Observer-Based Fault Detection for Nuclear Reactors

by

Qing Li

Submitted to the Department of Nuclear Engineering
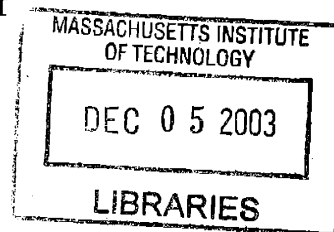in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[September 2001]
August 2001

Author . . . .
Department of Nuclear Engineering
August 8, 2001

Certified by . . . .
John A. Bernard
Director, MIT Nuclear Reactor Laboratory
Thesis Supervisor

Certified by . . . .
Michael W. Golay
Professor, MIT Nuclear Engineering Department
Thesis Supervisor

Accepted by . . . .
Jeffrey A. Coderre
Chairman, Departmental Committee on Graduate Students

Observer-Based Fault Detection for Nuclear Reactors

by

Qing Li

Submitted to the Department of Nuclear Engineering
on August 8, 2001, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This is a study of fault detection for nuclear reactor systems. Basic concepts are derived from fundamental theories on system observers. Different types of fault - actuator fault, sensor fault, and system dynamics fault can be detected and localized by studying the asymptotic response of an error signal constructed from the system inputs, system outputs, and observer outputs.

False alarm and failure to detect a fault are two decision errors when noise is considered. The goal here is to achieve a reasonable compromise. The two types of decision errors can be characterized by their respective first hitting time of a decision threshold. This in turn is dependent on the design of the observer and the decision rule. Costs corresponding to these two types of decision error are defined by cost functions that are in turn constructed based on experience and knowledge of the system operation. A method has been developed in this research to find an optimal design of the observer, the design of a frequency-dependent output filter, and a decision rule that could achieve the desired economic goals.

This technique is applied to nuclear reactor systems and simulations are carried out. The one-group linear nuclear reactor model is used in the observer. The system is modeled by a one-group linear model and by a six-group non-linear model. Results show that this fault detection method can not only detect a fault but also localize it at the same time by constructing specially targeted fault detection filters. These fault detection filters are robust against measurement noise and modeling errors.

Thesis Supervisor: John A. Bernard
Title: Director, MIT Nuclear Reactor Laboratory

Thesis Supervisor: Michael W. Golay
Title: Professor, MIT Nuclear Engineering Department

Acknowledgments

I would like to thank the MIT Nuclear Reactor Lab for providing me with a wonderful environment to conduct my research. I thank all its staff, especially Mr. Ed Block, Eddy Lau, Frank Warmsley, Etienne Fonteneau, Rob Drysdale and Dr. Lin Wen Hu, for their support. I would like to thank Kathy for her assistance during the thesis writeup.

I am grateful to Dr. John A. Bernard, who has been my mentor for five years, for his encouragements, aid, and guidance.

I would like to express my sincere appreciation to Prof. Michael Golay for his support and for being my co-supervisor and thesis reader, whose guidance in the later stage of this report is critical to its success; to Prof. Sidney Yip, Prof. Jacquelyn Yanch and Dr. Richard C. Lanza for serving on my thesis committee.

I thank my mate Ju for his excellent advices and unfailing support especially during the difficult periods. I am grateful to my parents for their unconditional love from the day I was born.

Dedicated to my dear dear daughter, Clara

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The objective of this report is to design and evaluate a model-based fault detection method that could be used to detect incipient system faults under closed-loop situations in nuclear reactors.

This chapter begins with some general information on fault diagnostics and detection. The specific objectives of the research are then enumerated and the relation of the work to the on-going MIT program on reactor control is delineated.

1.1 Guidance on Automated Diagnostics

Two sources of guidance were considered for this research. These were the human approach to process control and the experience of the nuclear industry in devising procedures for responses to alarms.

1.1.1 Human Approach to Process Control

The human approach to process control has been extensively studied. The material presented here is drawn from several studies. The first are works by Sheridan [1] and Kelly [2] which are applicable to all fields of control. The second is work by Bernard which is specific to the nuclear industry [3][4][5].

Figure 1-1 is a block diagram of the human approach to process control[6]. Four

Planning　　　Prediction　Assessment　Implementation

```
┌──────────────┐    ┌──────────────┐   ┌──────────────┐       ╭──────────╮
│Specified State│◄──│  Objectives  │   │Basis of Model│       │ Training on│
│      or       │    └──────────────┘   │ - Theory     │       │ Simulators │
│ Sequence of   │                       │ - Simulation │       ╰──────────╯
│  Substates    │         ╭──────────╮  │ - Drills     │   ╭──────────╮ ╭──────────────╮
└──────────────┘         │Searches for│  │ - Experience │   │  Alarm   │ │Safety Parameter│
                          │ Technical  │  │ - Plant Layout│  │Prioritization││Display System │
                          │Requirements│  └──────────────┘   ╰──────────╯ ╰──────────────╯
                          ╰──────────╯
┌──────────────┐                        ┌──────────────┐    ┌──────────────────┐
│Constraints   │                        │ Adaptation   │◄───│ - Equipment Failures│
│ - Procedures │                        │    of        │    │ - Trends          │
│ - Engineering Limits│                 │   Model      │ Outer└──────────────────┘
│ - Regulations│                        └──────────────┘ Loop    ╭──────────────╮
└──────────────┘        ╭──────────╮                             │Signal Validation│
                        │ Procedure │                            │ Fault Detection │
                        │ Prompting │    ┌──────────────┐         ╰──────────────╯
┌──────────────┐        ╰──────────╯    │ Operator's   │    ┌──────────────┐ ┌──────────┐
│Identification │                        │   Model      │    │   Sensors    │ │Controlled │
│and Evaluation │                        │    of        │◄──│(Current Status)│◄│ Output   │
│ of Options    │                        │  System      │    └──────────────┘ │Variables │
└──────────────┘                        │  Dynamics    │                     └──────────┘
                                         └──────────────┘       Inner
                                                                Loop
┌──────────────┐        ┌──────────────┐   ┌──────────────┐  ┌──────────┐
│Determination  │        │Prediction of │   │  Observed    │  │ Process  │
│ of Desired    │        │ Subsequent   │   │  System      │  │ System   │
│  System       │        │  System      │   │  Behavior    │  │          │
│  Response     │        │  Behavior    │   └──────────────┘  └──────────┘
└──────────────┘        └──────────────┘
                  ⊕      ┌──────────┐ ┌──────────────┐
                         │   Time   │ │Error Observed in│
                         │Difference│ │  Prediction    │
                         └──────────┘ └──────────────┘
      ┌──────────┐                                    ┌──────────┐ ┌──────────┐
      │Significant│                                   │ Control  │ │ Actuator │
      │ Deviation │──────────────────────────────────►│ Action  │ │          │
      │ Predicted │                                   └──────────┘ └──────────┘
      └──────────┘
```

Figure 1-1: Block diagram of the human approach to process control.

subtasks are involved. These are planning, prediction, implementation, and assessment. The planning phase involves determination of the desired system response. This is accomplished by first noting the operational objectives that have been promulgated by managerial personnel. The operator then determines the desired plant state and the most efficient means for achieving that state given the confines of approved procedures, engineering limitations, and regulatory requirements. If a complete change in plant status (i.e., shutdown to operating) is scheduled, then the result of the planning process may be a sequence of desired operational sub-states, each corresponding to a specific phase in the startup.

The predictive aspect of the human approach to control involves the estimation of expectation values. An operator is capable of controlling a complex plant in proportion to his (or her) ability to anticipate that plant's response given the application of any of the available control options. The accuracy and range (in time) of these predictions are in turn dependent on the validity of the operator's mental model of the plant's dynamics. These models are a composite of the operator's training including theoretical understanding, simulated drills, and actual experience. Using such a model and knowing both the plant's current status and any information on trends, the operator forms an estimate of the future behavior of the system. This predicted response is then compared to the desired one. If a significant deviation is anticipated, the operator will initiate corrective control action.

Implementation of the specified control action often requires the simultaneous application of signals to several plant subsystems. While this process may be entirely automated, the extent to which this has been done varies widely depending on the industry in question.

The fourth phase of the control process, assessment, is characterized by inner and outer feedback loops. Note that each of these loops is dependent on information derived from the same sensors. The inner loop is the simplest. Feedback from the controlled variable is used to ascertain that the required control action is being implemented. Loops of this type are often fully automated. The outer loop fulfills two functions. First, it provides information for use with the model. Data on plant

15

trends and on the availability of equipment are preferred to mere status information because the former facilitate prediction. Second, this outer loop permits operators to evaluate their understanding of the plant. Specifically, if their estimates of the controlled parameters are in error, then their model may require revision.

Adaptation of these mental models is often arduous because the only means available to the operator for detecting an imperfection in his (or her) understanding of the plant's dynamics is to compare current observations with previously made predictions. The difference in the time scale may create a problem. A further complication is that discrepancies between observation and prediction may reflect valid changes in the plant. For example, suppose the indication of condenser vacuum is lower than normal. Is this occurring because the plant is running at full power on an unusually hot day? Or is there a small air leak in one of the condensate return lines? If the former is true, then the operator's model requires revision. If the latter is the proper explanation, then corrective action should be initiated. Unfortunately, the operator has no means of selecting between these alternatives. Further observation, for which there may not be time if an actual abnormality is developing, is the likely course of action.

1.1.2   Experience of Nuclear Industry

The nuclear industry has used both event and symptom-based reasoning in an effort to facilitate the response of licensed personnel to abnormal situations. The event-based method was the one adopted during the early years of the industry and it remained in use until the Three Mile Island incident in 1979. Under the event-based method, all possible causalities are enumerated and a unique alarm or sequence of alarms is identified as being indicative of each casualty. The plant operators are then required to memorize the identifying alarm sequences or to deduce the plant state from the indicated alarms. Problems associated with this approach include:

- There is no guarantee that every conceivable casualty has been identified.

- It is difficult to instrument a plant so that there is a one-to-one correspondence

between every given casualty and a single sensing device. Many casualties can only be distinguished from one another by alarm sequences. Thus, temporal information becomes important.

- The burden of memorization on the licensed personnel is excessive.

- The event-based approach with its emphasis on memorization utilizes a person's rule-based skills instead of his or her pattern recognition skills.

The Three Mile Island accident made obvious the deficiencies of event-based analysis. In its place, symptom-based procedures were developed. The intent is to recognize and correct the symptoms of a plant casualty even though the exact cause may not have been determined. For example, a symptom might be inadequate heat removal as evidenced by an increasing coolant temperature. The cause could be loss of secondary coolant pumps, or overpower transient, a problem with the cooling towers, etc. The symptom-based approach reduces the need for rote memorization. However, it increases the complexity of procedures because there will now be multiple entry points to emergency plans. One result has been an effort to develop computer-assisted procedures. Another was the effort in the late 1980s to develop expert systems that would offer advice to plant operators [5].

1.1.3 Relevance of Guidance

Automated diagnostics is a form of artificial intelligence. The most fundamental precept of that field is that "information is power." The capability to diagnose an event as to its cause is a function of the available information. As noted above, it is not practical and probably not possible to instrument a plant so that every conceivable casualty can be uniquely identified at the moment of its initiation. Accordingly, one lesson to be gathered from the guidance is that an automated diagnostics system should, at least initially, focus on the accurate identification of the symptoms as opposed to determination of the actual initiating event. That approach is adopted here. The initial objective is detection of the abnormality, then its confinement to a subsystem, and finally its assignment to a specific component.

A second lesson to be recognized in the guidance is the need for expectation values. Humans achieve control by predicting the outcome and initiating control action only if the predicted outcome is not expected to be as desired. The analog of the mental model that humans utilize to form there predictions are computer-based numerical models of plant process systems. These numerical models can be used to generate the needed expectation values.

A third relevant factor is that humans do preplan their actions, and hence have some idea of the demanded sequence of plant states. In the event that an outcome is not as expected, this information may be of use in the subsequent analysis.

1.2 Scientific and Engineering Objectives

This report describes the development and demonstration of a methodology for the automated diagnostics of nuclear reactors under conditions of closed-loop reactivity control.

The focus is on the capability to identify failures in either the reactor itself or in the actuator and to distinguish between those two types of failures. The motivation for this work was a series of experiments performed in 1988-1991 by Bernard and Wyant in which the importance of automated reasoning to the autonomous control of spacecraft nuclear reactors was demonstrated [3]. Experiments were crafted to illustrate the need for automated diagnostics in each of the four stages (planning, prediction, implementation, and assessment) of the control process. For example, among the faults illustrated were ones involving failure to achieve a demanded rate of power rise because the actuator was in the wrong initial position and hence its available rate of change of reactivity was too low (improper planning), inability to halt a transient at the specified power level because the actuator had been withdrawn too far and hence its available rate of change of reactivity was again too low (incorrect planning), and difficulties in the attainment of the correct power level because of a saturated sensor (invalid assessment). Automated diagnosis can resolve these types of problems by both identifying the fault and by providing suggested corrective actions.

The work reported here was performed jointly by the MIT Nuclear Reactor Laboratory (NRL) and the MIT Department of Nuclear Engineering (NED).

The control research reported here is one component of an on-going MIT program to develop and experimentally evaluate an integrated, fault-tolerant methodology for the closed-loop, digital control of power and temperature in nuclear reactors. In particular, the overall controller should provide signal validation, instrument fault detection, a supervisory algorithm that precludes challenges to the safety system, a number of selectable control laws, a means of reconfiguring those control laws depending on the mission, a method for verifying signal implementation, and some means for on-line performance assessment such as automated reasoning. In addition, there should of course be an independent safety system. Details of this approach, as applied to nuclear reactors, are given in five earlier reports. The first was prepared by MIT for the National Science Foundation under grant CPE-831787 [4]. It describes the formulation and experimental evaluation of the "reactivity constraint approach" which is a means of precluding challenges to a reactor's safety system as the result of any action initiated by an automated controller. The second report was prepared by MIT for Sandia National Laboratories (SNL) as part of the U.S. Department of Energy's Multi-Megawatt Reactor Program[7]. It describes the derivation and experimental assessment of the "MIT-SNL Period-Generated Minimum Time Control Laws" which are a trajectory tracking technique suitable for the rapid maneuvering of reactor power. The third report, which was also prepared for Sandia National Laboratories, concerns applications of the MIT-SNL laws including their use for automated power increases from sub-critical [8]. The fourth report, which was prepared for the U.S. Department of Energy, concerns the extension of the reactivity constraint approach to the closed-loop digital control of reactors characterized by spatial dynamics [9]. The fifth report, which was also prepared for the U.S. Department of Energy, addressed the closed-loop digital control of mutli-modular reactors [10]. In addition to these five reports on reactor control, MIT has also conducted an extensive study of the use of expert systems within the nuclear industry [5]. That material bears on the issue of autonomous control.

1.2.1 Automated Diagnosis

Modern industrial processes are becoming more and more complex. For such systems to function properly, i.e. to implement their desired mission, the integrity of the plant has to be maintained. In the case of a component or system fault, not only may the system be damaged but also catastrophic events such as an explosion, fire, or radiation release may occur. Thus, the detection of any fault and its repair before accidents occur is of great importance. Because of the complexity of modern processes and the limited response time available, even an experienced operator has difficulty in accomplishing the task of identifying the faults and implementing appropriate corrective responses. Automated diagnostics is aimed at aiding the operator in such tasks by using advanced techniques with the aid of modern computers.

Fault diagnosis includes fault detection, fault localization, and fault isolation. Fault detection refers to the identification of any abnormality in the system. Fault localization refers to the confinement of the fault to a subsystem for possible cause. Fault isolation refers to the identification of a specific component or subsystem that is directly responsible for the fault. For a very simple system such as one for which every state variable can be measured, the three concepts can be accomplished by one process. However, for most systems, the three are usually implemented separately.

Traditionally, techniques for automated diagnostics are divided into quantitative and qualitative approaches with each of these two categories dealing with a different domain. Quantitative techniques make use of the system dynamics and filter/estimators. They are used more for fault detection. In contrast, qualitative techniques make use of modern reasoning and are used more for fault localization and fault isolation.

Quantitative Methods

There are two main categories of quantitative methods, namely parameter estimation methods and comparative model methods. However, this categorization has not been universally adopted.

1. Direct estimation approaches: These include all kinds of estimation techniques for extracting information using current and recorded data. In general, these methods are combined with statistical tests to perform diagnosis.

2. Comparative models: Because of the difficulty of simultaneously estimating many process parameters by direct estimation, the problem is often solved by comparing the outputs of comparative models. This is also called the residual-generation method. One example is the observer-based fault detection filter that was first developed by Beard at MIT in 1970's [11].

Quantitative Methods require mathematical process models to describe the system dynamics. The effectiveness of the fault detection is directly dependent on the accuracy of the descriptive models. For some systems such a model is available during the design phase. But for some other systems, such a model can not be established easily. For example, some mechanical systems may belong to the first category, for which mathematical models are available handily, while most biological systems belong to the second category, where an accurate mathematical description is hard to establish based on our current knowledge.

For a nuclear reactor system, knowledge about the dynamics that has been developed during the past fifty years makes the use of quantitative methods possible. The work done in this report employs observer-based fault detection techniques together with decision theories and optimization techniques.

Qualitative Methods

Qualitative Methods include both shallow diagnostic reasoning techniques and deep diagnostic reasoning techniques.

1. Shallow diagnostic reasoning techniques: Shallow diagnostic reasoning is used in problems where all the faults are predefined (typically in medical diagnosis). This technique usually consists of a fault dictionary or a diagnosis tree. Both use look-up tables. The disadvantage of this technique is that the diagnosis is confined by the á priori knowledge[12].

2. Deep diagnostic reasoning techniques: Deep diagnostic reasoning is based on functional or structural models of the problem domain[12]. Methods include:

   (a) Constraint-based methods[13]: These methods deal with constraint satisfaction problems. The technique consists of a constraint network that comprises a set of variables, each with some domain and a set of constraints. By simulation, this technique tries to find a subset of variables that violates the constraints. Then comparison of the results with the known facts allows the identification of conflicts that may suggest the fault.

   (b) Causality - based methods: A causality model that is based on the system structural or functional design is used for inference. The causality model could be in the form of signed directed graphs or fault trees.

   (c) Boolean logic inference: This is based on the basic true or false reasoning with "and/or" operations.

   (d) Non-Boolean logic inference (reasoning under uncertainty)

      - Bayesian inference: Probabilistic reasoning based on Bayes's theory.

      - Shafer-Dempster evidence theory: Evidence theory proposed by Shafer in 1976 that is a variation of Bayesian inference.

      - Fuzzy logic: A logic that enables computers to convert human linguistic terms to arithmetical expressions. It was initiated by Lotfi A. Zadeh in 1965. The key element in fuzzy logic is the fuzzy set which is a set of ordered pairs. Each pair contains an element and the degree of membership of that element in the set. With fuzzy logic, notions like "fast" or "very fast" can be formulated mathematically and processed by computers. A summary of the concept as applied to reactor control has been given by Bernard [14].

The cause-consequence trees (CCT) that are the basis for these qualitative methods require a great deal of effort for their construction, e.g., a CCT for a nuclear plant requires approximately ten man-years [15]. In addition, a CCT is very plant specific,

and does not allow feedback models [15]. But for systems where an accurate mathematical model is not available and for which quantitative methods can therefore not be carried out, qualitative methods are the only alternative.

1.2.2 MIT Program on Reactor Control

The MIT program on advanced instrumentation and reactor control originated in the late 1970s under the direction of Professor David D. Lanning of the MIT Nuclear Engineering Department. Studies at that time principally concerned the development of accurate, real-time models of various plant components such as pressurizers, steam generators, and condensers. The existence of accurate models is, of course, fundamental to the construction of a controller. In 1980, Dr. John Deyst, Dr. John H. Hopps, and Dr. Asok Ray, who were all from the Charles Stark Draper Laboratory (CSDL), initiated an experimental program at the 5-MWt MIT Research Reactor (MITR-II) to demonstrate signal validation and instrument fault detection. The "parity space approach," which was subsequently used on both test and commercial reactors, was one result of this effort [16].

Once both validated signals and real-time models were available, it became possible to consider closed-loop digital controllers. A supervisory algorithm that precluded challenges to the reactor's safety system was developed and demonstrated on the MITR-II by Dr. John A. Bernard in February 1983. Designated as the "MIT-CSDL Non-Linear Digital Controller," or NLDC, this methodology uses reactivity constraints to determine if a change should be made to the present control signal in order to avoid a power overshoot at some future time [17]. The "reactivity constraint approach" was licensed by the U.S. Nuclear Regulatory Commission for general use on the MIT Research Reactor in April 1985 and soon became the basis of an experimental protocol for the on-line testing of novel control strategies under conditions of closed-loop digital control. Funding for the MIT control program was provided by the National Science Foundation (Dr. Royal Rostenbach, Division of Energy and Energetics) from 1984 to 1986. Other significant results achieved with NSF support included the design and implementation of a rule-based controller, an on-line demon-

stration of control law reconfiguration, and the derivation by Professor Allan F. Henry of the alternate formulation of the dynamic period equation [14][18][19]. In July 1986, a project was initiated with Sandia National Laboratories (SNL) to develop control strategies for reactor-powered spacecraft. One result of that effort was the derivation by Dr. Bernard of the "MIT-SNL Period-Generated Minimum Time Control Laws" which are closed-form expressions that permit a reactor's neutronic power to be raised by many orders of magnitude both without overshoot and within a few seconds [20]. With the support of Dr. Kwan S. Kwok and Mr. Paul T. Menadier of MIT and Mr. Frank V. Thome and Mr. Francis J. Wyant of SNL, both these laws and the NLDC were subsequently demonstrated experimentally on SNL's Annular Core Research Reactor [21]. In September 1985, a project was begun under the sponsorship of the U.S. Department of Energy to extend the reactivity constraint approach to large, spatially-dependent reactors. That work involved both the development of "faster-than-real time" reactor models that use super-nodal methods to describe the neutron flux behavior and fast running thermal-hydraulic codes to characterize reactivity feedback effects [22]. In April 1988, a program was begun under the sponsorship of Sandia National Laboratories to investigate use for the MIT-SNL laws for the rapid startup and control of reactor-power spacecraft [23]. Also, in 1988, joint research was initiated with the Oak Ridge National Laboratory and the U.S. Department of Energy concerning the closed-loop digital control of power and temperature in multi-modular reactors[24]. A complete, yet concise, summary of the MIT approach toward reactor control was given at the ANS/NIST Conference "Fifty years with Nuclear Fission" in April 1989 [25] and in several other, more recent publications [26][27].

1.3 Impact of Closed-Loop Control on Automated Diagnostics

The objective of the work reported here is to automate the diagnosis of closed-loop controllers. This in turn limits certain aspects of the problem's complexity. Namely,

1. Controller Response: The response of the controller is to adjust an actuator. This adjustment is normally measured accurately, often with redundant sensors. Hence, one can assume the availability of accurate information relative to the actuator.

2. Controller Model: Many controllers are model-based. Moreover, even if a model is not used in the controller itself, one was often developed in order to design the controller. Thus, some sort of mathematical representation of the system's dynamics is usually available.

3. Controller Demands: The purpose of the controller may be to maintain system output at a steady-state value or to cause that output to track a specified trajectory. In either case, a sequence of demanded control signals is known and hence available for use in diagnosing system response.

1.4 Process for Automated Diagnostics

The material summarized in Sections 1.1 and 1.3 of this report suggests the following approach to automated diagnostics:

1. Comparison of the measured system output with that which was demanded.

2. Observation of the system's state variables and comparison of those quantities with their expectation values.

3. Performance of a failure detection analysis to determine if the origin of any detected problems is in the plant dynamics or in the controller.

4. Identification of the cause of the failure. (Not within the scope of this report because this action is largely a question of the adequacy of instrumentation.)

Advantages of this approach are that it is sequential with each step increasing the complexity of the quantitative analysis, that its initial focus is on the identification of symptoms so that corrective action can be quickly implemented, and that it takes

advantage of the various demanded signals that are either known or determinable from the closed-loop controller.

The first three steps in the above sequence are mathematically implementable provided that a model-based controller is being utilized. These steps are the focuses of this report. The fourth step is a matter of the available data. That is, has the plant been sufficiently instrumented as to permit a meaning full causal analysis? That issue is not considered here.

1.5 Structure and Content of the Thesis

There are two main parts of the report. One is the theoretical development of the observer-based fault detection approach and the other is the application of this approach to a nuclear reactor system. The former is addressed in Chapters 2 and 4, the latter in Chapters 3 and 5.

Theories of observability and observer-based fault detection filter are the starting points of this research, and are reviewed in Chapter 2. There are two issues that are not covered by the current theories on observer-based fault detection, they are the influence of noise and how to design the filter. These are the concerns of this research. In Chapter 4, a brief review of optimal decision theory is provided to lay the groundwork for designing a fault detection filter and corresponding decision rule that are robust to noise influences. Then, a method that integrates the design of the fault detection filter, a post-process filter, and the decision rule are proposed. To achieve a desired economic goal, a nonlinear cost function is constructed and optimized for finding the optimal design. These are presented in Chapter 4. The construction of the cost function is the key point that links the optimization tool with our objectives.

To apply this approach to a nuclear reactor, its dynamics and representation models are presented in Chapter 3. Certain empirical models specific to the MIT Research Reactor (MITR-II) are also introduced so that simulation studies can be carried out directly. Application to a one-group linearized nuclear reactor model is

presented in Chapter 5. The optimal design is carried out for this observer and tested on systems using both a one-group linearized model and a six-group nonlinear model. In both situations, the simulation results are satisfactory in the criteria that our goals - prompt fault detection and low false alarm rate - are achieved. The fault detection filter and decision design are robust against noise and modeling errors.

This research proposes an integrated method for designing a fault detection filter gain and for constructing a decision rule to detect and localize a component failure for a linear-time-invariant (LTI) system. This is an extension of the work done by Beard in early 70's. The method proposed in this research addresses the noise effect and proposes a fault detection mechanism that includes a fault detection filter, a post process output filter, and a decision rule. This research has also applied this method to a reactor system using simulation. The work reported here is one component of an on-going MIT program to develop and experimentally evaluate an integrated, fault-tolerant methodology for the closed-loop, digital control of power and temperature in nuclear reactors.

The framework of this research is shown in Figure 1-2. There are two main branches, one the general theory development, the other is the application to nuclear reactor. The circles in the chart are the existing theories that have been applied to this research and the shaded squares are the original contribution of this research.
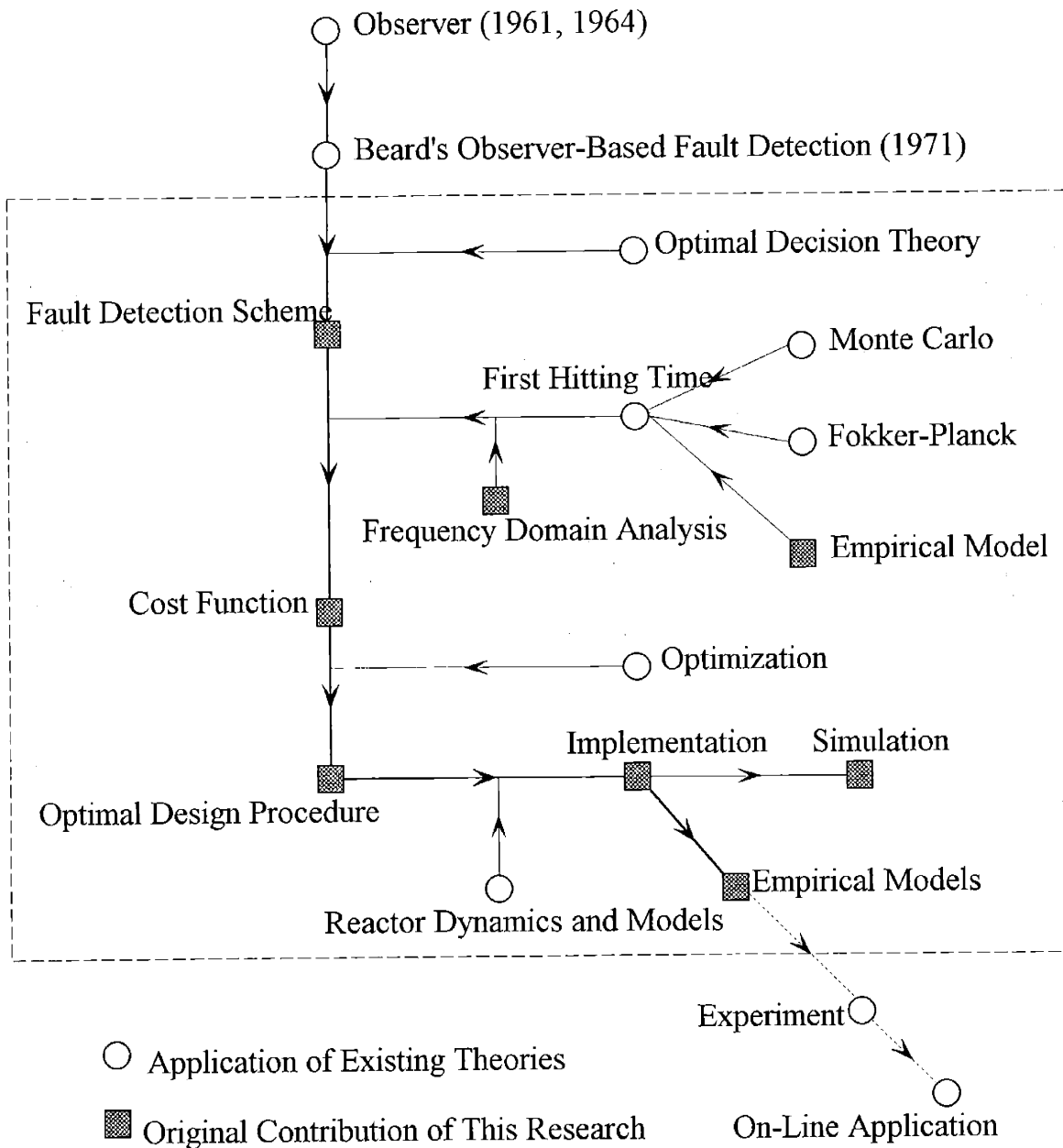
Figure 1-2: The framework of this research.

Chapter 2

Theories on Observability and Observer-Based Fault Detection

System observability and theories of observer-based fault detection are reviewed in this chapter. The objective is to provide background material needed for the development of a fault detection technique that is robust against noise and for application to a nuclear reactor. A brief overview of the origin and broad meaning of the basic concepts is included.

The chapter begins with a brief description of process representation, which is the mathematical language that is used in the rest of this report. It then introduces the concept of a system observer, which is a dynamic system (a nominal system) with state variables that are estimates of the state variables of another system (the target system of interest). For an observer to be stable, i.e. for the observer state variables to converge to those of the nominal system, certain criteria have to be satisfied. Not every system can have a stable observer. Thus, the concept of observability is introduced because it determines whether or not a stable observer can be constructed. The relationship between observability of a system and the construction of a system observer is discussed and relevant definitions and derivations are included. Once an observer is constructed, the estimated state variables can be compared with measurements and the differences between them can then serve as the basis for fault detection. In the last part of the chapter, an explanation is given of how fault analysis can be

carried out based on information from the observer and measurements.

This chapter is focused on the general concepts of observability and observer-based fault detection. The use of these techniques in a real system setting and the accommodation of issues such as noise and modeling errors are not addressed in this chapter but will be discussed in the following chapters.

2.1   Process Representation

A dynamic process can be described mathematically in several ways. For example, there are behavioral descriptions, input-output descriptions, and state-space descriptions. The state-space description is the most commonly used in modern control theory. Most dynamic processes, in this case that of a nuclear reactor, can be described, at least to a reasonable approximation, using standard state variable methods.

For a continuous, linear-time-invariant (LTI) system, the state-space representation is

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) \tag{2.1}$$

$$\dot{\mathbf{y}}(t) = C\mathbf{x}(t), \tag{2.2}$$

and for a discrete-time LTI system, the state-space representation is

$$\begin{aligned}
\mathbf{x}(k+1) &= A\mathbf{x}(k) + B\mathbf{u}(k) \\
\mathbf{y}(k) &= C\mathbf{x}(k).
\end{aligned} \tag{2.3}$$

Here $\mathbf{x}$ is an n-dimensional state vector, $\mathbf{u}$ is an r-dimensional control vector, $\mathbf{y}$ is an m-dimensional output vector, A is an $n \times n$ system matrix, B is an $n \times r$ control matrix, and C is an $m \times n$ output matrix. Equation 2.1 is a set of n first order differential equations that represent the plant while Equation 2.2 is a set of m linear algebraic equations that represent the plant output (i.e. the measurement observables.) Under this representation, the $\mathbf{x}_i$ are the state variables. These constitute the minimum set of information needed to describe a system, and may or may not correspond to

physical real parameters.

The process in question may or may not be controllable or observable. The former implies that it is possible, through application of the control $\mathbf{u}(t)$, to transfer the system from any initial state to any final state in finite time. The latter implies that, for the unforced system, it is possible to determine any initial state by use of a time sequence of the system output.

2.2   Luenberger Observer

An "observer" is defined as a dynamic system with state variables that are estimates of the state variables of another system. The concept was introduced in the early 1960s first by Kalman and Bucy[28], and then by Luenberger[29][30]. The concept is summarized here. Denote an estimated quantity by ˆ. Thus $\hat{\mathbf{x}}$ is the estimate of $\mathbf{x}$. The observer system is:

$$\dot{\hat{\mathbf{x}}} = \hat{A}\hat{\mathbf{x}} + \hat{B}\mathbf{u} + L\mathbf{y}, \tag{2.4}$$

and is thus driven by the control input $\mathbf{u}$ and the measurement vector $\mathbf{y}$. The error in the estimate is

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}. \tag{2.5}$$

Hence,

$$
\begin{aligned}
\dot{\mathbf{e}} &= \dot{\mathbf{x}} - \dot{\hat{\mathbf{x}}} \\
&= A\mathbf{x} + B\mathbf{u} - \hat{A}\hat{\mathbf{x}} - \hat{B}\mathbf{u} - L\mathbf{y} \\
&= A\mathbf{x} + B\mathbf{u} - \hat{A}(\mathbf{x} - \mathbf{e}) - \hat{B}\mathbf{u} - LC\mathbf{x} \\
&= \hat{A}\mathbf{e} + (A - LC - \hat{A})\mathbf{x} + (B - \hat{B})\mathbf{u}. \tag{2.6}
\end{aligned}
$$

The matrices $\hat{A}$, $\hat{B}$ and K are chosen so that the error goes to zero regardless of $\mathbf{x}$ and $\mathbf{u}$. Hence,

$$\hat{A} = A - LC, \tag{2.7}$$
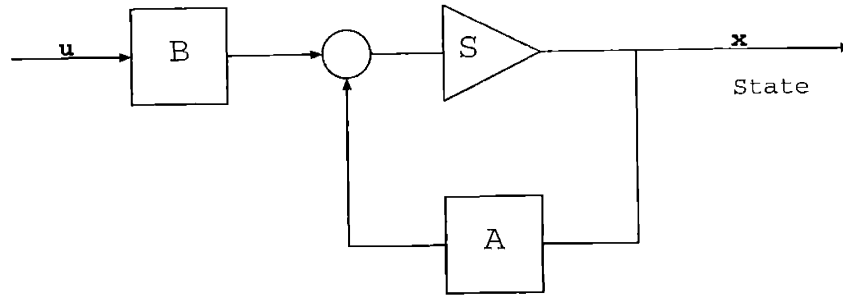
Figure 2-1: Block diagram of a system.

and

$$\hat{B} = B. \tag{2.8}$$

Equation 2.6 therefore becomes:

$$\dot{e} = \hat{A}e$$
$$= (A - LC)e. \tag{2.9}$$

The matrix L is undetermined. However, if the error is to go to zero, i.e. if the observer is to be stable, the real part of the eigenvalues of $(A - LC)$ must all be negative (i.e., lie in the left-half plane). This provides a criterion for selecting L.

Upon substitution of Equations 2.7 and 2.8, Equation 2.4 becomes:

$$\dot{\hat{x}} = A\hat{x} + \hat{B}u + L(y - C\hat{x}). \tag{2.10}$$

Thus, the observer has the same form as the original system (Equations 2.1 and 2.2) except for the presence of the term $L(y - C\hat{x})$. Figures 2-1 and 2-2 are block diagrams of the original system and the observer. Both diagrams are of the same form with the residual $(y - C\hat{x})$ in the lower figure corresponding to the error.

In order for the residual to go to zero, the eigenvalues of $(A - LC)$ must all be negative. But can we always find an L that satisfies this condition? The answer is no. Not all systems can have a stable observer. Also some systems may have a partially stable observer. That is, some components of the residual converge to zero while some don't. Then what conditions does the system have to satisfy in order for
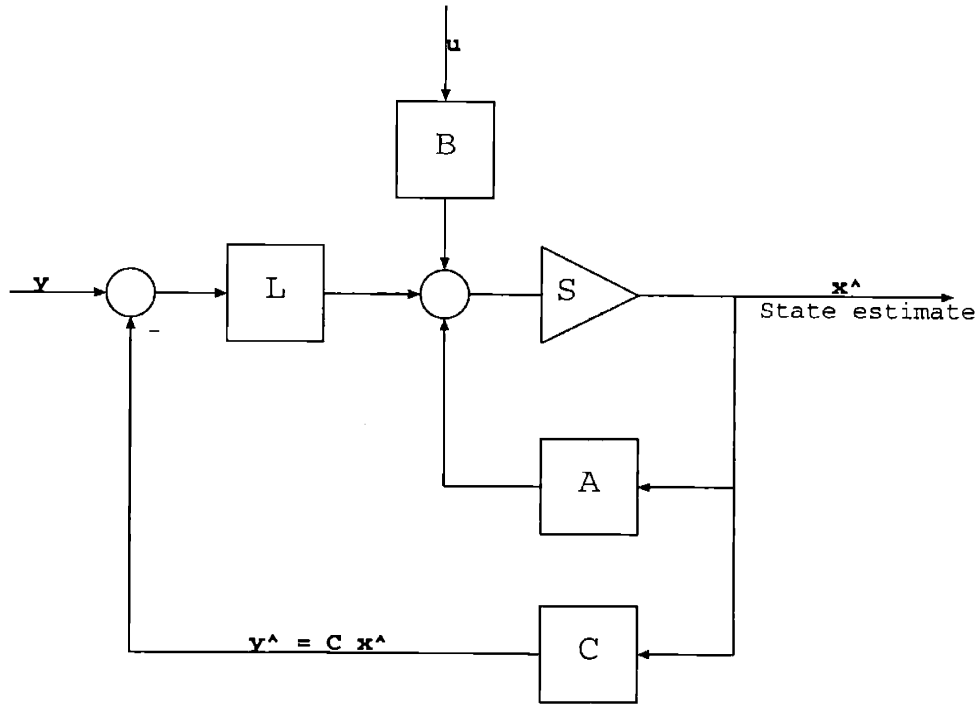
Figure 2-2: Block diagram of a linear observer.

it to have a stable observer? This issue can be systematically investigated after the notion of observability is introduced.

2.3 System Observability

Observability and controllability (also called reachability) are dual concepts in control theory that are of the utmost importance. Let us start with a discussion of observability. Most properties of controllability can be proved or argued in a similar way.

For a system that is represented in the state space description by Equations 2.1 and 2.2 or 2.3, the initial state $x(0)$ of a system can be uniquely determined if and only if the system is observable. Here, the initial state includes every component. This is the definition for observability.

The discrete-time LTI system is used here as an example. If $u(k)$ and $y(k)$ are given for time steps from 0 to T, that is for all $0 \leq k < T$, then by arranging them

in a vector format, we have:

$$
\begin{pmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(T-1) \end{pmatrix} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{pmatrix} \mathbf{x}(0) +
$$

$$
\begin{pmatrix} D & 0 & \ldots & 0 & 0 \\ CB & D & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ CA^{T-2}B & CA^{T-3}B & \ldots & CB & D \end{pmatrix} \begin{pmatrix} \mathbf{u}(0) \\ \mathbf{u}(1) \\ \vdots \\ \mathbf{u}(T-1) \end{pmatrix} \quad (2.11)
$$

The left side is the measurements. The first term on the right side is the state evolution from the initial state and the second term on the right side is the forced response of the system to the control. Here, the measurements and the forced response of the system to the control are known quantities. The initial state $\mathbf{x}(0)$ is unknown and is what we want to determine. Once it is determined, all the states in later steps (or time) can be determined as well according to the dynamics of the system. This is done by using equations 2.3. Without loss of generality, we can subtract the forced response from the measurement, both of which are known. Equation 2.11 then becomes:

$$
\begin{pmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(T-1) \end{pmatrix} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{pmatrix} \mathbf{x}(0) = O_T \mathbf{x}(0), \quad (2.12)
$$

with matrix O_T defined by

$$
O_T = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{pmatrix}. \quad (2.13)
$$

This matrix is called the "T-step observability matrix". Here we introduce an im-

portant conclusion that will be proved in the following paragraphs. If a system is unobservable over n steps, where n is the dimension of state vector, then it is unobservable over any number of time steps. Equivalently, a system is observable if and only if it is observable in n steps. This statement is, in turn, equivalent to $rank(O_n) = n$.

The "if statement" in the above conclusion, i.e. a system is observable if $rank(O_n) = n$, can be proved using a contradiction argument. If O_n does not have full rank, i.e. $rank(O_n) < n$, then $C, CA, CA^2, \ldots$ and CA^{n-1} must be linearly dependent. Thus, there must exist a non-zero vector $\mathbf{d}$, such that

$$O_T \mathbf{d} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{pmatrix} \mathbf{d} = 0 \tag{2.14}$$

Thus, if an initial condition $\mathbf{x}^*(0)$ satisfies Equation 2.11, $\mathbf{x}^*(0) + \mathbf{d}$ must also satisfy Equation 2.11. That means the system initial condition $\mathbf{x}(0)$ can not be uniquely determined. Thus, by definition, the system is unobservable. So far, we have shown that a system is observable if $rank(O_n) = n$.

To prove the "only if statement", we need to introduce the Cayley-Hamilton theorem. The determinant of an $n \times n$ matrix A can be written in the form of a characteristic polynomial:

$$a(s) = det(sI - A) = a_0 + a_1 s + a_2 s^2 + \ldots + a_{n-1} s^{n-1} + s^n, \tag{2.15}$$

where "det" is the determinant, which is a function of s. The a_0 to a_{n-1} are the polynomial coefficients of the matrix A. Then, when s is chosen to be matrix A itself, Equation 2.15 becomes

$$a(A) = a_0 I + a_1 A + a_2 A^2 + \ldots + a_{n-1} A^{n-1} + A^n = 0. \tag{2.16}$$

Thus, the matrix A^n can be written as a linear combination of matrices $I, A, A^2,$ $\ldots, A^{n-1}$. Thus, adding A^n to $[I A A^2 \ldots A^{n-1}]$ will not change its rank, neither will multiplying it by a constant non-zero matrix C. So, we have $Rank(O_n) = Rank(O_{n+1}) = Rank(O_{n+2}) = \ldots$. Thus, if a system is unobservable over n steps, no matter how many steps the system goes through, it remains unobservable. The converse is also true. Namely, if a system is observable, it must be observable over n steps. We have proved the "only if" part. The same results hold for a continuous-time system.

To summarize, we have introduced three essentially equivalent notions for a state-space LTI system. They are

1. A system is observable, or a pair (C, A) is observable,

2. The initial condition of a system can be determined uniquely,

3. The matrix $O = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{pmatrix}$ has rank n.

This O matrix is called the observability matrix, and it is essentially the n-step observability matrix.

The importance of knowing whether or not a system is observable will become clear after we discuss the properties of an observer. Before proceeding to that, another so called "modal test" for the determination of observability is reviewed. This test method is more convenient in terms of implementation and it can also help to visualize the concept of observability.

Again, using a discrete-time system as an example, the "modal test" is illustrated. If an $n \times n$ matrix A is diagonalizable (most real system matrices are), then it can be decomposed into:

$$A = \sum_{i=1}^{n} \lambda_i v_i w_i^T, \tag{2.17}$$

where λ_i is the ith eigenvalue of A, v_i is the ith right eigenvector of A, and w_i the ith left eigenvector of A. Under the assumption of $u = 0$ that we previously proposed

without loss of generality, we then have that the kth step output equals:

$$
\begin{aligned}
\mathbf{y}(k) &= CA^k\mathbf{x}(0) \\
&= C\sum_{i=1}^{n}\mathbf{v}_i\lambda_i^k\mathbf{w}_i^T\mathbf{x}(0) \\
&= \sum_{i=1}^{n}C\mathbf{v}_i\mathbf{w}_i^T\mathbf{x}(0)\lambda_i^k.
\end{aligned}
\tag{2.18}
$$

If there exists a right eigenvector $\mathbf{v}_i^*$ of A that satisfies $C\mathbf{v}_i^* = 0$, then when we choose a vector $\mathbf{d} = \mathbf{v}_i^*$, it will satisfy:

$$
\sum_{i=1}^{n}C\mathbf{v}_i\mathbf{w}_i^T\mathbf{d}\lambda_i^k = 0,
\tag{2.19}
$$

which renders the system unobservable because any initial condition $\mathbf{x}(0)$ will result in the same measurements as $\mathbf{x}(0) + \mathbf{d}$.

The modal test can be formulated as follows: a system is unobservable if and only if

$$
\begin{pmatrix} \lambda I - A \\ C \end{pmatrix} \mathbf{v} = 0,
\tag{2.20}
$$

for some complex or real number λ and a non-zero vector $\mathbf{v}$.

Now we have discussed the definition of system observability and methods that can be carried out mathematically to determine it. In the following section the relationship between observability and the design of an observer will be covered.

2.4   Selection of Observer Gain

Now we come back to the question that was raised in section 2.2. The question was if and how we can make all the real parts of the eigenvalues of matrix $(A - LC)$ negative.

The eigenvalues of matrix $(A - LC)$ can be solved as the roots of the characteristic

polynomial equation for $(A - LC)$. The latter is given by:

$$det|sI - (A - LC)| = a_0 + a_1 s + a_2 s^2 + \ldots + a_{n-1}s^{n-1} + s^n = 0. \qquad (2.21)$$

The polynomial coefficients on the right side depend on the matrices A, C, and L. Matrices A and C are known constants. Thus, the roots of the above polynomial equation, which are also the eigenvalues of the matrix A, will depend on the choice of L. If the desired eigenvalues are $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$, then we have,

$$(s - \alpha_0)(s - \alpha_1)\ldots(s - \alpha_{n-1}) = 0. \qquad (2.22)$$

By comparing the coefficients in equations 2.21 and 2.22 we have n^2 equations. Because the matrix L has n^2 elements, we also have n^2 unknowns to solve. But is there always a solution for this L and if there is one, is it unique?

As we did in the previous section, here we introduce a theorem first and then prove it. The theorem here states that "there exists a matrix L such that the eigenvalues of $(A - LC)$ can be any arbitrary self-conjugate set of complex numbers if and only if the pair (C, A) is observable."

If (C, A) is unobservable, from the modal test of Equation 2.20, then we have at least one unobservable mode v_i^* and a corresponding λ_i, such that $Cv_i^* = 0$ and $Av_i^* = \lambda_i v_i^*$. So,

$$
\begin{aligned}
(A - LC)v_i^* &= Av_i^* - LCv_i^* \\
&= \lambda_i v_i^*. \qquad (2.23)
\end{aligned}
$$

This means that this particular eigenvector v_i^* of A is also an eigenvector of $(A - LC)$, and both the corresponding eigenvalues are λ_i. Because this eigenvalue λ_i of $(A - LC)$ is determined by matrix A only and does not depend on the choice of L, we can not place this eigenvalue that is associated with eigenvector v_i^* of $(A - LC)$ arbitrarily by selecting L. In summary, when the pair (C, A) is unobservable, there must exist at least one eigenvalue of $(A - LC)$ whose corresponding eigenvalue can not be placed

38

arbitrarily.

If (C, A) is observable, then as we have shown in the previous section, it is equivalent to stating that the observability matrix $O = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{pmatrix}$ must have rank n.

We know that under a similarity transformation, a system described by (A, B, C) as in Equations 2.1 and 2.2, or in Equation 2.3 can also be described in the same way by $(\tilde{A}, \tilde{B}, \tilde{C})$, where the similarity transformation is carried out using any proper matrix T as,

$$\tilde{A} = T^{-1}AT, \ \tilde{B} = T^{-1}B, \ \tilde{C} = CT, \ \tilde{\mathbf{x}} = T^{-1}\mathbf{x}, \ \tilde{\mathbf{u}} = \mathbf{u}, \ \tilde{\mathbf{y}} = \mathbf{y}. \tag{2.24}$$

The system will not be affected by how we perceive it. A simple example of similarity transformation is the rotation of a set of coordinates.

Assume the system is a single output system, then the C matrix becomes a row vector. Using O as a similarity transformation matrix, we have,

$$OA = \begin{pmatrix} CA \\ CA^2 \\ \vdots \\ CA^n \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-2} & -a_{n-1} \end{pmatrix} \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}, \tag{2.25}$$

where $a_0, a_1, \dots a_{n-1}$ are the coefficients in the Cayley-Hamilton equation (Equation 2.16).

Denote the first matrix on the right side of the above equation as $\tilde{A}$. That is

$$\tilde{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{n-2} & -a_{n-1} \end{pmatrix}, \qquad (2.26)$$

then Equation 2.25 becomes

$$OA = \tilde{A}O, \qquad (2.27)$$

and

$$\tilde{A} = OAO^{-1}. \qquad (2.28)$$

The inverse matrix O^{-1} exists because matrix O has full rank n. It is easy to verify that the coefficients $a_0, a_1, \ldots a_{n-1}$ are also the coefficients of the characteristic polynomial of matrix $\tilde{A}$.

Similarly,

$$C = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} = \tilde{C}O, \qquad (2.29)$$

where $\tilde{C} = CO^{-1}$ and

$$\tilde{C} = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}. \qquad (2.30)$$

Thus, any observable pair (C, A) can be written as $(\tilde{C}, \tilde{A})$ with $\tilde{C}$ and $\tilde{A}$ defined as in Equations 2.30 and 2.26. It can be verified that the particular system with (C^*, A^*)

defined as

$$
A^* = \begin{pmatrix}
-a_0 & 1 & 0 & 0 & \cdots & 0 \\
-a_1 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & & & & & \\
-a_{n-2} & 0 & 0 & \cdots & 0 & 1 \\
-a_{n-1} & 0 & 0 & \cdots & 0 & 0
\end{pmatrix},
\tag{2.31}
$$

and

$$
C^* = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix},
\tag{2.32}
$$

is an observable system. Thus, it could also be written as $(\tilde{C}, \tilde{A})$ under a similarity transformation O^*. Because a similarity transformation can be performed in both directions, the results lead to the conclusion that any observable pair (C, A) can be transformed to (C^*, A^*) under a similarity transformation of OO^{*-1}.

It can be easily verified that the eigenvalues of $(A - LC)$ are the same as those of $(A^* - L^*C^*)$, where A^* and C^* are the pair after a similarity transformation $T = OO^{*-1}$ on (C, A), and $L^* = TL$. Suppose we would like the eigenvalues of $(A - LC)$ be placed at $(\alpha_0, \alpha_1, ..., \alpha_{n-1})$, then the eigenvalues of $(A^* - L^*C^*)$ will also be $(\alpha_0, \alpha_1, ..., \alpha_{n-1})$. For a single output system, the L and L^* are column vectors. Expressing L^* as

$$
L^* = \begin{pmatrix}
l_0^* \\
l_1^* \\
\vdots \\
l_{n-1}^*
\end{pmatrix},
\tag{2.33}
$$

then, $(A^* - L^*C^*)$ can be expressed as

$$
A^* = \begin{pmatrix}
-a_0 - l_0^* & 1 & 0 & 0 & \cdots & 0 \\
-a_1 - l_1^* & 0 & 1 & 0 & \cdots & 0 \\
\vdots & & & & & \\
-a_{n-2} - l_{n-2}^* & 0 & 0 & \cdots & 0 & 1 \\
-a_{n-1} - l_{n-1}^* & 0 & 0 & \cdots & 0 & 0
\end{pmatrix},
\tag{2.34}
$$

41

where $-a_i - l_i^*$ are the coefficients of the characteristic polynomial of $(A^* - L^*C^*)$, which can also be obtained by expanding the expression $sI - (A^* - L^*C^*) = (s - \alpha_0)(s - \alpha_1) \cdots (s - \alpha_{n-1}) = s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0$. If the desired eigenvalues $(\alpha_0, \alpha_1, ..., \alpha_{n-1})$ are determined, the L^* can be solved by comparing the coefficients, $l_i^* = a_i + b_i$. These are a set of (n) algebraic equations and the solutions are guaranteed for any selection of α_i. The vector L can then be solved by $L = T^{-1}L^*$.

In summary, we have shown that for a single output system, if a pair (C, A) is observable, then the eigenvalues of $(A - LC)$ can be placed arbitrarily (complex eigenvalues must be in conjugate pairs). For a multiple output system, the same conclusion holds, but the proof is more involved and will not be demonstrated here.

The method for determining the observer gain matrix L given eigenvalues of $(A - LC)$ has also been illustrated. This method can also be used to determine the feedback gain of a controllable system and is called Ackermann' formula. In Matlab, which is a widely used commercial computational mathematics package, this method is implemented to calculate the observer gain given desired eigenvalues. Although, theoretically, the solutions of the gain L are guaranteed and are unique, the implementation is prone to numerical errors because of the transformations that involve calculating inverse matrices. Here I will give an example to show how numerical errors could result in wrong answers when using Matlab's ACKER function.

Let A be a 7×7 matrix

$$A = \begin{pmatrix} -64.9000 & 0.0126 & 0.0301 & 0.1118 & 0.3014 & 1.1363 & 3.0137 \\ 2.1000 & -0.0126 & 0 & 0 & 0 & 0 & 0 \\ 14.2000 & 0 & -0.0301 & 0 & 0 & 0 & 0 \\ 12.7000 & 0 & 0 & -0.1118 & 0 & 0 & 0 \\ 25.7000 & 0 & 0 & 0 & -0.3014 & 0 & 0 \\ 7.5000 & 0 & 0 & 0 & 0 & -1.1363 & 0 \\ 2.7000 & 0 & 0 & 0 & 0 & 0 & -3.0137 \end{pmatrix},$$

$$(2.35)$$

Table 2.1: Desired and Calculated Eigenvalues of $(A - LC)$

Desired eigenvalues	Calculated eigenvalues	Relative error (%)
-1	-1.4248 + 0.2732i	50.5
-1	-1.4248 - 0.2732i	50.5
-1	-0.9490 + 0.4116i	41.5
-1	-0.9490 - 0.4116i	41.5
-1	-0.7394 + 0.1607i	30.6
-1	-0.7394 - 0.1607i	30.6
-1	-0.7756	22.4

and C be a 1×7 vector

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.36}$$

Suppose the eigenvalues of matrix $(A - LC)$ are to be all negative one ([-1, -1, -1, -1, -1, -1, -1]), the ACKER function of Matlab will give the corresponding L to be

$$L = 1.0e + 04 \times \begin{pmatrix} -0.00625039343443 \\ 4.29562501335193 \\ -2.09384827924070 \\ 0.08665238529537 \\ 0.00176883075166 \\ 0.00074990585454 \\ 0.00030361491611 \end{pmatrix}. \tag{2.37}$$

If we substitute this calculated values into $(A - LC)$, we will have the eigenvalues that are not exactly what we desired. The resulting eigenvalues and desired values are listed in Table 2.1. There is more than a 20% error between them. The stiffness of matrix A in this example causes a singularity problem while calculating the inverse matrices and results in the big error of the gain calculation. Kautsky and Nichols proposed improved algorithms to calculate the observer (or controller) gain for desired pole placement in 1985 [31]. In this research, we do not specify all the poles and then find a corresponding gain matrix L. For our purpose, pole placements are not the ultimate goal. Other elements in L are also important for enhancing the signal-to-

noise ratio. Thus, the procedure for obtaining an appropriate gain L is instead as follows:

1. Choose an initial gain matrix L.

2. Calculate the eigenvalues of $(A - LC)$.

3. Calculate a cost function that depends on the eigenvalues of $(A - LC)$.

4. Change L to lower the cost function.

5. Repeat steps 2 to 5 until the goal is met.

In step one, we still need to solve A^k where k is usually a number less than n as explained in the next section. It is usually the case that the matrix $(A - LC)$ is not stiff by choosing an appropriate L. I mention this because the differential equations describing the nuclear reactor kinetic are stiff due to two time scales in the process - prompt neutrons and delayed neutrons.

2.5 Observer-Based Fault Detection

Originally, a system observer was used to estimate system state variables where measurements are not available. These estimations are then used by a controller for system control. Beard was the first to introduce an observer for fault detection purpose in his PhD thesis that was done at MIT in 1971 [11]. The theories on applying an observer to fault detection are reviewed in this section.

It was shown in the previous sections that when the model for the observer and the real system match each other precisely, the estimation errors ($\mathbf{y} - \hat{\mathbf{y}}$ and $\mathbf{x} - \hat{\mathbf{x}}$) go to zero if all the eigenvalues of $(A - LC)$ are negative. But when the model differs from the real system, even if all the eigenvalues of $(A - LC)$ are in the left half plane, the estimation error may not go to zero. This estimation error contains information that can be used for fault detection. Two main causes can result in a non-zero estimation error.

One is changes in the dynamics of the system. These changes are of interest to us because they are usually indications of a component fault in the system and thus can be used for fault detection. The other is imperfections of the model. Because it is unusual for a model to reproduce a real system perfectly, this type of estimation error is present in almost all observers and poses a major difficulty for both system control and fault detectors that are based on observer estimations. Fortunately, by carefully designing the observer and by constructing appropriate decision rules, the effect of the latter will not mask totally the effect from the former and information about the system dynamics can still be extracted. In this section, fault detection based on a perfect model is discussed. The problem of the noise corruption and model imperfectness will be investigated in later chapters.

Because we only have access to $\mathbf{y}$, the residual (estimation error of measurements) refers to $\mathbf{y} - \hat{\mathbf{y}}$, where $\mathbf{y}$ is the measurement and $\hat{\mathbf{y}}$ is the estimated values for $\mathbf{y}$ i.e. the output from the observer.

Here let us adopt a convention that the system parameters are denoted by a subscript S, and those of the model with a subscript M. Estimated variables are superscripted by a hat () as before. Upon rewriting equations 2.3 using these notations, they become:

$$
\begin{aligned}
\dot{\mathbf{x}} &= A_S \mathbf{x} + B_S \mathbf{u}, \\
\dot{\mathbf{y}} &= C_S \mathbf{x},
\end{aligned}
\tag{2.38}
$$

and the model equations become

$$
\begin{aligned}
\dot{\hat{\mathbf{x}}} &= A_M \hat{\mathbf{x}} + B_M \mathbf{u}_d + L_M(\mathbf{y} - \hat{\mathbf{y}}), \\
\hat{\mathbf{y}} &= C_M \hat{\mathbf{x}}.
\end{aligned}
\tag{2.39}
$$

Notice that the subscript in the model for the input signal $\mathbf{u}$ are expressed as $\mathbf{u}_d$ meaning desired input which is to be distinguished from the real input that the system actually receives.

45

Then, the discrepancy between the system state variables and the observer state variables $\epsilon = \mathbf{x} - \hat{\mathbf{x}}$ satisfies,

$$\begin{aligned} \dot{\epsilon} &= \dot{\mathbf{x}} - \dot{\hat{\mathbf{x}}} \\ &= (A_S\mathbf{x} - A_M\hat{\mathbf{x}}) + (B_S\mathbf{u} - B_M\mathbf{u}_d) - L_M(C_S\mathbf{x} - C_M\hat{\mathbf{x}}). \end{aligned} \qquad (2.40)$$

When the observer parameters match those of the system, i.e. assuming we have perfect knowledge of the system, $A_S = A_M, B_S = B_M, C_S = C_M$, and also assuming that the actuator works perfectly such that $\mathbf{u} = \mathbf{u}_d$, then equation 2.40 is reduced to equation 2.9. Failure of those parameters to match each other could be the result of either model inaccuracy or an abnormality of the system.

Now consider three types of abnormal events,

1. <u>Actuator fault</u>: Deviation of $\mathbf{u}$ from the desired value which can be expressed as $\mathbf{u} - \mathbf{u}_d = \Delta\mathbf{u}$.

2. <u>Abnormality of system dynamics</u>: Deviation of A_S and/or B_S from those of the model, which can be expressed as $A_S - A_M = \Delta A$ and/or $B_S - B_M = \Delta B$.

3. <u>Sensor error</u>: There is a deviation in the measurements which can be expressed as $\mathbf{y} = C_S\mathbf{x} + \Delta\mathbf{y_m}$.

The rest of this section is focused on these three types of possible errors.

For the first case, actuator fault, where $\mathbf{u} - \mathbf{u}_d = \Delta\mathbf{u}$ while $A_S = A_M$, $B_S = B_M$ and $C_S = C_M$, the estimation for $\mathbf{x}$ becomes,

$$\dot{\epsilon}(t) = (A - LC)\epsilon(t) + B\Delta\mathbf{u}(t). \qquad (2.41)$$

The fault term $\Delta\mathbf{u}(t)$ can be decomposed into time-dependent scalars and time-

independent vectors. Considering a single actuator fault, we then have

$$\Delta \mathbf{u}(t) = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} f(t) = \mathbf{e}_i f(t), \tag{2.42}$$

where $\mathbf{e}_i$ is a time-independent vector with the ith element being one and all other elements being zero. Thus, $\mathbf{e}_i$ specifies which actuator goes wrong. The $f(t)$ in the above equation is a time-dependent scalar and it describes how the error in that particular actuator specified by $\mathbf{e}_i$ evolves in time.

If we substitute Equation 2.42 into Equation 2.41, the estimation error for state $\mathbf{x}$ under a single actuator fault can be represented as

$$\dot{\epsilon}(t) = (A - LC)\epsilon(t) + \mathbf{b}_i f(t), \tag{2.43}$$

which has a solution in the form of

$$\epsilon(t) = \epsilon_0 e^{(A-LC)t} + \int_0^t e^{(A-LC)(t-\tau)} \mathbf{b}_i f(\tau) d\tau. \tag{2.44}$$

Here b_i is the ith column of the matrix B. It is clear in Equation 2.44 that the settled $\epsilon(t)$ after the first exponential decay term dies off will be in the direction along vector $\mathbf{b}_i$ with magnitude changing in time according to a time convolution term that depends on $f(t)$.

We will discuss how to utilize the results obtained so far to detect an actuator failure in more detail later. Now we will proceed to derive similar results for changes in the dynamics and for sensor failures.

Consider a system dynamics error, where $A_S - A_M = \Delta A$ or $B_S - B_M = \Delta B$

while $\mathbf{u} = \mathbf{u}_d$ and $C_S = C_M$. Then, Equation 2.40 becomes,

$$\dot{\epsilon} = (A - LC)\epsilon + \Delta A\mathbf{x} + \Delta B\mathbf{u}, \tag{2.45}$$

which has a solution of the form

$$\epsilon(t) = \epsilon_0 e^{(A-LC)t} + \int_0^t e^{(A-LC)(t-\tau)}\Delta A\mathbf{x}(\tau)d\tau + \int_0^t e^{(A-LC)(t-\tau)}\Delta B\mathbf{u}(\tau)d\tau. \tag{2.46}$$

Again, decompose ΔA and ΔB into time-dependent scalars and time-independent vectors and consider only a single fault. We have,

$$\Delta A = \Delta a_{ij}\mathbf{e}_i\mathbf{e}_j', \tag{2.47}$$

and

$$\Delta B = \Delta b_{ij}\mathbf{e}_i\mathbf{e}_j', \tag{2.48}$$

where $\mathbf{e}_i$ and $\mathbf{e}_j$ are as defined in equation 2.42 and ′ means transpose. So, $\mathbf{e}_i$ is a column vector with the ith component being one and all other components being zero, and $\mathbf{e}_j$ is a column vector with the jth component being one and all other components being zero. Thus, $\mathbf{e}_i$ and $\mathbf{e}_j$ specify that it is the ijth component in the matrix A or B that goes wrong and the magnitude or the error is $\Delta a_{ij}(t)$ or $\Delta b_{ij}(t)$. Here $\Delta a_{ij}(t)$ or $\Delta b_{ij}(t)$ are time-dependent scalars.

Now substitute Equations 2.47 and 2.48 into 2.45 and separate the cases for error in A or B. We have, for error in A

$$\begin{aligned}
\dot{\epsilon} &= (A - LC)\epsilon + \mathbf{e}_i\Delta a_{ij}(\mathbf{e}_j'\mathbf{x}(t)) \\
&= (A - LC)\epsilon + \mathbf{e}_i\Delta a_{ij}x_j(t),
\end{aligned} \tag{2.49}$$

and for error in B

$$\begin{aligned}
\dot{\epsilon} &= (A - LC)\epsilon + \mathbf{e}_i\Delta b_{ij}(\mathbf{e}_j'\mathbf{u}(t)) \\
&= (A - LC)\epsilon + \mathbf{e}_i\Delta b_{ij}u_j(t).
\end{aligned} \tag{2.50}$$

Notice that $(\mathbf{e}'_j\mathbf{x})$ is $x_j(t)$, the jth element of the column vector $\mathbf{x}$ and $(\mathbf{e}'_j\mathbf{u})$ is $u_j(t)$, the jth element of the column vector $\mathbf{u}$. Both are time-dependent scalars. Thus, the settled estimation error ϵ for the case of error in A is along the direction $\mathbf{e}_i$ with magnitude proportional to $(\mathbf{e}'_j\mathbf{x})\Delta a_{ij}$. Similarly, the settled estimation error ϵ for the case of error in B is along the direction $\mathbf{e}_i$ with magnitude proportional to $(\mathbf{e}'_j\mathbf{u})\Delta b_{ij}$.

Finally, consider the case for sensor failure. Here $\mathbf{y} = C_S\mathbf{x} + \Delta\mathbf{y}_m$ while $A_S = A_M = A$, $B_S = B_M = B$, $C_S = C_M = C$ and $\mathbf{u} = \mathbf{u}_d$. If we only consider a single sensor failure, then Equation 2.39 becomes,

$$\mathbf{y} = C\mathbf{x} + \mathbf{e}_i f(t), \tag{2.51}$$

where $\mathbf{e}_i$ is similar to that in the cases of actuator fault and dynamics changes. It is a column vector with the ith component being one and all other components being zero and it specifies which sensor goes wrong. The $f(t)$ is a time-dependent scalar that denote the magnitude of the error.

Notice that the estimation error of state variable $\mathbf{x}$ under sensor error will satisfy

$$
\begin{aligned}
\dot{\epsilon} &= (A - LC)\epsilon - L\mathbf{e}_i f(t) \\
&= (A - LC)\epsilon - \mathbf{l}_i f(t), \tag{2.52}
\end{aligned}
$$

where $\mathbf{l}_i$ is the i th column vector of the matrix L.

In summary, all three types of fault events will result in an estimation error that can be expressed in an unified form of

$$\dot{\epsilon} = (A - LC)\epsilon - \mathbf{v}f(t), \tag{2.53}$$

where $\mathbf{v}$ is a time-independent column vector that is associated with the location of the fault while $f(t)$ is a scalar that characterizes the magnitude of the fault. Equation 2.54 has solution in the form of

$$\epsilon(t) = \epsilon_0 e^{(A-LC)(t)} - \int_0^t e^{(A-LC)(t-\tau)}\mathbf{v}f(\tau)d\tau. \tag{2.54}$$

The first term on the right side dies off if L can be chosen to make all the eigenvalues of $(A - LC)$ in the left half plane.

The fault event that is associated with the vector $\mathbf{v}$ is detectable if there exists a matrix L such that $C\epsilon$ is along a fixed direction in the output space and at the same time, all eigenvalues of $(A - LC)$ can be placed arbitrarily [11].

It is clear that to satisfy the first condition that $C\epsilon$ remains in a fixed direction, we must have

$$rk\ \ C[I, (A - LC), (A - LC)^2, \ldots, (A - LC)^{n-1}]\mathbf{v} = 1. \tag{2.55}$$

This result follows easily from Equation 2.54.

To satisfy the second condition that all eigenvalues of $(A - LC)$ can be placed arbitrarily, we have shown in the previous sections that we must have the observability matrix to have rank n.

Thus, to detect a fault event associated with vector $\mathbf{v}$, we must choose an appropriate matrix L to satisfy:

1. all eigenvalues of $(A - LC)$ have negative real parts, and

2. $rk\ \ C[I, (A - LC), (A - LC)^2, \ldots, (A - LC)^{n-1}]\mathbf{v} = 1$.

It has been proved that every vector in the state space is detectable if and only if (A, C) is an observable pair [11]. The proofs are skipped here and please consult [11] for more details. The procedure to find an appropriate L such that the two constraints can be satisfied are as follows:

Given (A, C), we define matrices C', K, and M' as:

$$C' = [I - C\mathbf{v}[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T]C, \tag{2.56}$$

$$K = A - A\mathbf{v}[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T]C, \tag{2.57}$$

50

and

$$M' = \begin{pmatrix} C' \\ C'K \\ \vdots \\ C'K^{n-1} \end{pmatrix}. \tag{2.58}$$

The maximal generator order ν is determined by

$$\nu = n - rank(M'). \tag{2.59}$$

The maximal generator g for vector $\mathbf{v}$ is determined by:

1. Find the null space of M', denoted by $Z = Null(M')$, such that $M' * Z$ has negligible elements, and $Z' * Z = I$.

2. Find a vector g in the null space of M' satisfying

$$\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{\nu-2} \end{pmatrix} g = \mathbf{0}, \tag{2.60}$$

and

$$CA^{\nu-1}g = CA^{\mu}\mathbf{v}, \tag{2.61}$$

where μ is defined by

$$CA^j\mathbf{v} = \mathbf{0} \quad for \quad j = 0 \dots \mu - 1,$$
$$CA^{\mu}\mathbf{v} \neq \mathbf{0}. \tag{2.62}$$

The matrix L constructed by

$$L = [p_0 g + p_1 Ag + \dots p_k A^{k-1} + A_k g] \cdot [(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T$$

$$+L'[I - C\mathbf{v}[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T], \tag{2.63}$$

will satisfy the two constrains. Here p_0 to p_k are the coefficients for the characteristic polynomial with the k roots being eigenvalues of $(A - LC)$. L' is an arbitrary matrix. Thus p_0 to p_k and L' are our choice in the design.

If we perform a Laplace transform on both sides of Equation 2.53, we can have the spectrum domain response of the measurement estimation error $\mathbf{E}(s)$ to the fault signal $F(s)$

$$\mathbf{E}(s) = \mathcal{L}[C\epsilon(t)] = C\mathcal{L}[\epsilon(t)] = C[sI - (A - LC)]^{-1}\mathbf{v}F(s) = G(s)F(s), \tag{2.64}$$

where $\mathcal{L}$ means the Laplace operation and $F(s) = \mathcal{L}f(t)$. $G(s)$ is the transfer function from $F(s)$ to $E(s)$. This relationship is very useful when we try to maximize the signal to noise ratio during design of L.

R.V. Beard has done a superb job in proving and deriving the construction of the matrix gain that can satisfy the two constraints for fault detection. But he didn't give explanations for how we should chose the arbitrary matrix L'. This may not be an easy task without the aid of a computer. In this research, freedom in choosing the matrix L' and the poles p_0 to p_k is essential for our design of an optimal fault detection filter that can function in the presence of noise.

Chapter 3

Nuclear Reactor Dynamics and MIT Research Reactor II

In this chapter, the reactor dynamics and its mathematical representation are reviewed. Linearized reactor model and various feedback mechanism models are introduced. The transfer function model that is based on the linearized state space representation is presented. Empirical models that would be needed to apply the fault detection technique to the MIT research reactor are developed based on MITR-II data. These models include a temperature response model, a temperature feedback worth model, a regulating rod reactivity worth model and a shim blade reactivity worth model.

3.1 Reactor Dynamics

A reactor's multiplication factor is defined as the ratio of the neutrons produced from fission to those lost by either leakage or absorption. The reactor is sub-critical, critical, or supercritical depending on whether the multiplication factor is less than, equal to, or greater than unity. The quantity most fundamental to the control of power in a nuclear reactor is the reactivity. It is defined as the fractional departure of the multiplication factor from unity. A reactivity of zero therefore corresponds to the critical condition. Adjustments of power in reactors described by space-independent

kinetics are generally accomplished by positioning neutron-absorbing control rods or rotating drums so as to temporarily alter the neutron multiplication rate. If a power increase is desired, a control rod would be withdrawn (or a drum rotated) so as to insert positive reactivity and thereby place the reactor on a period. (Note: Period is defined as the power level divided by the rate of change of power. Thus, a period of infinity corresponds to steady-state, while one equal to a small positive number indicates a rapid power increase.) Having established a period, the power is allowed to rise. Once the power level approaches the desired value, the control device is gradually returned to its original position in order to reduce the reactivity to zero and to level the power without overshoot. Reactor operation is characterized by several reactivity feedback mechanisms. One of the most important is that a rise in fuel temperature will result in the increased absorption of neutrons in reactions that do not lead to fission. Known as the "Doppler" effect, this is an inherent safety feature that limits the potential for an accident because any increase in the reactor power will cause the fuel temperature to rise which will in turn decrease the number of neutrons available to sustain the fission chain reaction. Another feedback mechanism is that a rise in the temperature of the moderator (which often also serves as the coolant) will cause there to be both more neutron leakage and less moderation. The net result is the generation of negative reactivity. The presence of these and other feedback effects makes reactor dynamics non-linear because the reactivity is influenced by the power level.

Figure 3-1 is a simplified schematic of the fission process. Relative to reactor control, the most significant feature is that there are three parallel but separate mechanisms for the production of neutrons.

Prompt neutrons appear directly following the fission event and have lifetimes that are quite short, typically 100 μs.

Delayed neutrons are produced following the decay by beta particle emission of fission products that are referred to as "precursors." The delay in the appearance of a delayed neutron relative to the fission is the result of the precursor half-life. About 20 different precursors have been identified and for convenience, they are combined
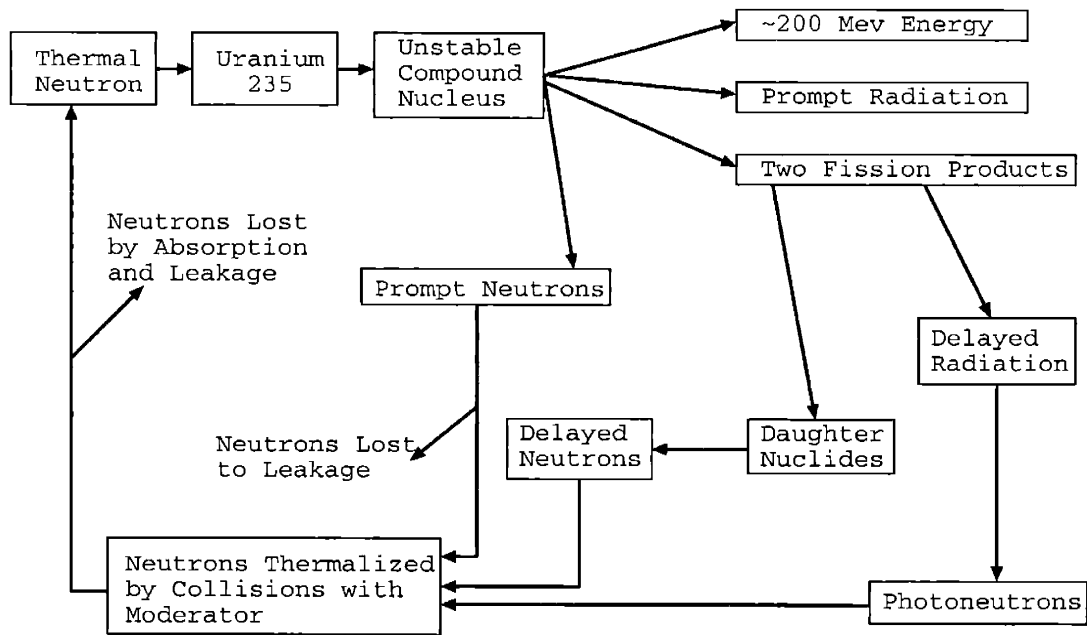
Figure 3-1: Fission process showing prompt, delayed, and photo neutron production.

into six groups with half-lives ranging from 0.23 to 55 s. The average value is 12.2 s. Table 3.1 lists the six-group precursor data for ^{235}U.

The third mechanism for neutron production is the interaction of fission-product gamma rays with certain moderating materials, most notably heavy water and beryllium. The appearance of photo-neutrons is delayed relative to the fission event because of the time required for the fission products to undergo radioactive decay and emit the needed gamma rays. Prompt-, delayed-, and photo-neutrons are all produced at high energies. In order to sustain the fission reaction, these neutrons must be slowed down or thermalized. This is accomplished by designing the reactor so that the neutrons will collide with the nuclei of a moderating material such as water, thereby giving up much of their kinetic energy. The efficiency of the thermalization process decreases as moderator temperature rises. This behavior, as noted above, contributes to safety. However, it is another factor that makes reactor dynamics non-linear. (Note: Photo-neutrons may be represented mathematically in the same way as delayed neutrons. Hence, the term "delayed" is subsequently used here to refer to both neutrons resulting from precursor decay and photo-neutrons.)

Table 3.1: Six-Precursor Group Half Lives and Delayed Neutron Fractions for Thermal Fission of ^{235}U [33]

Group	Half-life $T_{1/2}(s)$	Decay constant $(1/s)$	Delayed fraction β_i
1	55.0	0.0126	0.00021
2	23.0	0.0301	0.00142
3	6.2	0.1118	0.00127
4	2.3	0.3014	0.00257
5	0.61	1.1363	0.00075
6	0.23	3.0137	0.00027
Total	-	-	0.0065

The adjustment of reactor neutronic power requires planning because the respective appearances of prompt and delayed neutrons following fission occur on very different time scales. Specifically, at any given moment, the prompt neutron population will be proportional to the current power level while the delayed neutron population will be a function of the previous power levels or power history. This difference has no significance during extended steady-state operation because the previous and current power levels are the same. However, such will not be the case during power adjustments. For example, during power increases, the contribution of the delayed neutrons will always be less than it would be at equilibrium and the contribution of the prompt neutrons will therefore be correspondingly greater. Hence, upon attaining the desired power level, the rate of increase of the delayed neutrons, unlike that of their prompt counterparts, can not be immediately halted. Rather, the delayed neutron population will continue to rise until it attains equilibrium with the precursor population that corresponds to the reactor's power level. Hence, if power overshoots are to be averted, it is essential to limit the delayed neutron contribution so that, upon attainment of the desired power, the insertion of the control mechanism will make the rate of change of the prompt neutrons sufficiently negative so as to offset the continued increase in the delayed neutrons.

3.2 Mathematical Representation of Reactor Dynamics

As a starting point, it is assumed that the reactor's dynamics can be adequately described by the space-independent prompt-only kinetics equations. Then according to the neutron balance for a fission chain system, we have [32]:

$$\frac{dn}{dt} = \mu\Sigma_f\Phi - \Sigma_a\Phi - DB^2\Phi, \qquad (3.1)$$

where dn/dt is the change of neutron density n, μ is the number of neutron production per fission, Σ_f is the macro-scopic fission cross-section, or fission probability per unit length, Φ is the neutron flux, Σ_a is the macro-scopic total absorption cross-section, D is the diffusion coefficient, B^2 is the buckling defined as

$$B^2 = -\frac{\Delta^2\Phi}{\Phi}. \qquad (3.2)$$

The three terms on the right side of equation 3.1 are the neutron production rate from fission, the neutron absorption rate, and the leakage rate per unit volume respectively. Equation 3.1 can also be written as:

$$\frac{1}{v}\frac{d\Phi}{dt} = \mu\Sigma_f\Phi - \Sigma_a\Phi - DB^2\Phi. \qquad (3.3)$$

Upon dividing both sides by $\mu\Sigma_f$ and then rearranging the left side, we have:

$$\frac{1}{v\mu\Sigma_f}\frac{d\Phi}{dt} = (1 - \frac{\Sigma_a + DB^2}{\mu\Sigma_f})\Phi. \qquad (3.4)$$

Define the multiplication factor k, the reactivity ρ and the mean neutron generation time l^* as

$$k = \frac{\mu\Sigma_f}{\Sigma_a + DB^2}, \qquad (3.5)$$

$$\rho = \frac{k-1}{k}, \qquad (3.6)$$

and

$$l^* = \frac{1}{v\mu\Sigma_f}. \tag{3.7}$$

Equation 3.4 becomes:

$$\frac{d\Phi}{dt} = \frac{\rho}{l^*}\Phi. \tag{3.8}$$

The above equation describes prompt behavior. We now extend the description to include delayed behavior. Firstly, represent the delayed neutron fraction at thermal energies by the symbol $\bar{\beta}$. Then, the delayed neutron source dn_d/dt is

$$\frac{dn_d}{dt} = \bar{\beta}(\mu\Sigma_f\Phi), \tag{3.9}$$

and the prompt neutron source dn_p/dt becomes

$$\frac{dn_p}{dt} = (1 - \bar{\beta})\mu\Sigma_f\Phi, \tag{3.10}$$

where $\mu\Sigma_f\Phi$ is the total neutron source. Assuming six-group precursors, the delayed neutron source rate can also be written as

$$n_d = \sum_{i=1}^{6} \lambda_i C_i(t), \tag{3.11}$$

where λ_i is the decay constant for the ith group precursor and C_i is the ith group precursor density. The latter is determined by

$$\frac{dC_i}{dt} = \bar{\beta}_i\mu\Sigma_f\Phi(t) - \lambda_i C_i(t), \tag{3.12}$$

where the first term on the right side is the ith precursor production rate from fission process and the second term on the right is due to precursor decay. Substitute Equations 3.9, 3.10 and 3.11 into Equation 3.3, we have

$$\frac{1}{v}\frac{d\Phi}{dt} = (1 - \bar{\beta})\mu\Sigma_f\Phi(t) + \sum_{i=1}^{6} \lambda_i C_i - \Sigma_a\Phi(t) - DB^2\Phi(t). \tag{3.13}$$

Upon dividing both sides of the above equation by $\mu\Sigma_f$, it becomes

$$\frac{1}{v\mu\Sigma_f}\frac{d\Phi}{dt} = (1-\bar{\beta})\Phi(t) + \frac{1}{\mu\Sigma_f}\sum_{i=1}^{6}\lambda_i C_i - \frac{\Sigma_a\Phi(t) + DB^2\Phi(t)}{\mu\Sigma_f}. \qquad (3.14)$$

Substituting the definitions for k and l^*, we then have

$$l^*\frac{d\Phi}{dt} = (1-\bar{\beta} - \frac{\Sigma_a + DB^2}{\mu\Sigma_f})\Phi(t) + \frac{1}{\mu\Sigma_f}\sum_{i=1}^{6}\lambda_i C_i. \qquad (3.15)$$

Upon substituting the definition for ρ, moving l^* to the right side and dividing both sides by v, we then have

$$\frac{dn}{dt} = \frac{\rho - \bar{\beta}}{l^*}n(t) + \sum_{i=1}^{6}\lambda_i C_i(t), \qquad (3.16)$$

together with the equation that describes the behavior of delayed neutron precursors as

$$\frac{dC_i}{dt} = \frac{\bar{\beta}_i}{l^*}n(t) - \lambda_i C_i(t) \quad i = 1, 2, ...6, \qquad (3.17)$$

where

$n(t)$ is the amplitude function and is weighted integral of all neutrons present in the core,

ρ is the net reactivity,

$\bar{\beta}$ is the effective delayed neutron fraction,

$\bar{\beta}_i$ is the effective fractional yield of the ith group of delayed neutrons,

l^* Is the prompt neutron lifetime,

λ_i is the decay constant of the ith precursor group,

$C_i(t)$ is the concentration of the ith precursor group.

The fundamental assumption that underlies the validity of the point kinetics approach is that the shape (as distinct from the magnitude) of the neutron flux is

constant. If this is true, then the spatial and temporal behavior of the flux can be separated. Such will be the case for small reactors for which a local perturbation in the neutron flux is rapidly transmitted throughout the entire core. The amplitude of the flux therefore rises or falls uniformly and the flux shape remains constant.

Equations 3.16 and 3.17 are referred to as the six-group point kinetics equations. The precursors are divided into six groups, each with its own yield and decay parameter. It is also standard practice to formulate a one-group model. To do this, an effective multi-group decay parameter is defined as a time-dependent weighted average of the precursors. Thus,

$$\lambda_e(t) = \sum_{i=1}^{6} \lambda_i C_i(t) / \sum_{i=1}^{6} C_i(t). \tag{3.18}$$

The reactor point kinetics equations then become:

$$\dot{n}(t) = (\frac{\rho - \bar{\beta}}{l^*})n(t) + \lambda_e(t)C(t), \tag{3.19}$$

and

$$\dot{C}(t) = \frac{\bar{\beta}_i}{l^*}n(t) - \lambda_e(t)C(t), \tag{3.20}$$

where symbols not previously defined are:

$\lambda_e(t)$ is the effective multi-group decay parameter, and

$C(t)$ is the sum of the precursors in each group.

The value of the multi-group decay parameter varies with the precursor mix. Its value is typically 0.076 inverse seconds at steady state. If reactor power is rising, the short-lived precursor groups dominate and the value of the decay parameters increases. A figure of 0.1 s^{-1} is often assumed. Conversely, if reactor power is decreasing, the long-lived precursor groups dominate and the value of the decay parameter decreases. It is incorrect, but nonetheless common practice, to approximate $\lambda_e(t)$ as a constant. This is normally what is meant by a one-group model.

3.3 Reactor Models

Reactor dynamics are non-linear because the rate of change of reactor power is proportional to the product of reactivity and power. Also, the reactivity is dependent on the reactor power through various power-dependent feedback mechanisms such as fission product poisons (xenon), moderator/coolant temperature, Doppler effect, and fuel depletion. However, if a reactor's operation is limited to some small range about an operating point, then it is acceptable to linearize the point kinetics equations and use the result as the basis of a reactor model.

3.3.1 Linearized Model

Denote the values of the reactor power, precursor concentration, and reactivity at the operating point by n_0, C_0, and ρ_0 respectively and the deviation in power, precursor concentration, and reactivity as δn, δC, and $\delta \rho$ respectively. The six-group point kinetics equations, (Equations 3.16 and 3.17) become:

$$\frac{d(n_0 + \delta n)}{dt} = \frac{[(\rho_0 + \delta \rho) - \overline{\beta}](n_0 + \delta n)}{l^*} + \sum_{i=1}^{n} \lambda_i (C_{i0} + \delta C_i), \qquad (3.21)$$

$$\frac{d(C_{i0} + \delta C_i)}{dt} = \frac{\overline{\beta_i}(n_0 + \delta n)}{l^*} - \lambda_i (C_{i0} + \delta C_i). \qquad (3.22)$$

The deviations in the power, precursor concentration, and reactivity are defined as:

$$\delta n = n - n_0, \qquad (3.23)$$

$$\delta C_i = C_i - C_{i0}, \qquad (3.24)$$

$$\delta \rho = \rho - \rho_0. \qquad (3.25)$$

If the system is at the equilibrium, then

$$\frac{\overline{\beta_i}}{l^*} n_0 = \lambda_i C_{i0}. \qquad (3.26)$$

61

Expanding 3.21 and 3.22, while noting that the derivative of a constant is zero and neglecting second order terms, yields:

$$\frac{d(\delta n)}{dt} = \frac{\rho_0 n_0 + \delta \rho n_0 - \overline{\beta} n_0 + \rho_0 \delta n - \overline{\beta} \delta n}{l^*} + \sum_{i=1}^{n} (\lambda_i C_{i0} + \lambda_i \delta C_i), \qquad (3.27)$$

$$\frac{d(\delta C_i)}{dt} = \frac{\overline{\beta_i} n_0}{l^*} + \frac{\overline{\beta_i} \delta n}{l^*} - \lambda_i C_{i0} - \lambda_i \delta C_i. \qquad (3.28)$$

Substituting 3.26 into 3.27 and 3.28, noting that $\overline{\beta} = \sum_{i=1}^{6} \overline{\beta_i}$ and taking the value of the reactivity at the operating point as zero yields:

$$\frac{d(\delta n)}{dt} = \frac{\delta \rho n_0}{l^*} - \frac{\overline{\beta} \delta n}{l^*} + \sum_{i=1}^{n} \lambda_i \delta C_i, \qquad (3.29)$$

$$\frac{d(\delta C_i)}{dt} = \frac{\overline{\beta_i} \delta n}{l^*} - \lambda_i \delta C_i. \qquad (3.30)$$

Equations 3.29 and 3.30 were for a six-group model. A linearized one-group model can also be derived. It is:

$$\frac{d(\delta n)}{dt} = \frac{\delta \rho n_0}{l^*} - \frac{\overline{\beta} \delta n}{l^*} + \lambda_e \delta C, \qquad (3.31)$$

$$\frac{d(\delta C)}{dt} = \frac{\overline{\beta} \delta n}{l^*} - \lambda_e \delta C. \qquad (3.32)$$

In matrix forms, the linearized six-group point kinetics model and the linearized one-group point kinetics model become:

$$\begin{pmatrix} \delta \dot{n} \\ \delta \dot{C_1} \\ \delta \dot{C_2} \\ \delta \dot{C_3} \\ \delta \dot{C_4} \\ \delta \dot{C_5} \\ \delta \dot{C_6} \end{pmatrix} = \begin{pmatrix} -\frac{\overline{\beta}}{l^*} & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 \\ \frac{\overline{\beta_1}}{l^*} & -\lambda_1 & 0 & 0 & 0 & 0 & 0 \\ \frac{\overline{\beta_2}}{l^*} & 0 & -\lambda_2 & 0 & 0 & 0 & 0 \\ \frac{\overline{\beta_3}}{l^*} & 0 & 0 & -\lambda_3 & 0 & 0 & 0 \\ \frac{\overline{\beta_4}}{l^*} & 0 & 0 & 0 & -\lambda_4 & 0 & 0 \\ \frac{\overline{\beta_5}}{l^*} & 0 & 0 & 0 & 0 & -\lambda_5 & 0 \\ \frac{\overline{\beta_6}}{l^*} & 0 & 0 & 0 & 0 & 0 & -\lambda_6 \end{pmatrix} \begin{pmatrix} \delta n \\ \delta C_1 \\ \delta C_2 \\ \delta C_3 \\ \delta C_4 \\ \delta C_5 \\ \delta C_6 \end{pmatrix} + \begin{pmatrix} \frac{n_0}{l^*} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \delta \rho,$$

$$(3.33)$$

62

and

$$\begin{pmatrix} \delta \dot{n} \\ \delta \dot{C} \end{pmatrix} = \begin{pmatrix} -\frac{\bar{\beta}}{l^*} & \lambda_e \\ \frac{\beta}{l^*} & -\lambda_e \end{pmatrix} \begin{pmatrix} \delta n \\ \delta C \end{pmatrix} + \begin{pmatrix} \frac{n_0}{l^*} \\ 0 \end{pmatrix} \delta \rho. \qquad (3.34)$$

respectively.

3.3.2 Inclusion of Temperature Feedback Effects

The feedback effect that is most likely to be significant in terms of failure detection is that associated with the temperature of the moderator/coolant. This effect occurs for small changes of temperature and it occurs fairly quickly, in the time required for coolant to transit the primary system. Other feedback effects do not meet these criteria. The Doppler effect requires a very large increase in fuel temperature. Fission product poisoning and fuel burnup occur on time scales of hours or longer. Behavior of the moderator/coolant temperature can be modeled as [35]:

$$\frac{dT_k(t)}{dt} = Kn(t) - \alpha(T_k(t) - T_c), \qquad (3.35)$$

where $T_k(t)$ is the temperature of the moderator in the reactor core, T_c is the sink temperature, K is the reciprocal of the mass flow and heat capacity, and α is the reciprocal of the time constant for system heat removal.

3.4 MIT Research Reactor II

A distinguishing and perhaps unique characteristic of the MIT program on reactor control and also of the joint MIT-SNL research has been the performance of on-line experiments. Whenever possible, each control concept has been evaluated experimentally on either the 5-MWt Research Reactor (MITR-II) or the Annular Core Research Reactor (ACRR) that is operated by Sandia National Laboratories, or both.

In this research, experimental data from the reactor provided the source of information to construct a reactor model for simulation. The fault detection technique developed in this research has not been evaluated in a real experimental set up. It is

63

highly desirable if such evaluations could be carried out at some future time. However, simulation studies were performed in which models of the MITR-II were used. Also, a data acquisition capability was established for use in obtaining reactor data for the research reported here.

Here empirical models of MITR-II control mechanism and temperature worth are presented. These models are necessary for carrying out an experimental evaluation of fault detection methods. Frequency analysis of the measurements from normal operations are conducted to gain information on the noise characteristics of the system.

The MITR-II is a 5 MWt, light-water cooled and moderated, heavy-water reflected, tank-type reactor that uses plate-type, uranium-aluminide fuel. The fuel is enriched to 93% U-235. Energy is continuously removed by forced circulation of the primary coolant. The maximum permitted operating temperature is 55 degree C.

The nuclear instrumentation available for control experiments on the MITR-II consists of three neutron flux sensors and a gamma-ray sensor that correlates neutron power with the radioactivity (N-16) of the primary coolant. All four sensors are directly proportional to the power over the range of interest. Measurements are also available of the coolant flow, coolant temperature, and control mechanism position. Four independent measurements of primary coolant flow are obtained from the pressure differences across orifices. Primary coolant temperatures were measured as follows: two sensors for the hot leg, two sensors for the cold leg, and one sensor for the temperature difference between the legs. In effect, three measurements are available for the temperature difference. None of the sensors that form the MITR-II's safety system are used for experiments. Also, it should be noted that the noise and statistical characteristics of the MITR-II's flow, temperature, and neutron flux instrumentation are similar to those in commercial reactors.

3.4.1   Temperature Effect

The temperature in the core is not available directly. Instead, the cold and hot leg temperatures of the primary coolant are available. Thus, one alternative model is proposed here. Consider coolant into and out of the core tank together with the core

as the system of interest. Then, conservation of energy gives

$$C_k \frac{dT_k}{dt} V_k = Kn(t) - \alpha h(t), \tag{3.36}$$

where C_k (in the unit of $J/(m^3 \cdot K)$) is the heat capacity of the core, and V_k (m^3) is the volume of the core. The left side is the energy change rate of the core, and it comes from two contributors. One is the energy deposition from the fission process which can be expressed as Kn (J/s). The other is the energy loss due to heat exchange with the coolant h (J/s). Neglecting heat loss through air or other routes and assuming that what is lost by the core is entirely gained by the coolant, h can also be written as

$$\begin{aligned} h(t) &= \gamma(T_k - T_{in}) \\ &= C_c(T_{out} - T_{in})v = \alpha'\Delta T, \end{aligned} \tag{3.37}$$

where γ is a constant conversion factor, T_{out} is the coolant outlet temperature, T_{in} is the coolant inlet temperature, C_c is the heat capacity of coolant and v (m^3/s) is the coolant column velocity. The ΔT is defined to be $T_{out} - T_{in}$ and $\alpha' = C_c v$ is a constant. From Equation 3.37, we have that the quantity $(T_k - T_{in})$ is proportional to ΔT. Thus, we can write

$$T_k - T_{in} = \gamma'\Delta T, \tag{3.38}$$

with $\gamma' = \alpha'/\gamma$ being a constant. Now come back to Equation 3.36. Note that the derivative of a constant is zero. Here, we assume that the coolant inlet temperature is a constant. Then, we have

$$\begin{aligned} C_k \frac{d(T_k - T_{in})}{dt} V_k &= Kn(t) - \alpha h(t) \tag{3.39} \\ C_k \frac{d(\gamma'\Delta T)}{dt} V_k &= Kn(t) - \alpha\alpha'\Delta T \tag{3.40} \\ \frac{d(\Delta T)}{dt} &= K'n(t) - \zeta\Delta T. \tag{3.41} \end{aligned}$$

65

Here $K' = K/(C_k \gamma' V_k)$ and $\zeta = (\alpha \alpha')/(C_k \gamma' V_k)$ are new constants that can be determined empirically.

Simply using a different notions for the constants K' and ζ, we have our alternative temperature model written as

$$\frac{d(\Delta T(t))}{dt} = Kn(t) - \alpha \Delta T(t). \tag{3.42}$$

Note that the constants K and α are different from those in Equation 3.35. And be reminded that one assumption that we made to arrive at the above equation is that the coolant inlet temperature is constant. This may not be true in reality but it is a good approximation for small power changes or short time observations. The advantage of this model is that the constants can be easily determined experimentally. Experiments were designed and carried out as part of the work reported here on the MITR-II to determine the coefficients in the model.

Firstly, the quantity $\frac{K}{\alpha}$ was determined as the follows. At equilibrium (both neutron and thermal equilibrium), we have $\frac{d\Delta T}{dt}$ equals zero. Thus,

$$\Delta T_0 = \frac{K}{\alpha} n_0. \tag{3.43}$$

By measuring the equilibrium temperature (T_0) at different neutron levels (n_0) , we can estimate the parameter K/α as the slope of the curve of T_0 versus n_0. The results of five measurements are plotted in Figure 3-2. Here the neutron flux is measured in stead of neutron density because of the experiment set up. Also the measurement output has not been calibrated, but it is in a linear relationship with the neutron flux and thus the neutron density. Assume a linear factor of N_v, that is 1 measurement unit is equals to N_v neutron/cm^3. Thus, the fitted result has to be divided by N_v. The fitted slope of the curve in Figure 3-2 is 0.078. Thus, the value K is $0.078/N_v$.

Next, the parameter α can be determined as follows. After a change of neutron power level, the neutron density will attain equilibrium in about 1 minute, while the thermal system takes longer to reach equilibrium. For MITR-II, it is about 5 minutes. During this interval, the ΔT should change exponentially before reaching thermal
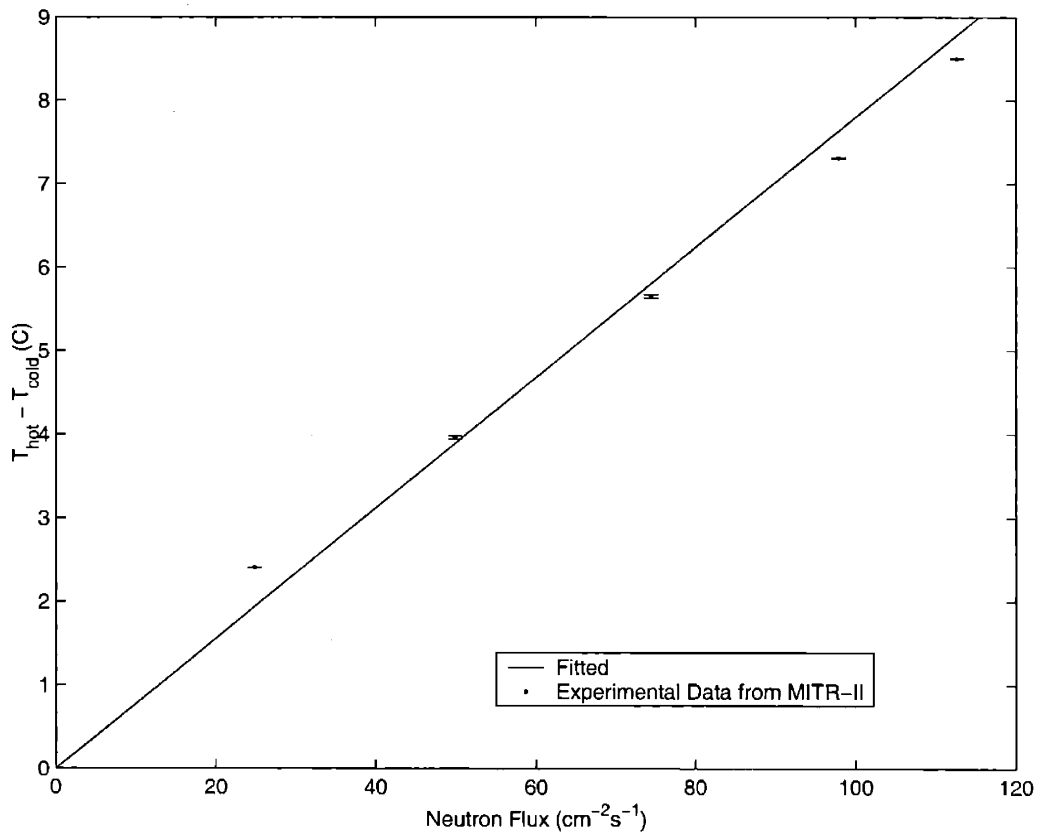
Figure 3-2: Experimental and fitted equilibrium ΔT (°C) vs neutron flux ϕ measurement. The slope is estimated to be 0.078.
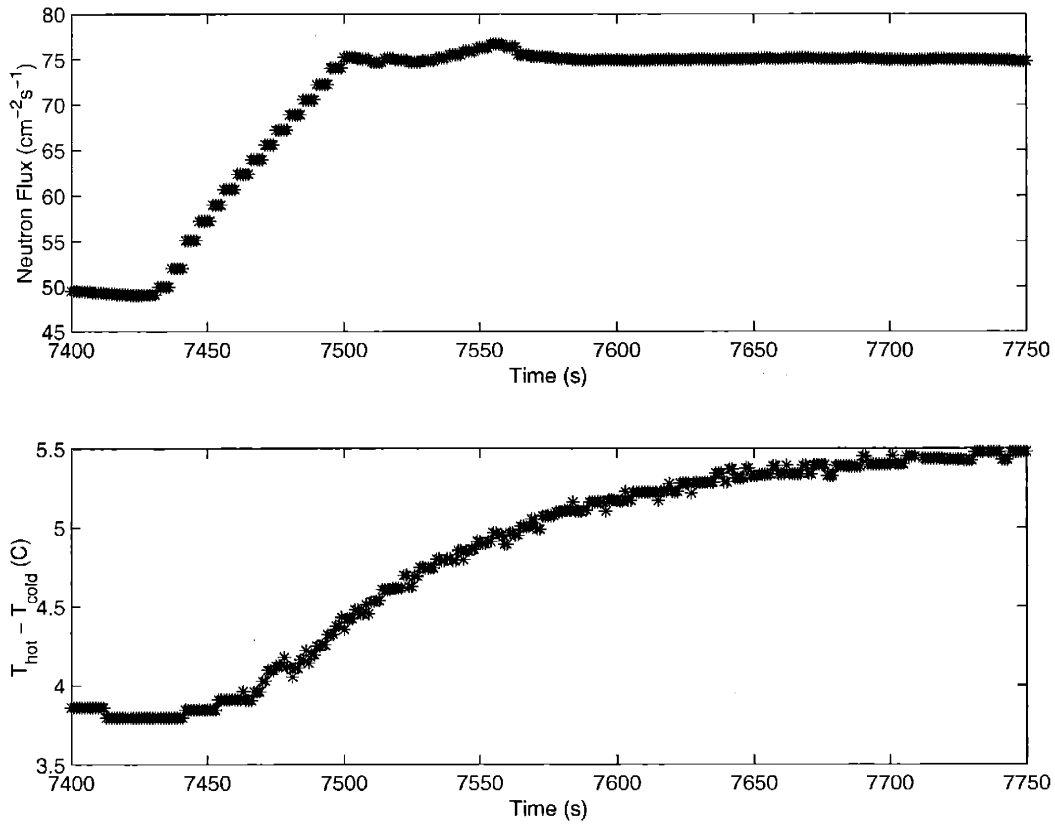
Figure 3-3: The profile of neutron flux (top) and ΔT after a power increase.

equilibrium according to Equation 3.42. Thus, the parameter α can be determined as the slope of the curve of $\log(\Delta T(t) - \Delta T_0)$ versus time. Here ΔT_0 is the target equilibrium temperature. An example of the profiles of neutron flux and ΔT after a power increase is presented in Figure 3-3.

The exponential part of the ΔT response is fitted for the slope. A total of 19 such slopes were obtained for independent power changes. Ten of those were power decreases while nine were power increases. The results are listed in Table 3.2.

Table 3.2: Calculated α from Exponential Interpolation of ΔT Response after Power Changes

α	0.0126	0.0114	0.0094	0.0107	0.0164	0.0108	0.0100
	0.0120	0.0101	0.0102	0.0094	0.0073	0.0058	0.0110
	0.0105	0.0047	0.0106	0.0134	0.0112		
mean	0.0104						
var	6.672e-06						

Thus, we have, for the MITR-II,

$$\frac{d(\Delta T(t))}{dt} = 0.078/N_v n(t) - 0.0104\Delta T(t).$$ (3.44)

So far, we have developed the model to describe the temperature changes (the change of the temperature difference between coolant outlet and inlet). Using this model to describe the effect of coolant/moderator temperature on neutron density has the following advantages over the model in Equation 3.35 from Weaver [35]. First, Weaver used the the core temperature to approximate the coolant/moderator temperature while we use $T_{in} + \frac{1}{2}\Delta T$. This alternative method is both more accurate and easier for experimental implementation. Second, Newton's law of cooling that Weaver adopted could lead to an incorrect result in that T_k could be less than the sink temperature. This violation of the second law of thermodynamics would lead to oscillatory behavior in the model. By describing the temperature difference between coolant outlet and inlet instead of the core temperature, this problem is avoided. The assumption in the alternative model that the coolant inlet temperature is constant is valid and realizable.

The relationship of the coolant temperature and the reactivity worth can be modeled as a linear function[35]. If we chose the operating point to be $(\Delta T_0, n_0)$, then Equation 3.42 becomes

$$\frac{d(\Delta T_0 + \delta(\Delta T))}{dt} = K(n_0 + \delta n) - \alpha(\Delta T_0 + \delta(\Delta T)).$$ (3.45)

Expressing the relationship between the average coolant temperature $\overline{T_c}$ and ΔT as

$$\overline{T_c} = T_{in} + \frac{1}{2}\Delta T,$$ (3.46)

the reactivity can be redefined to allow for temperature effects. For a small power variation around a equilibrium operating point, instead of requiring the ρ_0 be zero, we require that the quantity $(\rho_0 - \gamma'\overline{T_{c0}})$ be zero. Here, γ' is a constant. Hence, the

deviation in reactivity becomes:

$$\delta\rho = \rho - \rho_0 + \gamma'\delta(\overline{T_c}) \tag{3.47}$$

$$= \rho - \rho_0 + \gamma\delta(\Delta T), \tag{3.48}$$

and the equilibrium condition is

$$\rho_0 - \gamma\delta(\Delta T_0) + \zeta = 0. \tag{3.49}$$

Both γ and $\eta = \gamma T_{in}$ are constants that can be determined experimentally. The latter is different for different coolant inlet temperatures.

Thus, in matrix form, the one-group model including temperature feedback can be written as:

$$
\begin{pmatrix} \delta\dot{n} \\ \delta\dot{C} \\ \delta(\dot{\Delta T}) \end{pmatrix} =
\begin{pmatrix} -\overline{\beta}/l^* & \lambda_e & -\gamma n_0/l^* \\ \overline{\beta}/l^* & -\lambda_e & 1 \\ K & 0 & -\alpha \end{pmatrix}
\begin{pmatrix} \delta n \\ \delta C \\ \delta(\Delta T) \end{pmatrix} +
\begin{pmatrix} n_0/l^* \\ 0 \\ 0 \end{pmatrix} \delta\rho.
$$

A similar matrix form can be written for the six-group model that includes this temperature feedback effect.

Above, we have used a linear model for the reactivity worth of the temperature effect (Equation 3.47). Again, this is valid only for small variations. We also have a non-linear empirical model for the wide range of coolant temperature worth for the MITR-II. The temperature feedback worth curve has been measured at the MITR-II and is displayed in Figure 3-4. Here, the temperature is defined as the average of the primary coolant outlet temperature and the moderator outlet temperature in the units of degree C while the reactivity worth is in the units of $m\beta$. The experimental data was fitted, as part of the work performed here, using a fourth-order polynomial with the fitting parameters as [0.0001 -0.0173 0.5002 -4.7914 12.0944]. Thus, we have

$$\rho_T = 0.0001T^4 - 0.0173T^3 + 0.5002T^2 - 4.7914T + 12.0944. \tag{3.50}$$
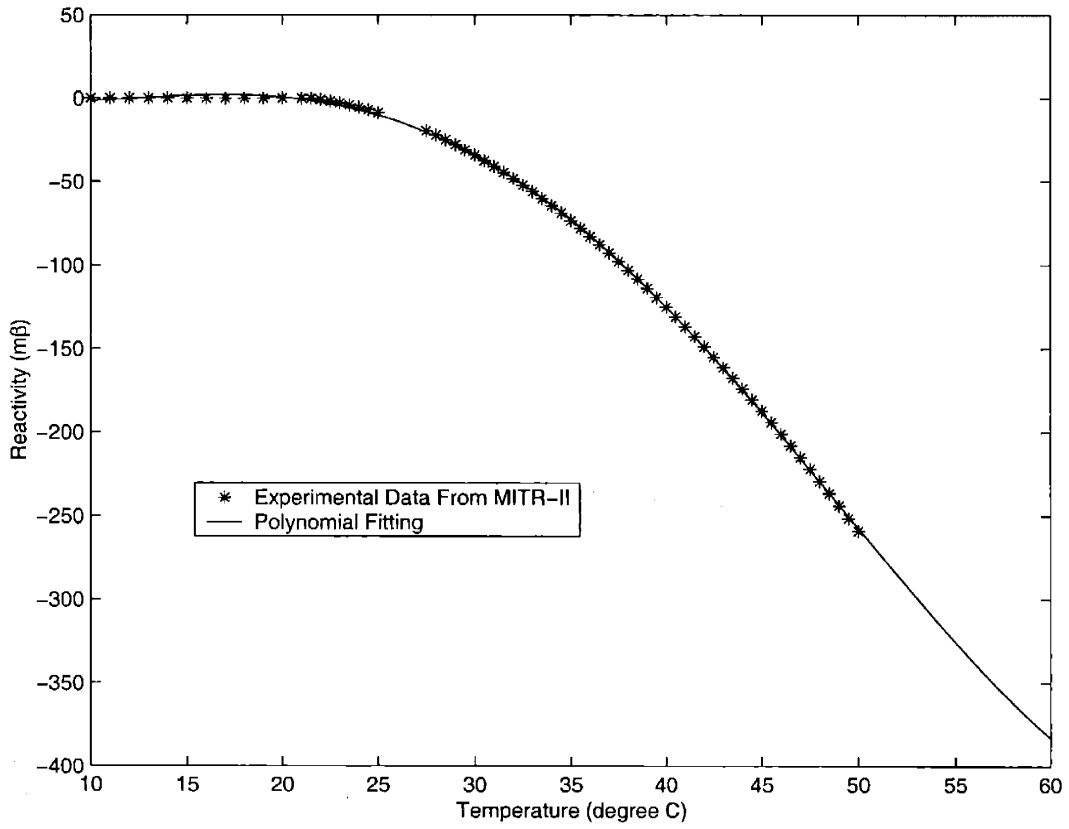
70

Figure 3-4: Temperature feedback worth curve for MIT Research Reactor-II.

3.4.2 Control Mechanism

Coarse control of the power in the MITR-II is achieved by positioning a bank of six shim blades. Once critical, the neutron flux is normally maintained constant by adjusting the position of a fine-control regulating rod. Both the regulating rod and the shim blades are available for experimental research. Each is normally moved at a fixed speed of 4.25 inches per-minute. However, each can be specially equipped with a variable speed stepping motor so that the rate of change of reactivity can be made to vary as specified by the control laws. The minimum allowed period on the MITR-II are 50-seconds steady and 30-seconds dynamic. There is a negative coefficient of reactivity associated with the fuel, coolant, and reflector temperatures. However, its magnitude is rather small, averaging $-8 \times 10^{-5} \Delta K / K / C$. The MITR-II's effective delayed neutron fraction and prompt neutron lifetime are $0.00786 \Delta K / K$ and 1200 microseconds, respectively.

The relationship between the regulating rod height and shim bank height and their respective reactivity worths are empirical models based on measurements done at MITR-II. Because the absorption material (boron) contained in these control devices burns up during operation, these models are continuously changing. However, they can be assumed to be stable for times in the range of weeks or even months. The measurements for these worth curves are standard calibration procedures at MITR-II and are performed once a year.

The regulating rod worth curve in Figure 3-5 was obtained from MITR-II on May 30, 2000 and the shim bank worth curve in Figure 3-6 was obtained from MITR-II on May 31, 2000. The data were fitted, as part of the work performed here, using fifth-order polynomials. For the regulating rod, the fitting results are:

$$\rho_{RR} = -0.0001 H_{RR}^5 - 0.0013 H_{RR}^4 + 0.0878 H_{RR}^3 - 3.4588 H_{RR}^2 + 48.3061 H_{RR} - 2.4395,$$
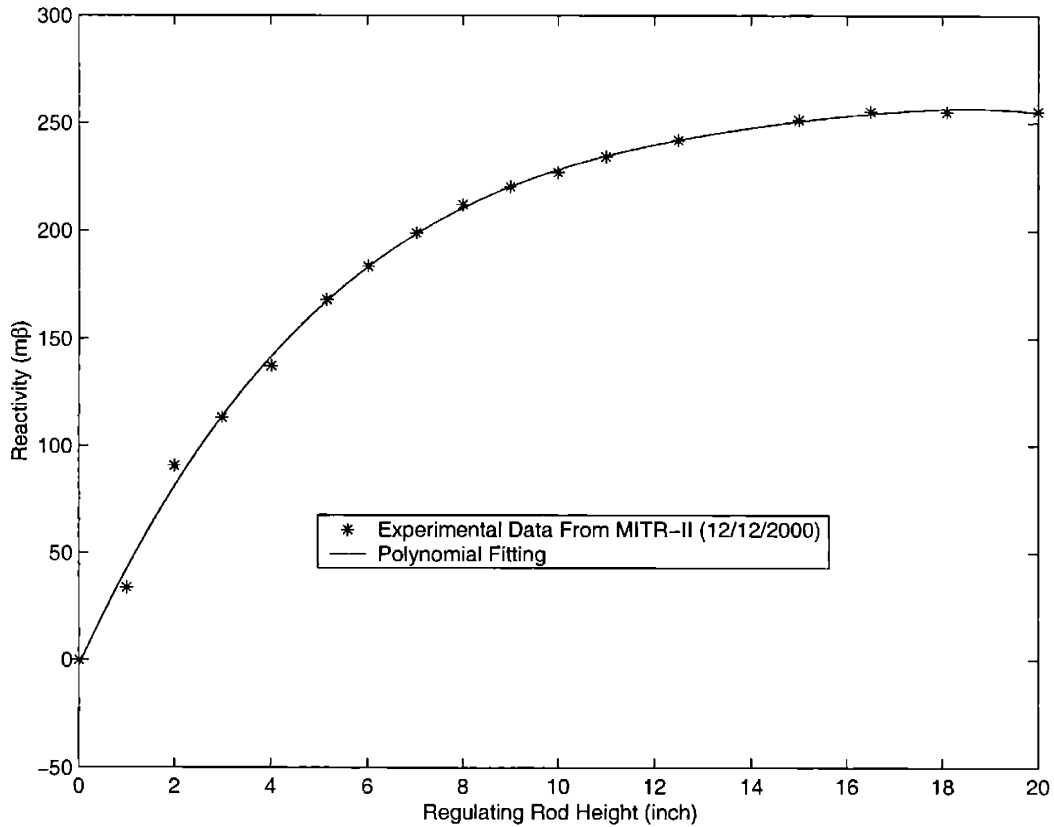
$$(3.51)$$

Figure 3-5: Reactivity worth curve of the regulating rod for MITR-II.

and for shim bank, the results are:

$$\rho_{SB} = 0.0064 H_{SB}^5 - 0.2437 H_{SB}^4 + 0.3426 H_{SB}^3 - 47.9409 H_{SB}^2 + 459.1145 H_{SB} - 78.0678.$$

$$(3.52)$$

3.4.3 Frequency Analysis

Equilibrium data for MITR-II was obtained on May 05 2001 with the thermal power at about 4.85 MW. The time domain and frequency domain data are shown in Figure 3-7.

The variance in the time domain is 0.1209. The integration of the power spectral density (PDS) in the frequency domain is 0.7595, which is 2π times the variance of 0.1209.

In figure 3-8, data is shown that was obtained from MITR-II on April 21, 2001,
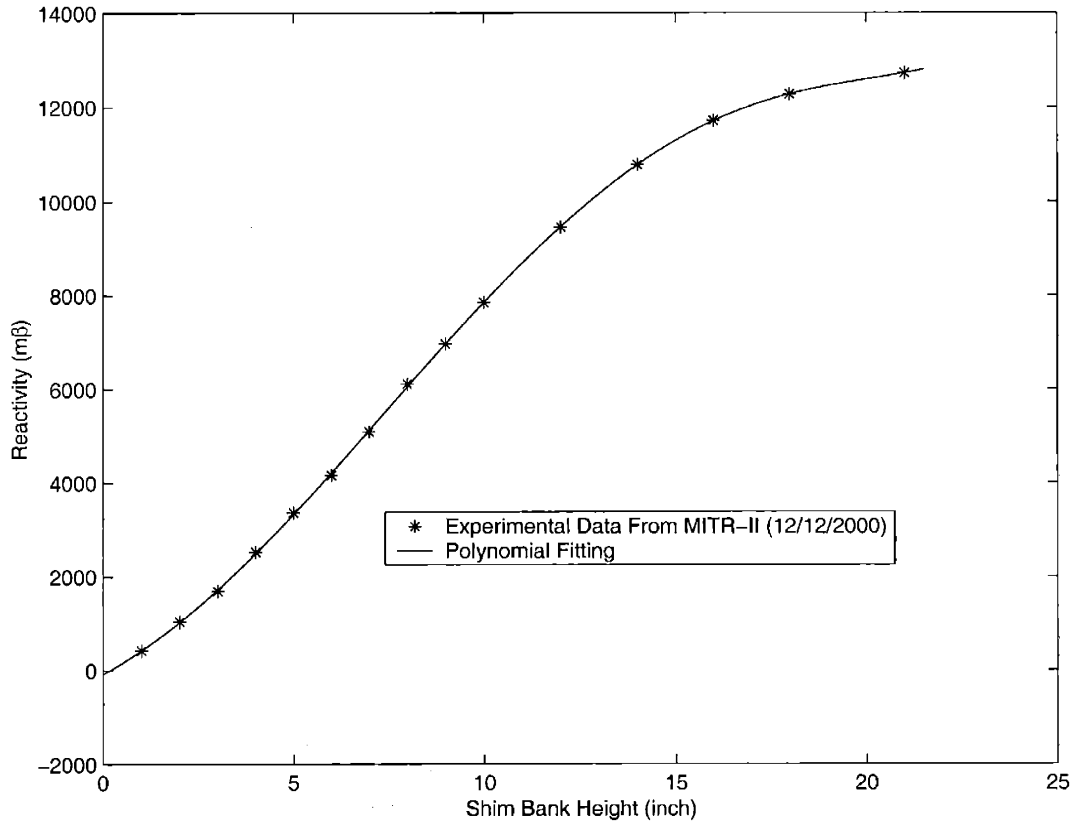
73

Figure 3-6: Reactivity worth curve of the shim bank for MITR-II.
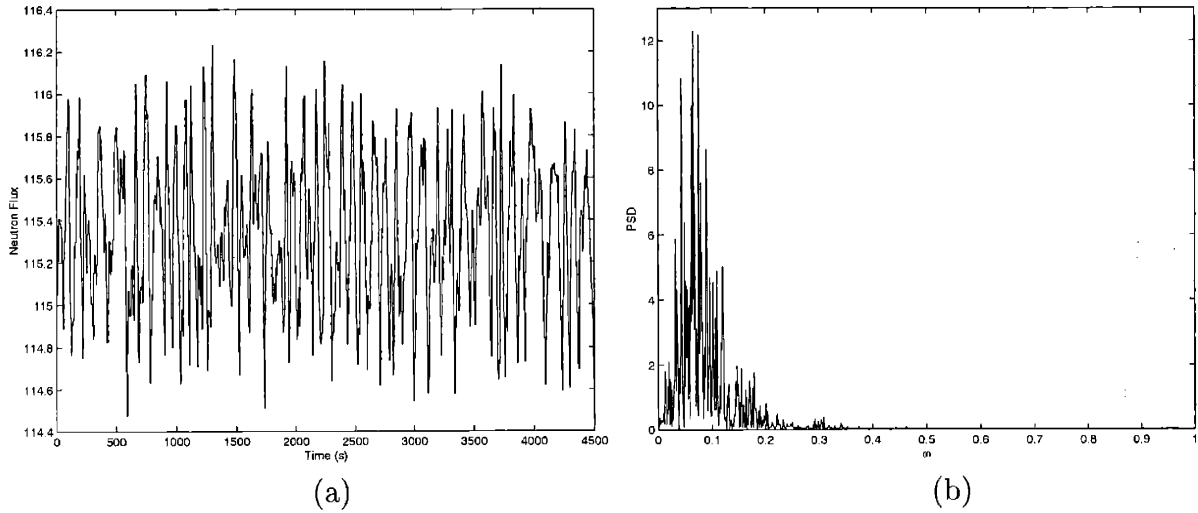

Figure 3-7: (a) Neutron flux vs. time. (b) Frequency domain response obtained from data in (a). The raw data was obtained from MITR-II on May 08, 2001. Reactor is at equilibrium with thermal power at about 4.85 MW.

Figure 3-8: Collected data from MITR-II on April 21, 2001.
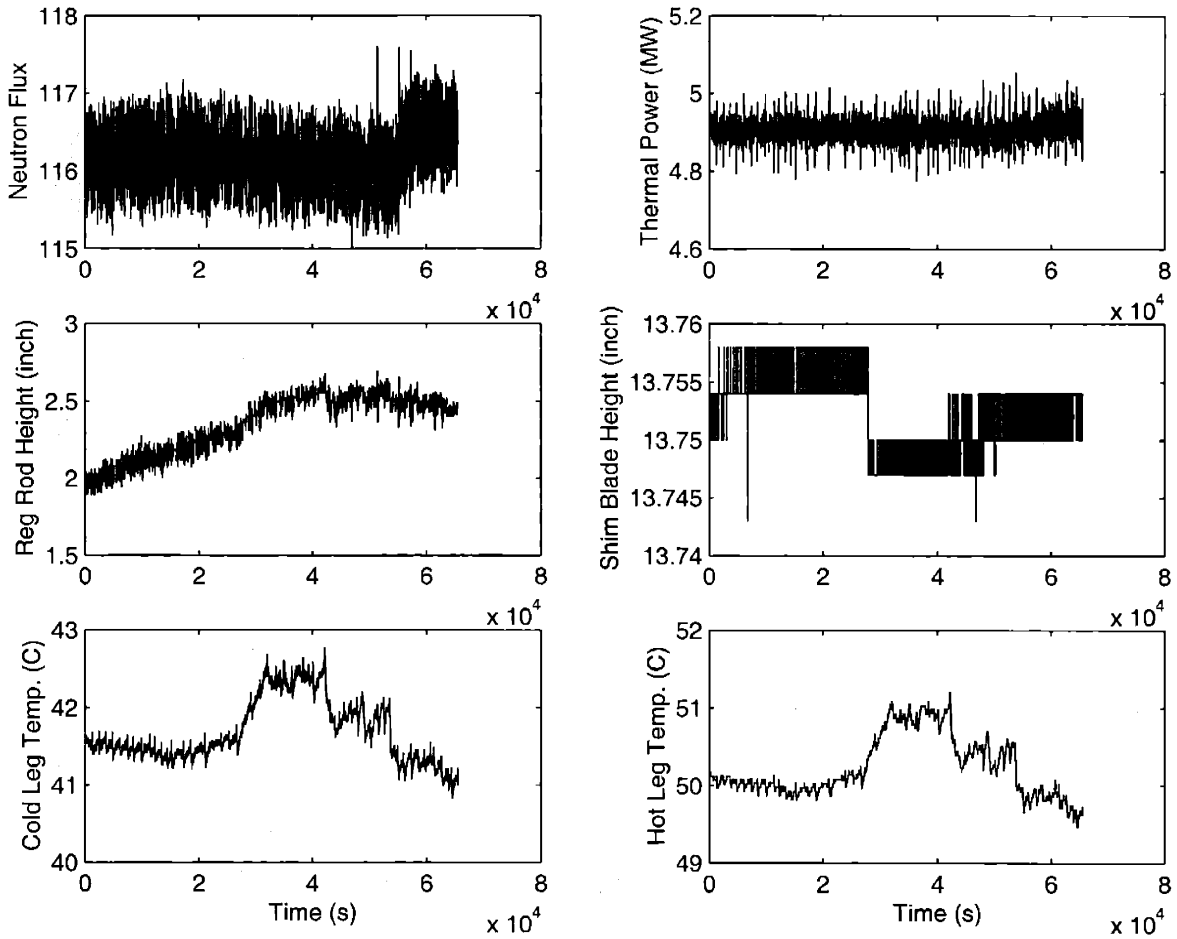
including neutron flux (top left), thermal power (top right), regulating rod position (middle left), shim blade position (middle right), primary cold leg temperature (bottom left), and primary hot leg temperature (bottom right). Figure 3-9 shows the power spectrum of the neutron flux and thermal power. Figure 3-10 shows the filtered power spectrum of the Figure 3-9.

Figure 3-9: (a) Power spectrum of neutron flux (b) Power spectrum of thermal power. The raw data was obtained from MITR-II on April 21, 2001. Reactor was at equilibrium with thermal power at about 4.85 MW.
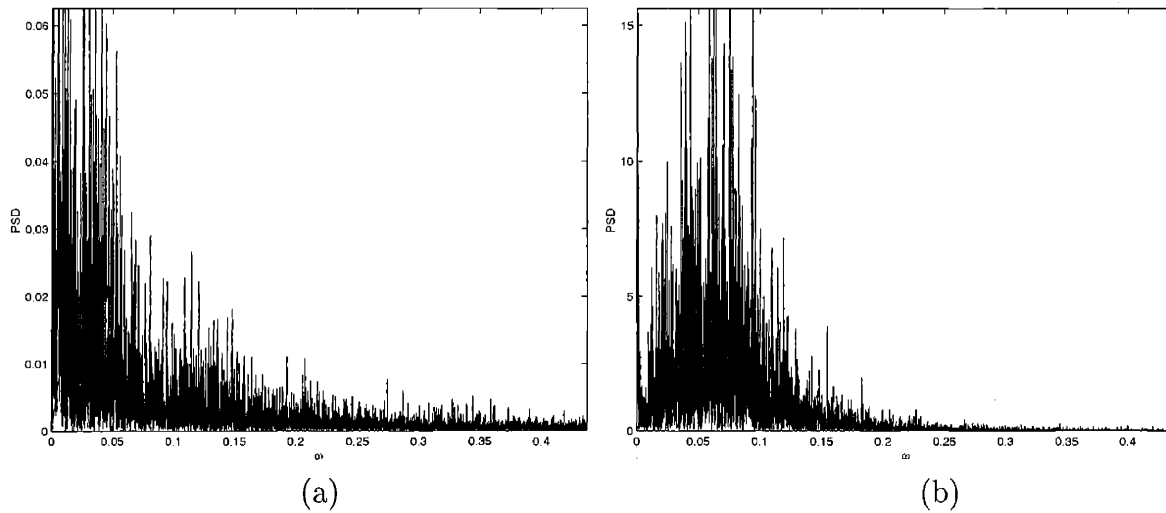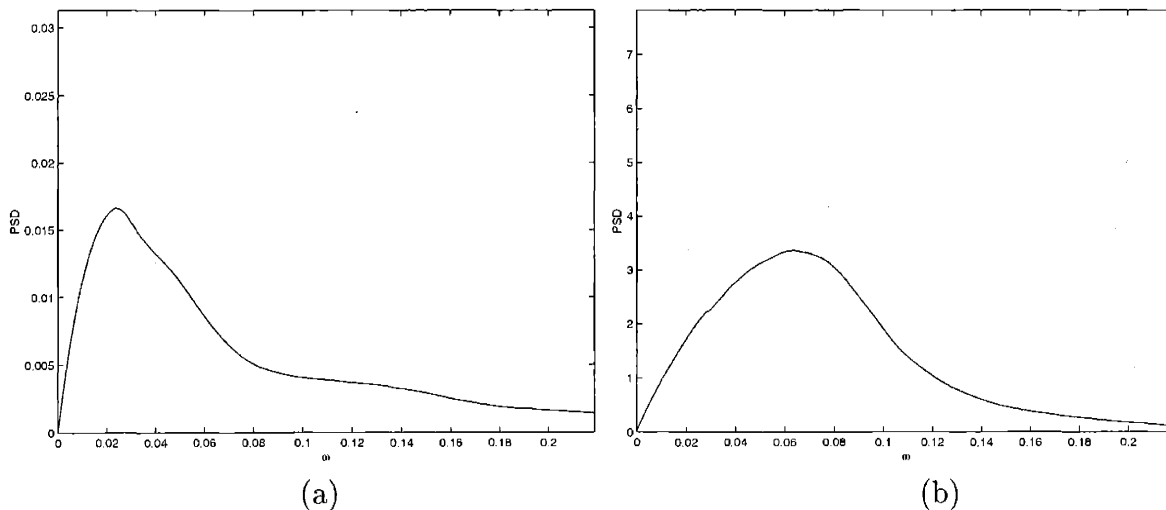

Figure 3-10: (a) Filtered power spectrum of neutron flux (b) Filtered power spectrum of thermal power. The raw data was obtained from MITR-II on April 21, 2001. Reactor was at equilibrium with thermal power at about 4.85 MW.

Chapter 4

Design of Optimal Observer-Based Fault Detector

In this chapter, a method for performing the fault detection based on observer and for finding the optimal gain for the observer-based fault detection is proposed. The objective is to extend Beard's method to address the problem of noise.

The fundamental idea for the optimal designing is optimal decision theory. This approach is necessary because of two conflicting requirements. On the one hand, we want to decrease the probability that a fault is not detected when it occurs. On the other hand, the probability that a false alarm is indicated when no fault has occurred should also be kept low. A compromise has to be made based on the respective costs of the two goals. In the following, a non-linear cost function, which is a function of the detector gain matrix L and the decision rule, is constructed. By minimizing such a cost function, the desired compromise for assurance of fault detection without excessive instances of false alarms is achieved.

This chapter is organized as follows. First, the basics of optimal decision theory are reviewed. Based on the results of optimal decision theory, a simple threshold decision rule is established. The value for the threshold is constructed conditioned on the choices of other degrees of freedom, namely the gain matrix L and a post-process filter frequency a. The objective is to minimize the total average cost due to both false alarms and non-detection.

Now we have our performance criteria, which are the false alarms rate and the delay in detection of a fault, and we want to use numerical optimization to design out filters. How can be bring them together? We need quantitative statistical descriptions for the two performance criteria. Here I propose that the false alarm rate is described by the first hitting time of the threshold by the random signal that contains noise only, and the speed of detecting a fault is described by the first hitting time of the threshold by the random signal that contains both noise and the fault signal. Then we can construct cost functions that take these two quantities as input and gives the cost for the two events – false alarm and delay in detecting the fault. Once we have the cost function, optimization can be carried out to obtain the optimal design for a given performance criteria.

To implement the proposed method, the problem for estimating the average first hitting time of the threshold level by a random process has to be solved. Two methods are introduced to estimate this time. One is the Fokker-Plank equation which gives an analytical solution. The other is Monte Carlo simulation. The derivation of the Fokker-Plank equation and its solution is reviewed and the results from solving the Fokker-Plank equation are compared with those from Monte Carlo simulations. The pros and cons of these two methods are also briefly discussed. In implementing the solution to the Fokker-Plank equation, frequency domain analysis is performed to extract the effective system response time and the signal variance. Finally, the issues that are involved in constructing cost functions are discussed. For this part, one has to rely on his/her knowledge of the prior distribution of the fault occurrence, system operation, and common sense. There is no right or wrong for a cost function. One can only say if it is reasonable.

4.1   Optimal Decision Theory

In this section the optimal decision theory is reviewed.

For a two choice situation, there are a total of four possible outcomes. Let us denote the two choices as H_0 and H_1. For example they can stand for the choices as:

H_0 : there is no fault; H_1 : there is a fault. Denote the truth as H and our judgment of the situation as $\hat{H}$.

Based on observations y, a decision $(\hat{H})$ is made such that either H_0 is true or H_1 is true. The four types of outcomes are:

1. The decision is that there is no fault and the fact is that there is no fault, whose probability can be written as $P(\hat{H} = H_0|H = H_0)$.

2. The decision is that there is no fault and the fact is that there is fault, whose probability can be written as $P(\hat{H} = H_0|H = H_1)$.

3. The decision is that there is a fault and the fact is that there is a fault, whose probability can be written as $P(\hat{H} = H_1|H = H_1)$.

4. The decision is that there is a fault and the fact is that there is no fault, whose probability can be written as $P(\hat{H} = H_1|H = H_0)$.

A cost can be assigned to each outcome and the objective is to minimize the expectation of the total cost. Assigning costs C_{00}, C_{01}, C_{11} and C_{10} to outcomes 1, 2, 3, and 4 respectively. The cost function is then written as:

$$C(\hat{H} = H_i, H = H_j) = C_{ij}, \quad i = 0, 1; \quad j = 0, 1. \tag{4.1}$$

The average cost $J(f)$, which is also called the "Bayes risk', is defined as,

$$J(f) = E\left[C(f(y), H)\right], \tag{4.2}$$

where H is the fact, and $f(y)$ is a decision function of observation y. The decision $\hat{H} = f(y)$ can be either H_0 or H_1. The function E gives the expectation value over both y and H. The optimal decision is then one that minimizes the average cost:

$$\hat{f} = arg \min_f J(f). \tag{4.3}$$

Thus, the average cost when we make a decision $\hat{H} = H_0$ is

$$E\left[C(H_0, H)\right] = C_{01}P(H = H_1|y) + C_{00}P(H = H_0|y), \qquad (4.4)$$

while the average cost when we make a decision $\hat{H} = H_1$ is

$$E\left[C(H_1, H)\right] = C_{11}P(H = H_1|y) + C_{10}P(H = H_0|y). \qquad (4.5)$$

Thus, the optimal decision rule f(y) is

$$\hat{H} = H_0,$$

$$if \quad C_{01}P(H = H_1|y) + C_{00}P(H = H_0|y) \quad < \quad C_{11}P(H = H_1|y) + C_{10}P(H = H_0|y);$$

$$\hat{H} = H_1,$$

$$if \quad C_{01}P(H = H_1|y) + C_{00}P(H = H_0|y) \quad > \quad C_{11}P(H = H_1|y) + C_{10}P(H = H_0|y).$$

Using Bayes' Rule,

$$P(H = H_i|y) = \frac{P_{y|H}(y|H_i)P_i}{P_{y|H}(y|H_1)P_1 + P_{y|H}(y|H_2)P_2}. \qquad (4.6)$$

If we substitute Eq.4.6 into Eq.4.6 and rearrange terms, we have

$$L(y) = \frac{P_{y|H}(y|H_1)}{P_{y|H}(y|H_2)} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\underset{<}{>}}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} = \eta. \qquad (4.7)$$

This rule is called the likelihood ratio test.

In reality, there is usually no cost for outcomes 1 and 3, because we made the correct judgment. Then C_{00} and C_{11} are zero.

Suppose the distribution of y given H_0: $P_{y|H}(y|H_0)$), and H_1: $P_{y|H}(y|H_1)$ are as depicted in Figure 4-1. Note that they do not necessarily have the same shape. If we assume that $P_0 = P_1$, and $C_{10} = C_{01}$ then the likelihood ratio test gives that there is a cutoff point. For y above this value, choose $\hat{H}(y) = H_1$; For y below this point, choose $\hat{H}(y) = H_0$. The cutoff point is where the two distributions $(P_{y|H}(y|H_0)$ and
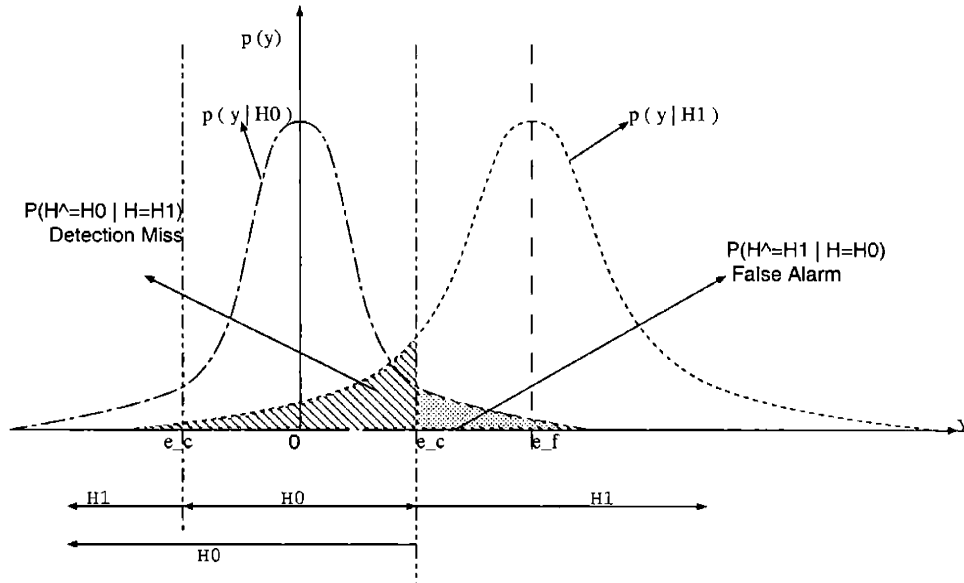
Figure 4-1: Diagram of probability distribution curves.

$P_{y|H}(y|H_1))$ intersect. But, in most cases, the two á priori probabilities may not be equal. Neither are the two costs C_{10} and C_{01}. Thus, a criterion has to be chosen based on the knowledge or assumptions of the relations of $P_{y|H}(y|H_0)$ and $P_{y|H}(y|H_1)$, C_{10} and C_{01}. Then, the probabilities of the two types of decision errors, i.e. the probability for failure to detect a fault $P(\hat{H} = H_0|H = H_1)$ and the probability for false alarm $P(\hat{H} = H_1|H = H_0)$, are the two shadowed areas in Figure 4-1 respectively. Moving the threshold e_c, the area of the two regions change accordingly. Thus, an e_c has to be chosen to balance these two competing factors such that a preset goal can be reached. If we have control of the distributions $P_{y|H}(y|H_0)$ or $P_{y|H}(y|H_1)$, then by narrowing both or either distribution widths, the two types of decision error can be decreased at the same time. Assume that the peak value is the signal and that the spread is caused by noise. Then, by increasing the signal-to-noise ratio, the two types of decision error can be decreased.

These results are the guidelines for designing the observer and fault detection filter. Our final signal is the estimation error which is the difference between the measurements and the estimated measurements. But this signal is always corrupted by noise and disturbances. The signal should be centered around zero if there is no fault, while it would be non-zero if there is a fault. Measurements are taken to

81

minimize the two kinds of mis-judgments, failure to detect a fault and false alarm.

To estimate the variance of the output signal, i.e. the spread of the peaks in figure 4-1, we can use frequency domain analysis, which is reviewed in the following section.

4.2 Frequency Domain Analysis

In this section, frequency domain analysis is performed. This tool is important in designing a noise rejecting filter gain. Analysis for noise rejection in the fault detection design is investigated.

Figure 4-2 is the block diagram of a system and its observer with noise and disturbance added. Noise is added to the measurement and usually has high frequency. Examples of noise include instrument noise, electronic noise and so on. Disturbance is added to the y signal and usually has lower frequency. If the observer is used for estimating state variables for a controller, the effects of noise and disturbance on the system output y would be different. For fault detection purposes, the signal that we are interested is the estimation error ϵ. The relationship of the noise or disturbance to the signal ϵ can be derived from figure 4-2 as

$$\mathbf{E}(s) = \left[I - C(sI - (A - LC))^{-1}L \right] \mathbf{D}(s). \tag{4.8}$$

Thus, both the noise and disturbance have the same transfer function. $\mathbf{W}(s)$ in the form of

$$\mathbf{W}(s) = \left[I - C(sI - (A - LC))^{-1}L \right]. \tag{4.9}$$

Thus, we will discuss only the rejection of noise. The same method can be used for rejection of disturbances.

The process for fault detection in frequency domain is expressed in figure 4-3. $D(s)$ is the noise or disturbance, $W(s)$ is the transfer function from $D(s)$ to $E(s)$. The latter is the estimation error. $F(s)$ is the fault signal, which is zero when there is no fault. $G(s)$ is the corresponding transfer function from $F(s)$ to $E(s)$. $W(s)$ and $G(s)$ are given in equations 4.9 and 2.64 respectively. Thus, $E(s)$ has two components.
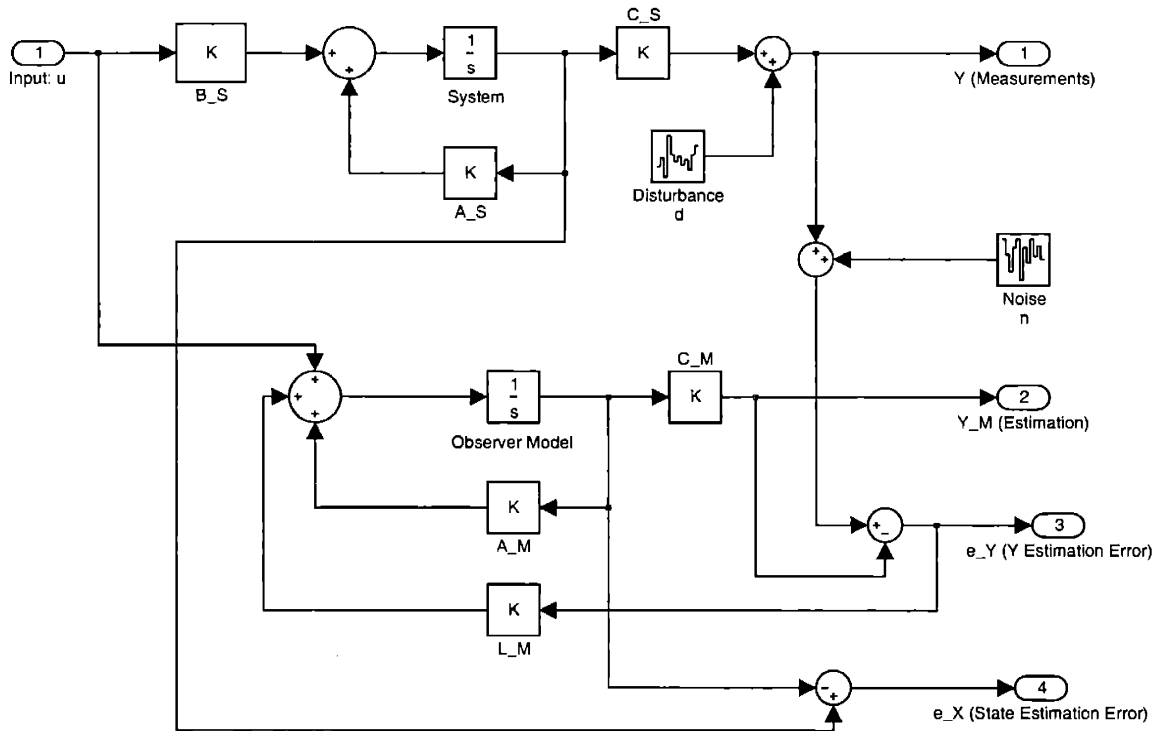
Figure 4-2: Block diagram of a system with noise and disturbance and its observer.

One is from the fault, the signal that we want to detect. The other is from the noise. Both $W(s)$ and $G(s)$ depend on the gain matrix L of the observer. The signal $E(s)$ is then passed through a filter that has a parameter a. The reason to add a filter after $E(s)$ is to ensure that the signal will have finite power. The post-processing filter parameter a is another parameter that we need to design in order to achieve desired goals. Lastly, the filtered signal is passed through a decision rule. In most cases, optimal decision theory gives a simple threshold decision rule as explained in the previous section. However, the threshold may not always be constant in time. Hence, the last step has to be performed in time domain. The threshold value is one more quantity that needs to be designed.

As shown in the previous section, the variance of the signal will affect the decision rule and the resulting judgment error. The variance in the time domain is connected to the power density in the frequency domain by Rayleigh's theorem:

$$\int_{-\infty}^{\infty} [e(t)]^2 = \int_{-\infty}^{\infty} |E(\omega)|^2 d\omega/(2\pi). \tag{4.10}$$
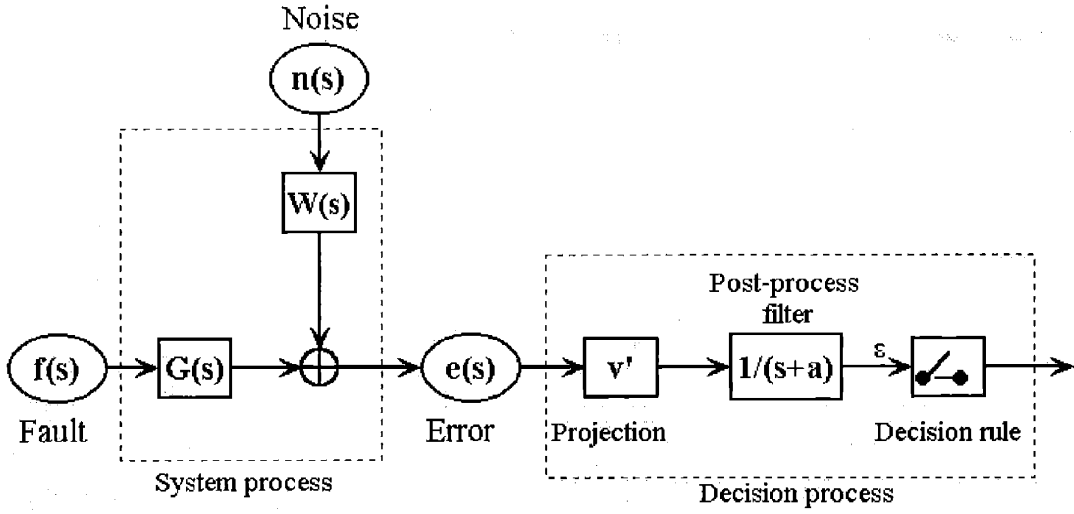
Figure 4-3: Block diagram of the fault detection process in frequency domain.

Thus, we can derive the relationship between the variance of $e(t)$ and the gain L and the a using frequency domain analysis.

Assume the ith eigenvalue of $(A - LC)$ to be λ_i, upon performing the eigenvalue de-composition, we have

$$sI - (A - LC)^{-1} = \sum_{i=1}^{n} \tilde{\mathbf{v}}_i \frac{1}{s - \lambda_i} \tilde{\mathbf{w}}'_i, \tag{4.11}$$

where $\tilde{\mathbf{v}}_i$ and $\tilde{\mathbf{w}}_i$ are the ith left and right eigenvectors respectively. Then,

$$\mathbf{W}(s) = I - \sum_{i=1}^{n} C\tilde{\mathbf{v}}_i \frac{1}{s - \lambda_i} \tilde{\mathbf{w}}'_i L \tag{4.12}$$

$$= I - \sum_{i=1}^{n} \mathbf{v}_i \frac{1}{s - \lambda_i} \mathbf{w}'_i, \tag{4.13}$$

where the vectors $\mathbf{v}_i$ and $\mathbf{w}_i$ are defined as

$$\mathbf{v}_i = C\tilde{\mathbf{v}}_i, \tag{4.14}$$

$$\mathbf{w}'_i = \tilde{\mathbf{w}}'_i L. \tag{4.15}$$

For the general case, $E(s)$ is a n-component vector. The αth component of $E(s)$

is

$$E_\alpha(s) = D_\alpha(s) - \sum_{i=1}^{n} v_{i\alpha} \frac{1}{s - \lambda_i} \sum_{\beta=1}^{n} w_{i\beta} D_\beta(s). \tag{4.16}$$

The notation here is that α and β are the indices to denote the number of the components of a vector, while i is the index to denote the number of the eigenvector. After the frequency dependent filter, $E(s)$ becomes $\tilde{\mathbf{E}}(s)$:

$$\tilde{\mathbf{E}}(s) = \mathbf{E}(s) \frac{1}{s + a}, \tag{4.17}$$

where a is a positive real number. Substitute equation 4.16 into equation 4.17 and the power spectral density of $\tilde{E}(j\omega)$ can be obtained by:

$$psd = \tilde{E}_\alpha(j\omega)\tilde{E}_\alpha^*(j\omega) = \frac{1}{a^2 + \omega^2} \cdot \left[D_\alpha - \sum_{i=1}^{n} v_{i\alpha} \frac{1}{j\omega - \lambda_i} \sum_{\beta=1}^{n} w_{i\beta} D_\beta \right] \cdot$$
$$\left[D_\alpha^* - \sum_{i'=1}^{n} v_{i'\alpha}^* \left(\frac{1}{j\omega - \lambda_{i'}}\right)^* \sum_{\beta'=1}^{n} w_{i'\beta'}^* D_{\beta'}^* \right]. \tag{4.18}$$

Assume that the noise vector consists of independent white noise sources. Therefore, we have

$$\overline{D_i(j\omega)D_j^*(j\omega)} = d_i \delta(i - j), \tag{4.19}$$

and

$$\overline{D_i(j\omega)} = 0. \tag{4.20}$$

Thus, the psd becomes:

$$psd(j\omega) = \overline{\tilde{E}_\alpha(j\omega)\tilde{E}_\alpha^*(j\omega)}$$
$$= \frac{d_\alpha}{a^2 + \omega^2} \cdot \left[1 - \sum_{i=1}^{n} \frac{v_{i\alpha}w_{i\alpha}}{j\omega - \lambda_i} - \sum_{i'=1}^{n} \left(\frac{v_{i'\alpha}w_{i'\alpha}}{j\omega - \lambda_{i'}}\right)^* \right]$$
$$+ \frac{1}{a^2 + \omega^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} v_{i\alpha}v_{i'\alpha}^* \frac{1}{j\omega - \lambda_i} \left(\frac{1}{j\omega - \lambda_{i'}}\right)^* \sum_{\beta=1}^{n} w_{i\beta}w_{i'\beta}^* d_\beta. \tag{4.21}$$

The integral of psd along ω from $-\infty$ to ∞ can be calculated by using the Cauchy Integral Formula for complex functions [36]. If C is a contour, $I(C)$ is region included

by C and $E(C)$ is the region outside $I(C)$, and the $p(z)/q(z)$ is a rational function for which all the zeros of $q(z)$ are inside $E(C)$, then we have

$$\frac{1}{2\pi j} \oint_C \frac{p(z)}{q(z)} \frac{dz}{z - \alpha} = \begin{cases} p(\alpha)/q(\alpha) & \alpha \in I(C) \\ \\ 0, & \alpha \in E(C). \end{cases} \qquad (4.22)$$

Using the above equation, we have

$$\int_{-\infty}^{\infty} \frac{d_\alpha}{a^2 + \omega^2} d\omega = \frac{\pi d_\alpha}{a}. \qquad (4.23)$$

Because the eigenvalues of $(A - LC)$ are all at left half plane, that is $Re(\lambda_i) < 0$, thus

$$\int_{-\infty}^{\infty} \frac{d_\alpha}{a^2 + \omega^2} \sum_{i=1}^{n} \frac{v_{i\alpha} w_{i\alpha}}{j\omega - \lambda_i} d\omega = -\sum_{i=1}^{n} \frac{\pi d_\alpha v_{i\alpha} w_{i\alpha}}{a^2 + a\lambda_i} = -\frac{\pi d_\alpha}{a} \sum_{i=1}^{n} \frac{v_{i\alpha} w_{i\alpha}}{a + \lambda_i}, \qquad (4.24)$$

and

$$\int_{-\infty}^{\infty} \frac{d_\alpha}{a^2 + \omega^2} \sum_{i'=1}^{n} \left(\frac{v_{i'\alpha} w_{i'\alpha}}{j\omega - \lambda_{i'}} \right)^* d\omega = -\frac{\pi d_\alpha}{a} \sum_{i=1}^{n} \left(\frac{v_{i\alpha} w_{i\alpha}}{a + \lambda_i} \right)^* + 2\pi d_\alpha \sum_{i=1}^{n} \frac{v_{i\alpha}^* w_{i\alpha}^*}{a^2 - \lambda_i^{*2}}, \qquad (4.25)$$

and

$$\int_{-\infty}^{\infty} \left[\frac{1}{a^2 + \omega^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} v_{i\alpha} v_{i'\alpha}^* \frac{1}{j\omega - \lambda_i} \left(\frac{1}{j\omega - \lambda_{i'}} \right)^* \sum_{\beta=1}^{n} w_{i\beta} w_{i'\beta}^* d_\beta \right] d\omega$$

$$= \sum_{i=1}^{n} \sum_{i'=1}^{n} v_{i\alpha} v_{i'\alpha}^* \left(\sum_{\beta=1}^{n} w_{i\beta} w_{i'\beta}^* d_\beta \right) \left[\frac{\pi}{a(a + \lambda_i)} \left(\frac{1}{a + \lambda_{i'}} \right)^* + \frac{2\pi}{(a^2 - \lambda_{i'}^{*2})(\lambda_{i'}^* + \lambda_i)} \right]. \qquad (4.26)$$

Thus, the sample variance of $e(t)$ over a sample time of T can be expressed by a function of the observer gain L and the filter frequency a as the sum of the right side of equations 4.23 to 4.26 divided by the sample time T.

Before proceeding to the next section, it may be helpful to summarize what we have achieved so far.

Our goal is to detect if there is a fault based on the estimation error between the output from an observer and the output from the system. Under normal conditions

the error should be zero. If there is a fault, the error would not be equal to zero. For a certain fault, the error would be along certain fixed direction. Thus, by studying the direction of the error, the fault can be localized at the same time. The problem is a little more complicated when noise is present, which is inevitable in any real system. Now, the estimation error is a random process. We can no longer set our criteria to be exactly zero, because if so, we will have false alarms probably every several seconds. And fault detection would become meaningless. Based on optimal decision theory, the decision rule is a simple threshold. If the signal exceeds a certain level, we believe that there is a fault. It can be imagined that, for a certain cutoff level, the bigger the variance of the signal, the more frequently the level would be exceeded even without a fault. The average time for the signal to first exceed such a level corresponds to how frequently false alarms occur. Longer would be better. On the other hand, when a fault does indeed occur, the average time for the signal to exceed the preset level corresponds to the average time for a fault to be detected. Thus, we would like it to be short so a fault can be detected before damage is done. Both goals rely on solving the first hitting time of a preset level by a random signal. Given a preset cutoff boundary and the variance of a random Gaussian process, how can we estimate the average first hitting time of the boundary by the random signal? Fortunately, results from statistical physics on the Wiener process or driftless Brownian motion can be borrowed here. The solution to this problem is presented in the following section. It can be solved both analytically by the Fokker-Planck equation and by Monte Carlo simulation. Each method has its advantages and disadvantages.

4.3   First Stopping Time and Fokker-Planck Equation

The first hitting time (or first stopping time) problem is illustrated in Figure 4-4. The $R(t)$ is the upper bound, and the $L(t)$ is the lower bound. They correspond to the upper and lower thresholds of the decision rule in our case. The $y(t)$ is a random
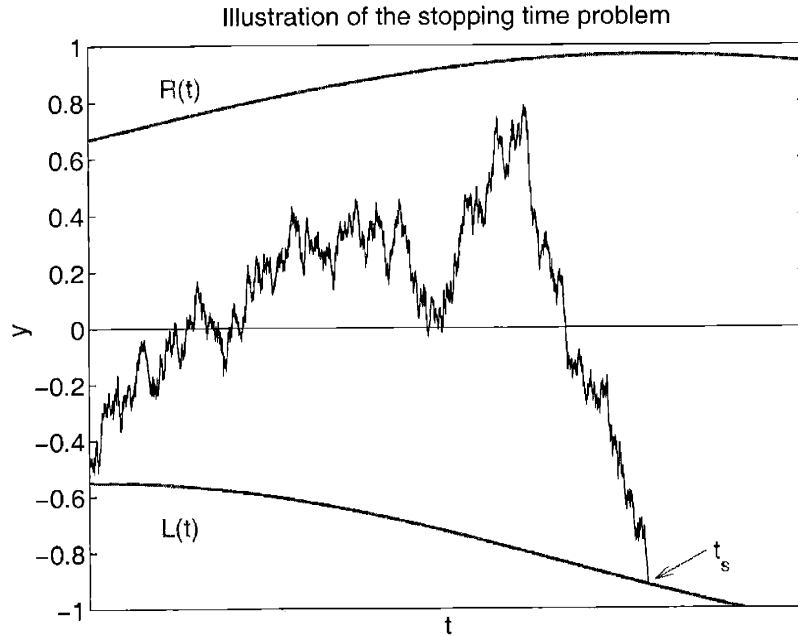
Figure 4-4: Illustration of the stopping time problem.

signal that has a Gaussian distribution. The first hitting time t_s is the first time that the signal touches the boundaries. It is obvious that t_s is also a random variable. Thus, we can not estimate the exact value for t_s, but its distribution and average value.

Let $R(t)$ and $L(t)$ be our decision thresholds, and $y(t)$ be our filtered estimation error ϵ as shown in Figure 4-3. When there is no fault, the signal ϵ contains only noise. If the boundaries are triggered by the signal, the decision rule will decide that a fault has occurred. This is a false alarm because the thresholds are not triggered by real fault, but noise. The average first hitting time $\overline{t_s}$ for this setting can then describe how often a false alarm occurs. On the other hand, if there is a fault, the filtered signal ϵ will contain both noise which is random and a deterministic fault signal. If the boundaries are triggered by this signal, the decision rule will decide that a fault has occurred. This is a true fault detection. The average first hitting time $\overline{t_s}$ for this setting can then describe how long it takes for a real fault to be detected. In the latter situation, we can subtract the deterministic component of the fault signal from the boundaries, and the problem becomes the same as the former except that now the thresholds are different. Thus, these two situations come down to the same

88

mathematical problem, which I call it the average first hitting time problem. The solution to this problem is important because it gives a quantitative description of the performance criteria that we are concerned. We then can assign costs to each of the performances and use optimization to achieve our goals.

In this section the derivation of the Fokker-Planck equation for solving the first hitting time and the solution to it are reviewed.

Consider prototype system:

$$\dot{y}(t) = -\frac{y(t)}{\tau} + n(t), \tag{4.27}$$

where $n(t)$ is Gaussian white noise,

$$\langle n(t)n(t+\tau) \rangle = \sigma^2 \delta(\tau). \tag{4.28}$$

When $\tau \to \infty$, $y(t)$ is called the Wiener process or driftless Brownian motion [37]. However, relaxations or dissipations exist in many systems, and the simplest way to handle their effects is by introducing a single relaxation time constant τ. If we regard $n(t)$ as an input to the system, its impulse response would be,

$$h(t) = e^{-\frac{t}{\tau}}\theta(t), \tag{4.29}$$

with Laplace transform,

$$H(s) = \frac{1}{s + \frac{1}{\tau}}, \tag{4.30}$$

and power spectrum modulation,

$$K(\omega) = \frac{1}{\omega^2 + \frac{1}{\tau^2}}. \tag{4.31}$$

Our intention is to fit the above form to a system's actual $K(\omega)$ which can be quite complicated, to extract an effective τ, and then using (4.27) to analyze the stopping time behavior of $y(t)$ driven by $n(t)$, under time-varying thresholds $[L(t), R(t)]$.

The nominal solution to (4.27) is,

$$y(t) = y(0)e^{-\frac{t}{\tau}} + \int_0^t e^{-\frac{t-t'}{\tau}} n(t')dt', \quad t > 0, \tag{4.32}$$

for a given $n(t'), t' \in [0, t]$. But, of course, $n(t')$ is random, so $y(t)$ can never be evaluated for sure, and the best one can obtain for the effects of $n(t')$ on $y(t)$ is in the probabilistic sense. Namely, suppose one knows the distribution of $y(0)$ as $P(y, t = 0)$, what would be the distribution of $y(t)$: $P(y, t)$? The fact that $P(y, t)$ is determinable by $P(y, t = 0)$ suggesting that $P(y, t)$ satisfies a time-evolution equation. [38]

To derive this time-evolution equation, let us take the small $t = \Delta t$ limit of (4.32) so that,

$$y(\Delta t) = y(0)\left(1 - \frac{\Delta t}{\tau}\right) + \int_0^{\Delta t} n(t')dt' + \mathcal{O}(\Delta t^2), \tag{4.33}$$

where terms $\mathcal{O}(\Delta t^2)$ are ignored. Thus, the increment in y,

$$\Delta y \equiv y(\Delta t) - y(0) = -y(0)\frac{\Delta t}{\tau} + w, \tag{4.34}$$

where

$$w \equiv \int_0^{\Delta t} n(t')dt', \tag{4.35}$$

is a random variable of zero mean, and variance

$$\begin{aligned}
\langle w^2 \rangle &= \left\langle \left(\int_0^{\Delta t} n(t')dt'\right)^2 \right\rangle \\
&= \left\langle \int_0^{\Delta t} n(t')dt' \cdot \int_0^{\Delta t} n(t'')dt'' \right\rangle \\
&= \int_0^{\Delta t} \int_0^{\Delta t} \langle n(t')n(t'') \rangle \, dt'dt'' \\
&= \int_0^{\Delta t} \int_0^{\Delta t} \sigma^2 \delta(t' - t'')dt'dt'' \\
&= \sigma^2 \Delta t. \tag{4.36}
\end{aligned}$$

Furthermore, because $n(t)$ is a Gaussian process, any linear time-integration of $n(t)$ is also a Gaussian process. Therefore $y(t)$ by virtue of (4.32) must be a Gaussian

process, so Δy, given $y(0)$, is just a Gaussian random variable with mean $-y(0)\frac{\Delta t}{\tau}$ and variance $\sigma^2 \Delta t$.

Therefore, the probability to find $y(\Delta t)$ between $(z, z + dz)$ is, under the conditional probability rule,

$$\Pr(z < y(\Delta t) < z+dz, \Delta t) = \int_{-\infty}^{\infty} dx P(x, t = 0) \cdot \Pr(z < y(\Delta t) < z+dz \mid y(0) = x) \tag{4.37}$$

and from our above derivations,

$$
\begin{aligned}
&\Pr(z < y(\Delta t) < z + dz \mid y(0) = x) \\
=\ &\Pr(z - x < \Delta y < z - x + dz \mid y(0) = x) \\
=\ &\Pr\left(z - x + x\frac{\Delta t}{\tau} < w < z - x + x\frac{\Delta t}{\tau} + dz \mid y(0) = x\right) \\
=\ &\frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left(-\frac{(z - x + x\frac{\Delta t}{\tau})^2}{2\sigma^2\Delta t}\right) dz. \tag{4.38}
\end{aligned}
$$

Therefore,

$$\Pr(z < y(\Delta t) < z + dz, \Delta t) = dz \int_{-\infty}^{\infty} dx \frac{P(x, t = 0)}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left(-\frac{(z - x + x\frac{\Delta t}{\tau})^2}{2\sigma^2\Delta t}\right).$$

So, by the definition of probability density,

$$P(z, \Delta t) = \int_{-\infty}^{\infty} dx P(x, t = 0) \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left(-\frac{(z - x + x\frac{\Delta t}{\tau})^2}{2\sigma^2\Delta t}\right). \tag{4.39}$$

We are interested in the small differences between $P(y, t = \Delta t)$ and $P(y, t = 0)$ for the purpose of deriving a differential equation. As $\Delta t \to 0$, (4.38) as a function of x behaves more and more like a delta-function centered at z, making it easier to analyze the above integral. However, a bit of caution is advised because we would like to keep all terms $\mathcal{O}(\Delta t)$ in the integral, not just the zeroth-order term. A systematic way to explore the small Δt asymptote is by a Taylor expansion of $P(x, t = 0)$ and

then carrying out saddle-point integration. Let us rewrite (4.39) as,

$$P(z, \Delta t) = \frac{1}{1 - \frac{\Delta t}{\tau}} \int_{-\infty}^{\infty} dx \frac{P(x, t = 0)}{\sqrt{2\pi\sigma^2\Delta t/(1 - \frac{\Delta t}{\tau})^2}} \exp\left(-\frac{\left(\frac{z}{1 - \frac{\Delta t}{\tau}} - x\right)^2}{2\sigma^2\Delta t/(1 - \frac{\Delta t}{\tau})^2}\right).$$

The reason for the above rewrite is that we would like to make the Gaussian function inside the integral a normalized density in x. Then, by expanding $P(x, t = 0)$ around $x_0 = \frac{z}{1 - \frac{\Delta t}{\tau}}$,

$$P(x, t = 0) = P\left(\frac{z}{1 - \frac{\Delta t}{\tau}}, t = 0\right) + P' \cdot \left(x - \frac{z}{1 - \frac{\Delta t}{\tau}}\right) + \frac{1}{2}P'' \cdot \left(x - \frac{z}{1 - \frac{\Delta t}{\tau}}\right)^2 + ...,$$

and carrying out the integration, we get,

$$P(z, \Delta t) = \frac{1}{1 - \frac{\Delta t}{\tau}}\left[P\left(\frac{z}{1 - \frac{\Delta t}{\tau}}, t = 0\right) + 0 + \frac{1}{2}P''\sigma^2\Delta t + o(\Delta t)\right], \quad (4.40)$$

because the first moment vanishes under the integration, the second moment yields $\sigma^2\Delta t$, and all higher moments are ignored because their contribution would be $o(\Delta t)$.

Therefore,

$$\begin{aligned} P(z, \Delta t) &= \left(1 + \frac{\Delta t}{\tau}\right)\left[P\left(z + z\frac{\Delta t}{\tau}, t = 0\right) + \frac{1}{2}P''\sigma^2\Delta t\right] + o(\Delta t) \\ &= P(z, t = 0) + \frac{\Delta t}{\tau}P + P'z\frac{\Delta t}{\tau} + \frac{1}{2}P''\sigma^2\Delta t + o(\Delta t), \quad (4.41) \end{aligned}$$

thus,

$$\frac{P(z, \Delta t) - P(z, t = 0)}{\Delta t} = \frac{P + P'z}{\tau} + \frac{1}{2}P''\sigma^2 + o(1),$$

so in the $\Delta t \to 0$ limit, we obtain,

$$\frac{\partial P(y, t)}{\partial t} = \frac{\partial[P(y, t)(y/\tau)]}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2 P(y, t)}{\partial y^2}. \quad (4.42)$$

This is the Fokker-Planck equation, a PDE of the convection-diffusion type. It is a significant result for us because $P(y, t)$, the time-evolution of the "survival probability

density" profile, fully determines the stopping time statistics of $y(t)$ from (4.27).

(4.42) can be easily generalized to multiple dimensions, in which case (4.27) becomes a general state-space realization of a LTI system. Therefore, the stopping time statistics of LTI systems driven by white noise is a solvable problem, at least on paper. Another direction is to extend (4.27) to nonlinear dynamical systems, in which case (4.42) becomes the so-called Smoluchowski equation [38], prescribing general phase space flow and diffusion. We will not pursue these topics here.

4.3.1 Steady State Behavior With No Absorption

Consider the steady state solution to (4.42),

$$\frac{\partial[P(y,t)(y/\tau)]}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2 P(y,t)}{\partial y^2} = 0,$$

or

$$\frac{\sigma^2\tau}{2}P'' + (yP)' = 0,$$

where we let the absorbing boundaries $[L(t), R(t)]$ to be $(-\infty, +\infty)$. So,

$$\frac{\sigma^2\tau}{2}P' + yP = C.$$

However, because $P(y)$ has the meaning of a probability density whose integral is ≤ 1, there should be $\lim_{y\to\infty} P' = 0$ and $\lim_{y\to\infty} yP(y) = 0$. Therefore, C must be 0. Thus,

$$\frac{\sigma^2\tau}{2}P' + yP = 0 \qquad \longrightarrow \qquad \frac{d\log P}{dy} = -\frac{2y}{\sigma^2\tau},$$

and so,

$$\log P = -\frac{y^2}{\sigma^2\tau} + D \qquad \longrightarrow \qquad P = \frac{1}{\sqrt{\pi\sigma^2\tau}}\exp\left(-\frac{y^2}{\sigma^2\tau}\right), \qquad (4.43)$$

if under normalization. Thus, at steady state, $y(t)$ conforms to a zero-mean Gaussian distribution with variance $\langle y^2 \rangle = \sigma^2\tau/2$, or in other words the power of $y(t)$ is $\sigma^2\tau/2$.

This is consistent with (4.31), because the power spectrum density of $n(t)$ is,

$$\text{psd}_n(\omega) \equiv \lim_{T \to \infty} \frac{\langle |n_T(\omega)|^2 \rangle}{T} = \sigma^2,$$

where $n_T(\omega)$ is the Fourier transform of $n(t)$ under time truncation T. Thus, the power spectrum density of $y(t)$ is,

$$\text{psd}_y(\omega) = K(\omega) \cdot \text{psd}_n(\omega) = \frac{\sigma^2}{\omega^2 + \frac{1}{\tau^2}},$$

and thus,

$$\langle y(t)^2 \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{psd}_y(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sigma^2}{\omega^2 + \frac{1}{\tau^2}} d\omega,$$

which by using a Cauchy integral in the upper half plane gives,

$$\langle y(t)^2 \rangle = \frac{1}{2\pi} \cdot 2\pi i \cdot \frac{\sigma^2}{\frac{i}{\tau} + \frac{i}{\tau}} = \frac{\sigma^2 \tau}{2}.$$

One may extend this result to,

$$\langle y(t) y(t') \rangle = \frac{\sigma^2 \tau}{2} \exp\left(-\frac{|t - t'|}{\tau}\right).$$

4.3.2 Time-Dependent Solution with Absorbing Boundaries

Consider Fig. 4-4. The particular trajectory shown can be interpreted as a realization of one microscopic particle diffusing in a 1D domain with absorbing boundaries, whose appropriate macroscopic description should be,

$$\frac{\partial P(y,t)}{\partial t} = \frac{\partial[P(y,t)(y/\tau)]}{\partial y} + \frac{\sigma^2}{2} \frac{\partial^2 P(y,t)}{\partial y^2},$$

$$(4.44)$$

$$P(y = L(t), t) = P(y = R(t), t) = 0, \quad P(y, t = 0) = P_0(y).$$

where $P(y,t)$ is taken to be the number density profile of such particles, with initial profile $P_0(y)$. Once $y(t)$ touches one of the boundaries, it is taken out of the con-

sideration for t_s statistics. Thus t_s statistics is completely determined by the $P(y,t)$ solution to (4.44) as,

$$\Pr\left(t < t_s < t + dt | L(0) < y(0) < R(0)\right) \equiv \tilde{\rho}(t)dt \quad = \quad \frac{dt \cdot -\frac{d}{dt}\left[\int_{L(t)}^{R(t)} P(y,t)dy\right]}{\int_{L(0)}^{R(0)} P_0(y)dy},$$

where the denominator counts the total number of particles at $t = 0$, the numerator counts how many particles hit the boundaries between $(t, t+dt)$ and thus contribute to the t_s statistics.

Define a dimensionless variable $x \in [0, \pi]$,

$$x \equiv \frac{\pi(y - L(t))}{R(t) - L(t)}, \quad y = \frac{(R-L)x}{\pi} + L, \tag{4.45}$$

and expand $P(y,t)$ as,

$$P(y,t) \quad = \quad \sum_{n=1}^{\infty} p_n(t)\sin(nx), \tag{4.46}$$

which automatically satisfies the boundary conditions. Because

$$\frac{\partial x}{\partial y} \quad = \quad \frac{\pi}{R-L},$$

$$\frac{\partial x}{\partial t} \quad = \quad \frac{-\pi\dot{L}}{R-L} - \frac{\pi(y-L)(\dot{R}-\dot{L})}{(R-L)^2} \quad = \quad \tilde{\mu}(t)x + \tilde{\nu}(t),$$

where

$$\tilde{\mu}(t) \equiv -\frac{\dot{R}-\dot{L}}{R-L}, \quad \tilde{\nu}(t) \equiv \frac{-\pi\dot{L}}{R-L},$$

are known functions, we then have,

$$\frac{\partial P(y,t)}{\partial t} \quad = \quad \sum_{n=1}^{\infty} \dot{p}_n \sin(nx) + p_n n \cos(nx)(\tilde{\mu}x + \tilde{\nu}),$$

$$\frac{\partial [P(y,t)(y/\tau)]}{\partial y} \quad = \quad \frac{P + yP'}{\tau} \quad = \quad \frac{\sum_{n=1}^{\infty} p_n \sin(nx) + \left(x + \frac{\pi L}{R-L}\right)\sum_{n=1}^{\infty} p_n n \cos(nx)}{\tau},$$

$$\frac{\sigma^2}{2}\frac{\partial^2 P(y,t)}{\partial y^2} \quad = \quad -\frac{1}{2}\left(\frac{\sigma\pi}{R-L}\right)^2 \sum_{n=1}^{\infty} p_n n^2 \sin(nx).$$

95

Collecting terms in (4.44), we obtain,

$$\sum_{n=1}^{\infty} \left\{ \dot{p}_n + \left[\frac{n^2}{2} \left(\frac{\sigma \pi}{R-L} \right)^2 - \frac{1}{\tau} \right] p_n \right\} \sin(nx) = \sum_{m=1}^{\infty} p_m m \cos(mx)(\mu x + \nu), \quad (4.47)$$

with

$$\mu \equiv \frac{1}{\tau} - \tilde{\mu} = \frac{1}{\tau} + \frac{\dot{R} - \dot{L}}{R-L}, \quad \nu \equiv \frac{\pi L}{\tau(R-L)} - \tilde{\nu} = \frac{\pi(L/\tau + \dot{L})}{R-L}.$$

When $n \neq m$, there is,

$$\int_0^{\pi} \sin(nx) \cos(mx) dx$$
$$= \int_0^{\pi} \frac{\sin(nx + mx) + \sin(nx - mx)}{2} dx$$
$$= \frac{1 - (-1)^{n+m}}{2(n+m)} + \frac{1 - (-1)^{n-m}}{2(n-m)}. \quad (4.48)$$

When $n = m$, the integral is 0. So there is,

$$\cos(mx) = \sum_{n \neq m}^{\infty} C_{nm} \sin(nx),$$

$$C_{nm} \equiv \frac{1 - (-1)^{n+m}}{\pi(n+m)} + \frac{1 - (-1)^{n-m}}{\pi(n-m)} = \frac{2n(1 - (-1)^{n+m})}{\pi(n^2 - m^2)}.$$

When $n \neq m$, there is,

$$\int_0^{\pi} x \sin(nx) \cos(mx) dx$$
$$= \int_0^{\pi} \frac{x \sin(nx + mx) + x \sin(nx - mx)}{2} dx$$
$$= \frac{\pi(-1)^{n+m+1}}{2(n+m)} + \frac{\pi(-1)^{n-m+1}}{2(n-m)}. \quad (4.49)$$

When $n = m$, there is,

$$\int_0^{\pi} x \sin(mx) \cos(mx) dx = -\frac{\pi}{4m}.$$

So there is,

$$x\cos(mx) \;=\; -\frac{\sin(mx)}{2m} + \sum_{n \neq m}^{\infty} X_{nm}\sin(nx), \qquad X_{nm} \equiv \frac{2n(-1)^{n+m+1}}{n^2 - m^2}.$$

Therefore, the solution to (4.47) is,

$$\dot{p}_n + \left[\frac{n^2}{2}\left(\frac{\sigma\pi}{R-L}\right)^2 - \frac{1}{\tau}\right]p_n \;=\; -\frac{\mu p_n n}{2n} + \sum_{n \neq m}^{\infty} p_m m(\mu X_{nm} + \nu C_{nm}),$$

or,

$$\dot{p}_n \;=\; \sum_{m=1}^{\infty} K_{nm}p_m,$$

with

$$K_{nm} \;\equiv\; \delta_{nm}\cdot\left[\frac{1}{\tau} - \frac{n^2}{2}\left(\frac{\sigma\pi}{R-L}\right)^2 - \frac{\mu}{2}\right] + (1 - \delta_{nm})\cdot m\left(\mu X_{nm} + \nu C_{nm}\right),$$

$$\mu = \frac{1}{\tau} + \frac{\dot{R}-\dot{L}}{R-L}, \quad \nu = \frac{\pi(L/\tau + \dot{L})}{R-L}, \quad C_{nm} = \frac{2n(1-(-1)^{n+m})}{\pi(n^2-m^2)}, \quad X_{nm} = \frac{2n(-1)^{n+m+1}}{n^2-m^2}.$$

And the total particle number at t would be,

$$N(t) \equiv \int_{L(t)}^{R(t)} P(y,t)dy \;=\; \frac{R-L}{\pi}\int_0^{\pi} P dx \;=\; \frac{2(R-L)}{\pi}\sum_{n=1,3,5,\ldots}^{\infty}\frac{p_n(t)}{n}.$$

If the particle density is (4.43) at $t = 0^-$, the total particle number would be reduced to,

$$N(t=0^+) \;=\; \int_{L(0)}^{R(0)} \frac{1}{\sqrt{\pi\sigma^2\tau}}\exp\left(-\frac{y^2}{\sigma^2\tau}\right) \;=\; \frac{\operatorname{erf}\left(\frac{R(0)}{\sigma\sqrt{\tau}}\right) - \operatorname{erf}\left(\frac{L(0)}{\sigma\sqrt{\tau}}\right)}{2}, \qquad (4.50)$$

at $t = 0^+$, because those outside of $[L(0), R(0)]$ are immediately caught. The surviving particles has a truncated Gaussian distribution that has the following expansion,

$$P(y, t = 0^+) \equiv P_0(y) \;=\; \sum_{n=1}^{\infty} p_n(0)\sin(nx),$$

$$p_n(0) = \frac{2}{\pi} \int_0^\pi \frac{1}{\sqrt{\pi\sigma^2\tau}} \exp\left(-\frac{\left(\frac{(R-L)x}{\pi} + L\right)^2}{\sigma^2\tau}\right) \sin(nx)dx. \qquad (4.51)$$

It turns out that there is no special function representation for the above integral, which then has to be accounted for by numerical quadrature.

Therefore, the hitting time distribution is simply,

$$\rho(t) = (1 - N(0^+))\delta(t) + N(0^+)\tilde{\rho}(t) = (1 - N(0^+))\delta(t) - \frac{d}{dt}\left[\frac{2(R-L)}{\pi} \sum_{n=1,3,5,..}^{\infty} \frac{p_n(t)}{n}\right],$$

with the initial values for p_n given by (4.51).

4.3.3 Fokker-Planck versus Monte Carlo simulation

The first hitting time problem can also be solved by Monte Carlo simulation. The process and results are presented below.

An N point random sequence was generated to represent a discretization of a signal $y(t)$ over time T whose continuous Fourier transform $y(w) = \int y(t)e^{-jwt}dt$ is in fact $y(w) = 1/(jw + 1) * n(w)$. $n(w)$ is the frequency domain white noise and is expressed as $n(w) = R(w) + j * I(w)$ with the real part $R(w)$ in normal distribution $N(0, .5 * T)$, imaginary part $I(w)$ also in normal distribution $N(0, .5 * T)$, and $R(w)$ and $I(w)$ are independent. The variance of $y(t)$ can be calculated from the Rayleigh's theorem.

The simulation was first run for $N = 100000$ and $T = 2000$. A portion of $y(t)$ is displayed in Figure 4-5. The probability density function of $y(t)$ is indeed a Gaussian distribution with a mean of zero and a variance of 0.5, which is plotted in figure 4-6.

Chose a threshold level to be 1.5, the first hitting time distribution is plotted as a histogram in figure 4-7. The probability for τ to be zero is the tail probability that $y > threshold$, and it equals to $D = 0.0339$ in this example. Thus, the probability density at $\tau = 0$ is a δ function,

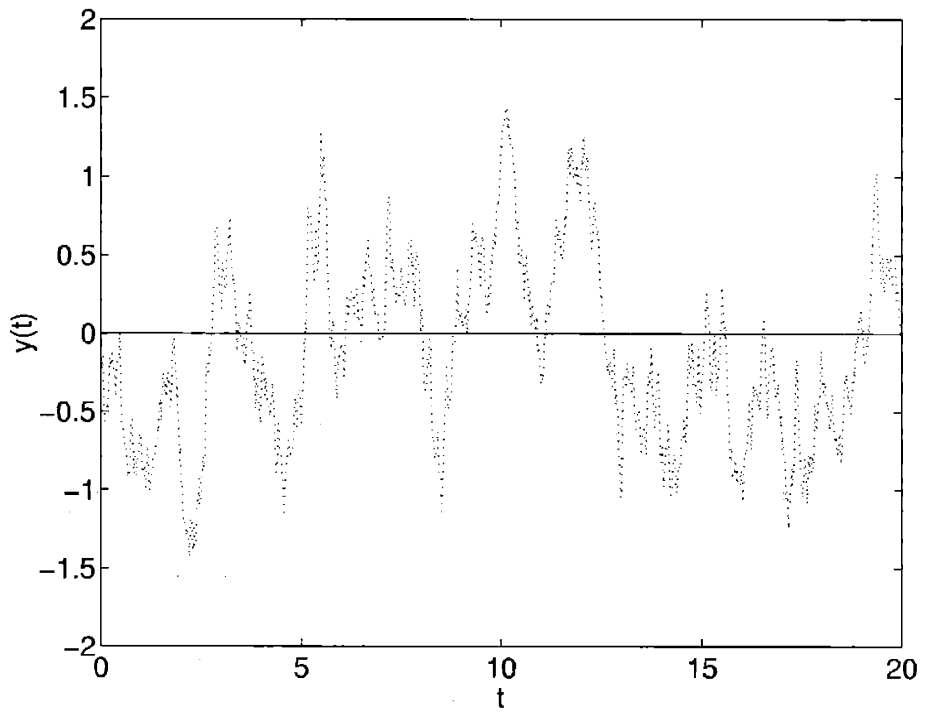$$f(\tau = 0) = D\,\delta(0) \qquad (4.52)$$

Figure 4-5: Sample $y(t)$ generated from MC simulation.
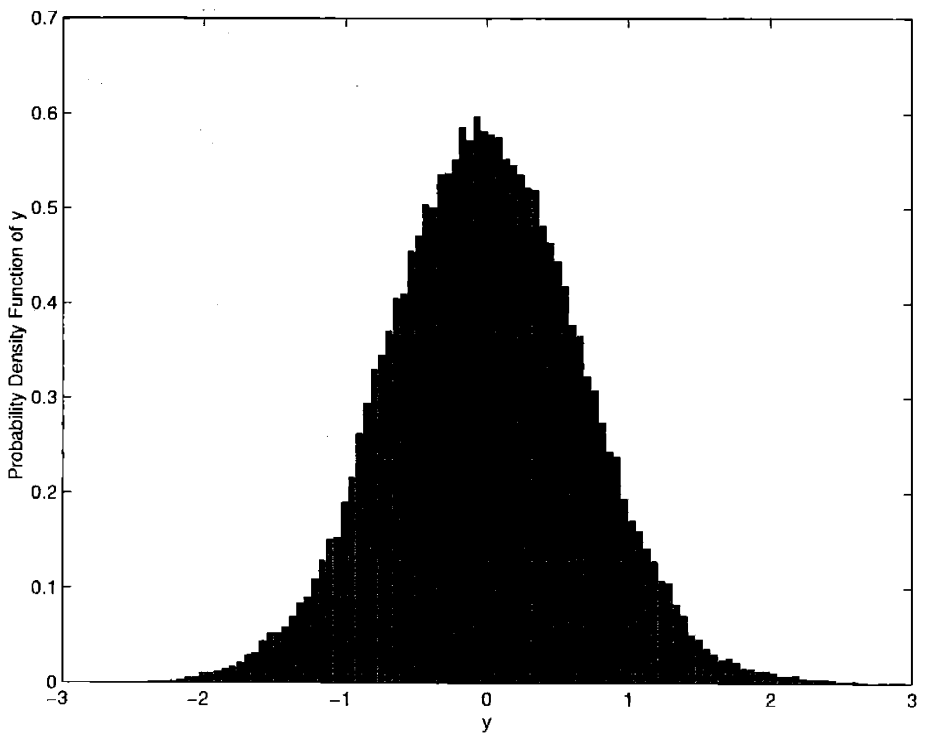

Figure 4-6: Probability density function of the $y(t)$ generated from Monte Carlo simulation.
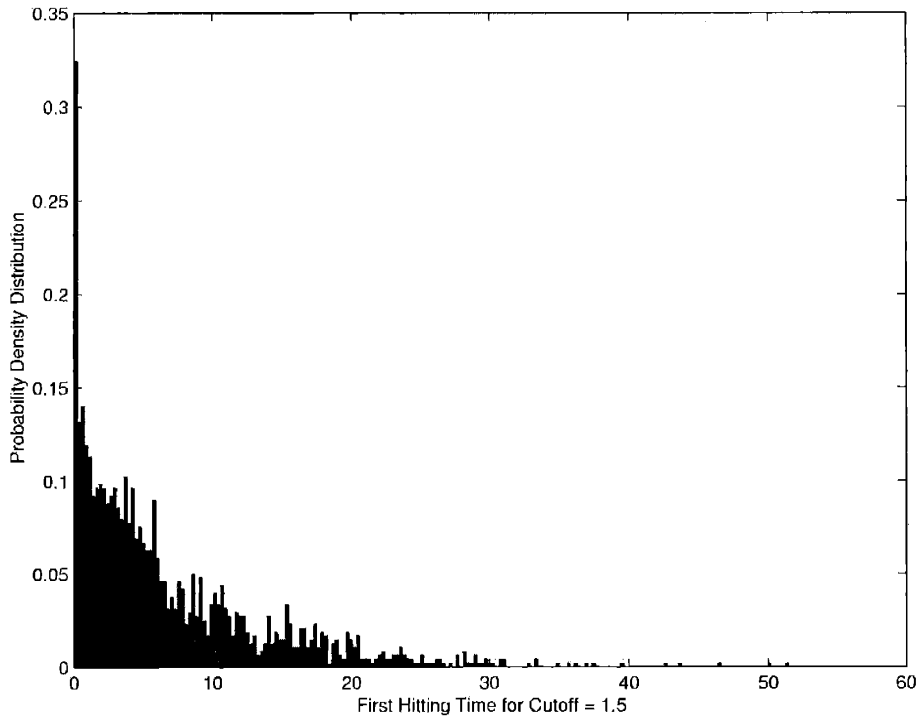
99

Figure 4-7: Distribution of the first hitting time t_s generated from Monte Carlo simulation.

If we remove the δ function at time equals zero, we get a more accurate result. The new pdf is plotted as a histogram in Figure 4-8.

The average first hitting time is thus calculated by

$$\overline{\tau} = \int_0^\infty t f(t) dt \tag{4.53}$$

and for this example, $\overline{\tau} = 6.7419$.

Although theoretically, the Fokker-Planck equation and the Monte Carlo simulation should give exact solutions, during numerical implementation, approximations are inevitable. For Fokker-Planck equation, the error comes from the truncation of the sine wave expansion. While the accuracy of the Monte Carlo simulation depends on the number of simulations, the number of sample points N and the sample time dT. Sensitivity studies of the accuracies of the Fokker-Planck and Monte Carlo results on these parameters were performed and the results are listed in tables 4.1 and 4.2. Both calculations were for a constant threshold of [-1 2]. In Table 4.1, N (wave) is the
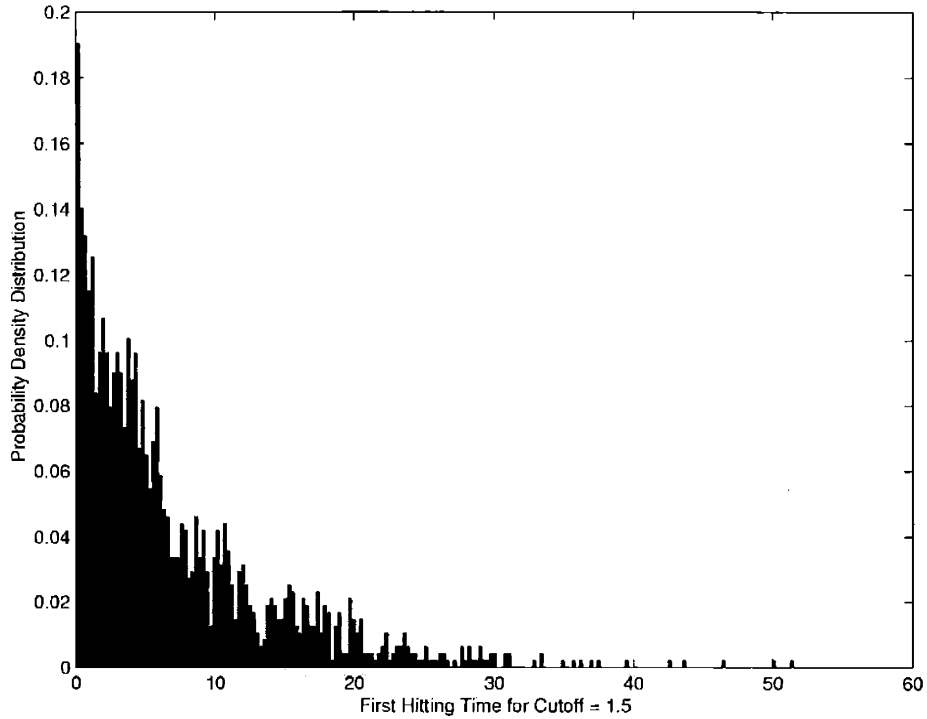
Figure 4-8: Distribution of the first hitting time t_s generated from Monte Carlo simulation after adjustment.

number of sine waves used to for expansion. The execution time is in unit of second and the program is run on a computer with an AMD Athlon 1.33 GHz CPU. The results converge after $N > 100$. But for $N = 20$, the result of the average first hitting time $\bar{\tau}$ is already fairly good in the sense that the relative error of this calculation and the converged result is 0.04%. and the execution time is only about 2 seconds. In Table 4.2, the first column is the sample points that were generated, the second column is the sample rate. The third column, the execution time is also in unit of second and the program is run on a computer with an AMD Athlon 1.33 GHz CPU. The last column, points for τ, is the number of points that was used for calculation of τ. It is different from the sample point which is in terms of time. A combination of $N = 2^{21}$ and $dt = 0.01$ gives good accuracy while keeping the execution time reasonable.

The numerical results from solving the Fokker-Planck equation and from Monte Carlo simulation are presented in Figure 4-9 to 4-14.

101

Table 4.1: Sensitivity studies of Fokker-Planck Calculation

N (wave)	Execution Time (s)	$\overline{\tau}$	Relative Error %
200	103.2762	3.1993	0
100	27.2300	3.1993	0
40	5.3913	3.1994	0.003
30	3.1202	3.1996	0.01
20	1.8667	3.2006	0.04
10	0.4894	3.2098	0.33
5	0.4649	3.2170	0.55

Table 4.2: Sensitivity studies of Monte Carlo simulation (threshold [-1 2], 10 simulations)

Sample Points	Time Step dt	Execution Time (s)	$\overline{\tau}$	Points for τ
2^{21}	0.001	171.1235	2.7195	5772
	0.01	249.2458	3.1014	52715
	0.05	1503.7	3.8142	223348
Time Step dt	Sample Points	Execution Time (s)	$\overline{\tau}$	Points for τ
0.01	2^{19}	44.5272	3.0288	13421
	2^{21}	171.1235	2.7195	5772
	2^{22}	647.8998	3.1023	105458

For a symmetric constant threshold of $[-2, 2]$, the survival probability as a function of time is displayed in figure 4-9. The circles are results from the Fokker-Planck equation while the stars are from the Monte Carlo simulation. The probability density for the first stopping times obtained from the Fokker Planck equation (circle) and Monte Carlo simulation (star) are shown in figure 4-10. The results agree very well. The calculated average first stopping time $\overline{\tau}$ are 23.7163 and 25.2955 from Fokker Planck equation and Monte Carlo simulation respectively with a relative error of about 6.66%.

For asymmetric constant threshold boundaries $[-1, 2]$, the survival probability results are shown in figure 4-11 and the probability density functions are shown in figure 4-12. The average first stopping time $\overline{\tau}$ are 3.2006 and 3.0649 from Fokker Planck equation and Monte Carlo simulation respectively with a relative error of about 4.24%.

For asymmetric time varying boundaries that are depicted in figure 4-13, the survival probability results are shown in figure 4-14. The average first stopping time
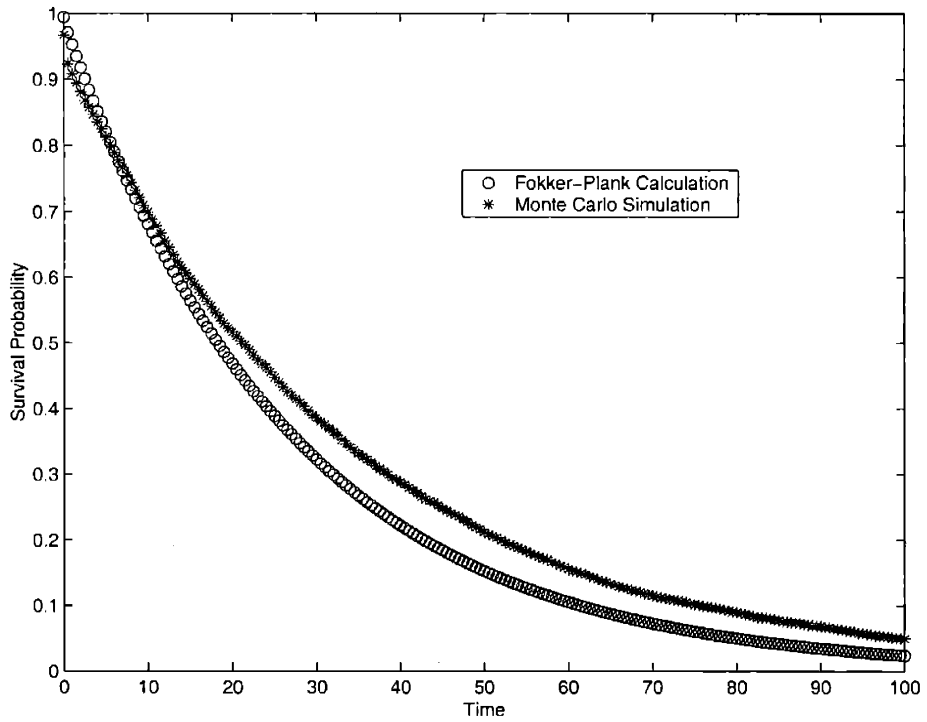
Figure 4-9: Survival probability as functions of time from Fokker Planck equation (circle) and Monte Carlo simulation (star) with a threshold of $[-2, 2]$.

Figure 4-10: Probability density for first stopping time obtained from Fokker Planck equation (circle) and Monte Carlo simulation (star) with a threshold of $[-2, 2]$.

Figure 4-11: Survival probability as functions of time from Fokker Planck equation (circle) and Monte Carlo simulation (star) with a threshold of $[-1, 2]$.
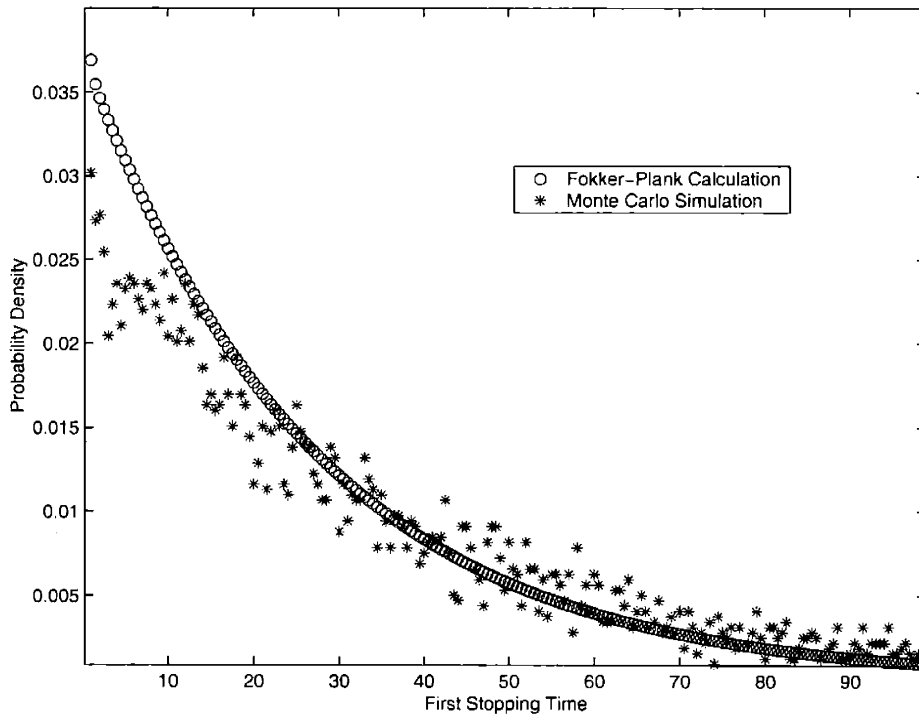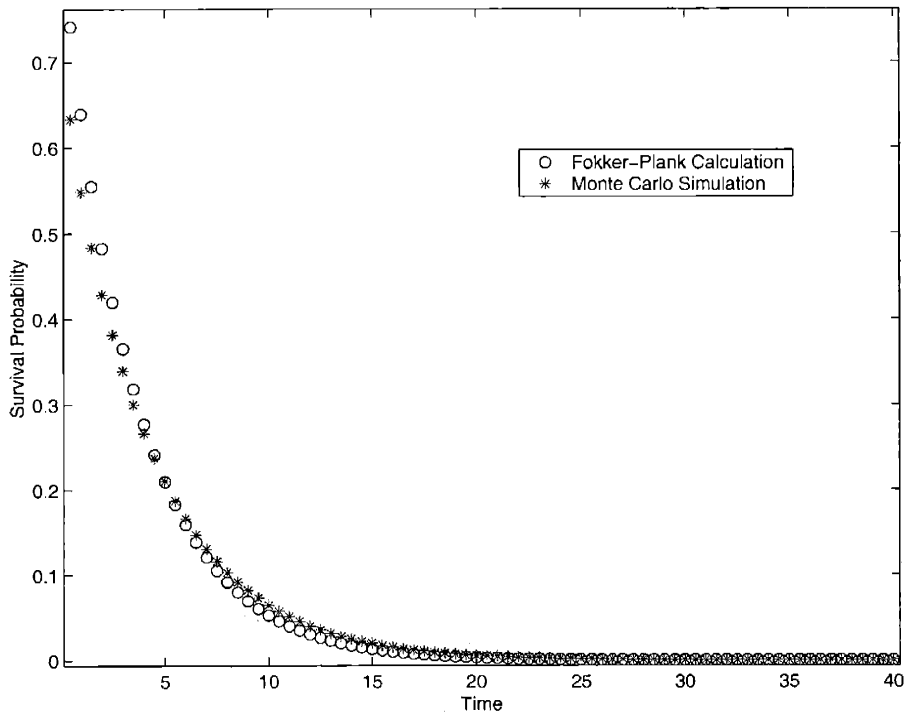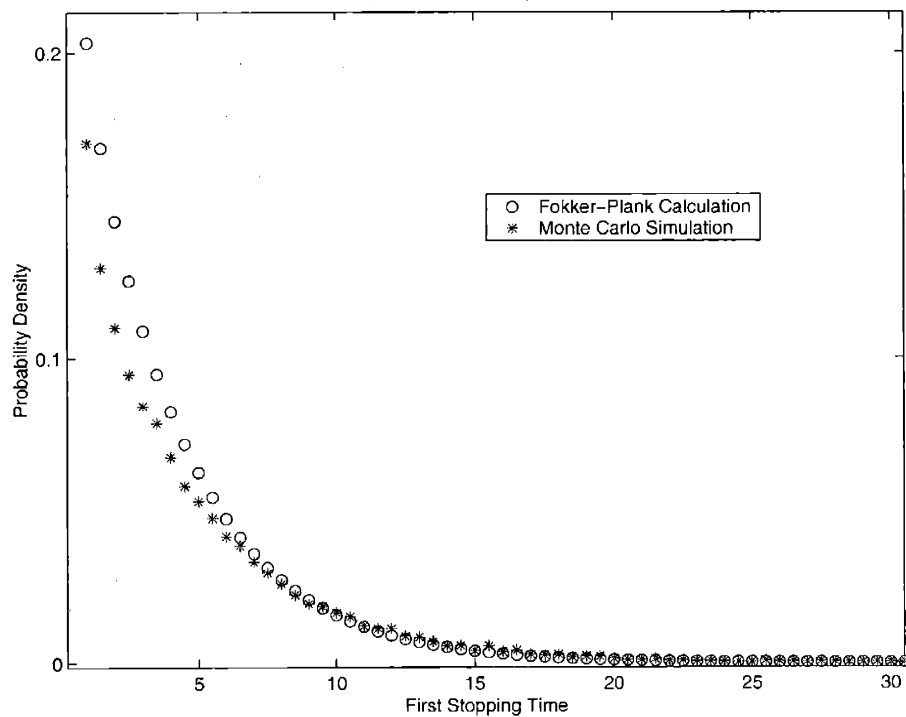

Figure 4-12: Probability density for first stopping time obtained from Fokker Planck equation (circle) and Monte Carlo simulation (star) with a threshold of $[-1, 2]$.
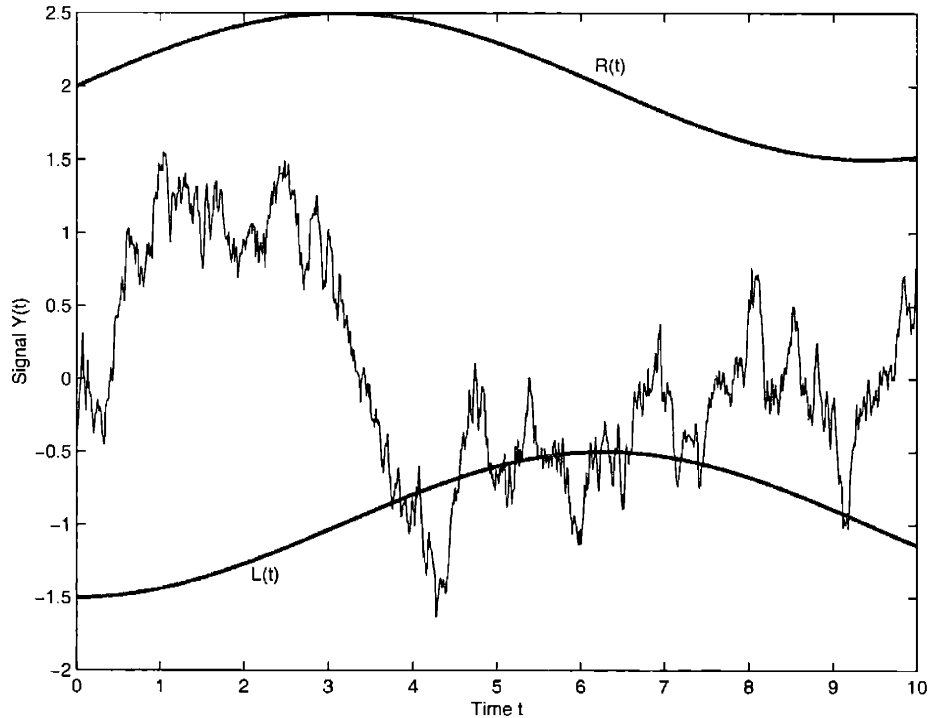
Figure 4-13: Illustration of the time varying boundary stopping time problem.

$\bar{\tau}$ are 3.5112 and 3.7512 from Fokker Planck equation and Monte Carlo simulation respectively. The relative error of the two is about 6.83%.

Both the Monte Carlo method and the Fokker-Planck equation can be used to solve the first stopping time problem. As illustrated above, the results from these two methods agree to each other within a relative error of 10% . The Monte Carlo method is more flexible and can deal with systems whose power spectrum modulation does not satisfy equation 4.31 easily. But, for a long first stopping time, it needs a long simulation to get a good statistics. The Fokker-Planck method is faster (more than 100 times faster than a comparable Monte Carlo simulation). But it is limited to a system whose power spectrum modulation can be approximated by equation 4.31. For a general LTI system, the Fokker-Planck equation can be generalized to be multidimensional. But the computation would be much more involved.
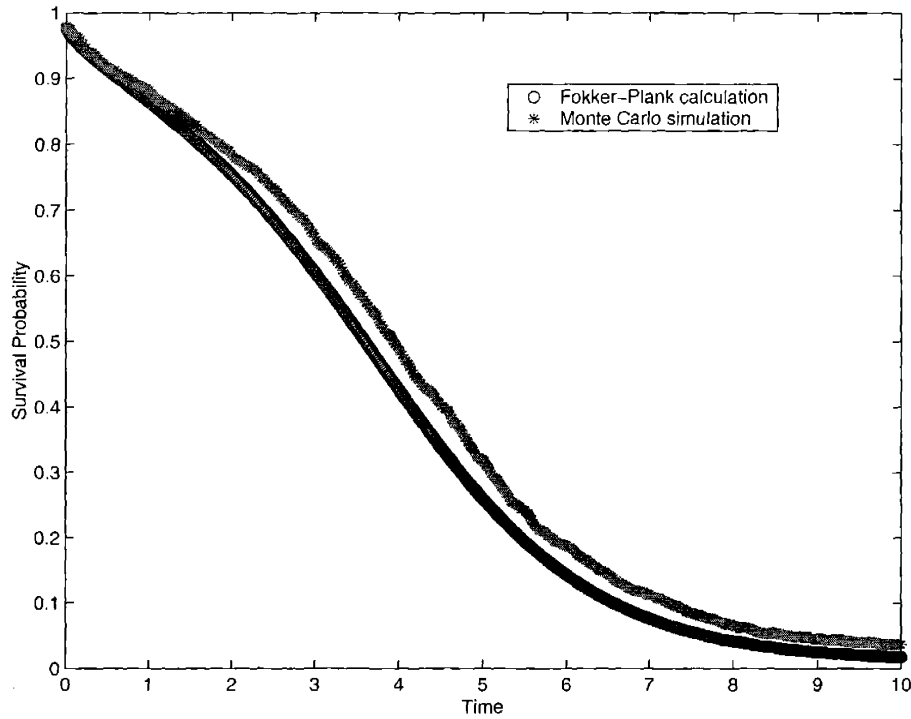
Figure 4-14: Survival probability as functions of time from Fokker Planck equation (circle) and Monte Carlo simulation (star) with time varying boundaries.

4.3.4 Empirical Model

For constant symmetric boundaries, an empirical model can be used for estimating the average first hitting time for a normally distributed random process. It is formulated as

$$
\begin{aligned}
\bar{\tau} &= \frac{\tau}{1 - \int_{-e_c}^{e_c} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(x-\mu)^2/(2\sigma_y^2)} dx} \\
&= \frac{\tau}{1 - erf\left(\frac{e_c - \mu}{\sqrt{2}\sigma_y}\right)}
\end{aligned}
\tag{4.54}
$$

where $\bar{\tau}$ is the average first hitting time, τ is the effective system response time, e_c is the threshold, μ is the mean, and σ_y is the variance of the random process. Note that σ_y is different from the σ in MC and FP, the later is the pre-system white noise variance. The two are related by $\sigma = \sigma_y * \sqrt{2/\tau}$. The effective system response time τ can be obtained by fitting the power spectrum to the Lorentzian spectrum function in equation 4.31. If we know the transfer function of the system, it can also be obtained
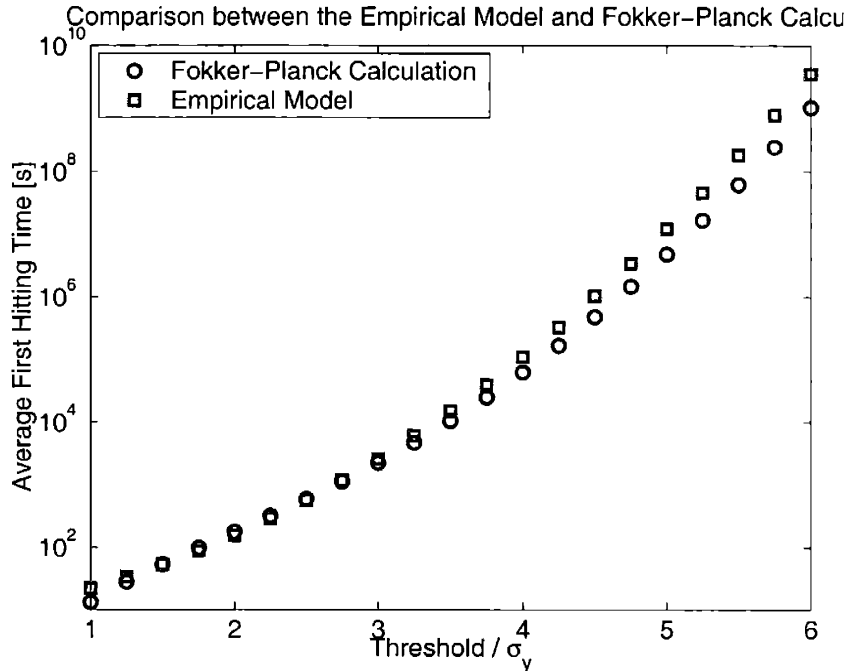
106

Figure 4-15: Comparison between the empirical model and Fokker-Planck calculations.

directly from the frequency domain calculation.

Although this empirical model is very simple, the results from this model agree with that from the Fokker Planck very well, especially in the range from $e_c/\sigma_y = 1.3$ to $e_c/\sigma_y = 5$. A comparison is shown in Figure 4-15.

4.4   Construction of Cost Function

After we have the average first stopping times for false alarms and fault detection, costs should be assigned to each case. The construction of cost functions should rely on one's knowledge of the system and common sense. There is no absolute right or wrong for cost functions. One can only say if it is reasonable or not.

For false alarms, it is intuitive to think that a small τ_1 would render the fault detection method useless. For example, from the standpoint of a reactor operation, a false alarm rate should be measured in the time scale of weeks or even months. One false alarm in a month may be acceptable. The lower limit of τ is the data acquisition sample rate, and for MITR-II it is 1 second. Scale the cost to a range of 0 to 100. An
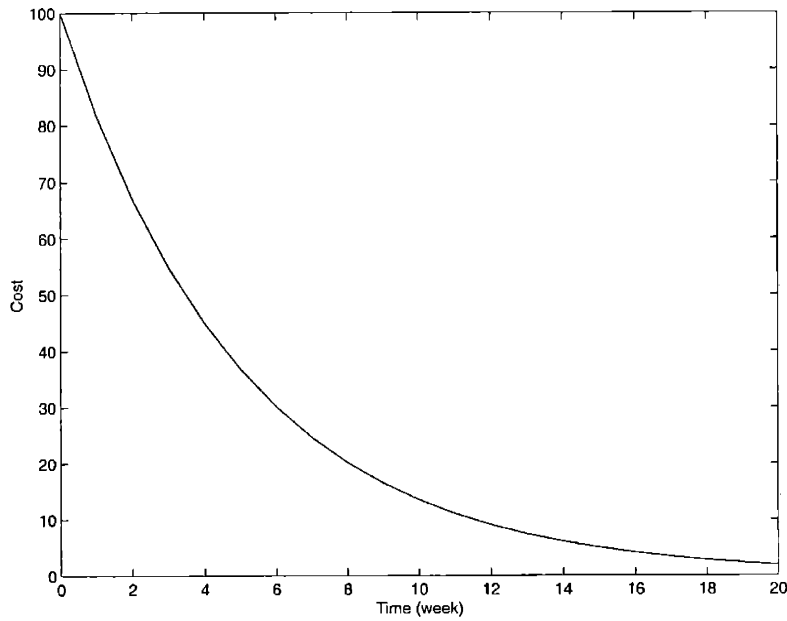
Figure 4-16: Cost function for false alarm.

example of a cost function for false alarms as a function of the average first hitting time is plotted in figure 4-16. The function can be generated using equation 4.55.

$$If \ \ \bar{\tau} <= 1 \ s, \ \ cost = 100;$$
$$Else \ \ cost = 100 * e^{-(\bar{\tau}-1)/5};$$

(4.55)

On the contrary, the cost for not detecting a real fault should be large for a long $\bar{\tau}$. Ideally, a fault should be detected within seconds or minutes of its occurrence. Thus, the cost function for the non-detection should be measured on the time scale of a second. The cost should also level off for long $\bar{\tau}$. It is intuitive to think that the cost would not go up forever. After all the damage has been done, it will not cost extra after that. For example, for a reactor, an upper limit of $\bar{\tau}$ may be 3 hours. That is if a fault is not detected for 3 hours, probably the maximum damage would have already resulted. The lower limit of $\bar{\tau}$ is also the data acquisition sample rate and it is 1 second for MITR-II. Scale the cost function to a range of 0 to 100, and it can have the shape plotted in figure 4-17, which is generated using equation 4.56.
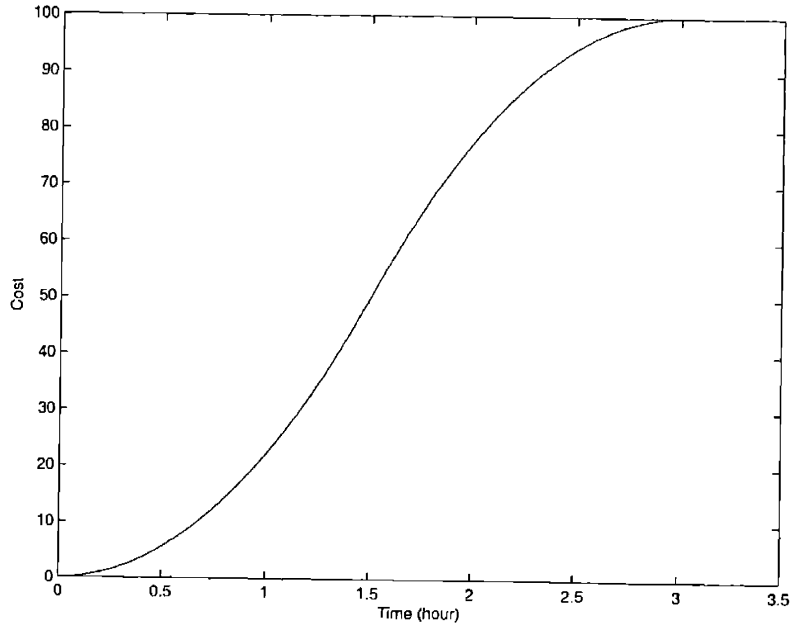
Figure 4-17: Cost function for non-detection.

The 162 h is a chosen to make the two pieces of curves be connected smoothly.

$$If \ \ \bar{\tau} >= 3 \ h, \ \ cost = 100;$$

$$Elseif \ \ \bar{\tau} >= 1.5 \ h, \ \ cost = 100 - (3 \ h - \bar{\tau})^2/162 \ h;$$

$$Else \ \ cost = \bar{\tau}^2/162 \ h; \quad\quad (4.56)$$

4.5 Fault Detection Filter Design Process

The overall objective is to detect a fault that belongs to the event ensemble listed in chapter 2, section 2.5 using a model-based observer. The estimation error of the observer is zero for normal conditions and non-zero when a fault occurs. For some faults, the error would be in a fixed direction, thus fault localization could also be achieved for this set of faults. In realty, the estimation error would be a random process. How can we make a decision under the influence of noise and disturbance becomes complicated.

Thus, the objective becomes to design a fault detection filter and a decision rule that would achieve desired goals. False alarms and failure to detect a fault are two in-

109

evitable errors that would exist. The goal is to achieve a reasonable compromise. The two types of decision errors can be characterized by their respective first hitting times of a boundary by a random process. The average first hitting times are dependent on the design of the observer and the decision rule. Costs corresponding to these two types of decision error are defined by the cost functions constructed based on experience and knowledge of system operations. The total average cost is minimized over the observer gain and the decision rule until an optimal solution is found.

The optimization process can be formulated as follows.

The objective is to find optimal values for the observer gain L, the filter frequency a and an optimal decision policy characterized by a threshold value $\mathbf{e}_c$ by minimizing the total average cost J.

There are two additional constraints on the observer gain L. For the observer to be stable and for it to be able to generate an estimation error that would be in a fixed direction for a fault vector $\mathbf{v}$, it has to satisfy that:

1. convergence: eigenvalues of $(A - LC)$ must have negative real parts

2. detectable: $rkC[I, (A - LC), \ldots, (A - LC)^{n-1}]\mathbf{v} = 1$

By constructing the gain matrix L using Equation 2.63, the second constraint can be met automatically. Then, the choice of L becomes the choice for L' and p_i. The first constraint has to be incorporated into the cost function so that it can be satisfied by the optimization process itself.

The procedure for optimization is to pick a set of design of the parameters L', p_i, a, and $\mathbf{e}_c$, and evaluate it based on the corresponding cost until a minimal of the cost is found.

The total average cost J can be formulated as:

$$J = w_{10}C_{10}(\overline{\tau_{10}}) + w_{01}C_{01}(\overline{\tau_{01}}) + w_{eig}\sum_{i=1}^{n} C_{eig}(\lambda_i) \tag{4.57}$$

where w_{10}, w_{01} and w_{eig} are weighting factors, C_{10}, C_{01} and C_{eig} are cost functions, $\overline{\tau_{10}}$ and $\overline{\tau_{01}}$ are the average first hitting times, and λ_i is the ith eigenvalue of the matrix

$(A - LC)$. Here the subscript (10) denotes false alarm and the subscript (01) denotes failure to detect a fault. J is a function of L, a, and $\mathbf{e}_c$. Thus, the optimization problem is solved by:

$$[L \quad a \quad \mathbf{e}_c] = arg \min_{L,a,\mathbf{e}_c}; J(L, a, \mathbf{e}_c) \qquad (4.58)$$

The procedure for the optimization scheme is as follows.

1. Given system matrices A, C, and fault vector $\mathbf{v}$, find a fault detection gain generator g such that the estimation error would be along a fixed direction in the output space. The gain L can be written as $L = [p_1 g + \ldots + p_k A^{k-1} g + A^k g] \cdot [(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T] + L'[I - C\mathbf{v}[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T]$ where L' is an arbitrary matrix with proper dimensions. The algorithm for determining g and k in the above expression is described in Appendix A.

2. Choose arbitrary numbers p_1 to p_k and an arbitrary matrix L' and obtain the gain matrix L. By this construction, the L is guaranteed to satisfy the constraint that $rkC[I, (A - LC), \ldots, (A - LC)^{n-1}]\mathbf{v} = 1$, so that the corresponding fault signal will along the fixed direction of $C\mathbf{v}$ in the output space.

3. Given matrices A,C, and L, obtain the transfer function from noise to output $W(s)$ and that from the fault signal to the output $G(s)$. Here $W(s) = [I - C(sI - (A - LC))^{-1}L]$ and $G(s) = C[sI - (A - LC)]^{-1}\mathbf{v}$.

4. Given $W(s)$, estimate $\overline{e^2}$, the sample variance of the output and the system relaxation time τ of the filtered signal.

5. Given $G(s)$ and a decision threshold $\mathbf{e}_c$, calculate the first hitting time boundaries $R_{01}(t)$ and $L_{01}(t)$ for non-detection of the filtered signal.

6. Given $\overline{e^2}$, τ and $\mathbf{e}_c$, estimate the average first hitting time $\overline{\tau_{10}}$ for a false alarm using the Fokker-Planck equation (or Monte Carlo simulation).

7. Given boundaries $R_{01}(t)$ and $L_{01}(t)$, $\overline{e^2}$ and τ, estimate the average first hitting

time $\overline{\tau_{01}}$ for non-detection using the Fokker-Planck equation (or Monte Carlo simulation).

8. Given $\overline{\tau_{10}}$, calculate the average cost from false alarm $cost_{10}$.

9. Given $\overline{\tau_{01}}$, calculate the average cost from non-detection $cost_{01}$.

10. Given $cost_{10}$, $cost_{01}$, w_{10} and w_{01}, obtain the total average cost J

11. Minimize J. If the minimum is found, then exit. Otherwise, return to step two.

Step one and two are to give a random try of a set of design and step three to ten are to evaluate this design. Step three, four and six are to evaluate the first hitting time for false alarm. Step three, four, five and seven are to evaluate the first hitting time for fault detection. Step eight, nine and ten are to evaluate the total average cost for this particular design. The details in each step have been covered above. The flow chart is in Figure 4-18.

In the process of implementing the Fokker-Planck equation for estimating the average first hitting time, two problems have to be solved. One is to extract an effective system response time constant τ and the signal variance. The other is to estimate the boundaries for fault detection. These are addressed as follows.

To extract an effective system response time constant: fit the power spectrum $psd(\omega)$ to the Lorentzian spectra function in the form of

$$Lz(\omega) = \frac{b}{\omega^2 + 1/\tau^2}. \tag{4.59}$$

We can either do a curve fitting or by comparing the zeroth and secondth moments. The zeroth moment of $Lz(\omega)$ is $b\tau^2$ and the secondth moment is $\int_{-infty}^{infty} Lz(\omega)d\omega = \pi b\tau$. Thus, we have

$$b\tau^2 = psd(0)$$
$$\pi b\tau = \int_{-\infty}^{\infty} psd(\omega)d\omega \tag{4.60}$$

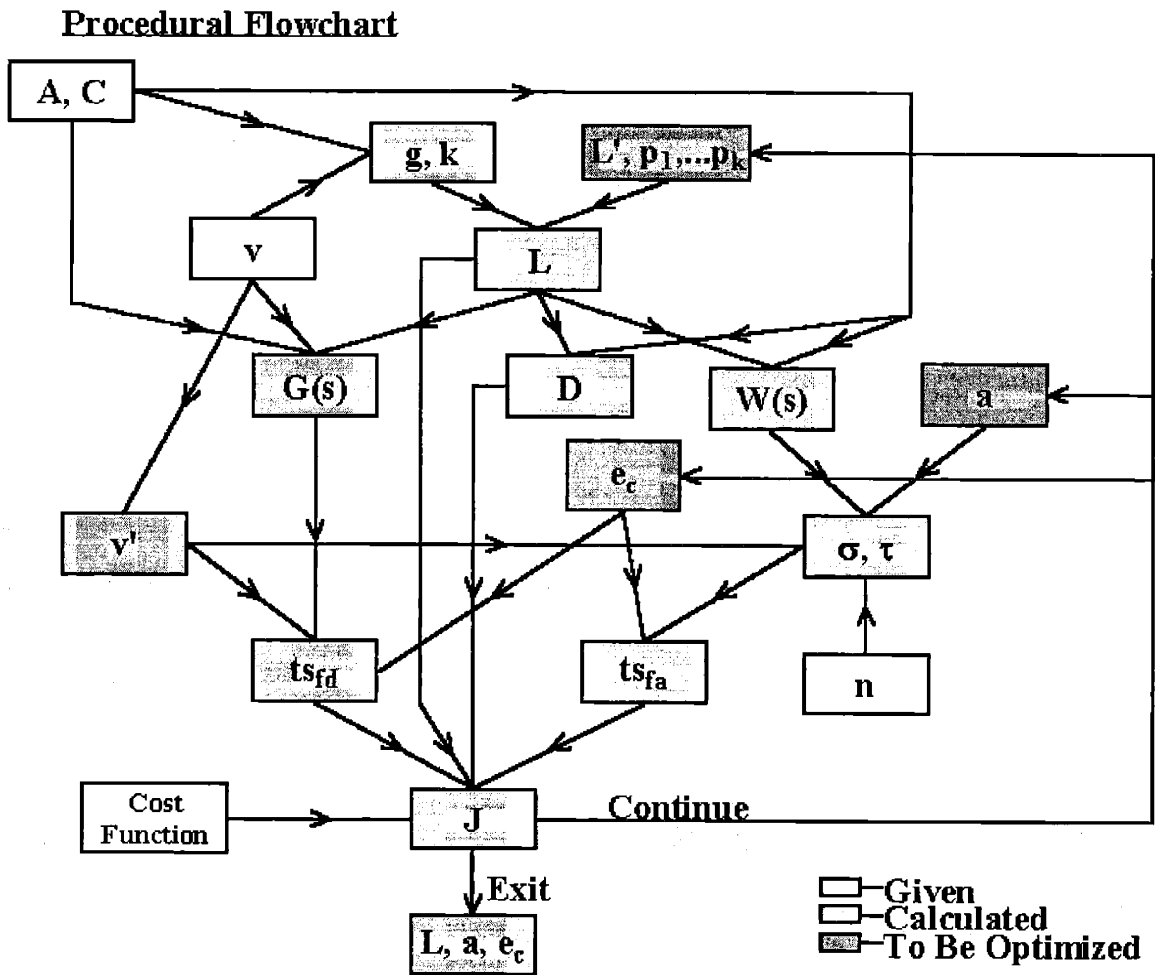$$\tag{4.61}$$

Procedural Flowchart

Figure 4-18: Procedural flow chart for optimization.

113

Solving those two equations, we can have the effective parameters of the system based on it power spectrum:

$$\tau = \frac{\pi psd(0)}{\int_{-\infty}^{\infty} psd(\omega)d\omega}$$

$$b = \frac{\int_{-\infty}^{\infty} psd(\omega)d\omega}{\pi\tau} \tag{4.62}$$

$$\tag{4.63}$$

One example of the fitting result is shown is figure 4-19. The example uses $d = \begin{pmatrix} .0001 \\ 100 \end{pmatrix}$ as the noise power, $a = 0.0105$ as the frequency dependent filter, and

$A = \begin{pmatrix} -65.0 & 0.0768 \\ 65.0 & -0.0768 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $L \begin{pmatrix} -64.9824 & -0.0635 \\ 65.0000 & 0.0475 \end{pmatrix}$. The

fitted b and τ are 45.5329 and 102.9786 respectively. We can see that the Lorentzian spectra approximation is good in this example.

To estimate the boundaries for fault detection: In the frequency domain, the fault component of the estimation error can be written in the transfer function form of $C\mathbf{e}(s) = C(sI-(A-LC))^{-1}\mathbf{v}f/s$ here f is a constant corresponding to the magnitude of the assumed step fault. Upon application of a fault detection filter and a frequency dependent filter to the output, we have the final signal as a scalar in the form of

$$\epsilon(s) = \mathbf{v}'C\mathbf{e}(s)\frac{1}{s+a}$$

$$= \mathbf{v}'C(sI - (A - LC))^{-1}\mathbf{v}\frac{f}{s(s+a)} \tag{4.64}$$

Write the matrix $(A - LC)$ in the eigenvalue form,

$$A - LC = \sum_{i=1}^{n} v_i\lambda_i w_i' \tag{4.65}$$

where n is the dimension of $A - LC$, v_i is its ith right eigenvector, λ_i is the ith eigenvalue, and w_i is the ith left eigenvector. Substitute equation 4.65 into equation
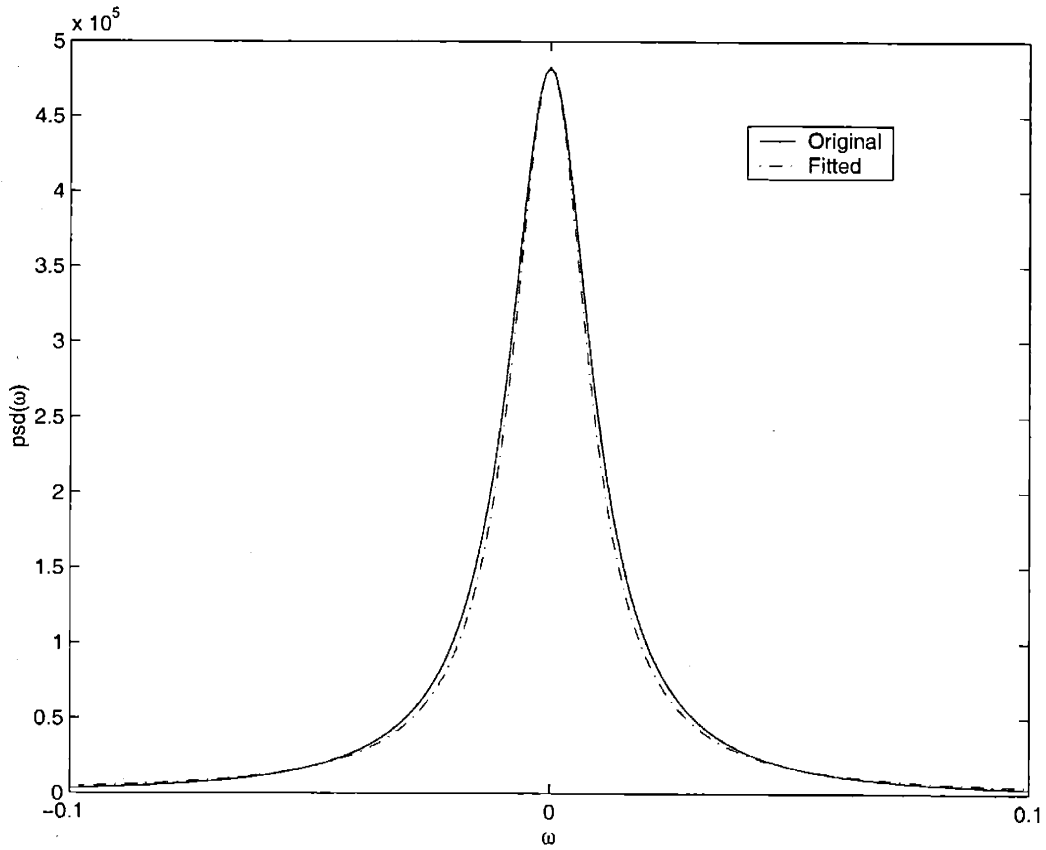
Figure 4-19: Calculated and fitted power spectrum of the noise component of the filtered estimation error. The dash-dot line is the calculated data while the solid line the fitted data using Lorentzian spectra function.

4.64, we have

$$
\begin{aligned}
\epsilon(s) &= \sum_{i=1}^{n} \frac{\mathbf{v}'Cv_i w_i' \mathbf{v} f}{s(s+a)(s-\lambda_i)} \\
&= \sum_{i=1}^{n} \frac{k_i}{s(s+a)(s-\lambda_i)} \\
&= -\sum_{i=1}^{n} \left(\frac{k_i/(a\lambda_i)}{s} + \frac{k_i/(a(a+\lambda_i))}{s+a} + \frac{k_i/(\lambda_i(\lambda_i+a))}{s-\lambda_i} \right) \quad (4.66)
\end{aligned}
$$

Notice that $k_i = \mathbf{v}'Cv_i w_i' \mathbf{v} f$ is a scalar, and so is the final signal $\epsilon(s)$. Thus, in the time domain the signal ϵ is still a sum of a constant component and exponentially decaying components. The constant term is $\sum_{i=1}^{n} k_i/(a\lambda_i)$ while the exponentially decaying terms have decay constant $a, \lambda_1, \ldots \lambda_n$. Also note that the filter pole a should not cancel the system poles λ_i. Otherwise, the filtered signal could be unstable.

$$
\epsilon(t) = \sum_{i=1}^{n} -\left(\frac{k_i}{a\lambda_i} + \frac{k_i}{a(a+\lambda_i)} e^{-at} + \frac{k_i}{\lambda_i(\lambda_i+a)} e^{-\lambda_i t} \right) \quad (4.67)
$$

Figure 4-20 shows the results calculated from equation 4.67 and from a simulation implemented in Matlab Simulink. The two results are exactly equal as they should be. The parameters used in this example are $A = \begin{pmatrix} -65.0 & 0.0768 \\ 65.0 & -0.0768 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $L = \begin{pmatrix} -60 & 50 \\ 65 & 50 \end{pmatrix}$, $f = -1$ and $a = 15$.

Up to here, the method for designing a fault detection filter and its implementation have been covered. In the next chapter, an example for applying this to a nuclear reactor model is presented.
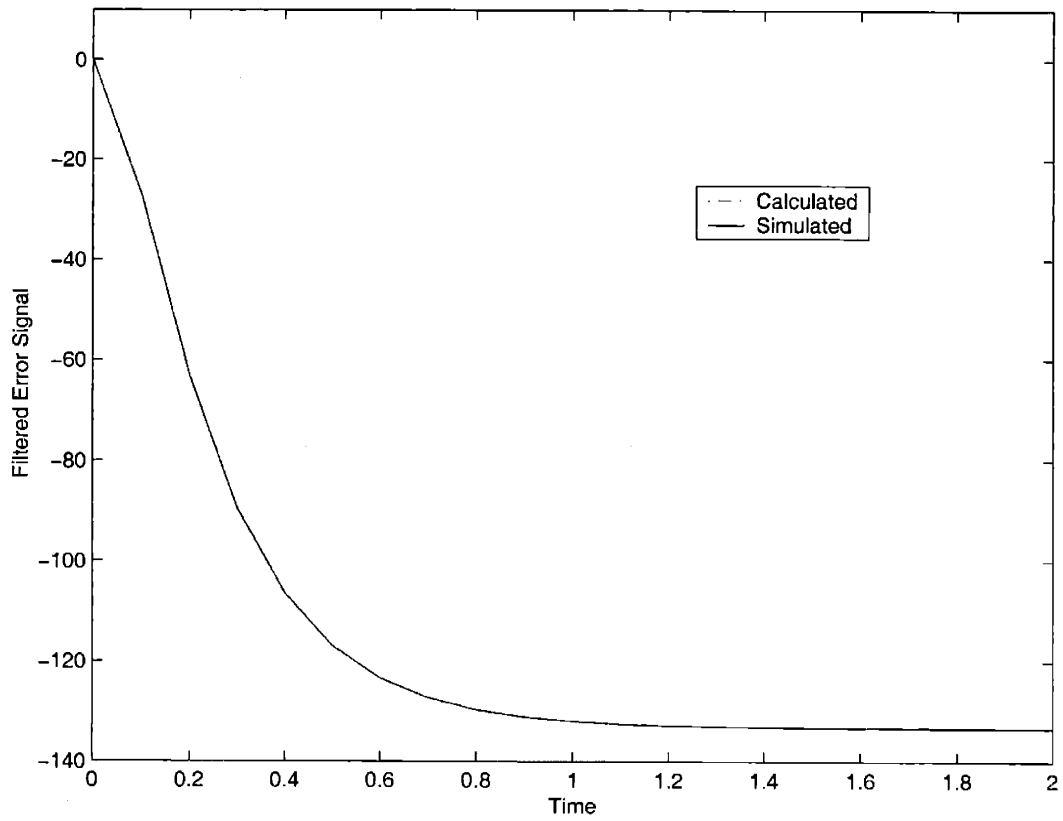
Figure 4-20: Time response of the filtered estimation error from calculation (dash-dot line) and simulation (solid line) for an unit input step fault and no noise.

Chapter 5

Application of the Fault Detection to a Nuclear Reactor

In this chapter the results from previous chapters are applied to reactor model. The fault types included are input fault and system dynamics fault. Sensor faults are not examined here because they can usually be detected by sensor redundancy. Faults are modeled as step functions because most continuous changes are normal deviations. Faults, on the contrary, tend to be abrupt changes and any abrupt changes can be decomposed into step functions. Faults that would change the system model structure are not within the scope of this research.

The observer is based on the one-group linear nuclear reactor model. The fault detection filter is designed based on this observer and the system is simulated using the same model. The resulted design is then applied to detect faults in the six-group nonlinear reactor model and the results are presented.

5.1 Application to One-group Reactor Model

5.1.1 Problem Analysis

For a one-group precursor reactor system, the effective decay constant λ_e should be treated as a time-dependent quantity. However, if we assume that the system is

operating about an equilibrium point, then the system can be linearized and λ_e can also be regarded as a constant. Thus,

$$C_{0i} = \frac{\overline{\beta_i} n_0}{l^* \lambda_i}.$$ (5.1)

The effective one-group decay constant can be calculated using equation 3.18.

$$\lambda_e(t) \simeq \lambda_e(0) = \frac{\sum_0^6 \overline{\beta_i}}{\sum_0^6 \overline{\beta_i}/\lambda_i}.$$ (5.2)

Note that λ_e only depends on $\overline{\beta_i}$ and λ_i, and is independent of n_0 and l^*. Using the parameters in Table 3.1, we have $\lambda_e = 0.0768 s^{-1}$. The prompt neutron life time l^* used in this research is $1 \times 10^{-4} s^{-1}$.

The system matrices according to our derivation given in Equation 3.33 are

$$\mathbf{A} = \begin{pmatrix} -65.0 & 0.0768 \\ 65.0 & -0.0768 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 10^4 \\ 0 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$ (5.3)

Note that the n_0 term in the B matrix is dropped because we are using n_0 as the normalization factor. All state variables including neutron density and delayed neutron precursor densities are multiples of n_0.

The observability matrix O given by

$$\mathbf{O} = \begin{pmatrix} C \\ CA \end{pmatrix},$$ (5.4)

can be verified to have a rank of two. Thus, the pair (A, C) is observable. As we discussed in chapter two, the eigenvalues of (A, C) can then be placed at any location.

119

For the observer to be stable, it is desired that all the eigenvalues have negative real parts.

According to the results in Chapter two on observer-based fault detection, the settled estimation error in the output space will be

$$
\begin{aligned}
C\epsilon(t) &= C \int_0^t e^{(A-LC)(t-\tau)} \mathbf{v} f(\tau) d\tau \\
&= \int_0^t e^{(A-LC)(t-\tau)} \mathbf{v} f(\tau) d\tau.
\end{aligned}
\tag{5.5}
$$

For a fault in the input, the vector $\mathbf{v}$ is b_i and in this example there is only one single input, thus $\mathbf{v} = b = \begin{pmatrix} 10^4 \\ 0 \end{pmatrix} = 10^4 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. $f(t)$ is the time characteristic of the fault modeled as a step function. Introduction of the unit step function as

$$
S(x) = \begin{cases} 1 & x>0 \\ 0 & x<0 \end{cases}.
\tag{5.6}
$$

allows $f(t)$ to be written as $f_c S(t)$ where f_c is an unknown constant.

Thus,

$$
C\epsilon(t) = f_c \int_0^t e^{(A-LC)(t-\tau)} S(\tau) d\tau \, \mathbf{v}.
\tag{5.7}
$$

If we have $rk\, C[I, (A-LC)]\mathbf{v} = 1$ then the settled $C\epsilon$ will remain in a fixed direction. In this case, the fixed direction is along $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, while the magnitude will approach a non-zero constant. Denote the constant as κ. We then have

$$
\begin{aligned}
\kappa &= f_c \times 10^4 \int_0^t e^{(A-LC)(t-\tau)} S(\tau) d\tau \\
&= f_c \times 10^4 \, V \begin{pmatrix} \int_0^t e^{\lambda_1(t-\tau)} S(\tau) d\tau & 0 \\ 0 & \int_0^t e^{\lambda(t-\tau)} S(\tau) d\tau \end{pmatrix} V^{-1} \\
&= f_c \times 10^4 \, V \begin{pmatrix} -\frac{1}{\lambda_1}[1 - e^{\lambda_1 t}] & 0 \\ 0 & -\frac{1}{\lambda_2}[1 - e^{\lambda_2 t}] \end{pmatrix} V^{-1}
\end{aligned}
$$

$$\rightarrow \quad f_c \times 10^4 \ V \left(\begin{array}{cc} -\frac{1}{\lambda_1} & 0 \\ 0 & -\frac{1}{\lambda_2} \end{array} \right) V^{-1}. \tag{5.8}$$

where λ_i is the ith eigenvalue of $(A - LC)$ and $V = \left(\begin{array}{cc} \mathbf{v}_1 & \mathbf{v}_2 \end{array} \right)$. Here, $\mathbf{v}_1$ and $\mathbf{v}_2$ are the right eigenvectors of $(A - LC)$. Thus, given κ, we can get the magnitude of the fault by

$$f_c = \kappa \times 10^{-4} \ V^{-1} \left(\begin{array}{cc} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{array} \right) V. \tag{5.9}$$

So far, we have discussed the characteristics of the error signal for a single step input fault. The output space error signal $C\epsilon$ remains in a fixed direction $\left(\begin{array}{c} 1 \\ 0 \end{array} \right)$ and the magnitude approaches a non-zero constant κ. The fault magnitude can be inferred from κ.

Next, the characteristics of the error signal for a single step system dynamics fault are discussed.

If we do not consider the physical meaning of the components of the system matrix A and B, then every term can be at fault arbitrarily. But, according to the derivation in Equation 3.33, the terms are linked by physics. It is more reasonable to look at the possible faults due to the changes in the physical parameters instead of each term in the matrices. Thus, possible faults of system dynamics include changes in the delayed neutron fraction $\overline{\beta}$, prompt neutron life time l^*, or effective delayed neutron precursor decay constant λ_e.

A fault in $\overline{\beta}$ will result in errors in a_{11} and a_{21}. Assume a step fault in $\overline{\beta}$ such that $\overline{\beta}_S = \overline{\beta}_M + f_c S(t)$. Then, the estimation error signal have the form of

$$\begin{aligned} \dot{\epsilon} &= (A - LC)\epsilon + \mathbf{e}_1 \Delta a_{11}(\mathbf{e}_1' \mathbf{x}) + \mathbf{e}_2 \Delta a_{21}(\mathbf{e}_1' \mathbf{x}) \\ &= (A - LC)\epsilon - (f_c \times 10^4) \left(\begin{array}{c} 1 \\ 0 \end{array} \right) n + (f_c \times 10^4) \left(\begin{array}{c} 0 \\ 1 \end{array} \right) n \end{aligned}$$

121

$$= (A - LC)\epsilon + (f_c \times 10^4) \begin{pmatrix} -1 \\ 1 \end{pmatrix} n. \qquad (5.10)$$

Thus, in the output space, we have

$$C\epsilon = Ce^{(A-LC)t} + (f_c \times 10^4) C \int_0^t e^{(A-LC)(t-\tau)} n(\tau) d\tau \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\rightarrow (f_c \times 10^4) \Psi(t) \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \qquad (5.11)$$

where $\Psi(t)$ is defined as

$$\Psi(t) = C \int_0^t e^{(A-LC)(t-\tau)} n(\tau) d\tau. \qquad (5.12)$$

Here $n(t)$ is a time-varying quantity and is not always directly accessible. But, in this case, it is directly measured. Thus, the time-varying quantity $\Psi(t)$ can be determined.

In summary, a single step fault in $\overline{\beta}$ results in an estimation error settled in a fixed direction along $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ with a magnitude proportional to $(f_c \times 10^4) \Psi(t)$.

A fault in λ_e will result in errors in a_{12} and a_{22}. Express a step fault in λ_e as $\lambda_{eS} - \lambda_{eM} = f_c S(t)$. We then have

$$\dot{\epsilon}(t) = (A - LC)\epsilon(t) + \mathbf{e}_1 \Delta a_{12}(\mathbf{e}_2' \mathbf{x})(t) + \mathbf{e}_2 \Delta a_{22}(\mathbf{e}_2' \mathbf{x})(t)$$

$$= (A - LC)\epsilon(t) + f_c \begin{pmatrix} 1 \\ 0 \end{pmatrix} C(t) - f_c \begin{pmatrix} 0 \\ 1 \end{pmatrix} C(t)$$

$$= (A - LC)\epsilon(t) + f_c C(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad (5.13)$$

and

$$C\epsilon(t) = Ce^{(A-LC)t} + f_c C \int_0^t e^{(A-LC)(t-\tau)} C(\tau) d\tau \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\rightarrow \; f_c \Phi(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{5.14}$$

where the $\Phi(t)$ is defined by

$$\Phi(t) = C \int_0^t e^{(A-LC)(t-\tau)} C(\tau) d\tau. \tag{5.15}$$

Because we can measure $C(t)$ directly in this case, $\Phi(t)$ is a known quantity. A step fault in λ_e thus results in an estimation error along a fixed direction $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ with a magnitude settled at $f_c \Phi(t)$

A fault in l^* will result in errors in a_{11}, a_{21}, and b_1. Express a step fault in l^* as $1/l_S^* - 1/l_M^* = f_c S(t)$. We then have

$$\begin{aligned}
\dot{\epsilon}(t) &= (A - LC)\epsilon(t) + \mathbf{e}_1 \Delta a_{11}(\mathbf{e}_1' \mathbf{x})(t) + \mathbf{e}_2 \Delta a_{21}(\mathbf{e}_1' \mathbf{x})(t) + \mathbf{e}_1 \Delta b_1 u(t) \\
&= (A - LC)\epsilon(t) - (\overline{\beta} f_c) \begin{pmatrix} 1 \\ 0 \end{pmatrix} n(t) + (\overline{\beta} f_c) \begin{pmatrix} 0 \\ 1 \end{pmatrix} n(t) + f_c \begin{pmatrix} 1 \\ 0 \end{pmatrix} u(t) \\
&= (A - LC)\epsilon(t) + (\overline{\beta} f_c) n(t) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + f_c u(t) \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{5.16}
\end{aligned}$$

This is not like the other cases that were discussed previously. Here, the direction of the error signal is not fixed.

In summary, faults result in different characteristics in the output space error signal as follows:

1. Step fault in input $u - u_d = f_c S(t)$ results in a settled error signal that has a fixed direction of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and a magnitude of ($f_c \times 10^4 \ V^{-1} \begin{pmatrix} -\frac{1}{\lambda_1} & 0 \\ 0 & -\frac{1}{\lambda_2} \end{pmatrix} V$).

2. Step fault $\overline{\beta}_S = \overline{\beta}_M + f_c S(t)$ results in a settled error signal that has a fixed direction of $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ with a magnitude of ($f_c \times 10^4$) $\Psi(t)$

3. Step fault $\lambda_{eS} - \lambda_{eM} = f_c S(t)$ results in a settled error signal that has a fixed

123

direction of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ with a magnitude settled at f_c $\Phi(t)$

4. Step fault $1/l_S^* - 1/l_M^* = f_c S(t)$ results in a settled error signal that has no fixed direction.

The quantity p here corresponds to the eigenvalues of the matrix $(A - LC)$.

The estimation error is first projected into the corresponding output direction that the targeted fault should be along. The projected signal is then passed through a filter. That is,

$$\epsilon(t) = \int_0^t e^{-a(t-\tau)d\tau} = \frac{1 - e^{-at}}{a} \tag{5.17}$$

If a is large, then the signal would be settled at a non-zero value quickly, but the settled value would be small. On the other hand, if a is small, the settled value would be large, but the time for the signal to settle would be long. A larger value of a would result in a fast detection of the fault but a smaller possibility of detecting it. In contrast, a smaller a would result in slower detection of the fault but a larger possibility of detecting it. The step fault f in the frequency domain can be written as f/s.

5.1.2   Fault Detection Filter Design and simulation

As discussed previously, there are two constraints that the matrix L has to satisfy. They are,

1. All eigenvalues of $(A - LC) = A - L$ are in the left half plane.

2. $rk\ C[I(A - LC)]\mathbf{v} = [I(A - L)]\mathbf{v} = 1$

The optimization goal is to minimize the total average cost for false alarms and missed detections. A more thorough treatment would require two dimensional cost functions, in both delay time and the magnitude of the fault. That is, we would assign a cost function that varies according to how long the fault has occurred before it is detected and how severe the fault is. The total average cost would then need to

be averaged over both detection time and fault magnitude. The á priori probability of fault magnitude has to be established based on the understanding of the system and operational experience. Here, we use a fixed magnitude and average over time. This method can be generalized to two dimensional cost functions.

For the second constraint to be confirmed, the maximal generator algorithm [11] is used to find a set of L that will always satisfy the constraint condition. The Matlab source code was included in Appendix A.

For the matrices A and C in equation 3.33, and for the fault vector $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, the generator has been calculated to be

$$g = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{5.18}$$

Thus, any gain matrix L that is in form of

$$
\begin{aligned}
L &= [pg + Ag] \cdot [(C\mathbf{v})^T C\mathbf{v}]^{-1} (C\mathbf{v})^T + L'[I - (C\mathbf{v})[(C\mathbf{v})^T C\mathbf{v}]^{-1} (C\mathbf{v})^T] \\
&= p \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -65 & 0 \\ 65 & 0 \end{pmatrix} + \begin{pmatrix} l'_{11} & l'_{12} \\ l'_{21} & l'_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} p - 65 & l'_{12} \\ 65 & l'_{22} \end{pmatrix}
\end{aligned}
\tag{5.19}
$$

will satisfy the second constraint. Here l'_{12} and l'_{22} are arbitrary real numbers, p is an arbitrary positive real number and $-p$ is one eigenvalue of $(A - LC)$. The matrix $(A - LC)$ then becomes

$$A - LC = \begin{pmatrix} -p & 0.0768 - l'_{12} \\ 0 & -0.0768 - l'_{22} \end{pmatrix} \tag{5.20}$$

We can see that the two eigenvalues of $A - LC$ are $-p$ and $-0.0768 - l'_{22}$ which are totally selectable by the designer.

The fixed direction in the output space for fault vector $\mathbf{v}$ will be $C\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in this case.

The guideline in choosing the parameters l'_{12}, l'_{22}, a, and e_c is as follows. We want a high signal-to-noise ratio for the final filtered signal. At the same time, the shorter the dead time at the observer initialization period the better. This dead time is a result of the first term in equation 2.54. It is obvious that in order to enhance the signal-to-noise ratio, a small a is desirable because this means that the integration period is longer. Thus, the noise will be smoothed out. But if a is too small, the system becomes sluggish and the fault takes longer to be detected. The effect of the terms in L is complicated and there is no easy way to describe it. For this particular case, where C is an identity matrix, intuitively, we would like the L non-diagonal terms to be close to zero. Thus the $C(sI - (A - LC)^{-1}L$ can be close to I. For the dead time of the observer to be shorter, the magnitude of the real parts of the eigenvalues of $(A - LC)$ should be larger. Remember that all the eigenvalues of $(A - LC)$ should have negative real parts.

As an example, the measurement noises are assumed to have powers of 1×10^4 and 100 for measurements that are normalized by the equilibrium neutron density. These correspond to about 1% noise-to-signal ratio.

Assume a 1 $m\beta$ unit step fault in the actuator. The results for different fault detection filter designs are illustrated in below.

First try a choice of $l'_{12} = 0$, $l'_{22} = 0.1$, $p = 0.1$ $a = 0.1$ which would satisfy all our constraints. The eigenvalues of $(A - LC)$ are -0.0020 and -0.0788. They have negative real parts as they should be. Their absolute values are not big. Thus, the signal's standard deviation is 9. The signal for the fault component is depicted in Figure 5-1 and the settled value is around 6.5. Thus, it is almost impossible to detect the fault because, even with a threshold of 6.5, the average time rate for a false alarm due to noise is about 2 seconds.

Adjust the parameters to $l'_{12} = 0$, $l'_{22} = 0.002$, $p = 0.002$ $a = 0.002$ and calculate the corresponding response. The results are that the eigenvalues of $(A - LC)$ are
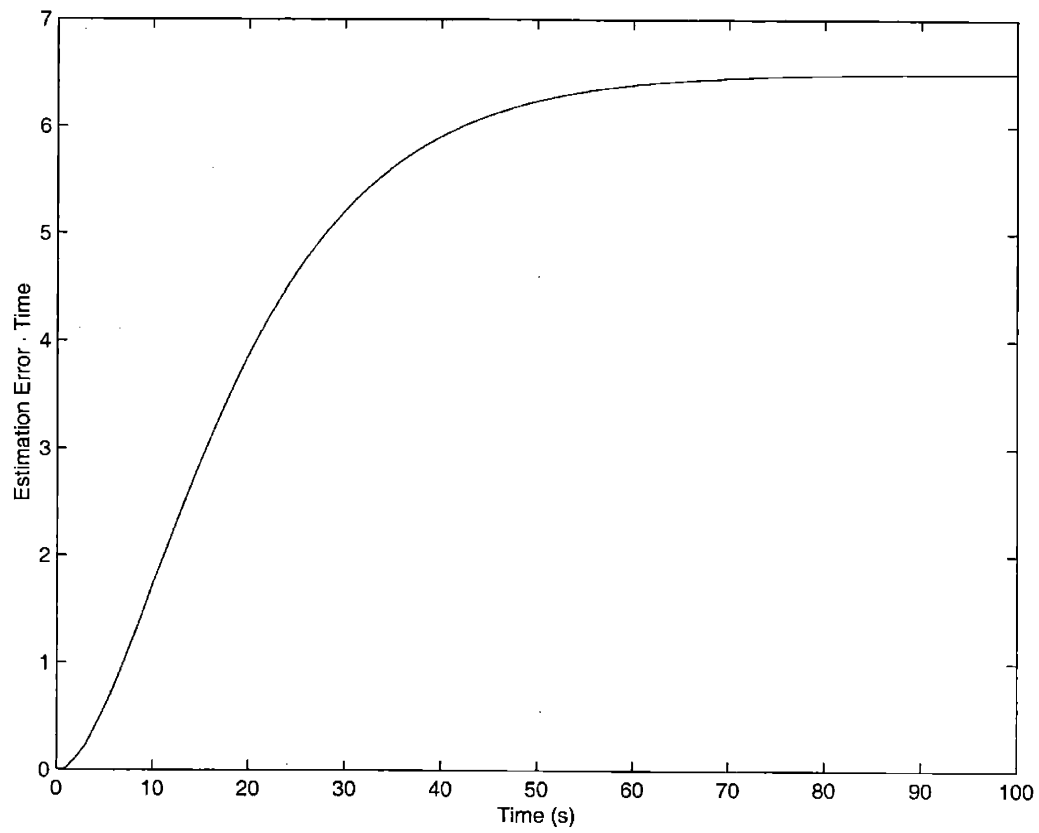
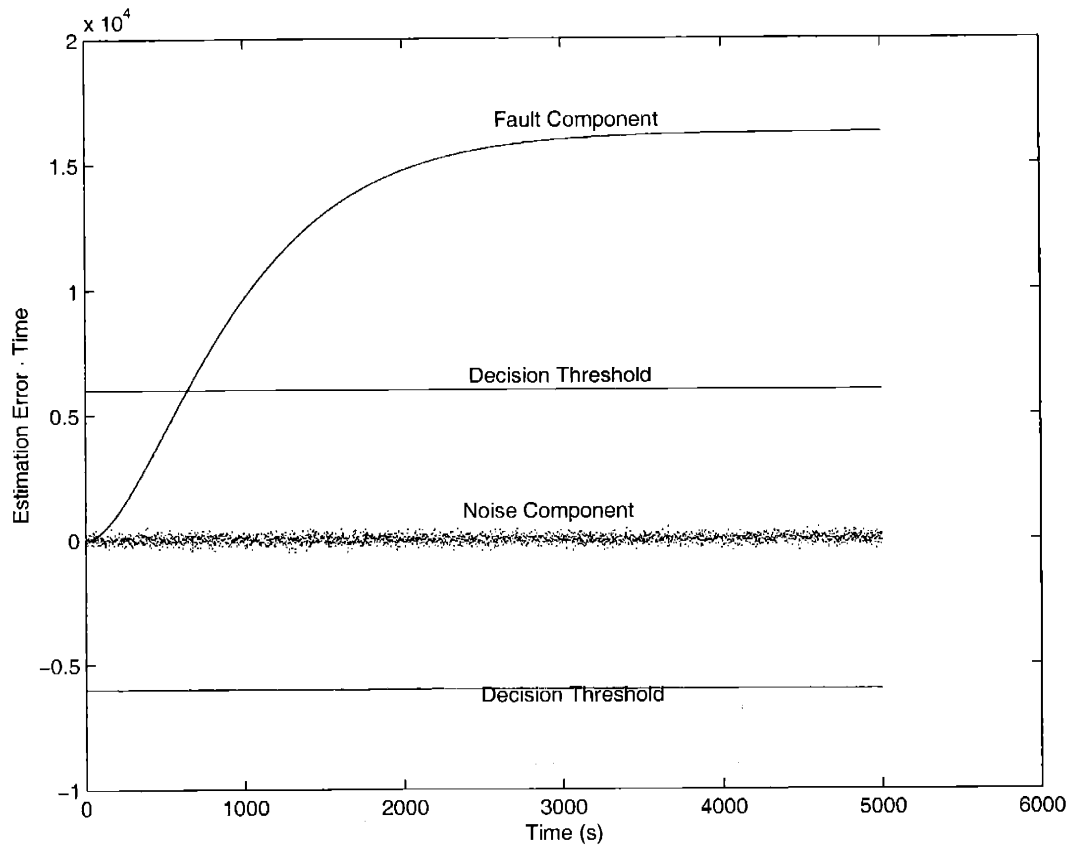Figure 5-1: The response of fault component of the filtered signal.

Figure 5-2: The response of the two components of the filtered signal for the design that $l'_{12} = 0$, $l'_{22} = 0.002$, $p = 0.002$ $a = 0.002$.

-0.0020 and -0.0788. The signal standard deviation is 169. The signal for the fault component is depicted in Figure 5-2 and the settled value is around $1.62e4$. Thus, if we choose a threshold of 6000, the fault can be detected within 645 seconds (about 10 minutes) while the false alarm rate is once in $4.3385e + 274$ weeks, which is a highly unlikely event. The cost is 0.7562. But this result is not optimal.

Using the optimization that has been implemented in Matlab, the results give $cost = 0.0213$, and the average false alarm rate is once per 64 weeks, average fault detect time is 108 seconds. The design parameters are $l'_{12} = -0.0084$, $l'_{22} = 0.0106$, $p = 0.0135$ $a = 0.0160$, and $e_c = 141.5$. The standard deviation for the filtered signal is about 22. Simulation implemented in "simulink" has been run using this design and the results are shown in Figure 5-3. The fault is detected in 73 seconds.
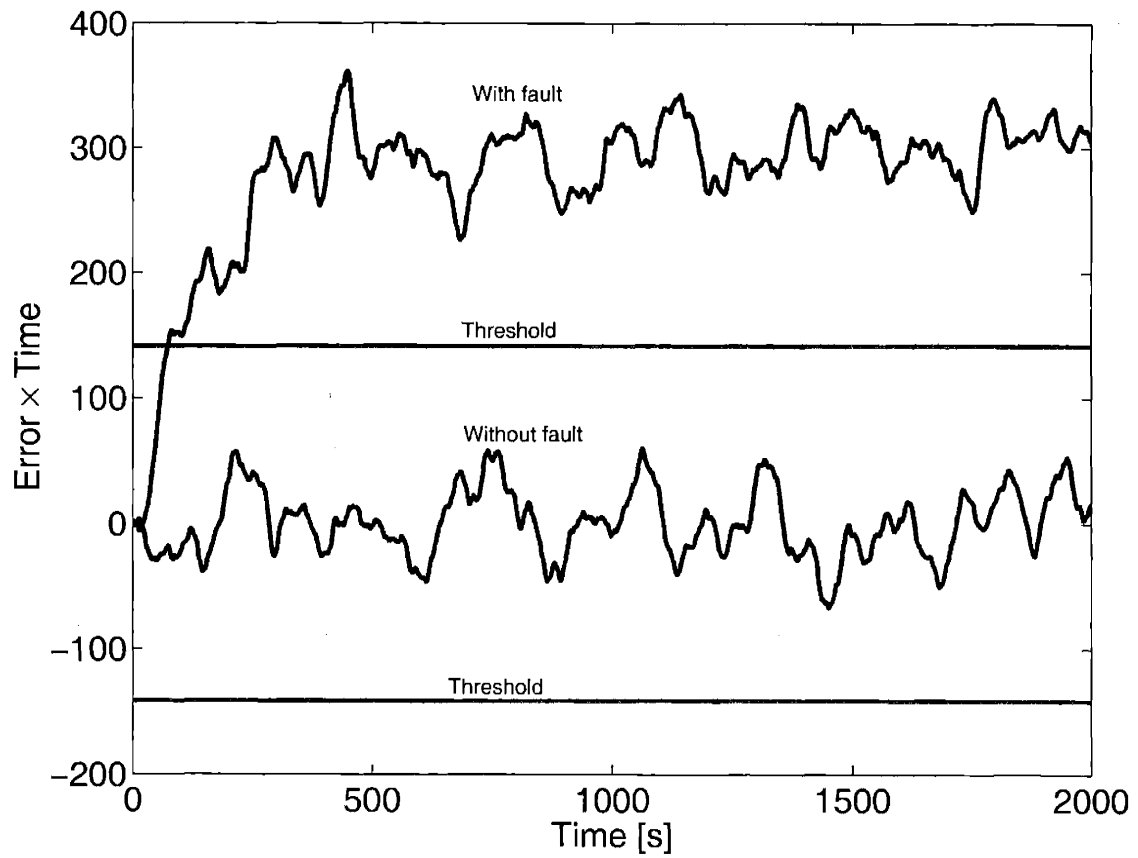
Figure 5-3: The response of the filtered estimation error signal. The upper curve is for simulation with a 1 $m\beta$ actuator fault. The lower curve is for simulation with no fault.

This fault detection design can not only detect a fault of the input, but also it is only sensitive to the type of fault in the output signal that would be along the fixed detection $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Thus, the fault detection filter also localizes the fault at the same time. If the fault is in λ_e instead, the resulting estimation error will have a fixed direction of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ as analyzed previously. Thus, the filter that has been designed to detect actuator faults will be insensitive to a fault in λ_e. The simulation results for a step error of 1.3% in λ_e such that the A matrix becomes

$$\mathbf{A} = \begin{pmatrix} -65.0 & 0.0767 \\ 65.0 & -0.0767 \end{pmatrix} \tag{5.21}$$

are displayed in figure 5-4. It is clear that it will not trigger the fault detection alarm. When both error present, the fault is detected in 160 seconds.

For a fault that is associated with $\mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, the generator is $\mathbf{g} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ with order of $k = 1$. Then the matrix L can be written as:

$$
\begin{aligned}
L &= [pg + Ag] \cdot [(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T + L'[I - (C\mathbf{v})[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T] \\
&= \frac{p}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + 32.5 \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} l'_{11} & l'_{12} \\ l'_{21} & l'_{22} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} p - 65.0768 + l'_{11} + l'_{12} & -p + 65.0768 + l'_{11} + l'_{12} \\ -p + 65.0768 + l'_{21} + l'_{22} & p - 65.0768 + l'_{21} + l'_{22} \end{pmatrix}.
\end{aligned}
\tag{5.22}
$$

As in the previous case, P, l'_{11}, l'_{12}, l'_{21}, l'_{22}, are arbitrary numbers we need to decide. Here we can see that the effect of l'_{11}, l'_{12}, l'_{21}, and l'_{22} are very complicated and it is difficult for human intelligence to figure out these numbers without the help of computers.

Optimization was carried out for this direction. The optimal results are $L' = \begin{pmatrix} -31.0 & 0.01 \\ 0.0 & 33.0 \end{pmatrix}$, $p = 33.0$, $a = 0.002$, $e_c = 160$. The simulation was run and results are shown in Figure 5-5. In Figure 5-5, both curves are for one-group observer and
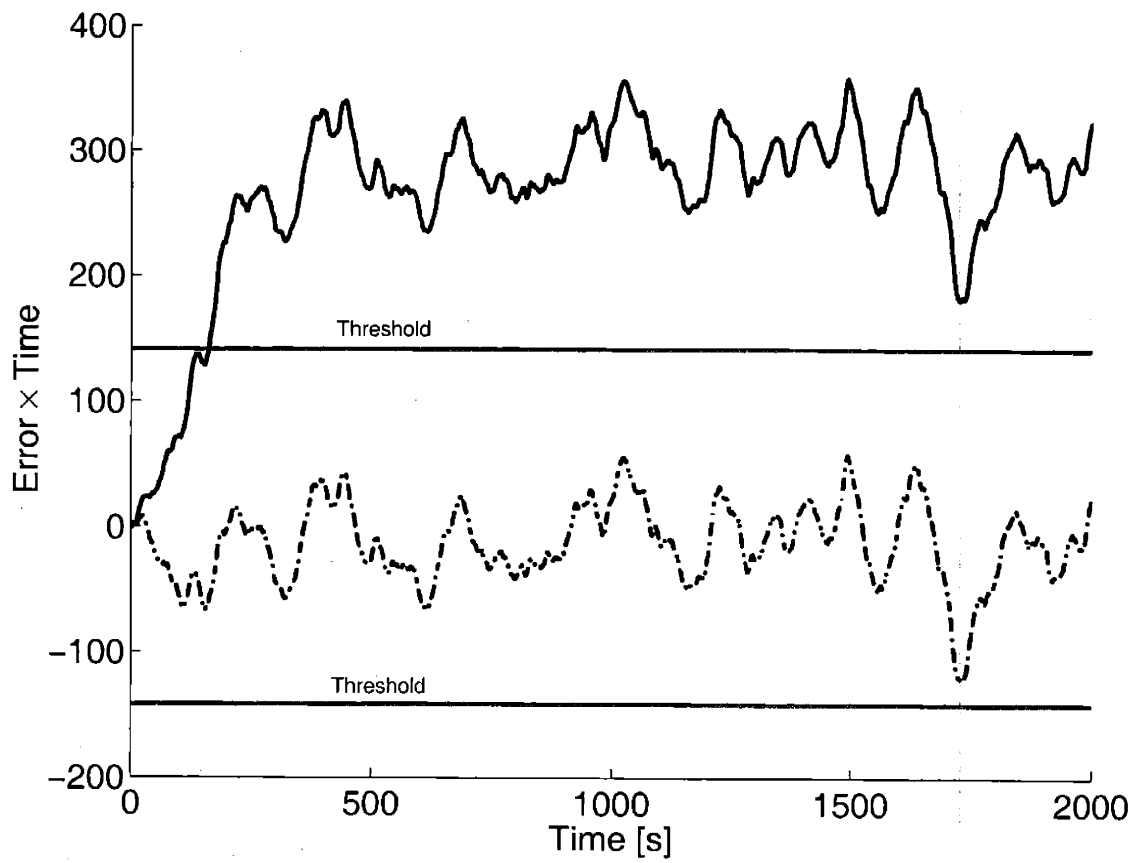
Figure 5-4: The response of the filtered signal. Both curves have a 1.3% error in λ_e. The upper curve also has a 1 $m\beta$ actuator fault.

Figure 5-5: The response of the filtered signal. The system is a one-group linear nuclear reactor model while the observer is a one-group linear nuclear reactor model. The lower curve is when there is no fault. The upper curve is with a 1% λ step fault. Both with noise added to the measurements.

one-group system. The upper curve has a step fault of 1% in λ_e, while the lower curve do not have any fault. We can see that when there is no fault in this direction, the fault detection signal is a random signal with mean of zero and standard deviation of about 36. With a decision threshold of 160, the average false alarm rate is once per 15 weeks, while the average detection time for a 1% fault in λ_e is 125 seconds.

5.2 Application to Six-group Non-linear Reactor Model

To test the robustness of the fault detection filter against modeling error, the designs that have been obtained based on a one-group linear model are applied to a system

simulated by a six-group nonlinear model. The implementation of the simulations is done, again, using "simulink" of the Matlab package.

For a fault in the actuator that will result in the output fault signal in the fixed direction of $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, the filtered signal standard deviation σ without fault is 27, while the settled signal is around 300. The simulation is run for 1e5 seconds, with and without a $1m\beta$ actuator step fault. In the former run, there is no false alarm reported. In the latter run the fault is detected in 105 seconds. The two responses are shown in Figure 5-6. This confirms our belief that the modeling error would behave similarly to a measurement noise on the effect on the detection of a actuator fault. Thus, by constructing a targeted fault detection filter, the modeling error would not affect our ability to detect the fault signal that is along the targeted direction of an actuator fault.

In Figure 5-7, both curves are for one-group observer and six-group nonlinear system. The upper curve has a step fault of 1% in λ_e, while the lower curve do not have any fault. We can see that, even without a true fault, the signal is not centered around zero. Remember that we are trying to detect a fault in the six-by-six system matrix A_S using a two-by-two matrix in the observer. There is the modeling error due to the increase of dimension in the system matrix. Also the system has a non-linear term due to the coupling between input reactivity and neutron density. Thus, these modeling errors do affect our ability to detect a system dynamics fault. If we still use the same design as in the previous section, we will get false alarms almost once every ten minutes. That is not acceptable in real world. However, if we take this modeling error into account when choosing the threshold, we can revise the threshold to 260. The average false alarm rate then becomes once per tens of weeks while the fault detection time is about 300 seconds. This performance is acceptable.
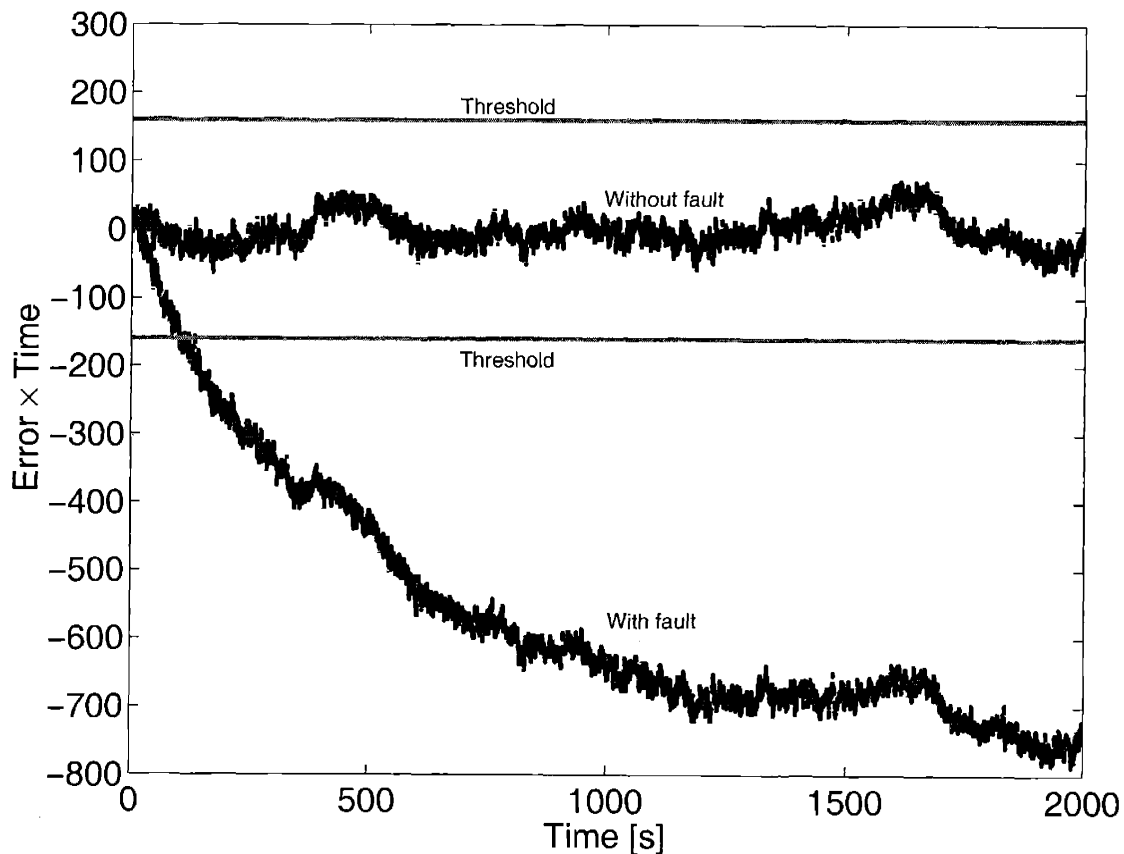
Figure 5-6: The response of the filtered signal. The system is a six-group nonlinear nuclear reactor model while the observer is a one-group linear nuclear reactor model. The lower curve is when there is no fault. The upper curve is with a $1m\beta$ actuator step fault. Both with noise added to the measurements.
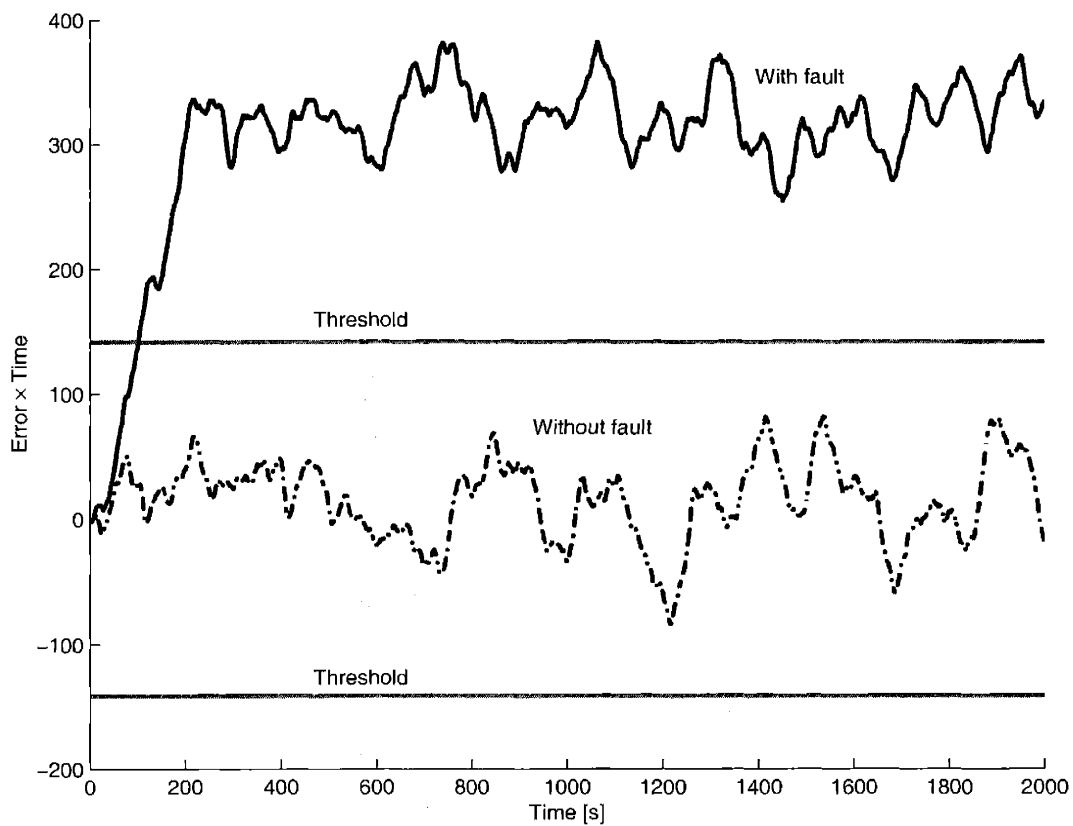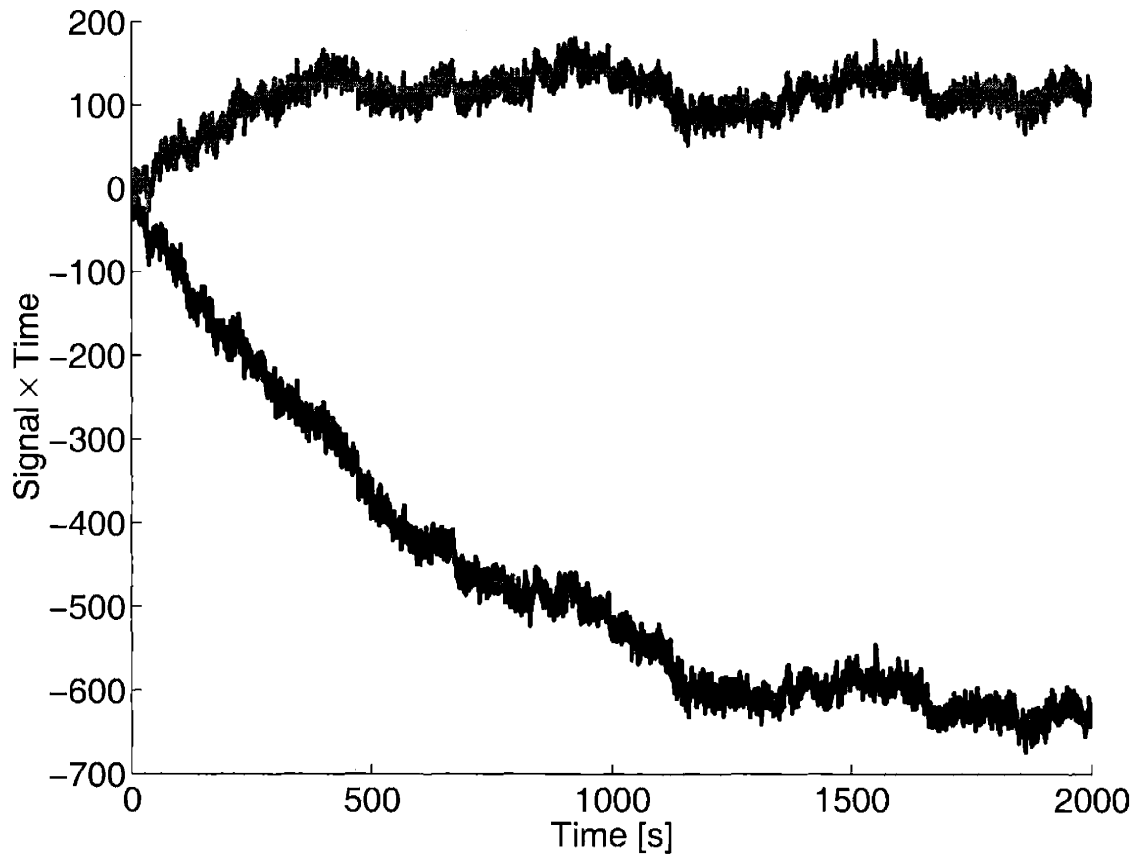
Figure 5-7: The response of the filtered signal. The system is a six-group nonlinear nuclear reactor model while the observer is a one-group linear nuclear reactor model. The lower curve is when there is no fault. The upper curve is with a 1% λ step fault. Both with noise added to the measurements.

Chapter 6

Summary

The motivation of this research was a series of experiments that were done by Dr. John A. Bernard and Francis J. Wyant during 1988 to 1992, in which the need for automated fault detections for an autonomous spacecraft nuclear reactor was demonstrated [3].

The methods that were available are quantitative methods versus qualitative ones as listed in Section 1.2.1. For a system like nuclear reactor, the quantitative methods are more desirable for two reasons. First, a mathematical description of the process is already available during design phase which makes a quantitative method possible. Second, a qualitative method needs the cause-consequence trees (CCT) which requires a great deal of effort for their construction, e.g. a CCT for a nuclear plant requires approximately ten man-years [15]. In addition, a CCT is plant specific and does not allow feedback. Within the quantitative methods, I prefer the observer-based method which compares the outputs of the system to that from a model than the direct estimation approach because of the difficulty of simultaneously estimating many process parameters by direct estimation. Furthermore, because there is no accurate models for the actuators, the direct estimation approach can not detect faults in the actuators while the observer-based method has the potential to detect and localize all three types of faults - system dynamics fault, and sensor fault - based on the system process model only.

However, immediate application of the observer-based fault detection to a nuclear

reactor was not possible because of two unsolved problems, the noise effect and the modeling error effect. The existing theories on observer-based fault detection were only good for deterministic processes. While for any real application, all the signals are random due to the noise. To overcome these problems, my research involved the following tasks, as was shown in Fig. 1-2.

I first designed the overall fault-detection scheme for a linear invariant system (LTI) with measurement noises. To implement this scheme, I introduced the concept of first hitting time to quantitatively describe the performance criteria. I then constructed cost functions which are direct mapping from real world judgments into mathematical descriptions. Then, a optimization procedure was developed to minimize the total average cost and obtain an optimal design of the fault detection process. The method developed in the research for applying the observer-based method to noisy processes is in a general form that can be applied to any LTI system with noise. Only the constructions of cost functions are subject to specific requirements. Although in this report, the noise is modeled as Gaussian White noise, this is not a necessary requirement. To use the Fokker-Planck equation or my empirical model for estimating the first hitting time, the noise must be Gaussian but not necessarily White. For a more general case where the noise is not Gaussian, the first hitting time can be estimated using Monte Carlo simulation. Thus, the method developed in this research can be applied to general cases where the noise is not Gaussian White noise. But as mentioned before, using Monte Carlo simulation is costly in terms of execution time.

This general method was applied to a nuclear reactor model. Simulation results have shown that the desired performance can be met using the optimal design obtained for a nuclear reactor system described by the point-kinetic equations with measurement noises added. In addition, even with modeling errors that exist between a one-group linear reactor model and a six-group non-linear reactor model, it is still possible to detect the three types of faults. However, because a real system is much more complicated than the models we tested, unexpected modeling errors could exist. Thus, it remains to implement the method on an actual reactor and

this requires calibrated system models and a systematic investigation of the modeling error effects.

What have been accomplished in this research is illustrated as follows.

The starting point of this research are the theories on observer and observer-based fault detection filters that were developed in 1961 by Kalman and 1971 by Beard respectively. Only deterministic situations are discussed in these theories. My research extended this approach to address stochastic signals when noise is present. In the design of the fault detection filter, there are total of $n \times n$ degrees of freedom, only n of which have clear physical meaning. The other degrees of freedom have not been addressed at all in the current theories. But in fact, they are important in controlling the final signal's sensitivity to noise and true faults. There is no analytical solution to the determination of these "mysterious" numbers. I proposed to design them by using numerical optimization.

On one hand, because of the noise, there will be false alarms and delays in detecting true faults. Minimization of these two problems is our criteria for performance. On the other hand we have a total of $[(n \times n) + 2]$ degrees of freedom in the design that can be tuned by a numerical optimization. (The other two degrees of freedom are the post-processing filter frequency and the decision threshold.) How can we link them together? We have to have a mathematical description of our performance criteria so that those criteria can be implemented in the optimization procedure. In this research, I proposed to use the average first hitting time which is defined as the average first time of a threshold being triggered by a random process. The random process could be the filtered signal that was driven by noise only or by both noise and a fault.

The first hitting time problem can be solved either by Monte Carlo simulation or Fokker-Planck calculation. Both methods are implemented and tested against each other. The former is more flexible and more general, but requires more execution time, especially for a long hitting time. Because the first hitting time has to be evaluated for each optimization step, a slow algorithm would greatly slow down the optimization process. On the other hand, Fokker-Planck gives an analytical solution

138

which can be implemented quickly. But, it is only valid for certain systems whose power spectrum density curve can be approximated by a Lorentzian function. Finally, I developed an empirical model for symmetrical and constant boundaries and the results are compared with those from the Fokker-Planck calculation. The accuracy is satisfactory. Based on a particular problem, different methods can be used.

Cost functions that take in the average first hitting time and give cost are constructed. They are the key piece to link the design and the performance criteria.

This method was applied to a nuclear reactor. The model is a one-group linearized point kinetics model. Optimization was carried out by designing a fault detection filter that is targeted for only actuator fault. The optimal design obtained was implemented by simulation. When there is no actuator fault, the designed false alarm rate is once per 64 weeks. When there is a 1 $m\beta$ input fault, the average detecting time is 108 seconds. Simulation results confirmed that this filter is sensitive only to the actuator fault that it was designed to monitor. When there is a system dynamics fault (a fault in the system matrix A), it has no influence on either the false alarm rate or the detection time of this actuator fault filter. This fault detection filter was also tested against a non-linear six-group reactor model. The simulation results showed that the designated filter is also robust against this type of modeling error. These results are truly exciting, because the modeling error has long been a problem for model-based control. Our conclusion is that by designing an optimal fault detection filter that is targeted at only one type of fault, it can be robust against noise and modeling error.

Empirical models for the control mechanism and the temperature feedback effects were constructed for MITR-II.

139

Appendix A

Algorithm for Determining the Maximal Generator

A.1 Algorithm for Determining the Maximal Generator

The following algorithm [11] determines the The maximal generator g for vector $\mathbf{v}$ and system (A, C).

Given $n \times n$ system matrices A and C and the n-vector $\mathbf{v}$, matrices C', K, and M' are defined as

$$C' = [I - C\mathbf{v}[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T]C \tag{A.1}$$

$$K = A - A\mathbf{v}[(C\mathbf{v})^T C\mathbf{v}]^{-1}(C\mathbf{v})^T]C \tag{A.2}$$

$$M' = \begin{pmatrix} C' \\ C'K \\ \vdots \\ C'K^{n-1} \end{pmatrix}. \tag{A.3}$$

where I is the identity matrix.

The maximal generator order ν is determined by

$$\mu = n - rank(M').$$

<div align="right">(A.4)</div>

The maximal generator g for vector $\mathbf{v}$ is determined by:

1. Find the null space of M', denoted by $Z = Null(M')$, such that $M' * Z$ has negligible elements, and $Z' * Z = I$.

2. Find a vector g in the null space of M' satisfying

$$\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{\nu-2} \end{pmatrix} g = \mathbf{0}$$

<div align="right">(A.5)</div>

$$CA^{\nu-1}g = CA^{\mu}\mathbf{v}$$

<div align="right">(A.6)</div>

where μ is defined by

$$CA^{j}\mathbf{v} = \mathbf{0} \quad for \quad j = 0 \ldots \mu - 1$$

$$CA^{\mu}\mathbf{v} \neq \mathbf{0}$$

<div align="right">(A.7)</div>

A.2 Matlab Source Code

```
%%
% beard.m                                            %
% find the maximal generator for n-vector f and system (A,C) %
% g is the generator, qp is the rank of M.           %
% Qing Li 2001                                        %
%%


function [g,qp] = beard(A,C,f)


%   C*f can not be zero %
  if norm(C*f) < eps
    fprintf (1,'C*f = %f, stop\n', norm(C*f));
    break;
  end;


  Cp = C - C*f / ((C*f)'*C*f) * (C*f)' * C;
  K  = A - A*f / ((C*f)'*C*f) * (C*f)' * C;
  m = size(C,1);
  n = size(C,2);
  Omega = eye(n);
  qp = 0;
  for L = 0 : n-1,
    for j = 1 : m,
      v = (Cp(j,:) * K^L)';
      w = Omega * v;
      if norm(w) > 1000*eps,
        qp = qp + 1;
        Omega = Omega - w*w' / (w'*v);
      end;
    end;
  end;
  for L = 0 : n-qp-1,
    for j = 1 : m,
      v = (C(j,:) * K^L)';
```

```
    w = Omega * v;
    if norm(w) > 1000*eps,
        Omega = Omega - w*w' / (w'*v);
        jlast = j;
        wlast = w;
    end;
  end;
end;
if norm(Omega) > 1000*eps
    fprintf (1,'** beard: (A,C) is not an observable pair.\n');
    break;
end;
nu = n - qp;
mu = 0;
while norm(C*A^mu*f) < 1000*eps, mu = mu + 1; end;
%    normalize g %
g = C(jlast,:)*A^mu*f/(C(jlast,:)*A^(nu-1)*wlast)*wlast;
```

40

50

143

Appendix B

Sample Matlab Codes for Finding the Optimal Solution to the Fault Detector Design and Decision Rule

```
%%%%%%%%%%%%
% cfuntest.m %
%%%%%%%%%%%%

clear all; more off;
global A C L D v d f e_c a sigma tau;

A = [-65. 0.0768; 65. -0.0768];
% x(1): neutron density [number/cm^3], usually 1e8 %
% x(2): delayed neutron precursor density, usually 1e11 %

C = [1 0; 0 1];
v = [1; 0];

beta = 0.0065; milli_beta = 1e-3 * beta;
delta_u = milli_beta; % control rod error %
f = delta_u * 1e4; % delta_K / K: reactivity error. f is around 0.065 %
```

% baseline sigma of noise is one percent of signal %

d = [1e−4; 100] * 1.0;   20

```
X0 = [
    0.              % L12 -> A12, less stiff %
    0.002           % L22 -> A22, modifies eigenvalue(2) %
    0.002           % -eigenvalue(1) (-p) fixed */
    0.002           % a is frequency scale of post-processing filter */
    6000            % proposed cutoff level [error*s] */
    ];
% cfun(X0); %
% break; %                                                                    30


X0 = [
    -0.503505 0.537805 0.020166 0.004179 233.163338
    ];


% cfun(X0); %


X = fminsearch(@cfun, X0, ...
            optimset('LargeScale', 'off', ...
                'Display', 'off', ...                                         40
                'MaxIter', 20000)...
        );
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%CE
% cfun.m : cost function                              %
% X(1) = l12, X(2)=l22, X(3)=p, X(4) = a, X(5)=e_c     %
% l12,  l22 arbitrary real number                      %
% X(3) arbitrary positive real number, -X(3) is a eigenvalue of (A-LC) %
% X(5) a positive real number                          %
% A, C are system matrices, v is fault vector          %
% d(s) is white noise power vector, which is a constant %
% f(s) is fault signal                                 %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%CE
```

```
function cost = cfun(X)
  global A C L D v d f e_c a sigma tau;


  L = [ X(3)-65 X(1)
        65  X(2) ];


  D = A-L*C;
  a = abs(X(4));
  e_c = abs(X(5));


  [DV DD] = eig(D);
  dd = abs([diag(DD); a]);
  maxw = 100 * max(dd);


  % Find the Lz spectral function parameters %
  psdn0 = eds(0);
  psdnint = quadl(@eds, -maxw, maxw);
  tau = psdn0 * pi / psdnint;
  bn = psdnint/tau/pi;
  y_sigma = sqrt(psdnint/2/pi);
  % sigma for the pre-system noise
  sigma = y_sigma * sqrt(2/tau);


  % calculate the average first hitting time for the false alarm %
```

10

20

30

146

```matlab
if e_c >= 1 * y_sigma,
  avg_ts = tau / max(eps,erfc(e_c/sqrt(2)/y_sigma)) / 2 / pi;
else
  Hbound = e_c;
  Lbound = -e_c;
  Ncut = 30;
  p0 = zeros(Ncut,1);
  for n = 1 : Ncut,
    buf = sprintf('exp(-(%f*x+(%f)).^2).*sin(%f*x)', ...
                (Hbound-Lbound)/pi/sigma/sqrt(tau), ...
                Lbound/sigma/sqrt(tau), n);
    p0(n) = 2 / pi / sqrt(pi*sigma^2*tau) * quadl(buf,0,pi);
  end;
  mu = 1 / tau;
  nu = pi * ( Lbound / tau) / (Hbound - Lbound);
  c = 1 : Ncut;
  N = c' * ones(1,Ncut);
  M = ones(Ncut,1) * c;
  M = M - diag(diag(M));
  NM1 = (-1).^(N+M+1);
  Kts = diag(1 / tau - (c * sigma * pi / (Hbound-Lbound)).^2/2-mu/2) + ...
      M .* ( mu * 2 * N .* NM1 + ...
          nu * 2 / pi * N .* (1 + NM1) ) ./ (N.^2 - M.^2);
  [VKts,DKts] = eig(Kts);
  WKts = inv(VKts);
  twocolumn_N = zeros(Ncut,2);
  avg_ts = 0;
  for n = 1 : Ncut
    for k = 1 : Ncut
      twocolumn_N(n,1) = twocolumn_N(n,1) + ...
        (1 - (-1)^k ) / k * VKts(k,n) * WKts(n,k) * p0(k);
    end
    twocolumn_N(n,2) = DKts(n,n);
    twocolumn_N(n,1) = twocolumn_N(n,1) * (Hbound-Lbound) / pi;
    if ( twocolumn_N(n,2) <= -10*eps )
      avg_ts = avg_ts - twocolumn_N(n,1) / twocolumn_N(n,2);
```

```
      end

    end

    avg_ts = real(avg_ts);

    if avg_ts <= 100*eps,

      avg_ts = 100*eps;

    end

  end


% calculate the average first hitting time for fault detection %            80
t0 = 0; tmax = 100. / min(dd); t = t0 : 0.1 : tmax;

twocolumn = twocolumn_with_a (C,v,D,a);

value = abs(f * twocolumn_value(twocolumn,t));

f_ts = 1e25;

for k = 1 : length(t),

  if value(k) > e_c

    f_ts = t(k);

    break;

  end

end                                                                          90


cost_miss = costmiss(f_ts);

cost_false = costfalse(avg_ts/604800);

cost_initial = .1 * (costmiss(abs(1/D(1,1))) + costmiss(abs(1/D(2,2))));

cost_cond = .01 * (cond(L)/1e4)^2;

cost = cost_miss + cost_false + cost_initial + cost_cond;


if (-D(1,1) <= 100*eps) | (-D(2,2) <= 100*eps) | (a <= 100*eps)

  cost = 1e20;

end                                                                          100


fprintf(1, 'X = %f %f %f %f %f\n', X);

fprintf(1, 'false alarm = %f (wk)\n', avg_ts/604800);

fprintf(1, 'fault detect = %f (s)\n', f_ts );

% fprintf(1, 'signal/noise = %f\n', f*twocolumn(1,1)/y_sigma ); %

fprintf(1, '--------------------------------\n');

fprintf(1, 'costmiss = %f\n', cost_miss);
```

148

```
fprintf(1, 'costfalse = %f\n', cost_false);
fprintf(1, 'costinitial = %f\n', cost_initial);
fprintf(1, 'cost_cond = %f\n', cost_cond);                          110
fprintf(1, 'total cost = %f\n\n', cost);
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%CE
% eds.m                                    %
% calculate the twice filtered error signal psd from noise  %
% W = I - C(sI-D)^(-1)L;                    %
% v is the fault detection filter          %
% 1/(s+a) is the frequency dependent filter %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%CE
function ed = eds(x)
  global C A L v d a;
  D = A - L*C;
  Imatrix = eye(size(C,1));
  for n = 1 : length(x)
    Ws = Imatrix - C*inv(x(n)*i*Imatrix-D)*L;
    Us = (C*v)' * Ws / (x(n)*i + a);
    ed(n) = abs(Us).^2 * d;
  end;
```

10

```
%%%%%%%%%%%%%%%%%%%%%%
% twocolumn_without_a.m %
%%%%%%%%%%%%%%%%%%%%%%%%

% twocolumn representation for (C*v)'*(C*(s-D)^-1/s/(s+a)*v) %
function twocolumn = twocolumn_with_a (C,v,D,a)
  twocolumn = zeros(size(D,1)+2,2);
  twocolumn(1,1) = (C*v)'*(C*inv(0-D)*v)/a;
  twocolumn(1,2) = 0;
  twocolumn(2,1) = (C*v)'*(C*inv(-a*eye(size(D,1))-D)*v)/(-a);
  twocolumn(2,2) = -a;
  [V,e] = eig(D); e = diag(e);
  W = inv(V);
  for n = 1 : size(D,1),
    twocolumn(n+2,1) = (C*v)'*C*V(:,n)*W(n,:)*v/e(n)/(e(n)+a);
    twocolumn(n+2,2) = e(n);
  end;
```

```
function value = twocolumn_value(twocolumn,t)
  value = t * 0;
  for n = 1 : size(twocolumn,1),
    value = value + twocolumn(n,1) * exp(twocolumn(n,2)*t);
  end;
```

```
%%%%%%%%%%%%
% costmiss.m %
%%%%%%%%%%%%

% the cost curve for non-detection %
% the non-detection rate is measured in the time scale of second %
% a fault should be detected in seconds or minutes %

function cost = costmiss(taumiss)
  tau_scale = 200;                                                    10
  cost = 1 - (taumiss/tau_scale+1).^-2.2;
```

```
%%%%%%%%%%%%
% costfalse.m %
%%%%%%%%%%%%

% the cost curve for false alarm %
% the false alarm rate be measured in the time scale of at weeks %
% for example one false alarm in a month is reasonable %

function cost = costfalse (taufalse)
  tau_scale = 8;                                                      10
  cost = (taufalse/tau_scale + 1).^-3;
```

Bibliography

[1] Sheridan, T.B. and W. R. Ferrell, *Man-Machine Systems: Information Control, and Decision Models of Human*, (the MIT Press, Cambridge, MA 1974.)

[2] Kelly, C.R., *Manual and Automatic Control*, (John Wiley and Sons, Inc., New York, NY 1968.)

[3] Bernard, J.A. and F.J. Wyant, "Experiments Illustrating the Importance of Automated Reasoning," IEEE Control Systems Magazine, Vol. 12, No. 2, Apr. 1992, pp 84-92.

[4] Bernard, J.A. and D.D. Lanning, "Fault-Tolerant Systems Approach Toward Closed-Loop Digital Control of Nuclear Power Reactors," CPE-8317878, National Science Foundation, Washingtwon, D.C., Jan. 1988.

[5] Bernard, J.A. and T. Washio, "Expert Systems Applications Within the Nuclear Industry", American Nuclear Society, La Grange Park, IL, Oct. 1989.

[6] Bernard, J.A., Ornedo, R.S., and D.D. Lanning, "Human Approach to Process Control and the Role of Digital Technology," Proceedings of the American Control Conference, Minneapolis, MN, Vol. 2, June 1987, pp 934-940.

[7] Bernard, J.A., "Formulation and Experimental Evaluation of Closed-Form Control Laws for the Rapid Maneuvering of reactor Neutronic Power," Report No. MITNRL-030, Massachusetts Institute od Technology, Cambridge, MA, Sept. 1989.

[8] Bernard, J.A., "Startup and Control of Nuclear Reactors Characterized by Space-Independent Kinetics," Report No. MITNRL-039, Massachusetts Institute od Technology, Cambridge, MA, May 1990.

[9] Bernard, J.A., Henry, A.F., Lanning, D.D., and J.E.Meyer, "Closed-Loop Digital Control of Nuclear Reactors Characterized by Spatial Dynamics ," Report No. MITNRL-041, Massachusetts Institute od Technology, Cambridge, MA, Mar. 1991.

[10] Bernard, J.A., Henry, A.F., Lanning, D.D., and J.E.Meyer, "Studies on the Closed-Loop Digital Control of Multi-Modular Reactors," Report No. MITNRL-049, Massachusetts Institute od Technology, Cambridge, MA, Nov. 1992.

[11] Beard, R.V., "Failure Accommodation in Linear Systems Through Self-Reorganization," PhD thesis, MIT, 1971.

[12] Ron Patton, Paul Frank and Robert Clark,"Fault Diagnosis in Dynamic Systems: Theory and Application," Prentice Hall International (UK) Ltd. 1989.

[13] William R. Simpson and John W. Sheppard, "System Test and Diagnosis", Kluwer Academic, 1994.

[14] Bernard, J.A., "Use of a Rule-Based System for Process Control," IEEE Control Systems Magazine, Vol 8, No. 5, Oct. 1988, pp3-13.

[15] William T. Scherer and Chelsea C. White, III, "A Survey of Expert Systems for Equipment Maintenance and Diagnostics," Singh, M.G., Hindi, K.S., Schmidt, G, and Tzafestas, S.G., "Fault Detection & Reliability - Knowledge Based & Other Approaches," International Series on Systems and Control, Vol 9, Pergamon Press, 1987.

[16] Ray, A. Bernard, J.A., and D.D. Lanning, "On-Line Signal Validation and Feedback Control in a Nuclear Reactor," Proceedings of the Fifth Power Plant Dynamics, Control & Testing Symposium, Knoxville, TN, March 1983, paper No. 38.

[17] Bernard, J.A., Henry, A.F., and D.D. Lanning, "Application of the Reactivity Constraint Approach to Automatic Reactor Control," Nuclear Science And Engineering, Vol. 98, No. 2, Feb. 1988, pp 87-95.

[18] Ornedo, R.S., Bernard, J.A., Lanning, D.D., and J.H. Hopps, "Design and Experimental Evaluation of an Automatically reconfigurable Controller for process plants," Proceedings of the American Control Conference, minneapolis, MN, Vol. 3, june 1987, pp 1662-1668.

[19] Bernard, J.A., Henry, A.F., and D.D. Lanning, "The Design and Experimental Evaluation of a Closed-Loop Digital Controller Based on an Alternate Formulation of the Dynamic Period Equation," Proceedings of the American Nuclear Society Topical meeting on Reactor Physics and Safety, Saratoga Springs, NY, Sept. 1986, pp 610-621.

[20] Bernard, J.A., "Evaluation of Period-Generated Control Laws for the Time-Optimal Control of Reactor Power," IEEE Transactions on Nucelar Science, Vol. NS-35, No. 1, Feb. 1988, pp 888-893.

[21] Bernard, J.A., Kwok, K.S., Thome, F.V., and F.J. Wyant, "The MIT-SNL Experiments Concerning the Closed-Loop Digital Control of Reactor Power," Proceedings of the ANS International Reactor Physics Conference, Jackson Hole, NY, Vol. IV, Sept. 1988, pp 193-206.

[22] Aviles, B.N., Lanning, D.D., and J.A. Bernard, "Design of a Multi-Tiered, Digital Controller for the Supervisory, Global, and Local Control of a Pressurized Water Reactor," proceedings of the ANS Topical Meeting on Advances in Reactor Physics, Charleston, SC, Mar. 1992, pp2-117-2-128.

[23] Bernard, J.A., "Demonstration of Feedback Using the MIT-SNL Minimum Time Control Laws for the Rapid Maneuvering of Reactor Power," IEEE Transactions on Nuclear Science, Vol. NS-38, No. 2 Apr. 1991, pp 838-844.

[24] Kim, K.K. and J.A. Bernard, "Considerations in the Control of Multimodular Reactor Plants," IEEE Transactions on Nuclear Science, Vol. T-NS-41, No. 6, Part II, Dec. 1994, pp 2686-2697.

[25] Bernard, J.A., "Progress Toward a Genetic methodology for the Closed-Loop Digital Control of Nuclear Reactor Power," Proceedings of the ANS/NIST Conference "Fifty Years with Nuclear Fission", Behrens, J.W. and A.D. Carlson, Eds., American Nuclear Society, La Grange Park, IL, Apr. 1989, pp 262-273.

[26] Bernard, J.A. and Lanning, D.D.,"Considerations in the Design and Implementation of Control Laws for the Digital Operation of Research Reactors," Nuclear Science and Engineering, Vol. 110, No. 4, Apr. 1992, pp 425-444.

[27] Bernard, J.A., "Non-Linear Control of Neutronic Power in Reactors Described by Space-Independent Kinetics," International Journal Control-Theory and Advanced Technology (C-TAT), Vol. 8, No.3, Sept. 1992, pp 495-511.

[28] Kalman, R.E. and Bucy, R.S., "New Results in Linear Filtering and Prediction Theory," Trans. ASME (J. Basic Engineering), vol. 83D, no.1, March 1961, pp95-108.

[29] Luenberger, D.G., "Observing the State of a Linear System," IEEE Trans. on Military Electronics, vol. MIL-8, April 1964, pp74-80.

[30] Luenberger, D.G., "An Introduction to Observers," IEEE Trans. on Automatic Control, vol. AC-16, no.6, December 1971, pp596-602.

[31] Kautsky, J. and N.K. Nichols, "Robust Pole Assignment in Linear State Feedback," Int. J. Control, 41 (1985), pp1129-1155.

[32] Ronald Allen Knief, *Nuclear Energy Technology*, (McGraw-Hill Book Company, 1981.)

[33] G. R. Keepin, T. F. Wimett, and R. K. Zeigler, "Delayed Neutrons from Fissionable Isotopes of Uranium, Plutonium, and Thorium," *Phys. Rev.*, Vol. **107**, 1957, pp.1044-1049.

[34] Bernard, J.A., "Issues Regarding the Design and Acceptance of Intelligent Support Systems for Reactor Operators," IEEE Transactions on Nuclear Science, Vol. **NS-39**, No. **5**, Oct. 1992, pp 1549-1558.

[35] Lynn E. Weaver, *Reactor Dynamics and Control* (American Elsevier Publishing Company, INC., New York, 1968).

[36] Watson Fulks, "Complex Variables", Marcel Dekker, NY, 1993.

[37] Sheldon M. Ross, "Introduction to Probability Models", 7th edition (Academic Press, 2000).

[38] N.G. Van Kampen, "Stochastic Processes in Physics and Chemistry", revised edition (North-Holland, 1992).

[39] "6430 and 6460 Series Electromax©III Controllers", Leeds & Northrup Company, North Wales, PA, 1978.