveness of the subset of m features. We have

$$\varepsilon(m) = E\{\left\| \Delta X(m) \right\|^2\} = E\{\sum_{i=m+1}^{n} \sum_{j=m+1}^{n} (y_i - b_i)(y_j - b_j) \varphi_i^T \phi_j\} = \sum_{i=m+1}^{n} E\{(y_i - b_i)^2\} \tag{5.59}$$

The optimum choice for b_i, is obtained by minimizing $\overset{-2}{\varepsilon}(m)$ with respect to b_i, as follows:

$$\frac{\partial}{\partial b_i} E\{(y_i - b_i)^2\} = -2[E(y_i) - b_i] = 0 \tag{5.60}$$

Solving (5.60) for b_i,

$$b_i = E(y_i) = \phi_i^T E[X] \tag{5.61}$$

Now, the mean-square error can be written as

$$\overline{\varepsilon}^{2}(m) = \sum_{i=m+1}^{n} E\{(y_i - E[y_i])^2\} = \sum_{i=m+1}^{n} \phi_i^T E[(X - E\{X\})(X - E\{X\})^T]\phi_i$$

$$= \sum_{i=m+1}^{n} \phi_i^T \Sigma_X \phi_i \tag{5.62}$$

We shall show that the optimum choice for the $\phi_i S$ is those which satisfy

$$\Sigma_X \phi_i = \lambda_i \phi_i \tag{5.63}$$

That is, the eigenvectors of Σ_X. Thus, inserting equation (5.63) into equation (5.62), and the minimum mean-square error becomes

$$\overline{\varepsilon}^{2}(m)_{opt} = \sum_{i=m+1}^{n} \lambda_i \tag{5.64}$$

The expansion of a random vector in the eigenvectors of the covariance matrix is called the discrete version of the Karhunen-Loeve (K-L) expansion.

In the context of pattern recognition, the coefficients $y_1,..., y_n$ in the expansion are viewed as feature values representing the observed vector X in the feature space. The feature space has several attractive properties which we can list [21].

1. The effectiveness of each feature, in terms of representing X, is determined by its corresponding eigenvalue. If a feature, say ϕ_i, is deleted, the mean-square error increases by λ_i. Therefore, the feature with the smallest eigenvalue should be deleted first, and so on. If the eigenvalues are indexed as $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_n \geq 0$ (Ascending Order), the features should be ordered in the same manner.

The feature values are mutually uncorrelated, that is, the covariance matrix of Y is diagonal. This follows since

$$\Sigma_Y = \Phi^T \Sigma_X \Phi = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ & \lambda_2 & & 0 \\ & & . & \\ & & . & 0 \\ & & & . \\ 0 & . & . & \lambda_n \end{bmatrix} = \Lambda \tag{5.65}$$

2. The set of m eigenvectors of Σ_X. Which correspond to the m largest eigenvalues, minimizes $\overset{-2}{\varepsilon}(m)$ over all choices of m orthonormal basis vectors.

$$\Sigma_Y = \begin{bmatrix} \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & \lambda_m \end{bmatrix}^{m*m} & & 0 \\ & & & \\ 0 & & & \lambda_n \end{bmatrix} \tag{5.66}$$

And transformation matrix is

$$A = [\phi_1 \quad . \quad . \quad . \quad \phi_m]_{n*m} \tag{5.67}$$

Then, reduced-dimension vectors are

$$Y_{m*1} = A^T_{n*m} X_{n*1} \tag{5.68}$$

The fault detection and isolation method presented above was applied in the previously described steam separator system. Relevant signals (water level, water flow and steam flow) were recorded over long time intervals to identify the system in all four operating modes (nominal mode and three sensor fault modes).

For each of these modes, the identification procedure was repeated for N=500 different measurement sequences such that each of the four classes is represented by N=500 nine-dimensional parameter vectors. K-L expansion is used to reduce dimensions, from the initial dimensions n=9 to the reduced dimensions m=2.As a result, the following, figure 5.15 represents two-dimensional (2D) space.
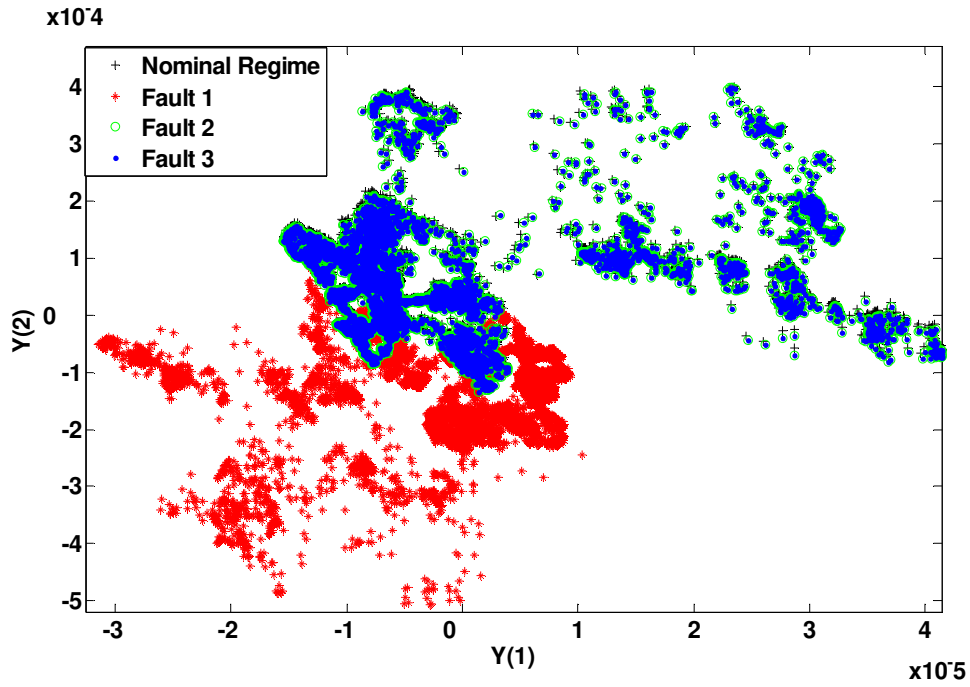
Figure 5.15: dimension reduction based on K-L expansion

The purpose of this result is whether there are separation between nominal mode and three sensor fault mode. The figure shows that are no separation between nominal mode and three sensor fault mode (they are interfere with each other). The conclusion is that, this method is not valid of this work. The next is scattering matrices method.

II. Scattering Matrices and Separability Criteria

We should aim to select features leading to large between-class distance and small within-class variance in the feature vector space. This means that features should take distant values in the different classes and closely located values in the same class.

In discriminant analysis of statistics, within-class, between-class, and mixture scatter matrices are used to formulate criteria of class separability.

A within-class scatter matrix shows the scatter of samples around their respective class expected vectors, and is expressed by

$$S_w = \sum_{i=1}^{L} P_i E\{(X - M_i)(X - M_i)^T | \omega_i\} = \sum_{i=1}^{L} P_i \sum_i \qquad (5.69)$$

Where P_i the a priori probability of class ω_i, that is $P_i \cong {n_i}/{N}$, where n_i is the number of samples in class ω_i, out of a total of N samples [28].

On the other hand, a between-class scatter matrix is the scatter of the expected vectors around the mixture mean as

$$S_b = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T \tag{5.70}$$

Where M_0 represents the expected vector of the mixture distribution and is given by

$$M_0 = E\{X\} = \sum_{i=1}^{L} P_i M_i \tag{5.71}$$

The mixture scatter matrix is the covariance matrix of all samples regardless of their class assignments, and is defined by

$$S_m = E\{(X - M_0)(X - M_0)^T\} = S_w + S_b \tag{5.72}$$

The covariance matrix of X equal

$$\Sigma_X = S_w^{-1} S_m \approx S_w^{-1} S_b \tag{5.73}$$

We apply the same properties as in the Karhunen-Loeve (K-L) expansion to get transformation matrix A and reduced-dimension vectors Y_{m*1} [88].

The following figure shows the current mode represented by a point in two-dimensional (2D) space Y1-Y2 for (nominal mode and three sensor fault modes) by using Scattering Matrices method.
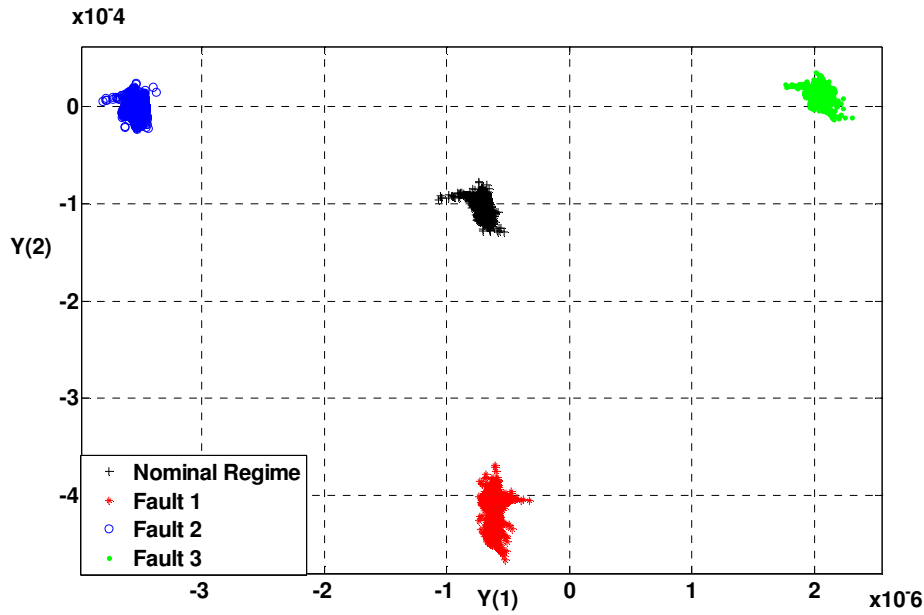
Figure 5.16: Two dimensional representation of the healthy/faulty operating mode

From the figure, we got that, dimension reduction from the initial dimensions n = 9 to the reduced dimensions m = 2 parameters and the separations between nominal mode and three sensor fault modes are obvious. From this result, we conclude that this method is the appropriate method for our work.

The next step of the proposed approach is the selection and design of classifiers, which the vector of reduced dimension Y will assign to one of four classes. Class H_0 will denote the nominal operating mode, H_1 the water-level sensor fault mode, H_2 the water flow meter fault mode, and H_3 the steam flow meter fault mode.

5.4.2 Classifier Selection and Design

Pattern classification (recognition) methods are widely used for fault detection and isolation. The statistical pattern recognition theory offers a wide spectrum of possible solutions which may be grouped into classifiers based on hypotheses testing as described in chapter three. Each of these approaches has its advantages and shortfalls.

Given the nature of the problem addressed in this work where one fault is often more important than another. Namely, it is especially important to detect a fault upon occurrence and not to classify a data sample belonging to a fault mode as a data sample of the nominal mode (the classification of a nominal-mode data as a fault-mode data has

134

less serious consequences). So if we select the first Pattern classification method (the Bayes decision rule for minimum error) in chapter three, here this classifier is designed to minimize the probability of error (the Bayes error $\varepsilon = P_1\varepsilon_1 + P_2\varepsilon_2$). ε_1 The probability of error of the first type, concerns the probability that a sample which belongs to the first class will be assigned to the second class, and vice-versa for the probability of error of the second type ε_2, so this pattern classification method Inappropriate of this situation.

Then, the best Pattern classification method is appropriated of this situation is the Neyman-Pearson test. The Neyman-Pearson test requires that the probability of a false alarm is set at a given value $\varepsilon_2 = \varepsilon_0$, minimizing the likelihood of misclassification as described in chapter three (section 3.2.3). In practical applications, this threshold is selected by generating a graphic of $\varepsilon_0 = \varepsilon_0(\mu)$, from which, given that ε_0 is a monotonously increasing function of parameter μ, a corresponding value of this parameter can easily be estimated for the desired error probability ε_2 [89].

5.5 Statistical Change Detection and Isolation

Scattering matrices method was used to reduce dimensions, from the initial dimensions $n=9$ to the reduced dimensions $m=2$. As a result, the current mode can be represented by a point in two-dimensional (2D) space Y_1-Y_2. Given that the original form of the Neyman-Pearson (NP) classifier was defined for two classes, and since our case involves four operating modes, three independent classifiers were designed to discriminate, in parallel, between the nominal mode and each of the fault modes.

As previously stated, when such a decision-making method is applied, more emphasis is placed on the probability of error of the first type (i.e. the probability of misclassification). Consequently, the probability of error of the second type ε_2 is specified as constant and, under such conditions, the probability of error of the first type is minimized. To design an NP classifier, probability density functions of the derived classes need to be estimated. For each of the classes, Gaussian distribution was assumed for dimension reduction and the mean vector and covariance matrix were estimated accordingly. This yielded a probability density function for each of the classes, keeping in mind the expression for the 2D Gaussian distribution written as

$$p_i(X) = \frac{1}{2\pi |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-M_i)^T \Sigma_i^{-1}(X-M_i)} \qquad , i = 1,...,4 \qquad (5.74)$$

Three Neyman-Pearson classifiers were designed, which distinguish the nominal mode from each of the three fault modes [99]. Thanks to class statistics, in terms of their mutual positions in the Y1-Y2 plane, it is apparent that in this specific case the probability of error of the first and second type is equal to zero (i.e. that the classifiers separate classes with no misclassified data points).

Figure 5.17 shows 2D points for 2000 different modes generated by the dimension reduction procedure (500 for each of the classes), together with the designed classifiers for four different probabilities $\varepsilon_0[\%] \in \{5,10,15,20\}$.
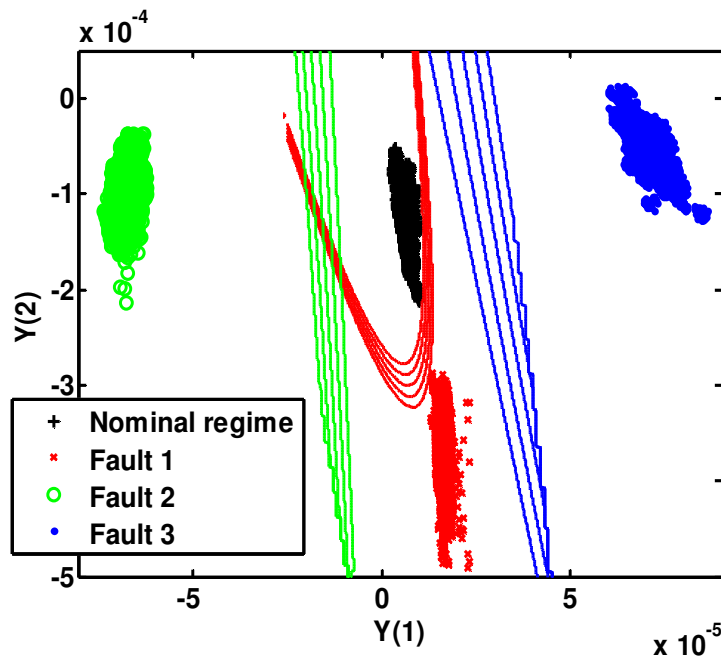
Figure 5.17: Neyman-Pearson Classifier

Figure 5.17 may be misinterpreted as showing that the selection of parameter ε_0 is not of key importance and that ideal classifiers, which classify with a zero error probability, are derived for a wide range of these values. However, analyses of transient states, immediately following a fault, show that a point in 2D space Y_1-Y_2 gradually moves from the nominal mode cluster to a fault mode cluster, and then the selection of the value of ε_0, or the position of the classification line, has a direct impact on classifier

delay in fault detection. To illustrate this phenomenon, Figure 5.18 shows the transition of the reduced vector from the nominal mode class to the steam flow meter fault class. The same figure also depicts the classification lines of the NP classifiers, which indicate that a lower probability of false alarm results in a larger time delay.
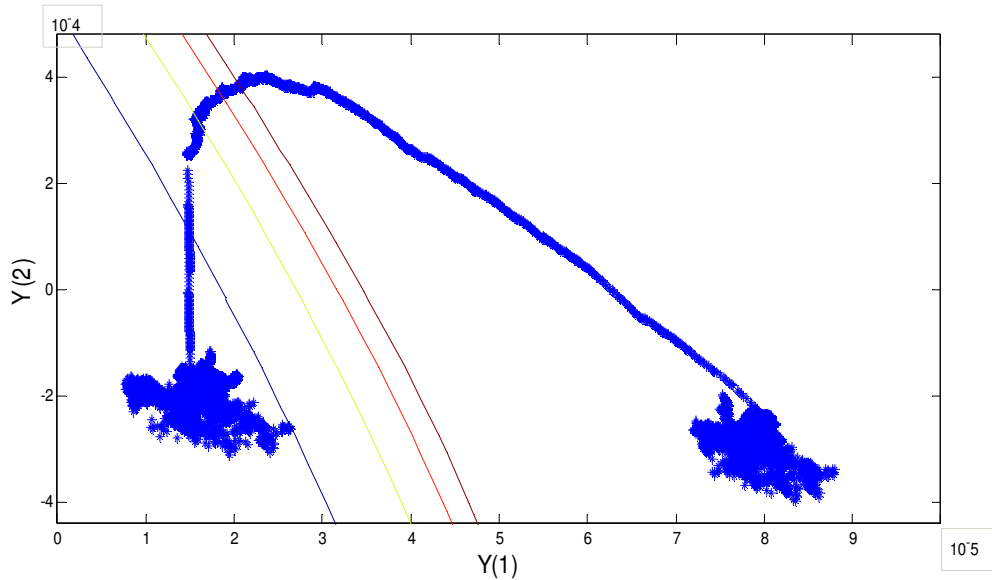
Figure 5.18: Parameter movement when there is a fault of the third type

For classifier design purposes, it is extremely important to interview the operator of the process to find out what an acceptable fault detection time delay would be. Delays which are too long do not leave enough time for the operator to react and prevent malfunction or even considerable damage to the plant. In the considered case, the physical quantities (such as the water level in the steam separator) are highly dynamic and the operator needs to respond quickly.

For example, at maximum feedwater pump loads, the water level could rise from the minimal 0m to the maximal 24m in some 80 seconds. As such, the 15s fault detection delay is acceptable for this type of process. Figure 5.20, which depicts a steam flow sensor fault, shows that this time delay results in the probability of false alarm of 0.94%, based on which the appropriate threshold in test, equation (3.30) can be calculated. The implementation of this approach in a real system showed that the mean time between two false alarms of more than 4.5 hours was considerably longer than the theoretical time.
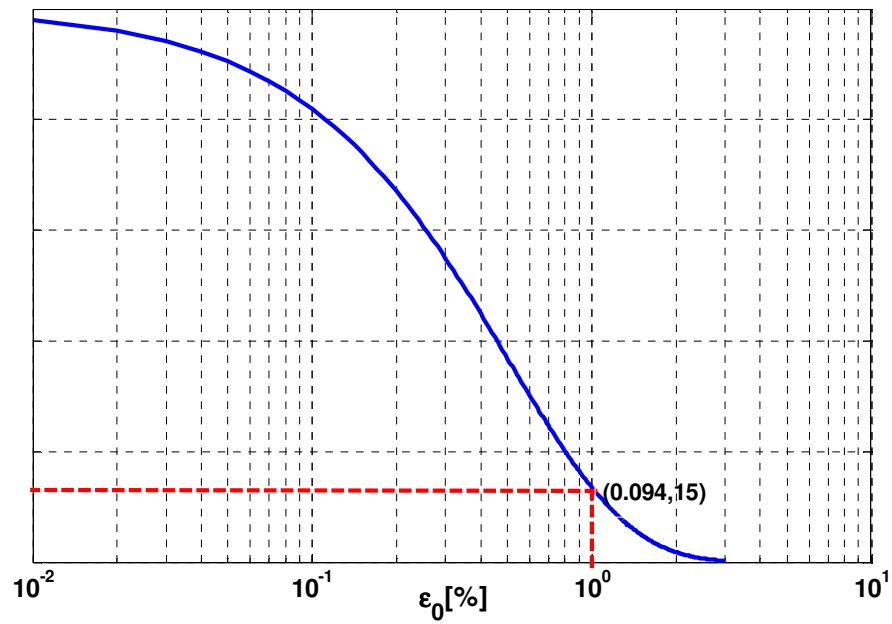
Figure 5.19: Required decision-making time vs. parameter μ

This fault detection and isolation approach, implemented in a real steam separation system at the TEKO B1 Unit of the Kostolac Thermal Power Plant, significantly increased the availability and safety of the entire system and extended the average time of trouble-free operation, given that failure of the water-and-steam circuit is one of the most frequent causes of downtime.

CHAPTER SIX

CONCLUSION

The FDI system is important topic in the modern control system design. During the last three decades, excessive work has been exerted in the field of FDI system. A robust FDI system is necessary to increase the overall system dependability [100]. A more dependable system is the system that has the ability to prevent faults from developing into failures at a subsystem or plant level.Designing Robust Adaptive Parameter Identification of steam separator parameters and fault detection based on these parameters are the main aspects of this work.

The first part of the thesis, Chapter 2 has been devoted to introduce some indispensable concepts inherent the model based FDI. The concepts of structural fault detectability, isolability and identifiability were reviewed to describe the structural property of a system from the FDI point of view. Furthermore, approaches to fault detection and diagnosis for model-free fault detection methods and model-based fault detection methods have been discussed. The main methods for residual generation and fault detection have been described with particular attention to parameter estimation.

Chapter 3 provides an overview of pattern recognition definition. Bayesian classification has been discussed; particular attention has been given to the Neyman - Pearson test that the error of one type is incomparably more serious with significant consequences and therefore its probability is minimize ε_1, while the probability of the error of the second, less important type, is adopted as a constant ε_2.

 In order to be useful for FDI purpose, in this chapter focus on parametric classifier that design classifiers do not based on probability density or probability functions. The chapter gives an explanation of some methods used for statistical density estimation. The common methods are the histogram, the Kernel Density estimate and the k-nearest neighbor-density estimate.

In Chapter 4, Fuel and Furnace Section of thermal power plant were presented, including coal conveyor, stoker and pulverizer. The second section is boiler and steam section, involving classification of boilers, superheater, reheater, economizer, condenser and air path. The next section is steam turbine section, by means of the steam turbine part the thermal energy from steam under pressure is converted into rotary motion, or mechanical work. Also stack gas path and cleanup and cooling towers have been described.

In Chapter 5, some definitions that used in the chapter were described. This chapter presents fault detection and isolation approach specifically designed for steam drums at the TEKO B1 Unit of the Kostolac Thermal power plant in Serbia, whose nominal power output, is 320MW.

Summarizing the conclusions if is possible to list the relevant contributions of the thesis:

1. Robust version of recursive identification was designed, which was proven to be rather insensitive to the presence of outliers.

2. The parameter vectors of the identified model were classified. In this step, the dimension of the initial parameter vector is reduced to 2D space, allowing the operating mode of the plant to be represented as a planar point at any given time. Three possible fault modes are analyzed in addition to the nominal mode, resulting in four separable clusters in 2D space.

3. The selection and design of classifiers was carried out. A suitable selection is the Neyman-Pearson hypotheses test, demonstrating the need for a trade-off between the probability of false alarm and the time delay between fault occurrence and detection.

Design a robust version of recursive identification has been presented in this thesis. However, it would be interesting to carry out the identification of the process using robust parity equation in the future work.

References

[1] Orani, N. Higher-Order Sliding Mode Techniques For Fault Diagnosis, PhD thesis, Dept. of Electrical and Electronic Engineering, University of Cagliari, March, 2010

[2] Abid M. Fault detection in nonlinear systems: An observer-based approach, PhD thesis, University of Duisburg, July, 2010

[3] http://www.websters-online-dictionary.org/definitions/SUPERHEATER

[4] Isermann R. Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance, Springer, 2006

[5] Abdel-Geliel S.M.A.M., Fault Diagnosis and Performance Recovery Based on the Dynamic Safety Margin, PhD thesis, University of Mannheim, 2006

[6] Paraskevopoulos P.N., Modern Control Engineering, NewYork, Marcel Dekker, 2002.

[7] Kilic E., Fault Detection and Diagnosis in Nonlinear dynamical systems, Master thesis, Middle east technical university Sayfa, August, 2005

[8] Rodriguez D. A. L., Time Misalignments in Fault Detection and Diagnosis, PhD thesis, University of Girona, October, 2008

[9] Nygren P., Implementation of Model-Based Dianosis Methods on an Inverted Pendulum, PhD thesis, Linkoping university, 1996

[10] Gertler, I. J., Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, 1998

[11] Pandey A., Fault detection of multivariable system using its directional properties, Master's thesis, Texas A&M University, 2004

[12] Giacomelli M. O., Fault Detection of an Aircraft and Development of a Contingency Control Strategy, Master's thesis, Luleå University of Technology, May, 2008

[13] Bai H., A generic fault detection and diagnosis approach for pneumatic and electric driven railway assets, PhD thesis, University of Birmingham, July, 2010

[14] Abadi I. I. N., Fault Diagnosis in Sampled-data Systems, PhD thesis, University of Alberta, 2006

[15] Simani S., Fantuzzi C. and Patton R.J ., Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques, Spinger, 2002.

[16] Zucchini W., Applied Smoothing Techniques, University of Göttingen, October, 2003

[17] Marques J. P. de Sa., Pattern Recognition Concepts, Methods and Applications, Springer, 2001

[18] Kvam P., Vidakovic B., Nonparametric Statistics with Applications to Science and Engineering, John Wiley & Sons Inc, 2007

[19] Akaike H., A new look at the statistical model identification, IEEE Transactions on Automatic Control 1974; 19(6):716–723.

[20] Andrew W., Statistical Pattern Recognition, John Wiley & Sons Ltd, 2002

[21] Silverman B. W., Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986

[22] Mei Y., Asymptotically Optimal Methods for Sequential Change-Point Detection, PhD thesis, California Institute of Technology, 2003

[23] Sever M., Lajovic J., Rajer B., Robustness of the Fisher's discriminant function to skew-curved normal distribution, Metodološki zvezki 241-242, National and university library of Slovenia, 2005

[24] Jeon B., Landgrebe D. A., Fast Parzen Density Estimation Using Clustering-Based Branch and Bound, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 9, pp 950-954. September, 1994.

[25] Burnham K., Anderson, D., Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer, 2002

[26] Isermann R., Fault-diagnosis systems: An introduction from fault detection to fault tolerance, Springer, 2006

[27] Isermann R., Fault diagnosis of machines via parameter estimation and knowledge processing, Automatica 1993;29(4):815–835.

[28]   Martinez L. W., Martinez R. A., Computational Statistics Handbook with MATLAB, Taylor & Francis, 2001

[29] Huber P. J., Robust Statistics, John Wiley, 1981;

[30] http://www.websters-online-dictionary.org/definitions/SUPERHEATER

[31] Ding S., Model-based Fault Diagnosis Techniques, Springer, 2008[32]

[32] Nady N. M., Thermal Power Plants, John Wiley & Sons Ltd, 2002.

[33] Bhuiyan M. D. S., Comparing and Evaluating Frequency Response characteristics of Conventional Power Plant with Wind Power Plant, Master's thesis, Chalmers University of Technology, Goteborg, Sweden, June, 2008

[34] Annaratone D.,Steam Generators Description and Design, Springer, 2008

[35]   Flynn D., Thermal Power Plant Simulation and Control, The Institution of Electrical Engineering, London, 2003

[36] http://www.tpub.com/fireman/43.htm

[37] Brkić LJ., Živanović T., Tucaković D, Parni kotlovi, Mašinski fakultet, Beograd, 2007

[38] Hailu A., Simulation of Modern Steam Power Plant, PhD thesis, Addis Ababa University Faculty of Technology Mechanical Engineering, June 2005

[39] Energy Efficiency Guide for Industry in Asia – www.energyefficiencyasia.org

[40]   http://www.powerelectricalblog.com/2007/03/thermal-power-plant-layout-and.html

[41]   Raja A. K., Srivastava A. P, Power Plant Engineering, New Age International, 2007

[42] Theodoridis S., Koutroumbas K., Pattern Recognition. Elsevier, USA, 2003

[43] Mićević Z., Kotlovska postrojenja, Energoprojekt Entel d.d., Beograd, 1999

[44] Ganapathy V., Industrial Boilers and Heat Recovery Steam Generators Design, Applications, and Calculations, Marcel Dekker, 2003

[45] http://engineering.myindialist.com/2009/lab-manual-to-study-about-various-types-of-steam-condensers/#.T-nEBZHpXQY

[46] http://en.wikipedia.org/wiki/Thermal_power_station

[47] Keystig E., Advanced Engineering Mathematics, JohnWiley, 2005

[48] Practical Boiler Plant Operation and Management for Engineers and Technicians www.idc-online.com

[49] Helgadótti Ó., Dynamic Behavior of Geothermal Power Plants Located at a Weak Point of a Transmission System, Master's thesis, Polytechnic University, June 2008

[50] http://www.economizersforboilers.com/

[51] Joseph C. W., Introduction to Statistical Methodology, Department of Mathematics, University of Arizona, 2011

[52] Zhu Y., Multivariable System Identification for Process Control. Elsevier Science & Technology Books, 2001

[53] Evans A. A., Maximum Likelihood Estimation, San Francisco State University, 2008

[54] Djurovic Z., Kovacevic B., QQ-plot approach to robust kalman filtering, International Journal of Control 1995; 61(4):837–857

[55] Kvascev G., Djurovic Z. and Kovacevic B., Adaptive recursive m-robust system parameter identification using the qq-plot approach, IET Control Theory Application 2011, 5(4):579– 593.

[56] Chen J., Patton R. J., Robust model-based fault diagnosis for dynamic systems: Kluwer, 1999

[57] P. F., Robust model-based fault detection in dynamic systems, IFAC Symposium on On-Line Fault Detection in the Chemical Process Industries, Newark, 1992

[58] Barnet V., Lewis T., Outliers in Stohastic Data, John Wiley, 1978

[59] Ding S., Model-based Fault Diagnosis Techniques, Springer, 2008

[60] Willsk S., A survey on design methods for failure detection in dynamic systems. Automatica 1976; 12(6):601–611.

[61] Gertler J., Residual generation in model-based fault diagnosis, Controltheory and advanced technology 1993; 9(1):259–285.

[62] http://www.paleniska.pl/portal/pub/view/0x4764b?lang=en

[63] http://www.ehow.com/facts_5836613_definition-turbines.html

[64] Matthews, J, S., ADAPTIVE CONTROL OF MICRO AIR VEHICLES, Master's thesis, Brigham Young University, December, 2006

[65] Zhongshan Wu., Simulation Study and Instability of Adaptive Control, Master's thesis, Louisiana State University, December, 2001

[66] Basseville, M., Nikiforov, I., Detection of abrupt changes: theory and application; vol. 15. Prentice Hall Englewood Cliffs, 1993

[67] Gertler J., Costin M., Fang, X., Hira R., Kowalczuk Z., Luo Q., Model-based on-board fault detection and diagnosis for automotive engines, Control Engineering Practice, 1993;1(1):3–17

[68] Dragan, D., Fault detection of an industrial heat-exchanger: A modelbased approach, Strojniški vestnik-Journal of Mechanical Engineering, 2011; 57(6):477– 484

[69] Verhagen, M. and Verdult, V. ,Filtering and System Identification: A Least squares Approach, (New York: Cambridge University Press, 2007)

[70] Viswanadham N., Srichander R., Fault detection using unknown input observers, Control–Theory and Advanced Technology, 1987;3(2):91–101.

[71] Ljung L., System Identification, Theory for the user, (New Jersey: Englewood Cliffs, Prentice Hall, 1987).

[72] Bosch, P. P. J. van den., and Klauw A.C. van der., Modeling, Identification and Simulation of Dynamical Systems, (Boca Raton: CRC Press, 1994)

[73] Støvring H., Q-Q plots, Aarhus University, Springer, 2010

[74] Willsky S., A survey on design methods for failure detection in dynamic systems, Automatica 12(6):601–611, 1976

[75] Chai D. and Bouzerdoum A. A., Bayesian Approach To Skin Color Classification in Ycber Color Space. Function to Skew-curved Normal Distribution, University of Wollongong, 2005

[76] Chiang L. H., Russell E. L. and Braatz R. D., Fault Detection and Diagnosis in Industrial Systems, Springer, 2001

[77] Pintelon R. and Schoukens J., System Identification: A Frequency Domain Approach, IEEE Press, 2001

[78] Heijde V. F., Duin R. P. W., de Ridder D. and Tax D. M. J., Classification Parameter Estimation and State Estimation, John Wiley, 2004

[79] Ljung L., Soderstrom T., Theory and Practice of Recursive Identification, MIT Press, 1983

[80] Tsypkin Ya. Z., Foundations of informational identification theory, Nauka, 1984

[81] Kinnaert M., Fault diagnosis based on analytical models for linear and nonlinear systems, a tutorial. In: Preprints of the 5th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS2003.:37–50, 2001

[82] Frank P., Ding S., Marcu T., Model-based fault diagnosis in technical processes. Transactions of the Institute of Measurement and Control 2002, 22(1): 57–101

[83] Solo V., The Convergence of AML', IEEE Trans Automat Contr 1979, 24, pp.958-963

[84] Filipović V. Ž., Kovačević B.D., On robust AML Identification Algorithms, 1994 Automatica, 11, (35), pp. 1775-1778

[85] Venables W. N., Ripley B. D., Modern Applied Statistics with S, Springer, 2002

[86] Huber P. J., Robust Statistics, John Wiley, 1981

[87] Bread R., Failure accommodation in linear systems through selfreorganization, Tech. Rep. Report MVT711; Massachusetts Inst. Of Technology, 1971

[88] Fukunaga K., Introduction to Statistical Pattern Recognition, Academic Press, London, 1990

[89] Barnet V, D. and Lewis T., Outliers in Stohastic Data, New York: John Wiley, 1978

[90] Martin R. D., Thomson D. J., Robust resistant spectrum estimation, Proc IEEE, 9, (70), pp. 1097-1114, 1982

[91] Lee C. H., On robust linear prediction of speech, IEEE Trans Acoust Speech Signal Processing, 2, (41), pp. 642-650, 1988

[92] Ding S.X., Model-based Fault Diagnosis Techniques, Springer, 2008

[93] Đurović Ž. and Kovačević B., QQ-plot approach to robust Kalman fitlering, Internationl Journal of Control, 4, (61), pp. 837-857, 1995

[94] Kovacevic B., Djurovic Z., Fundamentals of Stochastic Signals, Systems and Estimation Theory with Worked Examples, Springer, 2008

[95] Dixon W. J. and Tukey J.W., Approximate behavior of the distribution Winsorized/Trimming, Technometrics, 10, pp. 83-98, 1963

[96] Gertler, J., Survey of model-based failure detection and isolation in complex plants, IEEE Control Systems Magazine, 6:3–11, 1988

[97] Frank P., Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy a survey. Automatica, 26(3):459–474, 1990

[98] Keystig E. Advanced Engineering Mathematics, John Wiley, 2006

[99] Papoulis S. and Pillai U. Probability Random Variables and Stochastic Processes, McGraw-Hill, 2002

[100] Jones H., Fault detection in linear systems, Tech. Rep. Report T-608, The Charles Stark Draper Laboratory, Cambridge, 1973

CURRICULUM VITAE

PERSONAL INFORMATION

NAME: Nasar aldian Ambark Mohamed Shashoa   **DATE OF BIRTH:** 01/05/1975

PLACE OF BIRTH: Tamizawah Alshathi, Libya   **SEX:** Male

NATIONALITY: Libyan   **MARITAL STATUS:** Married

E-mail: Naser752006@yahoo.com

HOME ADDERESS: Alngela, Tripoli, Libya   **TELEPHONE:** 0926471075

CHILDREN: Four

SUMMARY OF QUALIFICATION

- Bachelor of Engineering, Electrical and Electronic Engineering department, Alfateh University, Tripoli, Libya
- Master degree in Electro technical sciences in the field of Electro-optical systems, Belgrade, Faculty of Electrical Engineering , University of Belgrade, Serbia

EDUCATION

1990-1992

- Secondary Education- Tamizawah School, Alshathi, Libya

1993-1998

- Primary Degree Alfateh University, Tripoli Libya

  Qualification: Bachelor of Engineering, Electrical and Electronic Engineering

2004-2005

- Master degree in Electro technical sciences in the field of Electro-optical systems, Faculty of Electrical Engineering, University of Belgrade, Serbia

EMPLOYMENT HISTORY

1999-2006

Electrical engineering in Electronic research centre

2006-2008

- Teacher in the field Signals and Systems, Communication Engineering department, Nasser Nation University, Soug Alahad, Libya.

TECHINICAL EXPERTISE

I have worked as a teacher for.

- Fundamental of control system and signals and systems in high electrical institute

, Gasar Ben gasher, Libya.

- Fundamental of electrical engineering in high electrical and mechanical technology institute, Angela, Tripoli Libya
- Electronic II in high technology institute, Soug Algoma, Tripoli Libya
- Fundamental of Communication engineering in high electrical and mechanical institute, Janzour, Tripoli Libya

PUBLICATION

1. Robust adaptive parameter separator identification steam boilers, Proceedings of

   the 55th ETRAN Conference, Banja Vrucica, June 2011.

2. Sensor Fault Detection and Isolation in a Thermal Power Plant Steam Separator,

   Control Engineering Practice, February 2013.

Referees

- Professor. Mousa Mohamed Mousa, Electrical and Electronic Engineering department, Alfath University, Tripoli, Libya

- Dr. Žarko Barbarić, Faculty of Electrical Engineering, University of Belgrade, Serbia

Прилог 1.

Изјава о ауторству

Потписани-а *Nasar Aldian Ambark Shashoa*

број уписа *806/2008*

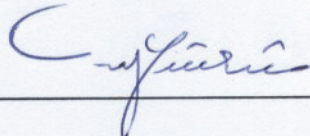<div align="center">Изјављујем</div>

да је докторска дисертација под насловом

Детекција и изолација отказа у сепаратору паре термо-енергетских постројења

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

<div align="right">Потпис докторанда</div>

У Београду, *10. 05. 2013.*

Прилог 2.

Изјава о истоветности штампане и електронске
верзије докторског рада

Име и презиме аутора *Nasar Aldian Ambark Shashoa*

Број уписа *806/2008*

Студијски програм *докторске студије елктротехнике*

Наслов рада *Детекција и изолација отказа у сепаратору паре термо-енергетских постројења*

Ментор *проф. Жељко Ђуровић*

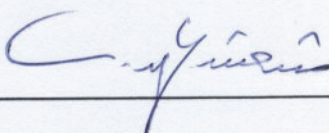Потписани *Nasar Aldian Ambark Shashoa*

изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, _10. 05. 2013._

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић" да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Детекција и изолација отказа у сепаратору паре термо-енергетских постројења

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство – некомерцијално
3. Ауторство – некомерцијално – без прераде
4. Ауторство – некомерцијално – делити под истим условима 5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, _10.05.2013._