ANOMALY DETECTION IN ELECTROMECHANICAL SYSTEMS
USING SYMBOLIC DYNAMICS

A Thesis in
Mechanical Engineering
by
Amol M Khatkhate

© 2006 Amol M Khatkhate

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2006
The thesis of Amol M Khatkhate was reviewed and approved* by the following:

Asok Ray
Distinguished Professor of Mechanical Engineering
Thesis Advisor, Chair of Committee

Eric Marsh
Associate Professor of Mechanical Engineering

Jeffrey Schiano
Associate Professor of Electrical Engineering

Qian Wang
Assistant Professor of Mechanical Engineering

Eric Keller
Senior Research Associate, Applied Research Laboratory
Special Member

H. J. Sommer III
Professor and Interim Head of the Department of Mechanical and Nuclear Engineering

*Signatures are on file in the Graduate School.
Abstract

An anomaly is defined as a deviation from the nominal behavior of a dynamical system and is often associated with parametric and non-parametric changes that may gradually evolve in time. Anomalies may manifest themselves with self excitation, or under excitation of certain exogenous stimuli. These anomalies may be benign or malignant depending on their impact on the mission objectives and operating conditions.

Major catastrophic failures in large scale engineering systems (e.g., aircraft, power plants and turbo-machinery) can possibly be averted if the malignant anomalies are detected at an early stage. This dissertation experimentally validates a novel method for anomaly detection in electromechanical systems, derived from time series data of pertinent measured variable(s). The core concept of the anomaly detection method is Symbolic Time Series Analysis (STSA) that is built upon the principles of Automata Theory, Information Theory, and Pattern Recognition.

The anomaly detection methodology, investigated in this dissertation, consists of two parts:

- **Forward problem** - The objective in the forward problem is to learn how the grammar of the underlying system dynamics evolves as the system parameters gradually change. The forward problem is that of learning where the value of a parameter is associated with an anomaly measure.

- **Inverse problem** - The inverse problem is that of inferring the system parameters based on the observed asymptotic behavior.

This dissertation deals with both the forward and inverse problems. The performance of this anomaly detection method is compared with that of other existing pattern recognition techniques from the perspectives of early detection of fatigue damage in polycrystalline alloy structures. The experimental apparatus, on which
the anomaly detection method is tested, is a multi-degree of freedom mass-beam structure excited by oscillatory motion of two electromagnetic shakers. The evolution of fatigue crack damage at one of the failure sites is detected from symbolic time series analysis of displacement sensor signal. Also, the dissertation quantifies the progression of damage using other mechanical sensors like accelerometers and load cells and damage sensing devices like ultrasonic flaw detectors. Another mechanical motion apparatus is designed and the accelerometer time series data, generated from this apparatus, has been utilized to detect the change in effective mass/moment of inertia of the system components. The dissertation also exemplifies usage of these techniques to an industrial application as described below.

**Industrial Application** - Critical components such as bearings, seals, and couplings are subjected to unbalanced axial/radial loads and excessive machine vibrations due to shaft misalignment in rotating machinery. The dissertation presents Symbolic Time Series Analysis (STSA) of bearing acceleration for detection and estimation of parametric changes in flexible disc/diaphragm couplings due to angular misalignment between shafts. The concept is validated on a simulation model, where the dynamic model of a flexible mechanical coupling is subjected to angular misalignment. Patterns of damage evolution are identified from symbolic time series analysis of data sets, generated for multiple torque inputs. Small changes in the coupling stiffness parameter(s) are detected and estimated for different torque inputs via inverse mapping based on ensemble of the statistical patterns, generated by STSA.
# Table of Contents

List of Figures viii  
List of Tables x  
Acknowledgments xi  

## Chapter 1 Introduction 1  
1.1 Background of the research . . . . . . . . . . . . . . . . . . . . . . . 1  
1.2 Objective of the research . . . . . . . . . . . . . . . . . . . . . . . 6  
1.3 Contributions of the research . . . . . . . . . . . . . . . . . . . . 7  
1.4 Organization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9  

## Chapter 2 Design of Experimental Testbeds 13  
2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13  
2.2 Design requirements . . . . . . . . . . . . . . . . . . . . . . . . . 13  
2.3 Test apparatus 1 - Resonance machine . . . . . . . . . . . . . . . 14  
2.3.1 Hardware implementation and software structure . . . . . . 18  
2.3.2 Sensors for damage detection and performance measurement 19  
2.3.3 Role of ultrasonics in damage sensing . . . . . . . . . . . . . 21  
2.3.4 Salient features and limitations . . . . . . . . . . . . . . . . 23  
2.4 Modelling of test apparatus 1 for anomaly detection . . . . . . . . . 24  
2.4.1 Physical modelling of test apparatus 1 . . . . . . . . . . . . 24  
2.4.2 System identification of test apparatus 1 . . . . . . . . . . . 35  
2.4.2.1 Open loop frequency domain identification . . . . . 35  
2.4.2.2 Construction of the excitation signal . . . . . . . . . . 36  
2.4.2.3 Frequency response of plant dynamics . . . . . . . . 37  
2.4.3 Data acquisition and estimation of transfer matrices . . . . . . 39
Chapter 3  Symbolic Time Series Analysis (STSA)  
3.1  Introduction ................................................................. 49
3.2  Wavelet space (WS) partitioning .......................................... 54
   3.2.1  Choice of scales and mother wavelet .......................... 55
3.3  The $D$-Markov machine construction ................................... 56

Chapter 4  Experimental Validation of Anomaly Detection  
4.1  Introduction ................................................................. 64
4.2  Experimental validation - test apparatus 1 ........................... 65
   4.2.1  Comparison of anomaly detection methods ....................... 71
   4.2.2  Anomaly detection using multiple sensors ....................... 75
   4.2.3  Sources of uncertainty ............................................... 80
   4.2.4  Ensemble of experimental results .................................. 81
4.3  Life extending control .................................................... 83
   4.3.1  Selection of anomaly measure threshold ........................ 84
   4.3.2  Selection of curvature threshold .................................. 84
   4.3.3  Remaining life prediction ........................................... 86
   4.3.4  Experimental results and discussion ................................ 88
4.4  Experimental validation - test apparatus 2 .......................... 89

Chapter 5  Industrial Application : Flexible Mechanical Coupling  
5.1  Introduction ................................................................. 95
5.2  Problem formulation for anomaly detection ........................... 101
5.3  Description of coupling simulation ....................................... 102
5.4  Simulation results .......................................................... 107
   5.4.1  Solution to the forward (analysis) problem ....................... 107
   5.4.2  Solution to the inverse (synthesis) problem ..................... 112
      5.4.2.1  Confidence interval ............................................. 115
      5.4.2.2  Confidence interval for the mean .......................... 118

Chapter 6  Summary, Conclusions and Future Work  
6.1  Summary and Conclusions ................................................ 120
6.2  Recommendations for future work ....................................... 122
Chapter A Appendix 125
   A.1 Existing pattern recognition techniques ............................... 125
       A.1.1 Principal Component Analysis (PCA) for anomaly detection 126
       A.1.2 Multi-Layer Perceptron Neural Network (MLPNN) 
           for anomaly detection ........................................ 128
       A.1.3 Radial Basis Function Neural Network (RBFNN) 
           for anomaly detection ...................................... 131
   A.2 Electronics and instrumentation of the test apparatus ........... 132

Bibliography 135
List of Figures

1.1 Photograph of test apparatus 1 ........................................ 9
1.2 Photograph of test apparatus 2 ........................................ 10

2.1 Schematic diagram for test apparatus 1 ............................... 16
2.2 Software implementation for test apparatus 1 ...................... 19
2.3 Sensor mounting for test apparatus 1 ................................. 21
2.4 Ultrasonic flaw detection system for 6061-T6, PhD dissertation, Eric Keller, 2001 ......................................................... 22
2.5 Free-body diagram ......................................................... 26
2.6 Side view of failure Site on the beam specimen ...................... 35
2.7 Frequency response of full state and reduced order model ............ 39
2.8 Output Positions v/s Input Voltage to Shaker 1 ....................... 42
2.9 Output Positions v/s Input Voltage to Shaker 2 ....................... 42
2.10 Schematic diagram for test apparatus 2 ............................... 43
2.11 Hydraulic circuit for test apparatus 2 ................................. 45
2.12 Control circuit for test apparatus 2 .................................... 46
2.13 Control circuit in Simulink for the test apparatus .................... 47
2.14 Specimen for fatigue failure ............................................ 48

3.1 Conceptual view of symbolic time series analysis ....................... 53
3.2 Finite state automaton with $D = 2$ and $\mathcal{A} = \{0, 1\}$ .................. 58

4.1 Power spectral density plot of time series data ....................... 67
4.2 Maximum entropy partitioning of scale series data .................... 68
4.3 Anomaly measure for alphabet size 6,8 and 10 .......................... 71
4.4 Performance comparison of anomaly detection methods ................ 74
4.5 Raw time series data for four different sensors ....................... 76
4.6 Normalised anomaly measure plots derived from accelerometer, load cell and displacement sensors ........................................ 79
4.7 Normalised anomaly measure derived from ultrasonic sensors .......... 80
4.8 Ensemble of results under identical loading conditions .................. 82
4.9 Anomaly measure plots of 10 similar specimens under identical loading 85
4.10 Top plate: Selection criteria of anomaly measure threshold; Middle plate: Selection criteria of curvature threshold; Bottom plate: plot of used life and remaining life at the point where curvature threshold is crossed. ................................................................. 87
4.11 Observed and Estimated remaining life after curvature threshold is crossed for 10 samples ............................................................ 88
4.12 Realtime life extension for two experiments with different levels of load reduction ................................................................. 90
4.13 Excitation signal generation ...................................................... 91
4.14 Time series data for the nominal condition .................................. 92
4.15 Time series at different positions of the mass .............................. 93
4.16 Performance of STSA method ................................................. 94

5.1 Examples of (a) offset (parallel) misalignment and (b) angular misalignment ................................................................. 98
5.2 Overall architecture of the symbolic time series analysis method for anomaly detection in flexible mechanical couplings ............... 101
5.3 Schematic of the disc/diaphragm coupling, Advanced Engineering Dynamics, J.H. Ginsberg ...................................................... 103
5.4 Family of anomaly measure plots for different levels of constant torque inputs in the range of 3000-6500 N-m vs diaphragm stiffness $K$ ................................................................. 111
5.5 Plots of actual probability distribution of diaphragm stiffness $K$ and corresponding Log-normal fit for four levels of anomaly measure ($\mathcal{M}$): a) 0.0445 b) 0.0954 c) 0.1591 d) 0.2227 .......................... 114
5.6 Plots of confidence intervals of the Log-normal distribution at 99%, 95% and 90% confidence levels. Bounds of confidence level 99% and mean are shown at $\mathcal{M}=0.111$ .................................................. 117
5.7 Plots of confidence intervals of the mean at 99%, 95% and 90% confidence levels. Bounds of confidence level 99% and mean are shown at $\mathcal{M}=0.111$ .................................................. 119

6.1 Hierarchical architecture for life extension and performance control 124
List of Tables

2.1 Structural dimensions of test apparatus 1 ................................. 16
2.2 Sensor mountings ................................................................. 21
4.1 Life extension for different levels of load reduction ....................... 89
5.1 Estimated values of diaphragm stiffness ($K$) for two different confidence levels for three different Torque Loads .......................... 117
5.2 Accuracy of the estimated diaphragm stiffness ($K$) ....................... 119
A.1 LVDT specifications ............................................................... 132
A.2 Load cell specifications ......................................................... 133
A.3 Inline amplifier for load cell .................................................. 133
A.4 Power amplifier specifications ............................................... 133
A.5 Shaker specifications ............................................................ 134
A.6 Scaling factors for calibrated sensors ..................................... 134
Acknowledgments

I would like to take this opportunity to thank my advisor and all time friend, Dr. Asok Ray who encouraged me and showed me the direction in which I should be heading. Also, I thank Dr. Eric Keller for being at my side guiding me in the design of the testbed and thanks to him I didn’t have to start from scratch. I would like to acknowledge my peer, Ishamu Chattopadhyay’s effort in the conceptual design and making of the mechanical motion apparatus which gave me an initial groundwork to build upon. I would also like to thank my peer Shalabh Gupta for his contribution of further development of the anomaly detection methodology that forms the theoretical tools that I have used in my dissertation. I would also like to thank Dr. Jeffrey Schiano for being patient in reading through my paper work for my MS in Electrical Engineering. His initial guidance in going about the system identification helped me a lot while planning out the experiments. I would also like to thank Dr. Eric Marsh and Dr. Qian Wang for their advice, support and encouragement in the pursuit of my dissertation work. I would really like to thank all of my fellow graduate students who made this journey a lot enjoyable in addition to being a learning one. I also thank Phil Irwin, Workshop superintendent in helping me make the test specimens and Larry Milinder, Inst. room in-charge in providing me with the needed instruments. I also thank my friends in India and abroad without their endless support this would have remained far from reality. I would like to thank the sponsors of this research effort: The U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. DAAD19-01-1-0646.

Finally, I would like to profusely thank my parents who have supported and been by my side in all the decisions I have taken in life and I would like to dedicate this work to them. Hope you enjoy reading through this piece of work.
Dedication

I dedicate my doctoral thesis to my parents and my younger brother Aditya who have been on my side all through the thick and thin of life.
1.1 Background of the research

An anomaly is defined as a deviation from the nominal initial behavior of a dynamical system and is often associated with parametric and non-parametric changes that may gradually evolve in time. Anomalies may manifest themselves with self excitation, or under excitation of certain exogenous stimuli. These anomalies may be benign or malignant depending on their impact on the mission objectives and operating conditions. The interest in the ability to monitor a structure and detect anomalous changes at the earliest possible stage is pervasive throughout the civil, mechanical, and aerospace engineering communities. Modelling a physical process with a nonlinear mathematical structure based solely on the fundamental principles of physics is often not feasible. In such complex dynamical systems, the physics of failure in an individual component in the system cannot explain the pathological behavior in the aggregated system. Complex macroscopic behaviors emerge from the nonlinear dynamics of the interacting components where,
based solely on the fundamental principles of physics, accurate and computationally tractable modelling of system dynamics becomes infeasible. Also, complex system behavior may range from strict order to chaos with great sensitivity to initial conditions [1]. A convenient way of learning the dynamical behavior is to rely on the (sensor-based) observed macroscopic behavior [2]. The need for quantitative global damage detection methods that can be applied to complex structures has led to the development and continued research of methods that examine changes in the vibration characteristics of the structure.

The increase in research activity regarding vibration-based damage detection is the result of the coupling between many factors [3]. The current state of aging infrastructure and the economics associated with its repair have been motivating factors for the development of methods that can be used to detect the onset of damage or deterioration at the earliest possible stage. Also, technological advancements including increases in cost-effective computing memory and speed, advances in sensors including non-contact and remotely monitored sensors and adaptation and advancements of the finite element method represent technical developments that have contributed to recent improvements in vibration-based damage detection.

Most currently used damage identification methods are included in one of the following categories: visual or localized experimental methods such as acoustic or ultrasonic methods, magnetic field methods, radiography, eddy-current methods or thermal field methods [4]. All of these experimental techniques require that the vicinity of the damage is known apriori and that the portion of the structure being inspected is readily accessible. However, the basic idea remains that com-
monly measured modal parameters (notably frequencies, mode shapes, and modal damping) are functions of the physical properties of the structure (mass, damping, and stiffness). Therefore, changes in the physical properties, such as reductions in stiffness resulting from the onset of cracks or loosening of a connection, will cause detectable changes in these modal properties. Because changes in modal properties or properties derived from these quantities are being used as indicators of damage, the process of vibration-based damage detection eventually reduces to some form of a pattern recognition problem. The need for quantitative global damage detection methods that can be applied to complex structures and procedures that can generate this information in real-time has led to the development and continued research of methods that examine changes in the vibration characteristics of the structure.

Some of the earlier approaches and the difficulties encountered have been cited here. The somewhat low sensitivity of frequency shifts to damage requires either very precise measurements or large levels of damage. However, recent studies have shown that resonant frequencies have much less statistical variation from random error sources than other modal parameters like damping and shapes (Farrar, et al. [5], Doebling, et al. [6]). For example, in offshore platforms damage-induced frequency shifts are difficult to distinguish from shifts resulting from increased mass from marine growth (Farrar, et al., [7]). The frequencies generally cannot provide spatial information about structural changes. Multiple frequency shifts can provide spatial information about structural damage because changes in the structure at different locations will cause different combinations of changes in the modal frequencies. However, there are often an insufficient number of frequencies
with significant enough changes to determine the location of the damage uniquely.

An alternative to using frequency shifts to obtain spatial information about sources of vibration changes is using mode shapes and its derivatives, such as curvature. Mayes [8] presents a method for model error localization based on mode shape changes. Ratcliffe [9] presents a technique for locating damage in a beam that uses a finite difference approximation of a Laplacian operator on mode shape data. Cobb and Liebst [10] present a method for prioritizing sensor locations for structural damage identification based on an eigenvector sensitivity analysis. Pandey, et al. [11] demonstrate that absolute changes in mode shape curvature can be a good indicator of damage for the FEM beam structures they consider. Chance, et al. [12] found that numerically calculating curvature from mode shapes resulted in unacceptable errors. They used measured strains instead to measure curvature directly, which dramatically improved results.

One issue of primary importance of these earlier mentioned methods is the dependence on prior analytical models and/or prior test data for the detection and location of damage. Many algorithms presume access to a detailed finite element modelling of the structure, while others presume that a data set from the undamaged structure is available. Often, the lack of availability of this type of data can make a method impractical for certain applications. While it is doubtful that all dependence on prior models and data can be eliminated, certainly steps can and should be taken to minimize the dependence on such information. This dissertation presents experimental validation of a novel approach for damage (defined as a anomaly) detection based on pertinent time series measurements, assuming no apriori modelling of the structure. The experimentally validated methodology of
anomaly detection in electromechanical systems is built on the concepts of Symbolic Dynamics, Pattern Recognition and Information theory. In particular, this research makes use of the D-Markov machine construction technique [13].

Another challenging task is the identification of anomalies and estimation of the fault parameters based on the pattern set of anomaly measures generated by the detection algorithm. For eg., it is essential to detect early changes in the stiffness of the coupling, subjected to misalignment due to dynamic movement, as a result of evolving fatigue damage, so that appropriate remedial action(s) can be taken before the onset of widespread damage. It is also necessary to relate changes in stiffness to the magnitude and location of incipient damage. The rationale is that loss of stiffness of the coupling may lead to further misalignment, excessive noise and vibrations causing rapid failure of other rotor components. Hence, the estimated change in stiffness during the operating period is crucial to allow for scheduled operation and maintenance.

Statistical techniques have been utilized for structural health monitoring and failure prognosis of mechanical systems in the past few decades [14]. For example, a commonly used tool is principal component analysis that relies on the estimation of eigenvalues and eigenvectors [15] [16] of the dynamical system. These parameters are dependant on the mass, stiffness and damping properties and as well as on the structural flaws in the system. They provide critical information about the characteristics and behavior of dynamical systems. In the approach described in this dissertation, the objective is to estimate the critical parameter(s) of the system, based on the observed time series data response. Statistical analysis of the pattern set of the anomaly measure profiles provides the estimate of the system
1.2 Objective of the research

The research experimentally validates a novel concept of anomaly detection [13] in electro-mechanical systems based on a combination the tools of Symbolic Dynamics and Pattern Discovery. The proposed anomaly detection methodology is divided into two parts:

- Forward problem and

- Inverse problem

The objective in the forward problem is to learn how the grammar of the underlying system dynamics changes as the system parameters change. The forward problem is that of learning where the value of a parameter is associated with an anomaly measure. In contrast, the inverse problem is that of inferring the system parameters based on the observed asymptotic behavior. In general, the inverse problem could be ill-posed and have no unique inverse. That is, it may not be possible to attribute a unique value to the parameter based on the observed behavior of the system. Nevertheless, the feasible range of parameter variation estimates can be narrowed down from the intersection of the information generated from responses under several stimuli chosen in the forward problem.
1.3 Contributions of the research

The research experimentally validates a novel approach to address the anomaly detection problem in electromechanical systems using D-markov machines. Injection of self excited stimuli is presented to detect small changes and anomalies at an early stage by taking advantage of symbolic dynamics. The novelty of the research and new contributions are summarized below:

- To tackle the **forward** problem ie., to identify how the system performance is affected due to gradually evolving anomalies and to classify the parametric and non-parametric conditions that affect the system behavior. Experimental validation of anomaly detection as proposed by Ray (2004) [13] in *real-time* for electromechanical systems.

- To deal with the **inverse** problem ie., based on the information derived from statistical pattern changes in the sensor data acquired in real-time, estimate the system parameters with a specified level of confidence. Patterns of damage evolution are identified from symbolic time series analysis of data sets, generated for multiple torque inputs using bearing accelerations from a coupling simulation model. Small changes in the coupling stiffness parameter(s) are detected and estimated via inverse mapping based on ensemble of generated patterns.

- Demonstration of **online life extending control** based on anomaly measure. The goal of life extending control is to detect the progression of malignant anomalies at an early stage and enable the system to take appropriate remedial action to circumvent damage. Also, control action should be de-
layed as far as possible so as to enable the system to perform at the desired level for maximum length of time.

There are two experimental testbeds that have been designed for validation of anomaly detection.

In the test apparatus shown in Figure 1.1, anomaly is the progression of fatigue damage in the critical components of the testbed. Detection is based on remotely placed sensor(s) like accelerometers and displacement transducers to quantify the effect of progressing damage on the global performance variables. Also, localized damage is captured by ultrasonic flaw detectors. However, the mechanical system in Figure 1.1 is highly resonant and needs to be excited in the region of resonance to cause sufficient stress in the test specimens and result in failure in reasonable amount of time.

To overcome this limitation, the test apparatus in Figure 1.2 was designed which is powered by hydraulic actuators. Also, this testbed is operated hydraulically providing with lot more power and over a wide band of operation. The objective is to capture the effects of parametric changes (i.e., the movement of mass / change in moment of inertia) on the dynamics of the system. Detection is based on statistical pattern changes in the time series of accelerometer sensors located at remote accessible locations.

The dissertation deals with the inverse problem as well. Bearing acceleration data for detection of parametric changes in flexible diaphragm couplings due to angular misalignment between shafts is derived from a coupling simulation model which encompasses the effects of angular misalignment in the coupling by translating the exerted forces (vibrations) to the bearings. Patterns of damage evolution
are identified from symbolic time series analysis of data sets, generated for multiple torque inputs. Small changes in the coupling stiffness parameter(s) are detected and estimated for different torque inputs via inverse mapping based on ensemble of generated patterns.

1.4 Organization

Chapter 2 focuses on the design of experimental test beds for validation of the anomaly detection methodology. The design requirements (see Section 2.2) spec-
ify the desired objectives of the test beds. Section 2.3 describes an apparatus that operates under resonating conditions. The system is excited near resonance to cause fatigue failure in one of the three failure sites and provides valuable information about the progression of damage in electromechanical systems. The modeling and system identification of the test bed is presented in detail in Section 2.4. The limitation of having to operate near resonance is overcome in the apparatus designed in Section 2.5. This apparatus is designed to detect the change in mass/moment of inertia of the system.

Chapter 3 provides a detailed description of the symbolic time series analysis (STSA). The forward and inverse problem have been detailed in this Chapter. Section 3.2 presents the maximum entropy partitioning of the wavelet space. Sec-
tion 3.3 presents the finite state machine construction and computation of anomaly measure.

Chapter 4 discusses the validation of the methodology provided in Chapter 3 for each of the experimental testbeds. The chapter also compares the efficacy of the STSA technique [13] with other pattern recognition tools described in Section A.1. Section 4.9 shows the collection of data under identical loading conditions to prove consistency of the results. Section 4.2.2 compares the performance of different sensors in damage detection: accelerometers, load cells, displacement transducers and ultrasonic transducers. Section 4.4 validates the anomaly detection methodology explained in Chapter 3 for detecting the change in mass/moment of inertia of the system.

Chapter 5 presents Symbolic Time Series Analysis (STSA) of bearing acceleration data for detection and estimation of parametric changes in flexible diaphragm couplings due to angular misalignment between shafts. The accelerometer readings are from a simulation, as shown in Section 5.3 which encompasses the effects of angular misalignment in the coupling by translating the exerted forces (vibrations) to the bearings. Section 5.4.1 solves the forward (analysis) problem to generate a pattern set which is representative of parametric changes under different excitation levels. The objective of the inverse problem solved in Section 5.4.2 is identification of anomalies and estimation of the parameters (i.e., the stiffness in case of the coupling) based on the pattern set of anomaly measures generated by the forward problem.

Chapter 6 summarizes the contributions of this thesis and makes recommendations for future work.
Section A.1 presents the theory of other conventional pattern recognition techniques such as Principal Component Analysis (PCA), Multilayer Perceptron Neural Network (MLPNN) and Radial Basis Function Neural Network (RBFNN). Section A.2 enlists the specifications of various sensors and actuators of the testbed used in conducting the experiments.
Design of Experimental Testbeds

2.1 Introduction

This chapter presents the design and modelling of laboratory test apparatuses that have been constructed to experimentally validate the concepts of anomaly detection [13] in electromechanical systems.

2.2 Design requirements

In contrast to the goal of reducing damage (i.e., to make damage as small as possible) in the design of an operating machinery (e.g., aircraft, power plant, and chemical plants), the test apparatus is designed to deliberately introduce fatigue damage in its critical component(s). These components are intentionally made to break in a reasonably short period of time to enhance the speed of conducting experiments. It is important that the damage in a critical component must not be strongly coupled with the plant dynamic performance. The rationale is that a strong coupling will preclude any application of life extending control to achieve
a large gain in structural durability without any significant loss of performance. Nevertheless, moderate coupling is necessary because the analyzed time series data of the displacement sensor signal must contain ample information on the system dynamics. If this assumption is relaxed, early detection by mechanical sensors may not be feasible and one would have to rely entirely on damage sensing devices like ultrasonic transducers used in non-destructive testing for anomaly detection. A focus of this research is to show the efficacy of real-time anomaly detection based on time series data from damage sensing devices like ultrasonics and mechanical sensors such as displacement sensors and accelerometers. From these perspectives, the requirements of the test apparatus are:

1. Operability under cyclic loading with multiple sources of input excitation;

2. Damage accumulation in test specimens (at selected locations) with negligible damage in other components of the test apparatus;

3. No strong coupling between the damage of test specimens and dynamic performance of the control system; and

4. Facilitate multiple failure sites to study comparative behavior of damage dynamics.

5. Facilitate change in mass/ moment of inertia of the system.

### 2.3 Test apparatus 1 - Resonance machine

The test apparatus is briefly described in this section. A mathematical model of the system dynamics is formulated for identification of these critical parameters
from the input/output time series data generated by persistent excitation. Details of the same are presented in Section 2.4. The test apparatus is designed to be complex in itself due to partially correlated interactions amongst its individual components and functional modules [17]. The experiments are conducted on the test apparatus to represent operations of mechanical systems where both dynamic performance and structural durability are critical.
Figure 2.1. Schematic diagram for test apparatus 1

Table 2.1. Structural dimensions of test apparatus 1

<table>
<thead>
<tr>
<th>Component</th>
<th>Material</th>
<th>Dimensions (mm),Mass (kg) L x W x T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass 1</td>
<td>Mild Steel</td>
<td>2.82</td>
</tr>
<tr>
<td>Mass 2</td>
<td>Aluminium 6061-T6</td>
<td>0.615</td>
</tr>
<tr>
<td>Mass 3</td>
<td>Mild Steel</td>
<td>3.87</td>
</tr>
<tr>
<td>Beam 1</td>
<td>Mild Steel</td>
<td>800 x 22 x 11</td>
</tr>
<tr>
<td>Beam 2</td>
<td>Aluminium 6061-T6</td>
<td>711.2 x 22.2 x 11.1</td>
</tr>
<tr>
<td>Specimens</td>
<td>Aluminium 6061-T6</td>
<td>203.2 x 22.2 x 11.1</td>
</tr>
</tbody>
</table>
In order to satisfy the design requirements as mentioned in Section 2.2, the test apparatus is designed as a multi-degree of freedom (DOF) mass beam structure excited by two shaker tables. A schematic diagram of the test apparatus and the instrumentation is shown in Figure 2.1 and dimensions of the pertinent components are listed in Table 2.1. The test apparatus system is logically partitioned into two subsystems: (i) the plant subsystem consisting of the mechanical structure including the test specimens that undergo fatigue crack damage, actuators and sensors; and (ii) the instrumentation and control subsystem consisting of computers, data acquisition and processing, and communications hardware and software. The sensors include two load cells for force measurement, two linear variable displacement transducers (LVDT) for displacement measurement; and two accelerometers for acceleration measurement.

The two actuators (i.e., Shaker #1 and Shaker #2 in Figure 2.1) directly control two of the three DOFs whereas the remaining DOF is observable via displacement measurements of the three vibrating masses as seen in Figure 2.1. The test specimens subjected to fatigue damage are representative of plant components suffering fatigue crack damage. The mechanical structure is excited at the resonant frequency so that the critical components can be excited at different levels of cyclic stress excitation with no significant change in the external power injection into the actuators. The excitation force vector generated by the two actuators serves as the inputs to the multi-DOF mechanical structure to satisfy requirement #1 in Section 2.2.
2.3.1 Hardware implementation and software structure

The fatigue damage test apparatus as shown in Figure 2.1 is interfaced with a Keithley Data Acquisition Board (DAS 17 STDA) having 6 A/D channels and 4 D/A channels. Data acquisition is carried out with a sampling rate at 500 Hz for monitoring and control. The time series data for statistical pattern recognition is decimated as needed. The real-time instrumentation and control subsystem of this test apparatus is implemented on a Pentium PC platform. The software runs on the Real-Time Linux Operating System and is provided with A/D and D/A interfaces to the amplifiers serving the sensors and actuators of the test apparatus.

The control software performs real-time communication tasks, in addition to data acquisition and built-in tests (e.g., limit checks and rate checks). The data acquisition is an Interrupt Service Routine (ISR) for the Direct Memory Access (DMA) completion interrupt. The A/D board is initialized to take 12 readings per frame. The DMA controller on the PC motherboard is programmed to read 12 single 16-bit words and store them sequentially in a given memory location for each transfer.

Additional hardware devices consist of ultrasonic sensors that provide information about the microstructural changes that take place as the fatigue crack damage evolves. The user space for ultrasonic data acquisition and processing consists of a C program that connects through the Data FIFO with the data acquisition kernel of the Real-Time Linux Operating System and also stores the data to a text file for off-line analysis. There is another program that connects through the Command FIFO and instructs the kernel to start or stop the system, increase or decrease the gain, and increase or decrease the excitation frequency. Other features of the soft-
ware are a graphical user interface (GUI) as shown in Figure 2.2 both for real-time data display as well as for connecting to the stage motion of the optical microscope.

![Software implementation for test apparatus 1](image)

**Figure 2.2.** Software implementation for test apparatus 1

### 2.3.2 Sensors for damage detection and performance measurement

The apparatus is equipped with multiple sensors for damage detection including ultrasonic flaw detectors and an optical microscope for localized damage sensing. The advantages of using ultrasonic transducers over optical microscope include the ease of installation at the desired damage site and detection of early anomalies before the onset of large fatigue crack propagation. Nevertheless, the optical microscope serves to calibrate the ultrasonic flaw detectors.

- **Ultrasonic flaw detector**- The ultrasonic flaw detector functions by emitting high frequency ultrasonic pulses that travel through the specimen to the receiver
transducer [18]. A piezoelectric transducer is used to inject ultrasonic waves in the specimen and a single receiver transducer is placed on the other side of the circular notch to measure the transmitted signal. The ultrasonic waves produced were 5MHz sine wave signals and they were emitted asynchronously almost once every cycle. The sender and receiver ultrasonic transducers are placed on two positions, above and below the hole, so as to send the signal through the region of crack propagation and receive it on the other side, as shown in Figure 2.4. Since material characteristics (e.g., voids, dislocations and short cracks) influence the ultrasonic impedance, minute damage in the specimen is likely to change the signature of the signal at the receiver end [19]. Therefore, the signal can be used to capture small changes during the early stages of fatigue damage, which may not be possible to detect by an optical microscope. Prior to the appearance of a crack on the specimen surface, deformations (e.g., dislocations and short cracks) inside the specimen may have already caused detectable attenuation and/or distortion of the ultrasonic waves. Therefore the ultrasonic sensors are utilized for generating the localized information of the growth of fatigue damage.

b) Optical microscope- The travelling optical microscope, shown as part of the schematic in Figure 2.2, provides direct measurements of the visible part of a surface crack. The microscope can be shifted from left to right side and vice versa to track the crack tip on the specimen.

c) Accelerometers, load cells and displacement transducers- These sensors are useful for analyzing the effects of local fatigue damage on the global performance of the vibratory system.

The locations of various sensors on the test apparatus is shown in Figure 2.3
and listed in Table 2.2.

![Figure 2.3. Sensor mounting for test apparatus 1](image)

**Table 2.2. Sensor mountings**

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDT</td>
<td>Mass 1, Mass 3</td>
</tr>
<tr>
<td>Accelerometer</td>
<td>Mass 1, Mass 3</td>
</tr>
<tr>
<td>Load Cell</td>
<td>Shaker 1, Shaker 2</td>
</tr>
<tr>
<td>Ultrasonic Transducers</td>
<td>Specimen between Mass 1 and Mass 3</td>
</tr>
</tbody>
</table>

### 2.3.3 Role of ultrasonics in damage sensing

In structures made of ductile alloys that are subject to fatigue failures, a large part of the service life is spent in crack initiation and in the presence of very small cracks. A secondary goal of this research is to have knowledge of damage state for
as much of the life of the structure as possible, and not simply to know when the life of the structure is used up. There are very few flaw detection methods that will work with flaws that are as small as the microstructural elements of ductile metals or even extremely small cracks. There are fewer still that are suitable for installations outside of the laboratory. Ultrasonic flaw detection is one of them [19].

Figure 2.4. Ultrasonic flaw detection system for 6061-T6, PhD dissertation, Eric Keller, 2001

The detector as shown in Figure 2.4 functions by emitting high frequency acoustic pulses and measuring the signal after it has had time to propagate through
the material being measured. Material features such as grain boundaries, inclusions, stress state, voids, or cracks along the path of the wave will cause additive and destructive interference [18]. A microscope is also fitted to calibrate the ultrasonic measurements with the actual observed crack length. However, most of the ultrasonic signal gets attenuated when the crack appears on the surface.

2.3.4 Salient features and limitations

The salient features of test apparatus 1 (see Figure 2.1) have been summarised as below:

- Sensing by mechanical sensors like displacement transducers, accelerometers and load cells to quantify the effect of progressive damage on the global behavior of the system.
- Localised damage sensing using ultrasonic flaw detectors.
- Use of optical microscope to measure surface crack length.
- Multiple failure sites for comparative study of progression of fatigue damage.

**Limitation**

The test apparatus in Figure 2.1 is highly resonant and needs to operate near resonance to cause sufficient stress for fatigue failure in the specimens. Also, the electromagnetic shaker tables provide very small bandwidth and limited power.
2.4 Modelling of test apparatus 1 for anomaly detection

This section focuses on design, modelling and system identification of an experimental test apparatus that has been recently fabricated for early detection of small anomalies like fatigue damage in polycrystalline alloys. A mathematical model of the system dynamics is formulated for identification of resonant frequencies from the input/output time series data generated by persistent excitation. (Note: Fatigue damage evolves at a time scale that is several orders of magnitude slower than the structural dynamics.)

2.4.1 Physical modelling of test apparatus 1

All physically realizable mechanical systems are continuous (distributed parameter) systems. This implies that when a mechanical system is excited it will vibrate in an infinite number of modes. If interest is confined to particular points in the system, one approach in modeling the movement is the use of infinite number of second order lumped parameter systems. The second order systems are chosen so that the natural frequencies match the modes of the continuous system and the superposition of their displacements matches the displacements of the points of interest.

A structural model of the mechanical system described above is formulated using a cubic interpolation polynomial to approximate the lateral displacement of the beams and a linear approximation for the lateral displacement of the masses [20]. The major assumptions in the model formulation include:
• Lumped representation of the beam masses. Beams subjected to pure bending (i.e., no deformation of the masses). The assumption that the masses, essentially behave as rigid bodies is justified by their relatively high stiffness.

• Neglect torsional effects as masses vibrate in a single plane or parallel planes.

• The actuator is assumed to force a point mass, $m_2$, at the tip of the Beam 2 as seen in Figure 2.1. The validity of the other assumption of negligible torsional effects is demonstrated later when the model is compared with actual experimental data.

Figure 2.5 shows the local co-ordinate system for each component along with the sign convention used for forces and bending moments. Physical modeling of the system gives us an idea of the resonant frequencies of the structure and helps in arriving at the appropriate dimensions for the various components. Also, in-order to cause fatigue failure in reasonable amount of time the specimens need to be subjected to loads near resonating conditions.

The force acting on the aluminum beam of length $L_5$ is $F_1$ acting on the free end of the beam. The other end of the beam is attached to a mass $m_3$ and corresponding values of the dimensions are shown in Table 2.1. A thorough derivation of the dynamic model is now presented.

As seen in Figure 2.1 and Figure 2.5, the mechanical system is highly resonant. Because damping elements are not included in the design it is expected that the mechanical system will admit a transfer function matrix with sharply peak resonances. However, all physical systems are passive and hence would have some inherent damping which would displace the resonant peaks from the undamped
ones. Also, damping would cause the 3 db bandwidth to increase and the peaks would not be that sharp as seen in the theoretical model.

The procedure for deriving a model follows. We first obtain a dynamic model for subsystem 1 consisting of mass $m_3$ and Beam #1 having Young’s modulus of elasticity E and moment of inertia I. This model uses a lumped parameter representation by obtaining a spring constant k as seen below which in turns gives us the resonant frequency for subsystem 1.

As seen in the Figure 2.5, bending moment acting at the end of mass $m_3$.
denoted by $M_0$ is given by,

$$M_0 = F_1(L_5 - z), \quad (2.1)$$

where $z$ is the distance of the section under consideration from the free end and $F_1$ is the force acting at the free end. By Castigliano’s Theorem which states that by taking the partial derivative of the strain energy $U$ with respect to an applied force, we can determine the deflection $\delta$ under that applied force, in the direction of that force [21], we obtain

$$U = \int_0^{L_4} \frac{M_0^2}{2EI} \, dz = \frac{F_1^2}{2EI} (L_4L_5^2 - L_5L_4^2 + \frac{L_4^3}{3})$$

$$\delta(\text{free}) = \frac{\partial U}{\partial F_1} = \frac{F_1L_4}{EI} [L_5^2 - L_5L_4 + \frac{L_4^2}{3}] \quad (2.2)$$

Then using the relation

$$k = \frac{F_1}{\delta},$$

we obtain

$$\Rightarrow k = \frac{3EI}{L_4(3L_5^2 - 3L_5L_4 + L_4^2)}$$

Hence, the natural frequency is

$$\omega_n = \sqrt{\frac{k}{m_3}} = \sqrt{\frac{3EI}{m_3L_4(3L_5^2 - 3L_5L_4 + L_4^2)}}, \quad (2.3)$$

The lumped parameter model for the forced vibration of the beam is as follows.

$$m_3\ddot{z} + kz = F_1 \quad (2.4)$$
\[
\dot{x} = A_1 x + B_1 u \\
y = C_1 x + D_1 u
\] (2.5)

where

\[
x = \begin{pmatrix} x \\ \dot{x} \end{pmatrix}; \\
A_1 = \begin{pmatrix} 0 & 1 \\ -k/m_3 & 0 \end{pmatrix}; \\
B_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \\
C_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}; \\
D_1 = \begin{pmatrix} 0 \end{pmatrix}; \\
y = \begin{pmatrix} x \end{pmatrix}
\]

where, the state variable \(x\) and \(\dot{x}\) represent respectively, the modal displacement of the mass \(m_3\) from the center of the beam or the rest position and its velocity in the \(x\)-direction. Also \(k\) and \(m_3\) have the meaning as shown in the Figure 2.5 and derived above.

A free-body diagram of System 2 is shown in Figure 2.5. For the 1-input 2-output system, mass \(m_2\) is assumed to be a point mass located at the tip of beam 2. The lateral displacements of beam 1 and 2, \(y_i(x_i)\), \(i = 1, 3\) are approximated by third order polynomials in \(x_i\) for beams 1 and 2 and the lateral displacement of mass \(m_1\), \(y_2(x_2)\), is approximated by a linear fit in \(x_2\) as seen in Figure 2.1. This is done as the moments and shear forces are the second and third spatial derivatives of the displacements. Expressions for \(y_i(x_i)\) are

\[
y_1(x_1) = C_1 x_1^3 + C_2 x_1^2 + C_3 x_1 + C_4 \\
y_2(x_2) = C_5 x_2 + C_6 \\
y_3(x_3) = C_7 x_3^3 + C_8 x_3^2 + C_9 x_3 + C_{10}
\]
where the coefficients $C_i$'s are computed making use of the boundary and compatibility conditions as stated later in the text.

The linear polynomial approximation is used as mass $m_1$ is assumed to be a rigid body with negligible deformation. This assumption is justified by the dimensions and material of the mass and the forces that it experiences. With the definitions

$$y_2 = y_2 \left( \frac{L_2}{2} \right)$$

$$y_3 = y_3 (L_3)$$

where, the boundary conditions at the end points and the compatibility conditions are given below.

Beam 1 is a cantilever supported beam and hence there can be no displacement or velocity at the fixed end of the beam

$$y_1(0) = 0$$

$$y'_1(0) = 0,$$

where $y'_1$ is the spatial derivative. Also, the cantilever beam is attached to a mass $m_1$ which in this case is modelled as a beam. Hence, by continuity principle, the displacement at the interface of beam 1 and mass $m_1$ should be the same

$$y_1(L_1) = y_2(0),$$

$$y'_1(L_1) = y'_2(0).$$

Writing the flexural bending equation for beam 1 with Young’s Modulus $E_1$ and
area moment of inertia $I_1$ yields

$$E_1 I_1 y_1''(L_1) = M_B^+$$

$$-E_1 I_1 y_1'''(L_1) = V_B^+,$$

where $M_B$ is the bending moment at the interface of beam 1 and mass $m_1$ and $V_B$ is the force acting at the interface. Convention used is + sign for clockwise moment and shear force producing clockwise moment.

Using similar analysis and convention as shown above for the interface between mass $m_1$ and beam 2 we get the following equations:

$$y_2(L_2) = y_3(0)$$

$$y_2'(L_2) = y_3'(0)$$

$$E_3 I_3 y_3''(0) = M_C^-$$

$$-E_3 I_3 y_3'''(0) = V_C^-$$

$$E_3 I_3 y_3''(L_3) = 0$$

$$E_3 I_3 y_3'''(L_3) = m_2 \ddot{y}_3 - F_2$$

where $M_C$ is the bending moment at the interface of beam 2 and mass $m_1$ and $V_C$ is the force acting at the interface. Convention used is + sign for clockwise moment and shear force producing clockwise moment. Also, $E_3$ is the Young’s modulus and $I_3$ is the area moment of inertia of beam 2. Also by Newton’s law,
we get the last equation with $F_2$ being the force applied by the Shaker #2.

Considering the equilibrium of each individual beam and mass in the single input, multi-output system, we get the following compatibility conditions:

\[ \sum \text{Bending Moments } @ B = 0 \text{ (taking clockwise +)} \]

\[ \Rightarrow -M_B^+ + M_C^- + V_C^- L_2 - m_1 \dddot{y}_2^2 \frac{L_2}{2} = 0 \quad (2.6) \]

\[ \sum \text{Forces } @ B = 0 \text{ (taking rightwards +)} \]

\[ \Rightarrow -V_B^+ + V_C^- - m_1 \dddot{y}_2 = 0 \quad (2.7) \]

Applying these boundary and compatibility conditions and solving manually for all constants in terms of parameters of Young’s modulus, moment of Inertia, force(s), length(s) and masses, we obtain the following expressions for the 10 constants in terms of physical quantities.

\[ C_1 = -\frac{1}{E_1 I_1} [F_2 - m_2 \dddot{y}_3 - m_1 \dddot{y}_2] \]

\[ C_2 = \frac{(L_1 + L_2 + L_3)}{E_1 I_1} [F_2 - m_2 \dddot{y}_3] + \frac{(2L_1 + L_2)}{2E_1 I_1} [-m_1 \dddot{y}_2] \]

\[ C_3 = C_4 = 0 \]

\[ C_5 = \frac{(L_1^2 + 2L_1 L_2 + 2L_1 L_3)}{2E_1 I_1} [F_2 - m_2 \dddot{y}_3] + \frac{(L_1^2 + L_1 L_2)}{2E_1 I_1} [-m_1 \dddot{y}_2] \]

\[ C_6 = \frac{(2L_1^3 + 3L_1^2 L_2 + 3L_1^2 L_3)}{6E_1 I_1} [F_2 - m_2 \dddot{y}_3] + \frac{(4L_1^3 + 3L_1^2 L_2)}{12E_1 I_1} [-m_1 \dddot{y}_2] \]
\[ C_7 = -\frac{1}{E_3 I_3} [F_2 - m_2 \ddot{y}_3] \]

\[ C_8 = -\frac{L_3}{E_3 I_3} [F_2 - m_\dot{y}_3] \]

\[ C_9 = C_5 \]

\[ C_{10} = \frac{(2L_1^2 + 6L_2^2 L_2 + 3L_2^2 L_3 + 6L_1 L_2 L_3) + [F_2 - m_2 \ddot{y}_3]}{6E_1 I_1} + \frac{(4L_1^2 + 9L_2^2 L_2 + 6L_1 L_2^2)}{12E_1 I_1} [-m_1 \ddot{y}_2] \]

Substituting for the constants in Eq. 2.6 and 2.7 we obtain the following dynamical equations:

\[ m_1 \ddot{y}_2 + P \dot{y}_2 + Q \ddot{y}_3 = 0 \]

\[ m_2 \ddot{y}_3 + R \dot{y}_2 + S \ddot{y}_3 = F_2. \]

The parameters P,Q,R,S are of the form given by the following expressions:

\[ P = \frac{a_4}{a_1 a_4 - a_2 a_3} \]

\[ Q = -\frac{a_2}{a_1 a_4 - a_2 a_3} \]

\[ R = \frac{a_3}{a_3 a_2 - a_1 a_4} \]

\[ S = -\frac{a_1}{a_3 a_2 - a_1 a_4} \]

where the constants \( a_1, a_2, a_3, a_4 \) are

\[ a_1 = \left[ \frac{4L_1^2 + 6L_2^2 L_2 + 3L_1 L_2^2}{12E_1 I_1} \right] \]
\[ a_2 = \left[ \frac{4L_1^3 + 9L_1^2L_2 + 6L_1L_2^2 + 6L_1^2L_3 + 6L_1L_2L_3}{12E_1I_1} \right] \]
\[ a_3 = \left[ \frac{4L_1^3 + 9L_1^2L_2 + 6L_1L_2^2 + 6L_1^2L_3 + 6L_1L_2L_3}{12E_1I_1} \right] \]
\[ a_4 = \left[ \frac{L_1^3}{3E_1I_1} + \frac{L_3^3}{3E_3I_3} + \frac{L_1^2L_2 + L_1L_2^2 + L_1^2L_3 + L_1L_2^2 + 2L_1L_2L_3}{E_1I_1} \right] \]

where \( L_1, L_2, L_3 \) are the lengths of beam 1, mass \( m_1 \) (modelled as a beam) and beam 2 respectively. Mass \( m_2 \) is considered to be a point mass as stated in the assumptions earlier. This completes the analysis of the single input, two output system and allows us to convert the physical model into state space form as shown below. Converting to SIMO state space form, we get

\[
\begin{align*}
\dot{x} &= A_2x + B_2u \\
y &= C_2x + D_2u \\
\end{align*}
\]

\( x = \begin{pmatrix} \ddot{y}_2 \\ \ddot{y}_2 \\ \ddot{y}_3 \\ \ddot{y}_3 \end{pmatrix} ;
A_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -P/m_1 & 0 & -Q/m_1 & 0 \\ 0 & 0 & 0 & 1 \\ -R/m_2 & 0 & -S/m_2 & 0 \end{pmatrix} ;
B_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1/m_2 \end{pmatrix} ;
C_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} ;
D_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ;
y = \begin{pmatrix} \dddot{y}_2 \\ \dddot{y}_3 \end{pmatrix}
\]

and \( \dddot{y}_2, \dddot{y}_2, \dddot{y}_3, \dddot{y}_3 \) represent the displacement and velocities of mass \( m_1 \) and \( m_2 \) respectively.
Equation 2.8 represents the dynamics of the mass-beam system where the damping terms are assumed to be insignificant. The displacement sensors are modelled as pure gains. The shaker table and the signal amplifier unit are fairly non-linear, and the dynamics of the shaker are influenced by the loading effects of the mass-beam system. (i.e., problem of determining the dynamic relationship between the applied voltage and the force generated in the actuator system was circumvented by using system identification techniques as shown in Section 2.4.2).

The two systems 1 and 2 are connected by a thin aluminium beam and the failure site is designed as shown in Figure 2.6. This connection serves as a gain between both the systems and is influential in motion transmission from one system to the other. As it also serves as a failure site, propagation of fatigue crack damage in this connection serves as special interest and can affect the performance of either systems. In the current configuration all three specimens are identical and manufactured from the same process so as to facilitate comparative behavior under fatigue failure.

Since the objective of this testbed is to investigate the influence of different control policies on specific modes of fatigue failure in a dynamic setting, a statistically weakened element which is representative of a critical plant component is introduced to ensure the occurrence of an observable failure. In the two-mass configuration, three parallel failure sites are introduced by drilling a 8.43 mm dia. hole located at 67.56 mm from center of Mass # 1, one in beam connecting Mass # 1 and Mass # 2 and other connecting Mass # 3 and Shaker # 1.
2.4.2 System identification of test apparatus 1

A set of open loop plant models is typically derived based on a priori information (e.g., fundamental laws of physics, plant operating conditions and physical dimensions) as seen in Section 2.4.1. The plant model parameters can be identified via time-domain or frequency domain techniques. In this section, a frequency-domain method of system identification based on sinusoidal sweep input excitation has been developed for the explicit purpose of identifying the resonant frequencies as accurately as possible.

2.4.2.1 Open loop frequency domain identification

Both the time-domain and frequency-domain approaches are expected to yield equivalent results [22], in general. However, since the mechanical system under consideration is highly resonant and interest lies in accurate assessment of the resonant frequencies of the system, a frequency-domain method of system identification [23] has been adopted based on sinusoidal sweep input excitation as shown in Section 2.4.2.2. Furthermore, the instrumentation of the test apparatus al-
allows acquisition of experimental data in the time domain. This is converted to a frequency response (function) from input to output $G(\omega)$ (with the transfer function evaluated on the unit circle or the imaginary axis). The frequency function $G(\omega)$ is a complex number, whose absolute value describes how a sinusoid of frequency $\omega$ is amplified by the system, and whose argument (phase) describes how the same signal is shifted (phase-lagged) by the system. These are estimates of the system's frequency function (frequency response), obtained as estimates from measured input-output data.

Frequency domain identification makes use of frequency function measurements via combination of a slowly swept sine. The reason for choosing the sinusoidal sweep as an input signal is to capture the resonant peaks that are the modes of the system without any undesired loss of accuracy. In this case, the system is characterized by measurements of the frequency response at a large number of discrete frequency points within the spectrum. In contrast, the parametric model is characterized by a number of selected parameters.

### 2.4.2.2 Construction of the excitation signal

The excitation signal needed to be a slowly swept sine and to be generated in real-time. With this in mind, it was decided to generate the solution of the van der Pol oscillator as in Eq. (2.9) which is an ordinary differential equation. This equation is given by

$$
\ddot{y} - \mu (1 - y^2) \dot{y} + \omega^2 y = 0
$$

(2.9)
where $\omega$ is the frequency of the sinusoid in rad/s and $\mu$ is a nonnegative constant. For every nonnegative value of the parameter $\mu$, the solution of van der Pol’s equation with $y(0) = 2$, $\dot{y}(0) = 0$ is periodic, and the corresponding phase plane trajectory is a limit cycle to which the other trajectories converge. $\mu = 0$ converges into a circle, however, a value of zero causes instability and gives floating point error, a very small nonnegative value of $\mu = 0.0001$ was used in computation. [24]

This equation is implemented in real-time by converting to the discrete setting with sampling time as $T_s = 0.0002$ seconds. On numerically solving the Eq.(3.2) the value of $y$ converges to a sine, and $\dot{y}$ converges to a cosine.

2.4.2.3 Frequency response of plant dynamics

The frequency domain modelling requires the plant output response at a set of discrete frequency points $\omega_k \in \{\omega_1, \omega_2, \ldots, \omega_N\}$ covering the bandwidth of interest (i.e., from 0-20 Hz). An application of curve fitting techniques [25] yields a rational transfer matrix to obtain a closed form autoregressive moving average (ARMA) model

$$H(z) = \frac{A_0 + A_1z^{-1} + A_2z^{-2} + \ldots + A_mz^{-m}}{1 + b_1z^{-1} + b_2z^{-2} + \ldots + b_nz^{-n}} \quad (2.10)$$

where the coefficient $b_0$ is normalized to unity; and $A_i$’s represent the $\ell \times p$ coefficient matrices of numerator polynomials; the number of inputs is $p$ and the number of outputs is $\ell$; and the $b_i$’s are the coefficients of the common denominator polynomial of degree $n$.

System identification is accomplished by using the frequency response data via the `invfreqz` algorithm of the `freqid` GUI package [26] under MATLAB. A sweeping
sinusoidal signal over discrete frequencies in the bandwidth of the actuator is applied and the resulting output is collected. Similar results could also be achieved by attaching a frequency analyzer to the system. This is done by individually applying an input excitation to each actuator successively and collecting the respective output. The identification is conducted with a sampling time of 2 ms for the data acquisition. The frequency $\omega$ of the excitation signal ranges from 0 to 113.1 rad/s (approx 18 Hz) with 176 points. The procedure followed in model identification is to minimize the frequency-weighted cost functional of the deviation of the model response from the generated plant data.

By default, \texttt{invfreqz} uses the principle of least squares to identify the model from the data series. This is accomplished by determining the transfer function coefficients by minimizing a cost functional in the following form:

\[
C(A_k, b_k) = \min \sum_{x=1}^{n} \sum_{y=1}^{n} \left( \frac{1}{2} \sum_{i=1}^{N} W_i \left| H(\omega_i) \left( \sum_{k=0}^{n} b_k z_i^{-k} - \sum_{k=0}^{m} A_k z_i^{-k} \right) \right|^2 \right) \tag{2.11}
\]

where $H(\omega_i)$ is the actual experimental frequency response at $\omega_i$; and $W_i$ is the weighting factor for each frequency point $\omega_i$. The weights are chosen as the magnitude response at $\omega_i$, putting higher weights on the frequencies closer to the resonant peaks. Furthermore, since Eq. (2.11) is a sum of quadratic terms, the cost functional $C(A_k, b_k)$ can be minimized using numerical techniques for solving nonlinear least-square problems. The analysis can be done in the complex field by ensuring that the coefficients $A_k, b_k$ are real.
2.4.3 Data acquisition and estimation of transfer matrices

Output data are collected at steady state (i.e., allowing the transients to die out) over a sufficient length of time.

![Figure 2.7. Frequency response of full state and reduced order model](image)

The results of the SIMO system experiments are presented in Figure 2.8 and 2.9. Resonant frequencies are seen to be located at approximately 71 rad/s and 91 rad/s, which are very close to 73.3 rad/s and 87.8 rad/s as predicted by the derived physical model using procedure in Section 2.4.1.

Individual transfer functions are then identified for each input output pair and then the models are combined. Furthermore, the combined model is reduced to a minimal realization removing the uncontrollable and unobservable states. The
minimal realization is further balanced using the Matlab function \textit{sysbal} to first isolate states with negligible contribution to the input/output response. While the order of state-space model was further reduced based on the magnitude of Hankel singular values [27] and by using the Matlab function \textit{modred}, it was ensured that the resonant peaks (poles) of the original higher order model are conserved in the reduced order model having 13 states.

2.4.4 Experimental results and discussion

This section presents the results of the derived model and compares them with the experimental data. Figures 2.8 and 2.9 exhibit comparison of the model responses with experimentally generated frequency response data under excitation of Shaker \#1 and Shaker \#2, respectively. The model fairly captures the resonant peaks that correspond to the modes of the mechanical structure. The identified poles at 71 rad/s and 91 rad/s are close to the resonant peaks obtained from the physical model (see Section 2.4.1). The small deviations in the resonant frequencies of the physical model from those obtained experimentally are possibly due to lumped parameter approximation and interactions between the two systems, which has not been incorporated in the physical model.

\begin{align*}
x(k + 1) &= Ax(k) + Bu(k) \quad (2.12) \\
y(k) &= Cx(k) + Du(k) \quad (2.13)
\end{align*}

where
2.5 Test apparatus 2 - Multi-DOF mechanical motion system

Figure 2.10 depicts a multi degree-of-freedom mechanical motion apparatus that is not resonance-dependent. This implies that the apparatus can be excited over a wide band of frequency range to produce amplitude of considerable magnitude.

This apparatus has been recently constructed and is available for investigation of fatigue damage behavior, where the damage site is not precisely known a priori. Each of the three actuators is moved through electro-hydraulic position feedback control and is capable of providing a force of 3,400 kgf. The apparatus consists
of one vertical beam pivoted on its base and a horizontal beam pivoted to the vertical beam that has freedom to move only in a vertical plane; the horizontal beam moves in both vertical and horizontal planes. The beams are made of hollow square sections with 6 mm thick steel plates. The base is manufactured from
The mechanical motion system apparatus in Figure 2.10 is logically partitioned into two subsystems: (i) the plant subsystem consisting of the mechanical structure including flexible hinges that connect the horizontal beam with the forward movable beam; the hydraulic system that is shown in Figure 2.11 and (ii) the control and instrumentation subsystem consisting of control computers, data acquisition and processing, communications hardware and software, and sensors and actuators. The sensors include: linear variable differential transformer (LVDT) devices for displacement measurement; and ultrasonic flaw detectors on the flexible hinges made of Al-2024 that undergo fatigue damage and serve as test specimens for multi-axial fatigue damage sensing as shown in Figure 2.14. The software is executed under DSpace running on a Windows platform.
2.5.1 Hydraulic system design

The main hydraulic system elements are the pump, the supply manifold, and the cylinder. A schematic of the hydraulic system is shown in Figure 2.11. The purpose of the supply manifold is to properly sequence the system pressure and to accommodate the accumulators. The accumulators help the performance of the system by maintaining the system pressure when the instantaneous demand from the servo valve/cylinder is higher than the flow rate available from the pump. This allows a smaller pump to be used than if the accumulators were not in the system. The disadvantage posed by the accumulators is the stored energy that must be dissipated while the system is shut down. A small solenoid valve is used under most shutdown conditions to bleed the accumulator pressure slowly. This prevents a high-pressure spike from causing the return filters to fail.

The feedback control system as shown in Figure 2.12 is implemented on a Pentium PC along with necessary A/D and D/A interface to the feedback amplifiers serving the sensors and actuators of the test apparatus.

2.5.2 Hardware implementation and software structure

The multi-DOF test apparatus as shown in Figure 2.10 is interfaced with a DSpace Data Acquisition Board having 16 A/D channels and 8 D/A channels. Data acquisition is carried out with a sampling rate at 1 KHz for monitoring and control. The time series data for statistical pattern recognition is decimated as needed. The real-time instrumentation and control subsystem of this test apparatus is implemented on a Pentium PC platform. The software runs on the Windows XP Operating System and is provided with A/D and D/A interfaces to the amplifiers...
serving the sensors and actuators through the ControlDesk front end.

The ControlDesk front end loads a Simulink (Matlab based) module as shown in Figure 2.13 on to the data acquisition card to perform real-time communication tasks, in addition to data acquisition and built-in tests (e.g., software limit checks and saturation checks). The main components of the module are

a) **Reference signal generation block** - The reference signal block takes the current time and feeds it to the sine-wave generator as shown in Figure 2.13. This block then generates a 90 degree phase shifted version of the same to be fed to the
other actuator. Also, the sine-wave generation block finally outputs a trapezoidal profile of the reference trajectory. The two actuators back and horizontal are operated out of phase with each other to minimize noise and power consumption from the hydraulic setup.

b) The three actuator subsystem: Back, Horizontal and Bottom- The reference trajectory is fed to the three actuators simultaneously which have a control circuit as shown in Figure 2.12. The proportional gain control stabilizes the system and provides a fairly good system response and tracking to the reference input. Implementation of a real-time robust controller to improve performance would be area of future work.

c) Serial port communication block- The reference signal generation block communicates with the computer operating the ultrasonic transducers through the serial port. This communication is done in every cycle of operation when the actuators are at their extreme positions.
Figure 2.13. Control circuit in Simulink for the test apparatus

2.5.3 Sensors for damage detection and performance measurement

The apparatus is equipped with multiple Ultrasonic Flaw Detectors for damage detection for localized damage sensing. The Figure 2.14 shows the mounting of the ultrasonic transducer on the specimen. A piezoelectric transducer is used to inject ultrasonic waves in the specimen and the same transducer acts as a receiver to measure the reflected portion of the transmitted signal. Reflections are primarily from material inclusions, voids and dislocations along the path of the signal. Also, reflected wave energy would increase with the incubation and progression of
crack. The transducer face area is made to cover the stress concentrated region of the compact specimen as seen in Figure 2.14. The study of fatigue failure in the specimens is an area of future research and has not been dealt with in this dissertation.

### 2.5.4 Salient features

The salient features of test apparatus 2 (see Figure 2.10) have been summarised as below:

- Remote sensing by mechanical sensors like accelerometers to quantify parametric changes (like change in mass/moment of inertia) on the global behavior of the system.

- Localised damage sensing using multiple ultrasonic flaw detectors.

- Multiple failure sites for comparative study of progression of fatigue damage.
3.1 Introduction

This chapter introduces the underlying concepts of Symbolic Dynamics for anomaly detection in electromechanical systems. This research considers a class of non-linear non-autonomous electromechanical systems, represented by the following dynamical equation:

$$\dot{x}(t) = f(x(t), \theta(t_s))$$  \hspace{1cm} (3.1)

where $\dot{x}(t)$ is the derivative of $x(t)$, the time series data at the fast time scale and $\theta(t_s)$ is the parameter of the system at the slow time scale. Anomalies occur at the slow time scale $t_s$ which is of several order larger than $t$, while observations of the dynamical behavior, based on which inferences are made, take place at the fast time scale $t$.

It is assumed that
• The system behavior is stationary (in the wide sense) at the fast time scale; and

• Any observable non-stationary behavior is associated with parametric changes evolving at the slow time scale.

Using existing methods [28], only sufficiently large changes in the slowly varying parameters may lead to detectable difference in the asymptotic behavior. The need to detect small changes in parameters and hence early detection of an anomaly motivate the proposed stimulus-response approach. In the stimulus-response approach, the dynamical system is perturbed with an appropriate known input excitation to observe the asymptotic behavior at the fast time scale. As a result of the combination of the input stimulus and perturbed parameter(s), it might be possible to observe a detectable change in the nature of asymptotic behavior that would otherwise remain un-perceivable over a long period of time.

From the above perspective, the problem of anomaly detection is categorized into two parts:

a) **Forward problem** : The primary objective of the forward problem is to identify how the system performance is affected by gradually evolving anomalies and to classify the parametric (e.g., change in effective system mass) and non-parametric (e.g., fatigue damage) conditions that affect the system behavior. The problem of anomaly detection focuses on identification of the patterns followed by the dynamical system as the anomaly develops slowly. Solution of the forward problem requires the following steps:

• Generation of time series data from an experimental apparatus under known exogenous stimuli.
• Partitioning of the phase space (or wavelet space) for generation of symbolic
  sequences (on the fast time scale) at different epochs of the slow time scale.

• Finite state machine construction from the symbol sequences and computa-
  tion of the respective state probabilities.

• Computation of anomaly measures the respective state probability vectors
  with reference to the state probability vector under the nominal condition.

b) **Inverse Problem**: The inverse problem focuses on inferring the changes
in system behavior from the anomaly measures based on the observed time series
data. Since this problem may be ill-posed, selection of appropriate stimuli is critical
for prediction of the anomaly range. Nevertheless, the feasible range of parameter
variation estimates can be narrowed down from the intersection of the information
generated from responses under several stimuli chosen in the forward problem.

Solution of the inverse problem requires the following steps:

• Excitation with known input stimuli selected in the forward problem.

• Generation of the asymptotic behavior for each input stimulus as time series
data.

• Embedding the time series data in the phase space determined for the corre-
  sponding input stimuli in Step 1 of the forward problem, if the phase space
dimensionality is small, or generate wavelet transform coefficients of time
series data.

• Generation of the symbol sequence using the same phase-space partition as in
  Step 3 of the forward problem or partition the wavelet coefficients to generate
the symbol sequence.

- $D$-Markov machine reconstruction using the symbol sequence and determining the anomaly measure.

- Identification of the parameter deviation based on the anomaly measure.

This dissertation concentrates on both the problems and validates the methodology with an experimental setup. The forward problem has been dealt with in Chapter 4. The inverse problem has been successfully dealt with in Chapter 5.

The concept of Symbolic Dynamics and its usage for encoding nonlinear system dynamics from observed time series data have been reported in literature [2] [13]. It serves as a tool for behavior description of nonlinear dynamical systems based on the concept of formal languages for transitions from smooth dynamics to a discrete symbolic description [29] [13]. Working with symbols has several important practical advantages [2], one of which is that the efficiency of numerical computations is greatly increased over what it would be for the original data. This efficiency may be the reduction of the need for computational resources as well as speed of computation. The latter may be important for real-time monitoring and control applications. Another advantage is that analysis of symbolic data is often less sensitive to measurement noise. Symbolization, in some cases, can be accomplished directly in the instrument by the appropriate design of low-resolution sensing elements. Such sensors combined with an appropriate analysis can reduce instrumentation cost and complexity. Hence, given the above advantages, applications of symbolic methods are favored in circumstances where robustness to noise, speed, and/or cost are essential.
Figure 3.1 elucidates partitioning of a compact (i.e., closed and bounded) region of the phase space and a mapping from the partitioned space into the symbol alphabet, which becomes a representation of the system dynamics defined by the trajectories. It also shows the construction of a hidden Markov model (HMM) from the symbol sequence as a finite state machine. The state probability vectors represent patterns that are indicative of the nominal (or reference) and anomalous behavior of the dynamical system, as explained in later sections.

Several methods of phase-space partitioning have been suggested in the literature [30] [29]. One such method based on symbolic dynamics is Symbolic false nearest neighbors (SFNN) [31] that finds a ”generating” partition for symbolic orbits to uniquely identify a continuous space orbit; thus, the symbolic dynamics become equivalent to the continuous-space dynamics. This is the main reason of using this method over other existing methods. The key criterion for SFNN partitioning is that short sequences of consecutive symbols ought to localize the
corresponding phase-space points as closely as possible. This is achieved by forming an embedding of the symbolic sequence under the candidate partition and minimizing the apparent errors in phase-space points. The nearest neighbor to each point in the embedding is found in terms of Euclidean distance of symbolic neighbors. In general, better partitions yield a smaller proportion of symbolic false nearest neighbors. For convenience of implementation, the partitions are parameterized with a relatively small number of free parameters. This is accomplished by defining the partitions with respect to a set of radial-basis “influence” functions. The statistic for symbolic false nearest neighbors is minimized over these free parameters [31]. However, this partitioning method may become computationally very inefficient when the dimension of the phase space is large or if the data set is contaminated by noise.

3.2 Wavelet space (WS) partitioning

This dissertation has adopted an alternative partitioning approach for construction of symbol sequences from time series data, which is particularly effective for noisy data from high-dimensional dynamical systems. In this method, called wavelet space (WS) partitioning [32] [13], the time series data are first converted to the wavelet transform data at different scales and time shifts. By using wavelets, one achieves dimensionality reduction of the phase space as only pertinent information is captured in the scales associated with the evolving anomaly. The arrangement of the resulting scale series data in the wavelet space is then partitioned with alphabet size $|\mathcal{A}|$ which indicates the number of partitions of the scale series data by horizontal lines such that the regions with more information are partitioned
finer and those with sparse information are partitioned coarser. In this approach, the maximum entropy is achieved by the partition that induces uniform probability distribution of the symbols in the symbol alphabet [32]. Shannon entropy is defined as

\[ H = - \sum_{i=1}^{|A|} p_i \log(p_i) \]  

(3.2)

where \( p_i \) is the probability of the \( i^{th} \) state and summation is taken over all possible states. Uniform probability distribution of states is a consequence of the maximum entropy partitioning. For a given stimulus, the partitioning of wavelet space must remain invariant at all epochs of the slow time scale.

### 3.2.1 Choice of scales and mother wavelet

For every wavelet, there exists a certain frequency called the center frequency \( F_c \) that has the maximum modulus in the Fourier transform of the wavelet [33]. The pseudo-frequency \( f_p \) of the wavelet at a particular scale \( \alpha \) is given by the following formula [33] [34]:

\[ f_p = \frac{F_c}{\alpha \Delta t} \]  

(3.3)

where \( \Delta t \) is the sampling interval. The power spectral density (PSD) of the signal provides the information about the frequency content of the signal. This information along with Eqn 3.3 can be used for scale selection. The procedure of selecting the scales is summarized below:

- Identification of the frequencies of interest through PSD analysis of time series data.
• Substitution of the above frequencies in place of $f_p$ in Eqn 3.3 to obtain the respective scales in terms of the known parameters $F_c$ and $\Delta t$.

So, before the generation of the wavelet coefficients, the power spectrum of the time series data at the nominal condition must be analyzed to identify the dominant frequencies in the signal. After doing this, one can get the associated set of scales corresponding to the center (resonant) frequency and thus extract information only in that particular frequency spectrum [35]. Since, wavelet coefficients would be sensitive to different stimuli, the system needs to be excited with an apriori known stimulus (e.g., excitation in ultrasonic flaw detectors). Also, choice of wavelet basis function is another open problem not dealt with much in wavelet literature. Hence, this dissertation relied on using different mother wavelets like daubechies [36], haar (or db1), gaus [33] and finally picking up the choice which most closely suits or matches the original signal [35]. Thus, choice of wavelet basis would essentially vary from application to application. The wavelet coefficients of the signal are significantly large when the pseudo-frequency $f_p$ of the wavelet corresponds to the locally dominant frequencies in the underlying signal.

### 3.3 The $D$-Markov machine construction

This section introduces the $D$-Markov Machine Model [13] for representing patterns in a symbolic process, which is motivated from the perspective of anomaly detection [37]. The core concept of the $D$-Markov Machine is succinctly presented below.

The partitioning as described in Section 3.2 is performed at time epoch $t_0$ of the
nominal condition that is chosen to be a healthy condition. A finite state machine is then constructed, where the states of the machine are defined corresponding to a given alphabet $\mathcal{A}$ and window length $D$. The alphabet size $|\mathcal{A}|$ is the total number of partitions of the wavelet space while the window length $D$ is the length of consecutive symbol words forming the states of the machine [13]. The states of the machine are chosen as all possible words of length $D$ from the symbol sequence, thereby making the number $n$ of states to be equal to the total permutations of the alphabet symbols within word of length $D$, (i.e., $n \leq |\mathcal{A}|^D$). The choice of $|\mathcal{A}|$ and $D$ depends on specific experiments, noise level and also the available computation power. A large alphabet may be noise-sensitive while a small alphabet could miss the details of signal dynamics. Similarly, a high value of $D$ is extremely sensitive to small signal distortions but would lead to larger number of states requiring more computation power. For machine construction, the window of length $D$ on the symbol sequence $\ldots \sigma_i \ldots \sigma_k \ldots$ is shifted to the right by one symbol, such that it retains the last $(D-1)$ symbols of the previous state and appends it with the new symbol $\sigma_i$ at the end. The symbolic permutation in the current window gives rise to a new state. The machine constructed in this fashion is called $D$-Markov machine [13] because of its Markov properties. A symbolic stationary process is called $D$-Markov if the probability of the next symbol depends only on the previous $D$ symbols, i.e.,

$$P(\sigma_{i_0}/\sigma_{i-1}, \ldots, \sigma_{i-D}, \sigma_{i-D-1}, \ldots) = P(\sigma_{i_0}/\sigma_{i-1}, \ldots, \sigma_{i-D})$$

The finite state machine constructed above has $D$-Markov properties because the probability of occurrence of symbol $\sigma_i$ on a particular state depends only on the
configuration of that state, i.e., previous \( D \) symbols. For example, if \( \mathcal{A} = \{0, 1\} \), i.e., \(|\mathcal{A}| = 2\) and \( D = 2 \), then the number of states is \( n \leq |\mathcal{A}|^D = 4 \); and the possible states are \( Q = \{00, 01, 10, 11\} \), some of which may be forbidden in the sense that some transitions may not be exist, Fig. 3.2.

![Finite state automaton](image)

**Figure 3.2.** Finite state automaton with \( D = 2 \) and \( \mathcal{A} = \{0, 1\} \)

Once the partitioning alphabet \( \mathcal{A} \) and word length \( D \) are determined at the nominal condition (time epoch \( t_0 \)), they are kept constant for all (slow time) epochs \( \{t_1, t_2, ..., t_k, ...\} \), i.e. the structure of the machine is fixed at the nominal condition. The states of the machine are marked with the corresponding symbolic word permutation and the edges joining the states indicate the occurrence of an event \( \sigma_i \).

The probability of transitions from state \( q_j \) to state \( q_k \) belonging to the set \( Q \) of states under a transition \( \delta : Q \times \mathcal{A} \rightarrow Q \) is defined as

\[
\pi_{jk} = P(\sigma \in \mathcal{A} \mid \delta(q_j, \sigma) \rightarrow q_k); \sum_k \pi_{jk} = 1; \tag{3.4}
\]

Thus, for a \( D \)-Markov machine, the stochastic matrix \( \Pi = [\pi_{ij}] \) describes all transition probabilities between states such that it has at most \(|\mathcal{A}|^{D+1}\) nonzero
entries. The left eigenvector $\mathbf{p}$ corresponding to the unit eigenvalue of $\Pi$ is the state probability vector under the (fast time scale) stationary condition of the dynamical system [13]. To ensure the convergence of the state probabilities, there exists a rule of thumb which is given as $k = N/\eta$ where $k =$ number of symbols in the sequence, $N$ is the number of states of the machine and $\eta > 0$ is a threshold usually taken to be $1 \times 10^{-3}$. On a given symbol sequence $\ldots \sigma_i \tau_i \ldots \sigma_i \ldots$ generated from the time series data collected at slow time epoch $t_k$, a window of length $(D)$ is moved by keeping a count of occurrences of word sequences $\sigma_{i_1} \ldots \sigma_{i_D} \sigma_{i_{D+1}}$ and $\sigma_{i_1} \ldots \sigma_{i_D}$ which are respectively denoted by $N(\sigma_{i_1} \ldots \sigma_{i_D} \sigma_{i_{D+1}})$ and $N(\sigma_{i_1} \ldots \sigma_{i_D})$. Note that if $N(\sigma_{i_1} \ldots \sigma_{i_D}) = 0$, then the state $q \equiv \sigma_{i_1} \ldots \sigma_{i_D} \in Q$ has zero probability of occurrence. For $N(\sigma_{i_1} \ldots \sigma_{i_D}) \neq 0$, the transitions probabilities are then obtained by these frequency counts as follows.

$$\pi_{jk} = P[q_k | q_j] = \frac{P[q_k, q_j]}{P[q_j]} = \frac{P(\sigma_{i_1} \ldots \sigma_{i_D} \sigma)}{P(\sigma_{i_1} \ldots \sigma_{i_D})} \approx \frac{N(\sigma_{i_1} \ldots \sigma_{i_D} \sigma)}{N(\sigma_{i_1} \ldots \sigma_{i_D})} \quad (3.5)$$

where the corresponding states are denoted by $q_j \equiv \sigma_{i_1} \tau_i \ldots \sigma_{i_D}$ and $q_k \equiv \sigma_{i_2} \ldots \sigma_{i_D} \tau_i$.

The time series data under the nominal condition (set as a benchmark) generates the state transition matrix $\Pi_{\text{nom}}$ that, in turn, is used to obtain the state probability vector $\mathbf{p}_{\text{nom}}$ whose elements are the stationary probabilities of the state vector, where $\mathbf{p}_{\text{nom}}$ is the left eigenvector of $\Pi_{\text{nom}}$ corresponding to the (unique) unit eigenvalue. Subsequently, probability vectors $\{\mathbf{p}^1, \mathbf{p}^2, \ldots\}$ are obtained at slow-time epochs $\{t_1, t_2, \ldots\}$ based on the respective time series data. The behavioral changes from nominal condition are described as anomalies which are characterized by a scalar called Anomaly Measure ($\hat{M}$).
The anomaly measure is based on the following assumptions [38]:

- **Assumption 1**: The evolution of damage is an irreversible process, i.e., with zero probability of self healing. This assumption implies the following conditions for all time $t \geq 0$.
  
  i) $\hat{M} \geq 0$;
  
  ii) $\frac{d\hat{M}}{dt} \geq 0$

- **Assumption 2**: The damage accumulation at a slow time epoch $t$, when the dynamical system has reached an equilibrium, is a function of the entire path taken to reach that state.

The amount of damage accumulation in fatigue study can be attributed to a macroscopic variable like crack length. Although the crack length is traditionally defined by a straight line joining the starting point to the tip of the crack, the actual crack follows a complicated path, possibly fractal in ductile materials, to reach a particular point. Therefore, the above assumption 2 implies that the anomaly measure should be determined from the actual path traversed and not just the end points. Accordingly, the path traversed also called the scalar-valued anomaly measure $\hat{M}_k$ at a slow-time epoch $t_k$ is defined in terms of a distance function $d(\bullet, \bullet)$ as:

$$
\hat{M}_k \equiv \hat{M}(t_k) \equiv \sum_{l=1}^{k} d(p^l, p^{l-1}) ; \quad (3.6)
$$

$$
d(p^l, p^{l-1}) = \left( \sum_{j=1}^{|A|} |p^l_j - p^{l-1}_j|^\alpha w(j) \right)^{\frac{1}{\alpha}} \quad (3.7)
$$
where the exponent $\alpha \in [1, \infty)$ depends on the desired sensitivity to small deviations. Small changes in the time series data, which might be due to spurious fluctuations, can be suppressed with large values of $\alpha$. Therefore, the choice of $\alpha$ is a trade-off between suppression of small spurious fluctuations due to noise and those due to the actual changes in the signal profile resulting from damage growth. In this section, the exponent is chosen to be $\alpha = 2.0$, which implies that $d(\bullet, \bullet)$ is the standard Euclidean distance. Also, the weights $w(j)$ are chosen to be the squared mid-points of the respective partitions of alphabet size $|A|$ such that higher levels are heavily weighted in order to capture small changes in the amplitude during the initial phases of fatigue damage. The state probability vector at any time instant corresponds to a singleton point on the unity-radius hypersphere. During fatigue damage evolution, the tip of the probability vector moves along a path on the surface of this hypersphere. The initial starting point of the path is the probability vector with uniform distribution obtained with maximum entropy partitioning (see Section 3.2). As the damage progresses, the probability distribution changes; eventually when a very large crack is formed, complete attenuation of the signal occurs and consequently the tip of the probability vector reaches a point where all states have zero probabilities of occurrence except the one which has a probability 1 (i.e. a delta-distribution); this state corresponds to the partition region where all data points are clustered. The distance travelled by the tip of the state probability vector is calculated from the starting point ($p^0 \equiv$ uniform distribution obtained with maximum entropy partitioning) to the current time epoch along the evolution of the probability vector, see Figure 3.1. The scalar-valued anomaly measure $\hat{M}$ is obtained as:
\[ \hat{M}_k \equiv \hat{M}(t_k) \equiv \sum_{i=1}^{k} d(p_i, p_i^{i-1}) \] (3.8)

The relative distance travelled from the nominal state (i.e., \( p^0 \)) to the current (possibly anomalous) state (i.e., \( p^k \)) can be different for different specimens because of the stochastic nature of the fatigue damage [39].

The following algorithm serves to mitigate possible accumulation of spurious fluctuations [38]:

i. \( \hat{p} = p^0; \)

ii. if \( \left( \frac{d(p^k, \hat{p})}{d(p^f, p^0)} \right) \geq \epsilon \)

\[ I(p^k, \hat{p}) = \frac{d(p^k, \hat{p})}{d(p^f, p^0)}; \quad \hat{p} = p^k; \]

else \( I(p^k, \hat{p}) = 0; \)

iii. \( \hat{M}_k = \hat{M}_{k-1} + I(p^k, \hat{p}); \)

iv. \( k = k + 1; \) Goto step ii;

where \( \epsilon \) is chosen at the start of the experiment without exciting the system. At this point measurement noise should not contribute to anomaly measure and as such the anomaly measure must stay zero. This choice of \( \epsilon \) eliminated the measurement noise and the increment in anomaly measure indicated progression of anomalies after the start of the experiment. \( \epsilon \) is indicative of the trade-off between damage increment and noise effect, and for any damage detection algorithm it will be a non-zero value. Also, choice of \( \epsilon \) will vary from application to application and must be obtained from experimental data.
Appendix A briefly describes the following three pattern recognition techniques, which use time series data as inputs, for comparison with the symbolic-time-series-based anomaly detection method:

- Principal Component Analysis (PCA)
- Multilayer Perceptron Neural Network (MLPNN)
- Radial Basis Function Neural Network (RBFNN)
Experimental Validation of Anomaly Detection

Chapter 2 demonstrated the design and modeling of the laboratory test apparatuses. Chapter 3 explained the theory of symbolic time series analysis (STSA). In this chapter, the experimentally validated results on both the testbeds are presented.

4.1 Introduction

The interest in the ability to monitor a structure and detect damage at the earliest possible stage is pervasive throughout the civil, mechanical, and aerospace engineering communities. For the purposes of this dissertation, damage is defined as changes introduced into a system, either intentional or unintentional, which adversely effect the current or future performance of that system. Most currently used damage identification methods are included in one of the following categories:
visual or localized experimental methods such as acoustic or ultrasonic methods, magnetic field methods, radiography, eddy-current methods or thermal field methods [4]. All of these experimental techniques require that the vicinity of the damage is known \textit{a-priori} and that the portion of the structure being inspected is readily accessible. The need for quantitative global damage detection methods that can be applied to complex structures has led to the development and continued research of methods that examine changes in the vibration characteristics of the structure [40]. The current state of aging infrastructure and the economics associated with its repair have also been motivating factors for the development of methods that can be used to detect the onset of damage or deterioration at the earliest possible stage. Finally, technological advancements including increases in cost-effective computing memory and speed, advances in sensors including non-contact and remotely monitored sensors and adaptation and advancements of the finite element method have contributed to recent improvements in vibration-based damage detection.

This dissertation experimentally validates the symbolic time series analysis (STSA) method as proposed by Ray [13] for damage detection based on time series of remotely placed sensors (see Section 4.2.1) and time series from sensors at the localized failure site (see Section 4.2.2).

4.2 Experimental validation - test apparatus 1

The problem of feature extraction from time series data for structural health monitoring has been recently addressed by many researchers [41] [42]. This chapter focuses on early detection of small anomalies, resulting from fatigue crack damage.
in polycrystalline alloys, which is a major source of failures in structural components of operating machinery [43]. This section makes a comparative assessment of the $D$-Markov machine method with different pattern recognition techniques for anomaly detection (see Appendix A). Time-series data, generated from experiments on the test apparatus in Figure 2.1, have been used for this purpose. Both actuators in the test apparatus were excited by a sinusoidal input of amplitude 0.85 V and frequency 11.39 Hz (approximately the resonance frequency) throughout the run of each experiment. The time series data of the displacement sensor on Mass#3, which serve as an indicator of the system performance, were collected from the beginning of the experiments until breakage of specimens. The ensemble of data were saved in a total of 35 files, with each file containing a minute of sensor time-series data. Based on the stopping rule (see Section 3.3), 12000 points were used in the analysis. Following the procedure outlined in Chapter 3 and Section A.1, the anomaly measure was obtained from the data at each 2 minute interval from the sensor data contained in each file. This separation between time epochs was chosen based on the fact that the average life span of the specimen under the loading condition was $\sim 30$ minutes. The time-series data sets were collected after the dynamic response attained the stationary behavior. The data set at the time epoch of 2 minutes was taken as the reference point representing the nominal behavior of the dynamical system.

The following anomaly detection approaches explained in detail in Appendix A are investigated by using the same set of time series data generated from the experiments on the test apparatus in Figure 2.1:

- Principal Component Analysis (PCA)
The next four paragraphs describe how anomaly measures are calculated based on the above five techniques of pattern recognition. The fourth paragraph addresses both SFNN and WS methods of partitioning in the $D$-Markov method.

Following the Principal Component Analysis (PCA) procedure described in
Section A.1.1, a block of sampled time series data, having length $\ell = 12000$, is divided into $n = 4$ segments of length $d = 3000$; these segments are arranged to form a $3000 \times 4$ data matrix. The resulting $4 \times 4$ (symmetric positive-definite) covariance matrix of the data matrix yields a monotonically decreasing set of eigenvalues, $\lambda_1...\lambda_4$, and the associated orthonormal eigenvectors $v^1, ..., v^4$. At the nominal condition, three eigenvalues are found to be dominant (i.e., $q = 3$) for a threshold of $\eta = 1.0 \times 10^{-2}$ such that

$$\frac{\sum_{i=3}^{4} \lambda_i}{\sum_{i=1}^{4} \lambda_i} = 0.0001 < \eta$$

The matrix $\tilde{M}_0$ in Eq. (A.3) is calculated from the data set at the nominal condition. Similarly, the matrices $\tilde{M}_2, \tilde{M}_4, ..., \tilde{M}_{32}$ are obtained corresponding to different time epochs, respectively. The anomaly measures at different time epochs are determined according to Eq. 3.8 relative to nominal matrix $\tilde{M}_0$.

**Figure 4.2.** Maximum entropy partitioning of scale series data
Following the Multilayer Perceptron Neural Network (MLPNN) procedure, described in Section A.1.2, the resulting pattern matrix $\mathcal{P}_0$ is made of $N = 100$ columns. Each column, having a length $\ell = 25$, is generated from the time series data at nominal condition to train the MLPNN that is chosen to have a input layer (with 30 neurons), 4 hidden layers (with 50 neurons in layer 1, 40 neurons in layer 2, 30 neurons in layer 3, and 40 neurons in layer 4), and the output layer (with 5 neurons): this structure of the MLPNN yields very good convergence for the data sets under consideration. The target corresponding to each input pattern vector is chosen to be $5 \times 1$ zero vector. The MLPNN is trained with the nominal data set at; the gradient descent back-propagation algorithm has been used for network training with an allowable performance mean square error of $1.0 \times 10^{-5}$. The input pattern matrices, $\mathcal{P}_0, \mathcal{P}_2, \mathcal{P}_4 \ldots \mathcal{P}_{32}$, each of dimension $(25 \times 100)$, are then generated from the anomalous data sets at various time epochs to excite the trained network. The resulting output matrices of the trained MLPNN are $\mathcal{O}_0, \mathcal{O}_2, \mathcal{O}_4 \ldots \mathcal{O}_{32}$, which yield the respective performance vectors, $u_0, u_2, u_4 \ldots u_{32}$. The anomaly measures at different time epochs are determined according to Eq. 3.8.

Following the Radial Basis Function Neural Network (RBFNN) procedure, described in Section A.1.3, the length of the sampled time series data is chosen to be $N = 12000$. The exponent $\alpha$ for standard RBF is usually chosen to be 2, for improved anomaly measure sensitivity. An estimate of the parameters, $\mu$ and $\theta_\alpha$, are obtained according to Eq. A.11 based on the data under the nominal condition, which yields the requisite radial basis function $f_0$ following Eq. A.10. The anomaly measures at different time epochs are determined according to Eq. 3.8.
Based on the time series data of the nominal condition, the first step is generation of wavelet scale series data. The wavelet basis function 'gaus2' [33] is chosen as it closely represented the sinusoid signals. Based on the PSD plot as shown in Figure 4.1, the scale associated with the center frequency was identified. The wavelet coefficients were computed at the selected scale and the resulting 1-D signal was partitioned by using the methods described in Section 3.2. Based on the time series data of the nominal condition, the first step in the the $D$-Markov machine method is to find a partition for symbol sequence generation. The partitioning methods, SFNN and WS, described in Chapter 3, have been investigated for efficacy of anomaly detection. For the given stimulus of this experiment, partitioning of the phase space/wavelet space must remain invariant at all epochs of the slow time scale. Absolute values of the wavelet scale series data (see Chapter 3) were used to generate the partition because of the symmetry of the data sets about their mean. The partitioning using maximum entropy can be seen in Figure 4.2. The finite state machine constructed with the choice of the parameters $|\mathcal{A}| = 8$ and $D = 1$ has only 8 states and it was able to capture early anomalies. Increasing the value of $|\mathcal{A}|$ further did not improve the results as seen in Figure 4.3 and increasing the value of depth $D$ created a large number of states of the finite state machine, many of them having very small or zero probabilities. Hence, the value of $D = 1$ was used for construction of the $D$-Markov machine for data at all time epochs. Following the procedure, described in Chapter 3, Section 3.3, the state machines are constructed and the connection matrix $\Pi \equiv [\pi_{jk}]$ and the state probability vector $\mathbf{p}$ for each phase plot. The state machines were constructed with a symbol alphabet of size 8. As seen in Eq. 3.8, $\epsilon$ is chosen at the start of the
experiment without exciting the system. At this point measurement noise should not contribute to anomaly measure and as such the anomaly measure must stay zero. This choice of $\epsilon = 0.005$ practically eliminated the measurement noise and the increment in anomaly measure indicated progression of anomalies after the start of the experiment.

![Figure 4.3. Anomaly measure for alphabet size 6, 8 and 10](image)

**4.2.1 Comparison of anomaly detection methods**

The anomaly detection method is validated on the laboratory test apparatus in Figure 2.1. Test runs are planned for exogenous stimuli at one of the natural frequencies of the mechanical structure to cause fatigue failure within reasonable length of time. Using the same set of time series data generated from the above experiments, the five plots in Figure 4.4 compare the anomaly measures obtained by using the previously described five anomaly detection approaches, WS, SFNN,
PCA, MLPNN and RBFNN, for the first 16 files (i.e., up to \(\sim 32\) minutes) when
the service life of the test specimen has expired, i.e., the specimen is about to
break. (Note: The estimated service life of the specimen under this load excitation
is \(\sim 30\) minutes.) The nominal condition is chosen at the time epoch of 2 minutes
to ensure that all transients have decayed. Each of the five methods shows a
sharp rise in the anomaly measure at about 22 minutes, which is indicative of the
transition from the crack incubation to crack propagation phase in fatigue damage.
The symbolic time series-based anomaly detection with both WS and SFNN yield
better performance than the three other existing pattern recognition tools and are
responsive to progressive damage in the short crack regime till the time epoch of
\(\sim 22\) minutes. At this time, cracks begin to form, indicating the transition from
crack incubation to crack initiation.

Figure 4.4 shows that symbolic time series analysis under both WS and SFNN
partitioning is able to capture the progressive accumulation of fatigue damage in
the crack initiation phase while the performance of PCA and RBFNN is some-
what inferior and MLPNN is the worst. This is indicated by the magnitude of the
anomaly measure as well as the change in slope starting at \(\sim 15\) minutes. Slope
of the anomaly measure curve is representative of the rate of damage progression.
The distributed nonlinearities in MLPNN may not be specifically suited to capture
these small parameter perturbations in the largely linear behavior of the dynamic
response of the vibrating structures. The WS-partitioning shows higher anomaly
measure as compared to PCA. The rationale is that the PCA method is depen-
dent on eigenvalues and eigenvectors of the covariance matrix that is sensitive to
measurement noise in the data acquisition process. In contrast, the symbolic time
series analysis, with both WS and SFNN partitioning, is much less sensitive to (zero-mean) measurement noise because of the inherent averaging due to repeated path traversing in the finite-state machine. All five anomaly detection methods exhibit a gradual increase in anomaly measure from the beginning up to the time epoch of \( \sim 22 \) minutes when the transition takes place from the crack initiation to crack propagation regime. RBFNN, MLPNN and PCA exhibit good performance in the long crack regime in the 24-28 minute interval. The specimen breaks at \( \sim 28 \) minutes. The various pattern recognition techniques discussed earlier are data driven and hence application based. The actual validity or generalization of the result cannot be claimed. The techniques have to be tested on off-line data acquired from the actual process and apriori knowledge of the physical process would help in efficient usage of these tools to the problem at hand. Some observations that can be drawn from the results in Figure 4.4 which can help in choice of technique:

1. PCA is dependent on eigenvalues and eigenvectors of the covariance matrix that is sensitive to measurement noise in the data acquisition process.

2. Distributed nonlinearities in MLPNN may not be specifically suited to capture small parameter perturbations in the largely linear behavior of the dynamic response. Also, choice of number of neurons, hidden layers has not been dealt with in literature.

3. STSA has less number of parameters to optimise and has shown considerably good performance in comparison with other pattern recognition techniques and has been effectively demonstrated in a previous publication [44] [45].
Figure 4.4. Performance comparison of anomaly detection methods
4.2.2 Anomaly detection using multiple sensors

This section focuses on integrating the information generated from both the ultrasonic sensors and the sensors measuring the global variables like accelerometers, LVDTs, and load cells for characterization of fatigue damage in the small crack regime. This section describes the experimental validation of symbolic dynamic tools for early detection and quantification of fatigue damage based on time-series data, generated from the sensors mounted on the test apparatus as shown in Figure 2.1. In the experiments, both shakers are excited by a sinusoidal input of amplitude 0.85 V and frequency 14.19 Hz (91 rad/s or the higher mode of the system) throughout the run of each experiment. The time series data from displacement, accelerometer, load cell and ultrasonic sensors, mounted on the experimental apparatus (see Table 2.2), were collected from the start of the experiments after the system response attained the stationary behavior. The results shown in Figure 4.6 are based on the analysis of data collected from sensors mounted on mass # 1 and Shaker # 1. The time series data for each sensor were saved for a total period of ~90 kilocycles in 110 files. One minute of time-series data was stored in each file for all sensors. The total life of the specimen was ~110 min. The data set at the beginning stage of experiments served as the reference point representing the nominal behavior of the dynamical system. The anomaly measure at the nominal condition was chosen to be zero and was subsequently updated at ~ 3 minute intervals.

Absolute values of the wavelet scale series data (see Section 3.2) were used to generate the partition because of the symmetry of the data sets about their mean. The alphabet size, window length, and the wavelet basis were chosen as $|\mathcal{A}| = 8$, ...
$D = 1$ (see Chapter 3) and the wavelet basis ’gaus2’ [33], respectively, for time series data sets of all sensors (displacement, accelerometer, load cell and ultrasonic). Increasing the value of $|A|$ further did not improve the results and increasing the value of depth $D$ created a large number of states of the finite state machine, many of them having very small or zero probabilities. The finite state machine constructed with the choice of the parameters $|A| = 8$ and $D = 1$ has only 8 states and was able to capture early anomalies. The wavelet basis of ’gaus2’ provided better results than many other wavelets of the Daubechies family [46] because it closely matches the shape of the sinusoidal signals.

The three plots in Figure 4.6 represent the evolution of anomaly measure de-
rived from the time series data of accelerometer, load cell and displacement sensors respectively. These sensors measure the changes in the global performance of the system due to evolving fatigue damage in a localized region of the system. Fatigue crack growth in the specimen connecting Mass # 1 and Mass # 3 (see Table 2.2) changes the stiffness of the material when the crack reaches the long crack region. Consequently, the dynamics of the system are altered because of coupling between different components of the apparatus (see Figure 2.1). As a result, the time series data of these sensors record changes in the statistical patterns as shown in Figure 4.6. The three plots in Figure 4.6 are normalized by dividing the anomaly measure values with the respective maximum for comparative evaluation.

Figure 4.7 exhibits the normalized anomaly measure plot derived from the time series data of ultrasonic sensors that are mounted on the specimen connecting Mass # 1 and Mass # 3 (see Table 2.2). This specimen is subjected to a relatively higher load because of resonance and synchronization of two shakers and therefore has higher probability of failure. Ultrasonic sensors capture the effects of fatigue growth in the localized region of the mechanical structure. The crack propagation stage starts as multiple small cracks coalesce into a single large crack. At this point, a rapid change is observed in the profile of the ultrasonic signal. Finally, the specimen breakage causes complete attenuation of the received ultrasonic signal.

As seen in Figure 4.7, ultrasonic sensors capture slow progression of fatigue damage in the localized region of the mechanical structure from the very beginning. The sharp change in the slope at \( \sim 58 \) minutes indicate the transition from the crack initiation phase to crack propagation phase. The results from displacement, load cell and accelerometer sensors, as seen in the three plots of Figure 4.6
show similar trends for all three sensors. These plots show little change in the anomaly measure during the crack initiation phase because small fatigue cracks do not significantly affect the global performance; subsequently, the effects of fatigue damage on the overall system behavior become noticeable. The three plots in Figure 4.6 exhibit a slope change at $\sim 45\text{min}$. These results for localized damage and its effect on the global performance of the system indicate that a control strategy can be built to take predictive actions, based on the derived damage information, for fault mitigation and control. In this experiment, the STSA method captured the gradual progression of faults much earlier than the occurrence of catastrophic failure. This is of paramount practical importance as it can provide ample time for the hierarchical supervisory control system to execute decision and control laws for life extension without significant loss of performance [47]. This is an area of future research.
Figure 4.6. Normalised anomaly measure plots derived from accelerometer, load cell and displacement sensors
4.2.3 Sources of uncertainty

As seen in Chapter 3, by Equation 3.8, damage is an irreversible phenomenon throughout life that ultimately leads to retirement or failure. We need to be able to model this type of cumulative damage (CD) in-order to access reliability and life cycle cost for systems. The CD stems from material behavior at the atomic or microscopic level. However, currently there are no physical models in the research literature in the short crack region (no surface visibility of the crack). Hence, CD models have to be phenomenological in nature, that is, they must be based on current understanding of the phenomena at the macroscopic level. Also, inferences have to be based on data from sensors. This dissertation has modelled the fatigue (under bending) phenomena using mechanical sensors like accelerometers, load
cells and displacement transducers (see Section 4.2.1) and pure damage sensing
devices like ultrasonic flaw detection system (see Section 4.2.2).

The following sources of uncertainty are the main causes for the variability \cite{48} in observed fatigue life \cite{49}:

- **Initial conditions** - The initial conditions of each specimen varies and the initial state of damage is uncertain for the components in the system. The initial damage in the component depends upon virgin material properties, manufacturing processes, inspection techniques, installation procedures, storage procedure before installation, and any prior usage.

- **Material Response** - This quantifies the change in behavior of the material under the fatigue loading and response to fatigue under identical loading conditions

- **Severity of loading** - The loading is found to vary only $\sim 0.5$ percent and its effect on fatigue life variation can be safely assumed to be insignificant.

- **Progression of damage and changing material properties during the fatigue process.**

To compensate for the above sources of uncertainty, ensemble of data has been collected under identical experimental conditions as seen in Figure 4.8.

### 4.2.4 Ensemble of experimental results

The model representing the damage evolution process must accurately describe and analyze data, and predict behavior under conditions not covered by data.
Consistent with these requirements, the model must have few parameters and also encompass the phenomena in a reasonable complete manner. This dissertation makes use of a hidden Markov model (HMM), derived from time series data of pertinent measurement(s) as shown in Section 4.2.1. Physical modelling of fatigue damage in the short crack region has not been dealt with significantly in the technical literature. The concept of symbolic time series analysis (STSA) [2] as described in Chapter 3 has been applied for early detection of slowly evolving anomalies [13] and a comparative evaluation of this novel analytical method shows its better performance relative to other existing pattern recognition tools in terms of early detection of small changes in dynamical systems [44]. This has been convincingly shown in Section 4.2.1.

Figure 4.8. Ensemble of results under identical loading conditions
4.3 Life extending control

The key idea of life extending control of mechanical structures is that substantial improvements in the useful service life of critical components can be achieved by insignificant reduction in the system dynamic performance (Ray et al. [50]). Zhang and Ray [47] and Zhang et al. [17] have demonstrated the efficacy of damage mitigation on a laboratory test apparatus, where intermittent overload pulses were deliberately injected as part of the feedforward control signal to cause crack retardation so that structural durability of critical structures is increased. However, damage-mitigation may not be meaningful unless the malignant anomalies leading to catastrophic failures are detected at an early stage. Also, control action should be delayed as far as possible so as to enable the system to perform at the desired level for maximum length of time.

The steps used in formulating the control policy are as follows and described in sections below:

1. Selection of anomaly measure threshold $\hat{M}_{th}$ based on stochastic data of ultrasonic sensors collected under identical loading conditions.

2. Calculation of curvature $\rho$ (i.e., change of slope) of anomaly measure plots after crossing the anomaly measure threshold $\hat{M}_{th}$. A sharp curvature indicates transition from crack initiation to crack propagation phase.

3. Selection of curvature threshold $\rho_{th}$ to decide the point of control action in *real-time.*
4.3.1 Selection of anomaly measure threshold

Experiments were conducted under identical loading conditions (sinusoidal pulses with peak to peak amplitude of 0.85 V) to generate a stochastic data set of 10 identical samples as shown in Figure 4.9. The threshold for anomaly measure is selected based on observation of the stochastic data. The anomaly measure value $\hat{M}^*$ for each data set at the point of transition from crack initiation to crack propagation phase was recorded (see Figure 4.10). The value of threshold $\hat{M}_{th} = 0.15$ (see Figure 4.9) was chosen with a conservative margin such that $\hat{M}_{th} < \min(\hat{M}^*) = 0.1695$. This threshold provides an early warning of impending failure and one can soon expect a sharp deviation from the nominal behavior. At this point, small discontinuities and multiple short cracks have originated inside the specimen. Beyond this point, the behavior transits from short to long crack regime. The short crack regime is also evident by initial rise in the anomaly measure. This threshold also ensures that the control policy is robust to spurious effects during the initial short crack regime. This threshold allows the system to operate at the desired performance for maximum amount of time before the onset of widespread fatigue damage.

4.3.2 Selection of curvature threshold

After the crossover of the anomaly measure threshold $\hat{M}_{th}$ in each sample of the stochastic data, curvature $\rho_k$ was calculated at every time epoch $t_k$ by the following equation

$$\rho_k \equiv (\hat{M}_k - \hat{M}_{k-1}) - (\hat{M}_{k-1} - \hat{M}_{k-2})$$
The value of threshold $\rho_{th} = 0.024$ (see Fig. 4.10) was chosen such that $\rho_{th} < \min(\rho^*) = 0.0243$. The point of control action is chosen at the time epoch where curvature threshold is reached indicating a transition from crack initiation to crack propagation phase. At this point, remaining life (with no control action) must be estimated (see Section 4.3.3) for a comparative evaluation with the observed life under the control action.
4.3.3 Remaining life prediction

The traditional approach to remaining life prediction is based on S-N curves of the test material. However, this approach is incapable of accurately identifying the rate at which fatigue cracks develop, and cannot be used for remaining life prediction in real-time [51]. Standard model based analysis [52] used to estimate fatigue crack growth rate are infeasible for real-time applications because of the stochastic nature of fatigue damage. Recently, stochastic modelling of fatigue crack propagation has been reported [39]. However, this model is based on long crack data and is incapable of issuing early warnings in the region of transitions from short crack to long crack regime. Therefore, this section has utilized ensemble of data collected from ultrasonic sensors which are capable of capturing evolving fatigue damage in the short crack regime. Also, the approach used in this section takes advantage of the real-time computation of anomaly measure $\hat{M}$ which is indicative of damage (relative to the initial nominal condition). At the point of control action, the remaining life can be estimated as follows:

\[
R_\rho = f(\hat{M}, N); \hat{M} > 0, N > 0 (4.2)
\]

where $R_\rho$ is the expected remaining life at the point where $\rho_{th}$ is crossed, $\hat{M}$ is the corresponding anomaly measure and $N$ is the used life at that time instant. The function is assumed to be a quadratic (second-order) polynomial model of the form:

\[
R_\rho = a_0 + a_1\hat{M} + a_2N + a_3\hat{M}^2 + a_4N^2 + a_5\hat{M}N (4.3)
\]
where the coefficients $a_i's$ are identified by least square multiple regression fit with the remaining life of the stochastic data (see Figure 4.10). The squared terms of the polynomial are called the quadratic effects and are used to model curvature in the response surface. The coefficients were identified as $a_0 = 1.555 \times 10^4$, $a_1 = 10.8 \times 10^4$, $a_2 = -2.03$, $a_3 = -21.16 \times 10^4$, $a_4 = 0.335 \times 10^{-4}$ and $a_5 = 1.876$. Figure 4.11 shows the distribution of estimated and observed remaining life of the stochastic data.
4.3.4 Experimental results and discussion

This section presents the experimental results describing the effects of control action on structural durability, i.e., life of test specimens. Experiments were conducted at a sinusoidal load with peak to peak amplitude of 0.85 V to collect the stochastic data for off-line analysis. Based on the collected data for 10 samples, the specimens have an expected total life of 27150 cycles with a standard deviation of 6450 cycles. The proposed life extending methodology (described in Section 4.3) has been implemented on-line and the resulting real-time plots are shown in Figure 4.12. For symbolic time series analysis (STSA) of the ultrasonic data, the finite state machine was constructed using alphabet size $|\mathcal{A}| = 8$, and depth $D = 1$. Increasing the value of $|\mathcal{A}|$ led to no further improvement in anomaly detection.

![Figure 4.11. Observed and Estimated remaining life after curvature threshold is crossed for 10 samples](image-url)
and increasing the depth $D$ created a large number of machine states, many of them having very small or zero probabilities. The finite state machine constructed with this choice of the parameters has only 8 states and was able to capture early anomalies. The wavelet basis function was chosen to be ‘$gaus2$’ as it closely fits the sinusoidal signal [33].

The two plates in Figure 4.12 represent the anomaly measure plots for experiments conducted on two identical specimens with different control actions (i.e., load reduction in the experiments that were taken in real time based on the control strategy in Section 4.3). The estimated life and the actual life after taking the control actions are presented in Table 5.1. For total life calculation, experiments were stopped when the ultrasonic signal was completely attenuated upon breakage of the specimen. With 20% load reduction, the percentage increase in life was about 108% of the expected total life. Also, with 10% load reduction, the increase in life amounted to about 32%. In-order to scale up the procedure to large scale structures, one needs to identify the important characteristics of the diagnosis curve and devise a control strategy accordingly.

<table>
<thead>
<tr>
<th>% Load Reduction</th>
<th>% Load Reduction</th>
<th>Estimated Life</th>
<th>Observed Life</th>
<th>Extended Life (% total life)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>28000</td>
<td>7000</td>
<td>18250</td>
<td>32</td>
</tr>
<tr>
<td>20</td>
<td>23500</td>
<td>6400</td>
<td>38900</td>
<td>108</td>
</tr>
</tbody>
</table>
Figure 4.12. Realtime life extension for two experiments with different levels of load reduction

Time-series data from the tip accelerometer, generated from experiments on the test apparatus in Figure 2.10, have been used for this purpose. The base accelerometer time series at all time epochs was found to be statistically stationary as the reference input excitation remain unchanged for all time epochs. Both the horizontal and back actuators in the test apparatus were excited by a trapezoidal reference input of amplitude 0.5 V and frequency 6.35 Hz as seen in Figure 4.13 throughout the run of each experiment.

The reference excitation signal fed to the closed loop plant is trapezoidal in nature as shown in Figure 4.13 and generated in a way such that either the horizontal/back actuator are in motion. This is achieved by a proper combination
of a rectified sine wave, its phase-shifted version and a pulsating sequence as described in Appendix A. This ensures less power consumption from the hydraulic unit, thereby reducing overheating of the hydraulic fluid limiting pressure spikes and also reduces noise. The time series of the accelerometer on tip of the beam, which serve as an indicator of the system performance, was collected from the beginning of the experiment for 2 minutes. The mass was shifted by 0.5 inches towards the tip and data collected after steady state was reached. The ensemble of data were saved in a total of 22 files, with each file containing 50 seconds of sensor time-series data as shown in Figure 4.14. Following the procedure outlined in Chapter 3, the anomaly measure was obtained from the time series data which represented the statistical pattern of the dynamics of the system with the mass at
that position. The time-series data sets were collected after the dynamic response attained the stationary behavior. The data set when the mass was at the farthest from the tip was taken as the reference point representing the nominal behavior of the dynamical system.

![Time series data for the nominal condition](image)

Figure 4.14. Time series data for the nominal condition

Based on the time series data of the nominal condition, the first step in the the D-Markov machine method is to find a partition for symbol sequence generation. The partitioning method described in Chapter 3, has been investigated for efficacy of anomaly detection. For the given stimulus of this experiment, partitioning of the phase space/wavelet space must remain invariant at all epochs of the slow time scale. Absolute values of the wavelet scale series data (see Chapter 3) were used to generate the partition because of the symmetry of the data sets about their mean.
The finite state machine constructed with the choice of the parameters $|\mathcal{A}| = 4$ and $D = 2$ has only 16 states and it was able to capture early anomalies. Increasing the value of $|\mathcal{A}|$ further did not improve the results and increasing the value of depth $D$ created a large number of states of the finite state machine, many of them having very small or zero probabilities. Hence, the value of $D = 2$ was used for construction of the $D$-Markov machine for data at all time epochs.

As seen in Figure 4.15, the time series of the Y-axis of the accelerometer mounted on the base does not change much in peak to peak magnitude but the texture of the signal changes. This is captured by the wavelet space partitioning and $D$-Markov machine as described in Chapter 3. The change in texture of the time series (after de-noising is performed to filter out high dimensional noise) as
mass moves from 2.5 inches to 7.5 inches is very well captured by STSA. From 7.5 inches to 10 inches, one can see a drastic change in magnitude of the accelerometer data at 9.5 inches which occurs due to matching of the excitation frequency with one of the natural modes of the beam - mass system. Furthermore, increase in magnitude in the accelerometer time series is due to significant changes in the effective mass/moment of inertia of the system components as the mass moves closer to the tip.
This chapter presents the Symbolic Time Series Analysis of bearing acceleration data for detection and estimation of parametric changes in flexible diaphragm couplings due to angular misalignment between shafts. The accelerations are from a simulation model which encompasses the effects of angular misalignment in the coupling by translating the exerted forces (vibrations) to the bearings. This work has been done in collaboration with Intelligent Automation, Inc. and has been supported in part by the U.S. Army Research Laboratory and the U.S. Army Office under Grant No. DAAD19-01-1-0646.

5.1 Introduction

Critical components such as bearings, seals, and couplings are subjected to unbalanced axial/radial loads and excessive machine vibrations due to shaft misalign-
ment in rotating machinery. Recent advances in laser alignment technology \cite{53,54} have made it possible to align the driver with the driven machine within a very fine tolerance. The objective of accurate shaft alignment is to increase the operating lifespan of rotating machinery. To achieve this goal, machinery components that are most likely to fail must operate within their design limits. Since the components that are most likely to fail are the bearings, seals, coupling, and shafts, accurately aligned machinery \cite{55} will achieve the following results:

1. Reduce excessive axial and radial forces on the bearings to insure longer bearing life and rotor stability under dynamic operating conditions.

2. Minimize the amount of shaft bending from the point of power transmission in the coupling to the coupling end bearing.

3. Minimize the amount of wear in the coupling components.

4. Eliminate the possibility of shaft failure from cyclic fatigue.

Despite these efforts, extremely precise alignment may not be possible due to rotor imbalance and lateral deflections of the rotating shafts. Consequently, harmonic forces are generated, which cause large stresses and premature failures in bearings, seals, and couplings.

Shaft misalignment is the deviation of relative shaft position from a collinear axis of rotation measured at the points of power transmission when equipment is running at normal operating conditions. Shaft misalignment is categorized into two classes \cite{56}.

- *Offset misalignment* that occurs if the centerlines of two shafts are parallel but do not meet at the power transfer point;
- *Angular misalignment* that occurs if the non-parallel centerlines of two shafts intersect at the power transfer point.

Figure 5.1 shows a typical misalignment situation on a driver and a driven machine. For a flexible coupling to accept both parallel and angular misalignment there must be at least two points where the coupling can 'flex' or give to accommodate the misalignment condition. By projecting the axis of rotation of the driver shaft toward the driven shaft (and conversely the driven shaft rotational axis towards the driver shaft) there is a measurable deviation between the projected axes of rotation of each shaft and the actual centerlines of each shaft where the power is being transmitted through the coupling from one 'flexing' point to another. Since we measure misalignment in two different planes (vertical and horizontal) there will be four deviations that occur at each coupling. When rotating equipment is started, the shafts will begin to move to another position.

Often misalignment in actual machinery exhibits a combination of both offset and angular misalignment. Coupling failure, bearing failures, distorted rotors, and bearing housing damage are the commonly encountered effects of misalignment [57]. Proper shaft alignment reduces axial and radial loads with the associated reduction of noise, vibration, and local temperature rise. It also decreases wear of mechanical components and practically eliminates the possibility of shaft failure due to cyclic fatigue. Thus, availability, reliability, and maintainability are enhanced with the attendant reduction of life cycle cost [55]. Angular misalignment produces oscillatory forces that, in turn, generate cyclic stresses in the coupling materials. However, recent research [58] reveals that angular misalignment has a much smaller impact on bearing life than offset misalignment.
Furthermore, misalignment is not easy to detect on machinery that is running. The radial forces transmitted from shaft to shaft are typically static forces (i.e. unidirectional) and are difficult to measure externally. There are no analyzers or sensors that one can place on the outside of a machine case to measure how much force is being applied to the bearings, shafts, or couplings. One has to rely on vibration characteristics of the entire rotor dynamical system as a whole to identify the reduction or parametric changes like the changes in stiffness of the coupling. Hence, changes in statistical patterns of the time series of accelerometer(s) mounted on the bearings would serve as an indication of change in effective stiffness and health status of the rotor dynamical system as a whole. The objective here is to investigate the effects of angular misalignment on the life of flexible couplings and infer the changes in coupling stiffness for prognosis of fatigue failure after prolonged usage under the following assumptions.

1. The rotor shaft is sufficiently rigid so that the offset misalignment at steady state can be neglected.
2. The initial angular misalignment is bounded within tolerance limits of 0.05 degrees (∼ 0.8 mills/inch).

3. The load torque is statistically stationary with a small coefficient of variation (i.e., ratio of standard deviation to the mean) and small fluctuations in the torque are filtered out due to inertial effects in the system dynamics.

4. The shaft speed is proportional to the driving torque (e.g., positive displacement pumps, mixers, and extruders).
The nomenclature used in the Chapter is as follows:

\( \beta \) - rotation of diaphragm about \( x \) axis
\( \gamma \) - rotation of \( y \) diaphragm about axis
\( \varepsilon \) - displacement of diaphragm center in \( x \) direction
\( \eta \) - displacement of diaphragm center in \( y \) direction
\( \Omega \) - rotation rate of the shaft about its bearing
\( \Pi_k \) - state transition matrix at time epoch \( t_k \)
\( C_b \) - bearing rotation damping coefficient
\( F_b \) - force vector applied on bearing
\( g \) - acceleration due to gravity
\( I_2 \) - inertia of the diaphragm along the diameter
\( J_z \) - shaft rotation inertia
\( K \) - diaphragm stiffness coefficient
\( k_b \) - bearing spring stiffness in \( x \) and \( y \) direction
\( k_\varepsilon \) - elastic constant corresponding to translation in the \( x \) direction
\( k_\eta \) - elastic constant corresponding to translation in the \( y \) direction
\( k_\beta \) - elastic constant corresponding to rotation about the \( x \) axis
\( k_\gamma \) - elastic constant corresponding to rotation about the \( y \) axis
\( m \) - mass of diaphragm
\( \mathcal{P} \) - pattern matrix generated in the analysis part
\( r_c \) - radius of diaphragm
\( r_b \) - radius of bearing
\( H \) - entropy of the symbol sequence
\( T \) - torque applied on the shaft
\( x_c \) - vertical displacement of diaphragm center relative to the fixed coordinate
\( y_c \) - horizontal displacement of diaphragm center relative to the fixed coordinate along the un-deformed shaft axis
\( x_b \) - vertical displacement of the bearing relative to the fixed coordinate
\( y_b \) - horizontal displacement of the bearing relative to the fixed coordinate

### 5.2 Problem formulation for anomaly detection

![Figure 5.2. Overall architecture of the symbolic time series analysis method for anomaly detection in flexible mechanical couplings](image)

The entire framework for the problem is depicted in Figure 5.2. The inputs to the model are: (i) The instantaneous torque that is assumed to be constant under steady state operation.; and (ii) Range of the bending stiffness coefficients, \( k_\beta \) and \( k_\gamma \), of the diaphragm. The outputs of the model are the bearing accelerometer readings that, in turn, are provided to the symbolic time series analysis (STSA) subsystem. The output of the STSA subsystem is anomaly measure (see Chap-
ter 3) curve which indicates the progress of anomaly (i.e., reduction in stiffness) in the disc/diaphragm coupling. Under multiple torque inputs in the range of 3000-6500 N-m, a family of anomaly measure curves is generated that provides a ensemble pattern of anomaly growth (see Section 5.4.1). This pattern set is utilized by the inverse analysis section to generate an inverse look up table (see Section 5.4.2). This table provides the estimates of mean, std. deviation and confidence intervals of diaphragm stiffness for any arbitrary value of anomaly measure obtained in real time. The entire architecture has been built in Simulink within the MATLAB environment. The proposed architecture in Figure 5.2 is application-independent and is portable to other applications.

5.3 Description of coupling simulation

This section presents nonlinear modelling of a flexible disc/diaphragm coupling, based on the whirling phenomena [56], to generate simulated data of accelerometer readings under different load excitation. The flexible coupling under consideration connects two shafts, end-to-end in the same line. The main objective is to transmit power/torque from one shaft to the other, causing both to rotate in unison, at the same rotational speed. Another objective is to compensate for minor amounts of misalignment and random movement between the two shafts. The misalignment is primarily caused due to dynamic movement (i.e., whirling) of shafts at high speeds.

Figure 5.3 elucidates the effects of whirling phenomena in turbo machinery in which a rotating shaft undergoes deformation like a beam. The shaft supports the diaphragm that is assumed to be of negligible mass relative to the shaft and is flexible in bending but rigid in extension and torsion. The diaphragm is welded to
the shaft, such that its center $C$ coincides with the center line of the shaft. The rotational speed of the shaft about its bearing is constant at $\Omega$.

Let $x$-$y$-$z$ be a reference frame that rotates at the constant angular speed $\Omega$, with its $z$ axis concurrent with the line connecting the bearings. Let the deflected position of the center of the diaphragm, relative to the $x$, $y$, $z$ axes, be denoted as $\vec{r} = \varepsilon \vec{i} + \eta \vec{j}$ where $\vec{i}$ and $\vec{j}$ denote the unit vectors along $x$ and $y$ axis respectively. Furthermore, let $x'$, $y'$, and $z'$ be the principal centroidal axes for the diaphragm, and let $\beta$ and $\gamma$ be the angular rotations about the $x$ and $y$ axes, respectively, of the $x$-$y$-$z$ frame. For a shaft of symmetric cross section and a rotor concentrically situated at the midpoint of the shaft, each deformation is resisted solely by a corresponding proportional elastic force or moment; the elastic constants are $k_{\varepsilon}$, $k_{\eta}$, $k_{\beta}$, and $k_{\gamma}$.

The main assumptions in the development of the disc/diaphragm model are
delineated below.

1. The shaft is rigid (i.e., $k_\varepsilon = \infty$ and $k_\eta = \infty$), implying no translation or offset misalignment under steady state (i.e., $\varepsilon = 0$ and $\eta = 0$).

2. The initial angular misalignment is non-zero within the tolerance limits of 0.05 degrees (approx. 0.8 mills/inch) [55] for the steady-state rotational speed range of 300 to 1000 RPM.

3. Changes in stiffness due to angular misalignment have the same effects along both the rotation axes (i.e., $k_\beta = k_\gamma = K$).

4. The bearing stiffness, $k_b$, and bearing rotation damping coefficient, $C_b$, remain constant.

5. Changes in the statistical patterns of the accelerometer readings are solely due to structural degradation of the coupling.

6. Bearing life remains unaffected due to the misalignments under consideration.

The nonlinear equations of translational and rotational motion for the coupling dynamics are derived as follows [56]:

\begin{align*}
\ddot{\varepsilon} + 2\Omega \dot{\eta} + k_\varepsilon/(m - \Omega^2)\varepsilon &= g \cos \Omega t \quad (5.1) \\
\ddot{\eta} - 2\Omega \dot{\varepsilon} + k_\eta/(m - \Omega^2)\eta &= g \sin \Omega t \quad (5.2) \\
\ddot{\beta} - 2\Omega \dot{\gamma} + k_\beta/(I_2 - \Omega^2)\beta &= 0 \quad (5.3) \\
\ddot{\gamma} + 2\Omega \dot{\beta} + k_\gamma/(I_2 - \Omega^2)\gamma &= 0 \quad (5.4)
\end{align*}
where $I_2 = \frac{mr^2}{4}$ is the polar inertia of the disc/diaphragm. The governing equations for deflection of the disc/diaphragm are given below.

$$
\begin{pmatrix}
    x_c \\
    y_c
\end{pmatrix} =
\begin{pmatrix}
    \cos \Omega t & -\sin \Omega t \\
    \sin \Omega t & \cos \Omega t
\end{pmatrix}
\begin{pmatrix}
    \eta \\
    \varepsilon
\end{pmatrix}
$$

The force transmitted to the bearing due to translation and rotation (i.e., offset and angular misalignment) of the coupling are given below.

$$
F_b = m \begin{pmatrix}
    \ddot{x}_c \\
    \ddot{y}_c
\end{pmatrix} + \frac{1}{r_b} \begin{pmatrix}
    \cos \Omega t & -\sin \Omega t \\
    \sin \Omega t & \cos \Omega t
\end{pmatrix}
\begin{pmatrix}
    I_2 \ddot{\beta} \sin \beta \\
    I_2 \ddot{\gamma} \sin \gamma
\end{pmatrix}
$$

$$
T = J_z \dot{\Omega} + B \Omega
$$

where $J_z = \frac{mr^2}{2} + m(\eta^2 + \varepsilon^2)$ is the effective inertia of the diaphragm coupling.

$$
C_b \begin{pmatrix}
    \dot{x}_b \\
    \dot{y}_b
\end{pmatrix} + k_b \begin{pmatrix}
    x_b \\
    y_b
\end{pmatrix} = F_b
$$

The above set of equations is expressed in the state-space form as follows:

$$
\begin{align*}
S_1 &= \varepsilon, \quad S_2 = \dot{\varepsilon} \\
S_3 &= \eta, \quad S_4 = \dot{\eta} \\
S_5 &= \beta, \quad S_6 = \dot{\beta} \\
S_7 &= \gamma, \quad S_8 = \dot{\gamma} \\
S_9 &= \Omega, \quad S_{10} = x_b \\
S_{11} &= y_b
\end{align*}
$$

$$
\dot{S}_1 = S_2, \quad \dot{S}_3 = S_4
$$
\( \dot{S}_5 = S_6, \ \dot{S}_7 = S_8 \)

\( \dot{S}_2 = -2S_9S_4 - k_\varepsilon/(m - S_9^2)S_1 + g \cos S_9t \)

\( \dot{S}_4 = 2S_9S_2 - k_\eta/(m - S_9^2)S_3 + g \sin S_9t \)

\( \dot{S}_6 = 2S_9S_8 - k_\beta/(I_2 - S_9^2)S_5 \)

\( \dot{S}_8 = -2S_9S_6 - k_\gamma/(I_2 - S_9^2)S_7 \)

\( \dot{S}_9 = \frac{T - BS_9}{\frac{mr^2}{2} + m(S_3^2 + S_1^2)} \)

\[
\begin{bmatrix}
\dot{S}_{10} \\
\dot{S}_{11}
\end{bmatrix} = \frac{k_b}{C_b} \begin{bmatrix}
S_{10} \\
S_{11}
\end{bmatrix} + \frac{m}{C_b} \begin{bmatrix}
T_{11} & T_{12} \\
T_{21} & T_{22}
\end{bmatrix} \begin{bmatrix}
S_3 \\
S_1
\end{bmatrix}
\]

\[
+ \frac{2m}{C_b} \begin{bmatrix}
-(S_9 + \dot{S}_9t) \sin S_9t & -(S_9 + \dot{S}_9t) \cos S_9t \\
-(S_9 + \dot{S}_9t) \cos S_9t & -(S_9 + \dot{S}_9t) \sin S_9t
\end{bmatrix} \begin{bmatrix}
S_4 \\
S_2
\end{bmatrix}
\]

\[
+ \frac{m}{C_b} \begin{bmatrix}
\cos S_9t & -\sin S_9t \\
\sin S_9t & \cos S_9t
\end{bmatrix} \begin{bmatrix}
\dot{S}_4 \\
\dot{S}_2
\end{bmatrix} + \frac{I_2}{r_b} \begin{bmatrix}
\cos S_9t & -\sin S_9t \\
\sin S_9t & \cos S_9t
\end{bmatrix} \begin{bmatrix}
\dot{S}_6 \\
\dot{S}_8
\end{bmatrix}
\]

where

\[
T_{11} = -(S_9 + \dot{S}_9t)^2 \cos S_9t - 2S_9 \sin S_9t
\]

\[
T_{12} = -(S_9 + \dot{S}_9t)^2 \sin S_9t - 2S_9 \cos S_9t
\]

\[
T_{21} = -(S_9 + \dot{S}_9t)^2 \sin S_9t - 2S_9 \cos S_9t
\]
$$T_{22} = -(S_9 + ˙S_9 t)^2 \cos S_9 t - 2 ˙S_9 \sin S_9 t$$

The bearing accelerations were generated using the following equations:

\begin{align*}
\ddot{x}_b &= \ddot{S}_{10} \\
\ddot{y}_b &= \ddot{S}_{11}
\end{align*} 

(5.9) (5.10)

5.4 Simulation results

This section presents the solution procedures for both the forward and the inverse problems using the symbolic time series analysis (STSA) method as described in Chapter 3.

5.4.1 Solution to the forward (analysis) problem

The objective of the forward (analysis) problem is to generate a pattern set which is representative of damage evolution under different excitation levels. The requirement of the forward problem is to create time series data of system responses under both healthy and anomalous conditions and to define an appropriate anomaly measure which can identify subtle changes in data sets and is capable of recognizing small deviations from the nominal behavior. To solve the forward problem, time series data sets are generated for constant torque (or speed) levels under both nominal condition (i.e., when the stiffness parameter has the nominal value and it is not degraded due to structural damage) and anomalous conditions (i.e., when the stiffness of the coupling has degraded from the nominal value due to structural damage). One of the major features of the proposed STSA technique is that the
anomaly is quantified relative to the nominal condition, which is assigned zero anomaly measure. Furthermore, uncertainties due to small (zero-mean) torque fluctuations and exogenous disturbances are largely attenuated due to the filtering effects of wavelet-based symbolic dynamic analysis [35].

In this chapter, the bending stiffness parameters along both the x and y axes are assumed to be affected to an equal extent. Initially, the disc is allowed to rotate freely at an angular speed of 40 rad/s (∼380 RPM). The steady-state response of the accelerometer data is used for analysis when the angular speed reaches a constant value after ∼ 0.5 seconds. In each run, 8000 data points are generated at the sampling frequency of 1.0 KHz. For a single excitation level, system responses are measured under different values of the stiffness parameter that is reduced by 5 percent for each run until 75 percent of the nominal value is reached (see Figure 5.4), where the system is considered to be no longer usable and hence out of service.

Anomaly measures are then calculated using the procedure described in Chapter 3 with respect to the nominal condition. Similarly, time series data are generated under both nominal and anomalous conditions for multiple torque levels in the operating range of 3000-6500 N-m, separated at an interval of 100 N-m, thereby generating an ensemble of data sets for multiple excitations. Anomaly measures for different stiffness values are then calculated for each excitation level as mentioned above. In practice, accelerometers are required to be routinely calibrated before usage for data acquisition. Furthermore, variations in the parameters (e.g., frequency bandwidth) between successive maintenance actions are usually insignificant. Therefore, having known that there is no significant change
in the load torque, any abrupt changes in the anomaly measure can be attributed to large failure(s) (e.g., due to sensor malfunction). The effects of bias errors (i.e., low-frequency disturbances) and unbiased (i.e., zero-mean) random noise can be detected and compensated by the STSA technique. For example, the system dynamics are likely to change due to bias errors and therefore will call for calibration, which is done in the forward problem. Unbiased random noise is practically eliminated in the STSA algorithm [35].

This chapter has utilized the resultant $\ddot{z}_b$ of the accelerometer measurements in the x- and y- directions, i.e. $\ddot{x}_b$ and $\ddot{y}_b$ respectively, for each torque level for generating the anomaly measures using the STSA algorithm.

$$\ddot{z}_b = \sqrt{\ddot{x}_b^2 + \ddot{y}_b^2} \quad (5.11)$$

The alphabet size, window length, and the wavelet basis function were chosen as $A = 8$, $D = 1$ and $gaus2$ [33] respectively (see Section 3.2), for all time series data sets collected for different stiffness values and torque levels. Increasing the alphabet size further did not show any appreciable improvement in the results and increasing the value of depth D created a much larger number of states of the finite state machine, many of them having very small or zero probabilities. The finite state machine constructed with the above choice of parameters has 8 states; the algorithm is computationally fast and can be used on inexpensive processors for real-time execution. The $gaus2$ wavelet basis [33] provided better results than many other wavelets of the Daubechie’s family. The reason for this choice is attributed to close matching of the shapes of the signal and the wavelet basis function.

Figure 5.4 exhibits the profiles of anomaly measures for varying stiffness under
different excitations, separated by 100 N-m in the range of 3000 to 6500 N-m. This ensemble of the anomaly measure plots as seen in Figure 5.4 forms a pattern set for anomaly detection under different excitations in the operating range of 3000 to 6500 N-m. The ensemble consists of $\ell = 42$ plots derived from the simulation data generated in the operating range. This pattern set has been used to solve the inverse problem for estimating the stiffness value at any particular instant and torque level in the operating range, as presented in the next section.
Figure 5.4. Family of anomaly measure plots for different levels of constant torque inputs in the range of 3000-6500 N-m vs diaphragm stiffness $K$. 
5.4.2 Solution to the inverse (*synthesis*) problem

The objective of the inverse problem is identification of anomalies and estimation of the fault parameters based on the pattern set of anomaly measures generated by the forward problem. It is essential to detect changes in the stiffness of the coupling, due to evolving fatigue damage, so that appropriate remedial action(s) can be taken before the onset of widespread damage. It is also necessary to relate changes in stiffness to the magnitude and location of incipient damage. The rationale is that loss of stiffness of the coupling may lead to further misalignment, excessive noise and vibrations causing rapid failure of other rotor components. Hence, the estimated change in stiffness during the operating period is crucial to allow for scheduled operation and maintenance. The following procedure for stiffness estimation is based on the statistical pattern changes in the observed time series data.

Statistical techniques have been utilized for structural health monitoring and failure prognosis of mechanical systems in the past few decades [14]. For example, a commonly used tool is principal component analysis that relies on the estimation of eigenvalues and eigenvectors [15] [16] of the dynamical system. These parameters are dependant on the mass, stiffness and damping properties and as well as on the structural flaws in the system. They provide critical information about the characteristics and behavior of dynamical systems. In the approach described in this section, the objective is to estimate the critical parameter(s) of the system, based on the observed time series data response. Statistical analysis of the pattern set of the anomaly measure profiles (see procedure outlined in Section 5.4.1) provides the estimate of the coupling stiffness.
This section describes the procedure for parameter (i.e., stiffness) estimation of flexible mechanical couplings subjected to exogenous or self excitation in the operating range. For solution of the inverse problem (i.e., estimation of stiffness at any instant under arbitrary load), time series data is generated under both nominal condition and arbitrarily chosen anomalous conditions (e.g., between 100% and 75% of the nominal value of stiffness). Input excitation can be arbitrarily chosen in the range of 3000-6500 N-m. Anomaly measure is then calculated with respect to the nominal condition based on the observed time series data as described in Chapter 3. It is considered here that the exact determination of the stiffness parameter based on the derived value of anomaly measure at any particular instant under arbitrary load is not possible due to the spread in its value as observed in the forward problem. Therefore, the stiffness parameter is treated as a random variable because of the uncertainty in determining its exact value. It is assumed here that the only known quantity is the anomaly measure value which is derived from the observed time series data.

The range of anomaly measure $M$ (i.e. the ordinate in Figure 5.4) is discretized into $n$ uniformly spaced levels. A pattern matrix $P$ of dimension $\ell \times n$ is then derived from the anomaly measure plots as shown in Figure 5.4. Each column of the pattern matrix $P$ corresponds to the stiffness values calculated for $\ell$ torque inputs at a particular anomaly measure. The elements belonging to a particular column of the pattern matrix $P$ are distributed over a certain range of stiffness values. For the elements of each column, a two-parameter lognormal distribution [48] [14] [59] is hypothesized, and its goodness of fit is examined by both $\chi^2$ and Kolmogorov–Smirnov tests [60]. Details are provided in Appendix A. Figure 5.5
Figure 5.5. Plots of actual probability distribution of diaphragm stiffness $K$ and corresponding Log-normal fit for four levels of anomaly measure ($M$): a) 0.0445 b) 0.0954 c) 0.1591 d) 0.2227

compares the lognormal-distributed probability density functions (pdfs) of the diaphragm stiffness $K$ with the corresponding histograms generated at four different anomaly measure values in the range. The number of bins were taken to be $r = 8$ for each data set. With $m = r - 1$ degrees of freedom, the $\chi^2$-test shows that, for each of the $n$ data sets, the hypothesis of two-parameter lognormal-distribution passed the 90% confidence level (C.L.) [60] which suffices the conventional standard of 95% confidence level. Also, for each of the $n$ data sets, the hypothesis passed
the 70% confidence level of the Kolmogorov–Smirnov test which again suffices the conventional standard of 95% confidence level.

5.4.2.1 Confidence interval

Once the fits of lognormal distribution are obtained, the bounds for different confidence intervals are computed [60] [61]. Confidence intervals (C.I.) are obtained at a certain confidence level such as 95%. This implies that at a particular value of the (real-time) anomaly measure, the confidence interval is a measure of the spread of the stiffness with the specified level of confidence. As such, for a confidence level of 95%, the probability that the actual stiffness will lie between the obtained confidence level is 95%. Figure 5.6 shows the bounds for 99%, 95% and 90% confidence levels.

For validation of the proposed methodology, time series data have been generated at three different arbitrary torque levels in the range of 3000-6500 N-m for the nominal value of the stiffness parameter. Also to compute the anomaly measure using STSA (see Chapter 3), time series data is generated at three randomly selected stiffness values between 75% to 100% of the nominal value, for all three torque levels, as seen in Table 5.1. The chosen torque levels in the inverse problem are deliberately made to be different from those used in the forward problem (see Section 5.4.1) but each of them lies within the operating range of 3000-6500 N-m. For the generated anomaly measure values, the mean diaphragm stiffness \( \hat{K} \), its standard deviation \( \hat{\sigma} \) and confidence intervals are estimated at three confidence levels of 99%, 95% using the inverse problem solution (see Section 5.4.2). Table 5.1 provides the actual stiffness values used for generation of anomaly measures, and
the estimated parameters obtained using the inverse problem solution. As seen in Table 5.1, the confidence intervals become tighter as the confidence level reduces.
Anomaly Measure
Estimated mean
99% Confidence interval
95% Confidence interval
90% Confidence interval

Figure 5.6. Plots of confidence intervals of the Log-normal distribution at 99%, 95% and 90% confidence levels. Bounds of confidence level 99% and mean are shown at $M=0.111$

Table 5.1. Estimated values of diaphragm stiffness ($K$) for two different confidence levels for three different Torque Loads

<table>
<thead>
<tr>
<th>Torque N-m</th>
<th>$M$</th>
<th>% K</th>
<th>$K$</th>
<th>$\sigma$</th>
<th>99% C.I. lower</th>
<th>99% C.I. upper</th>
<th>95% C.I. lower</th>
<th>95% C.I. upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>3800</td>
<td>0.0599</td>
<td>93</td>
<td>93.64</td>
<td>0.539</td>
<td>92.1</td>
<td>94.89</td>
<td>92.5</td>
<td>94.62</td>
</tr>
<tr>
<td></td>
<td>0.1167</td>
<td>87</td>
<td>88.33</td>
<td>0.736</td>
<td>86.29</td>
<td>90.08</td>
<td>86.8</td>
<td>89.69</td>
</tr>
<tr>
<td></td>
<td>0.1863</td>
<td>78</td>
<td>81.50</td>
<td>1.363</td>
<td>77.78</td>
<td>84.77</td>
<td>78.73</td>
<td>84.07</td>
</tr>
<tr>
<td>4800</td>
<td>0.0328</td>
<td>96</td>
<td>96.30</td>
<td>0.454</td>
<td>94.95</td>
<td>97.29</td>
<td>95.31</td>
<td>97.09</td>
</tr>
<tr>
<td></td>
<td>0.111</td>
<td>88</td>
<td>88.87</td>
<td>0.700</td>
<td>86.92</td>
<td>90.52</td>
<td>87.41</td>
<td>90.16</td>
</tr>
<tr>
<td></td>
<td>0.2038</td>
<td>79</td>
<td>79.88</td>
<td>1.498</td>
<td>75.65</td>
<td>83.37</td>
<td>76.73</td>
<td>82.59</td>
</tr>
<tr>
<td>5800</td>
<td>0.0181</td>
<td>98</td>
<td>97.96</td>
<td>0.251</td>
<td>97.21</td>
<td>98.51</td>
<td>97.41</td>
<td>98.39</td>
</tr>
<tr>
<td></td>
<td>0.1013</td>
<td>90</td>
<td>89.78</td>
<td>0.645</td>
<td>87.98</td>
<td>91.3</td>
<td>88.44</td>
<td>90.96</td>
</tr>
<tr>
<td></td>
<td>0.1882</td>
<td>82</td>
<td>81.41</td>
<td>1.380</td>
<td>77.53</td>
<td>84.62</td>
<td>78.52</td>
<td>83.91</td>
</tr>
</tbody>
</table>
5.4.2.2 Confidence interval for the mean

Figure 5.7 shows the bounds for 99% and 95% confidence levels for the mean. Confidence intervals (C.I.) for the mean are obtained at a certain confidence level such as 95%. This implies that if the population is sampled at multiple occasions and the interval estimates are made on each occasion, then the true parameter (i.e., the mean of the stiffness parameter in this case) would lie in the resulting intervals in approximately 95% of the cases. Figure 5.7 is derived from the lognormal distribution for the entire range of the anomaly measure. Confidence intervals for the mean of the stiffness parameter are plotted for two different levels of 99% and 95%. For any anomaly measure value, if \( \ell \) samples are collected and if the confidence intervals are computed several times, then these intervals would contain the true mean for 99% and 95% of all cases for the two intervals, respectively, in the long run. Thus, the confidence interval depicts prediction accuracy of the true mean for a collection of runs for an observed anomaly measure in the range. This information is generated online. The table also shows the confidence intervals for the estimated mean stiffness for the three levels. As seen in Table 5.2, the bounds become tighter as the confidence level reduces.
Figure 5.7. Plots of confidence intervals of the mean at 99%, 95% and 90% confidence levels. Bounds of confidence level 99% and mean are shown at \( M = 0.111 \)

Table 5.2. Accuracy of the estimated diaphragm stiffness (\( K \))

<table>
<thead>
<tr>
<th>Observed Anomaly Measure (( \mathcal{M} ))</th>
<th>( K )</th>
<th>99% C.I.</th>
<th>95% C.I.</th>
<th>90% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(( K ))_{min}</td>
<td>(( K ))_{max}</td>
<td>(( K ))_{min}</td>
<td>(( K ))_{max}</td>
</tr>
<tr>
<td>0.0599</td>
<td>93.64</td>
<td>93.35</td>
<td>93.93</td>
<td>93.42</td>
</tr>
<tr>
<td>0.1167</td>
<td>88.33</td>
<td>87.93</td>
<td>88.72</td>
<td>88.04</td>
</tr>
<tr>
<td>0.1863</td>
<td>81.59</td>
<td>80.85</td>
<td>82.31</td>
<td>81.04</td>
</tr>
<tr>
<td>0.0328</td>
<td>96.30</td>
<td>96.06</td>
<td>96.54</td>
<td>96.12</td>
</tr>
<tr>
<td>0.111</td>
<td>88.87</td>
<td>88.49</td>
<td>89.24</td>
<td>88.58</td>
</tr>
<tr>
<td>0.2038</td>
<td>79.88</td>
<td>79.06</td>
<td>80.66</td>
<td>79.27</td>
</tr>
<tr>
<td>0.0181</td>
<td>97.96</td>
<td>97.82</td>
<td>98.08</td>
<td>97.86</td>
</tr>
<tr>
<td>0.1013</td>
<td>89.78</td>
<td>89.42</td>
<td>90.12</td>
<td>89.52</td>
</tr>
<tr>
<td>0.1882</td>
<td>81.41</td>
<td>80.66</td>
<td>82.13</td>
<td>80.85</td>
</tr>
</tbody>
</table>
Summary, Conclusions and Future Work

This chapter summarizes the contributions of the research work and makes recommendations for future work.

6.1 Summary and Conclusions

The research experimentally validates a novel approach to address the anomaly detection problem in electromechanical systems using symbolic time series analysis (STSA). It is assumed that the dynamical system is stationary at the fast time scale and that any non-stationary behavior is observable only on the slow time scale. Injection of self excited stimuli is proposed to detect small changes and anomalies at an early stage by taking advantage of symbolic dynamics. The novelty of the research and potential new contributions are summarized below:

- Experimental validation for detection of evolving anomalies (i.e., the for-
forward problem) in real-time for mechanical systems proposed by Ray [13] and compares with existing pattern recognition techniques. To this end, a laboratory test facility (see Figure 1.1) has been designed and fabricated to detect fatigue damage well ahead of reaching a critical condition.

- Also, another laboratory test facility (see Figure 1.2) has been fabricated to detect changes in mass/moment of inertia of the system components derived from time series data of pertinent measured variable.

- To deal with the inverse problem of inferring the system parameters based on the observed asymptotic behavior. Usage of the inverse solution in a simulation model of a disc/diaphragm mechanical coupling subjected to angular misalignment.

- Demonstration of online life extending control based on anomaly measure. The goal of life extending control is to detect the progression of malignant anomalies at an early stage and enable the system to take appropriate remedial action to circumvent damage. Also, control action should be delayed as far as possible so as to enable the system to perform at the desired level for maximum length of time.

The following conclusions can be drawn from the analysis of experimental and simulation data:

- The results of STSA on the laboratory apparatus (see Figure 2.1) show that the evolution of fatigue damage is detected in advance of component failure. This is of paramount importance to health and usage monitoring
of machinery operations as it provides ample time to take remedial control actions for life extension [47] and reliability enhancement.

- The results on the laboratory apparatus (see Figure 2.10) show that STSA is very well able to capture the textural changes in the sensor time series due to change in mass/moment of inertia of the system components.

- Efficacy of the STSA method for identification of behavior patterns is demonstrated on a simulation model of mechanical structures, where the source of possible anomalies is the change in stiffness of the disc/diaphragm coupling due to angular misalignment. The damage information is extracted from accelerometer sensor signals mounted on bearings. The histograms on state probability distribution are generated from the observed time series data to serve as patterns of the evolving behavior change. Statistical analysis of the pattern set of the anomaly measure profiles provides the estimate of the system parameters.

### 6.2 Recommendations for future work

A major goal in the control of dynamical systems such as advanced aircraft and spacecraft systems is to achieve high performance with enhanced reliability, component durability, and maintainability. The concept of damage mitigating control, also referred to as life extending control, is built upon the hypothesis that substantial improvement in the service life of critical plant components can be achieved by insignificant loss of system dynamic performance. Ray, Zhang and Phoha [47] have proposed a hierarchically structured hybrid (i.e., combined continuously vary-
ing and discrete event) decision control system which integrates damage-mitigating control with the planning and scheduling of plant maintenance and operation. The major challenge here is to characterize the damage generation process and then utilize this information for synthesizing algorithms of robust control, diagnostics and risk prediction in complex mechanical systems.

The concept of hybrid damage-mitigating decision and control of mechanical structures follows a two-tier architecture as shown in Figure 6.1.

The author of this dissertation makes the following recommendations for future work:

- Multi-variate analysis to provide fusion of information from heterogenous sensors located at different spatial locations.

- Investigate failure or evolution of damage in multiple failure sites.

- Design a hierarchical discrete event controller as shown in Figure 6.1 to achieve substantial improvements in service life without significant loss of performance.

- Study of nonlinear dynamics using multiple sources of excitation in test apparatus 2.

- Study fatigue failure under multi-modal (axial and bending) loading with different excitations (reference trajectories).

- Facilitate life extending control [17] on test apparatus 2 without significant loss of performance.
Figure 6.1. Hierarchical architecture for life extension and performance control

- Fabrication of a disc/diaphragm coupling apparatus and installation of accelerometers at the site(s) of interest.
Appendix A

Appendix

A.1 Existing pattern recognition techniques

This section briefly describes the three pattern recognition techniques which use time series data as inputs, for comparison with the symbolic-time-series-based anomaly detection method:

- Principal Component Analysis (PCA)
- Multilayer Perceptron Neural Network (MLPNN)
- Radial Basis Function Neural Network (RBFNN)

While details of these methods are provided in a previous publication [44], they are briefly described in this chapter for the sake of completeness.
A.1.1 Principal Component Analysis (PCA) for anomaly detection

Feature extraction methods in statistical pattern recognition determine an appropriate subspace of dimensionality \( q \in \mathbb{N} \), where \( \mathbb{N} \) is the set of positive integers, using either linear or nonlinear methods in the original feature space of dimensionality \( n \) \((q \leq n)\). The best known linear feature extractor relies on the Principal Component Analysis (PCA) \([62]\). The eigenvectors of the \((n \times n)\) (positive semi-definite) covariance matrix of the time series data, corresponding to the \( q \) largest eigenvalues, form the \( n \)-dimensional patterns. The linear transformation is defined as

\[
Y = HX
\]  

(A.1)

where \( X \) is the given \((n \times d)\) pattern matrix, made of \( n \) row vectors; \( H \) is the \( q \times n \) linear transformation matrix whose rows represent \( q \) feature vectors of dimension \( n \); and \( Y \) is the derived \( d \times q \) pattern matrix. Since the PCA method uses the most expressive features (e.g., eigenvectors with the largest eigenvalues), it effectively approximates the data by a linear subspace using the mean squared error criterion.

To detect growth in anomaly from time series data, principal component analysis is performed for dimensionality reduction. If the time response of an appropriate process variable \( y(t) \) is sampled to generate a time series sequence \( y_k \), then data samples of large enough length \((\ell = d \ n)\) can be used to capture the dynamical characteristics of the observed process. The length \( \ell \) of time series data is partitioned into \( d \) subsections, each being of length \( n = \frac{\ell}{d} \), where \( d > n \). The resulting \((d \times n)\) data matrix is processed to generate the \((n \times n)\) covariance matrix that is
positive-definite or positive-semidefinite real-symmetric. The next step is to compute the orthonormal eigenvectors $v^1, v^2, \ldots, v^n$ and the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ that are arranged in decreasing orders of magnitude. The eigenvectors associated with the first (i.e., largest) $q$ eigenvalues are chosen as the feature vectors such that

$$\sum_{i=q+1}^{n} \lambda_i < \eta$$

(A.2)

where the threshold $\eta \ll 1$ is a positive real close to 0. The resulting pattern is the matrix, consisting of the feature vectors as columns,

$$\tilde{M} = \left( \sqrt{\lambda_1} v_1 \ldots \sqrt{\lambda_q} v_q \right)$$

(A.3)

The above steps are executed for time series data under the nominal (stationary) condition to obtain $\tilde{M}_0$. Then, these steps are repeated at subsequent slow-time epochs, $\{t_1, t_2, \ldots\}$, as the (possible) anomaly progresses using the same values of parameters, $\ell, d, n$ and $q$, used under the nominal condition, to obtain the respective pattern matrices $\tilde{M}_1, \tilde{M}_2, \ldots$. The anomaly measures at slow-time epochs $\{t_1, t_2, \ldots\}$ are obtained as

$$\hat{\mathcal{M}}_k \equiv d \left( \tilde{M}_k, \tilde{M}_0 \right)$$

where the $d(\bullet, \bullet)$ is an appropriately defined path dependant distance function as given by Eq. (3.8).

It should be noted that different metrics may be used as anomaly measures as stated in [13]. One may choose the metric that yields the most satisfactory result.
for the specified purpose. Along this line, different metrics could be chosen for other pattern recognition techniques.

A.1.2 Multi-Layer Perceptron Neural Network (MLPNN) for anomaly detection

The multilayer perceptron neural network (MLPNN) is the most commonly used family of feed-forward neural networks for pattern classification tasks [62]. The MLPNN is a collection of connected processing elements, called nodes or neurons [63] [64]. Its structure is fixed by choosing the number of layers as well as the (possibly different) number of neurons in each layer. The MLPNN is trained based on the information contained in a given set of inputs and target outputs. The training phase includes modelling of the input-output system architecture and identification of the synapsis weights. A set of inputs is passed forward through the network yielding trial outputs which are then compared to the target outputs to obtain the error (i.e., the deviation of the trial output from the target output). The network parameters (i.e., synapsis weights and biases) are adjusted until the error is within specified limits. If the specified bound is exceeded, the error is passed backwards through the net and the training algorithm adjusts the synapsis weights.

The back-propagation algorithm has been used in this dissertation. The simplest implementation of back-propagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly. The mean square error criterion is adopted in the recursive algorithm to update the weight vectors \( \{w_k\} \) as follows:
where \( g_n \) is the gradient and \( \alpha_n \) is the learning rate.

Different layers in MLP neural networks may contain different numbers of neurons. Time series signals enter into the input layer nodes, progress forward through the hidden layers, and finally emerge from the output layer. Each node \( i \) at a given layer \( k \) receives a signal from all nodes \( j \) in its preceding layer \((k-1)\) through a synopsis of weight \( w_{ij}^k \) and the process is carried onto the nodes in the following layer \((k+1)\). The weighted sum of signals \( x_{j}^{k-1} \) from all nodes \( j \) of the layer \((k-1)\) together with a bias \( w_{i0}^k \) produces the excitation \( z_i^k \) that, in turn, is passed through a nonlinear activation function \( f \) to generate the output \( x_i^k \) from the node \( i \) at the layer \( k \). This is mathematically expressed as

\[
\begin{align*}
    z_i^k &= \sum_j w_{ij}^k x_j^{k-1} + w_{i0}^k \\
    x_i^k &= f(z_i^k)
\end{align*}
\]  

Various choices for the activation function \( f \) are possible; the hyperbolic tangent function \( f(x) = \text{tanh}(x) \) has been adopted in this chapter.

For anomaly detection, the MLPNN is trained by setting a set of \( N \) input vectors, each of dimension \( \ell \), and a specified target output vector \( \tau \) of dimension \( q \). This implies that the input layer has \( \ell \) neurons and the output layer has \( q \) neurons. If the time series data are obtained from an ergodic process, then a data set of length \( N \ell \) can be segmented into \( N \) vectors of length \( \ell \) to construct the input and target pattern matrices, \( \mathcal{P} \). The input pattern matrix \( \mathcal{P} \in \mathbb{R}^{\ell \times N} \) is obtained from
the $N$ input vectors as

$$\mathcal{P} \equiv [p^1 \ p^2 \ \cdots \ p^N] \quad (A.7)$$

where $p^k \equiv [y_{(k-1)\ell+1} \ y_{(k-1)\ell+2} \ \cdots \ y_{k\ell}]^T$ and each $y_k$ is a sample from the ensemble of the time series data. The corresponding output matrix $\mathcal{O}$ is the output of the trained MLPNN under the input pattern $\mathcal{P}$.

$$\mathcal{O} \equiv [o^1 \ o^2 \ \cdots \ o^N] \quad (A.8)$$

where $o^i \in \mathbb{R}^q$ is the output of the trained MLPNN under the input $p^k \in \mathbb{R}^\ell$. The performance vector $u \in \mathbb{R}^q$ is obtained as the average of the $N$ outputs.

$$u \equiv \frac{1}{N} \sum_{k=1}^{N} o^k \quad (A.9)$$

The time series data under the nominal condition generates the input pattern matrix $\mathcal{P}_0$ that, in turn, is used to train the MLPNN with respect to a target output vector $\tau$. The resulting output of the trained MLPNN with $\mathcal{P}_0$ as the input is $\mathcal{O}_0$ and the performance vector is $u_0$. Subsequently, input pattern matrices $\{\mathcal{P}_1, \mathcal{P}_2, \ldots\}$ are obtained at slow-time epochs $\{t_1, t_2, \ldots\}$ and corresponding output matrices of the trained MLPNN are $\{\mathcal{O}_1, \mathcal{O}_2, \ldots\}$, which yield the respective performance vectors $\{u_1, u_2, \ldots\}$. The anomaly measures at slow-time epochs $\{t_1, t_2, \ldots\}$ are obtained as

$$\hat{M}_k \equiv d(u_k, u_0)$$

where the $d(\bullet, \bullet)$ is an appropriately defined path dependant distance function as given by Eq. (3.8).
A.1.3 Radial Basis Function Neural Network (RBFNN) for anomaly detection

The radial basis function neural network (RBFNN) is a commonly used tool for pattern identification [63], where the activation of a hidden unit is determined by the distance between the input vector and the prototype vector; the RBFNN is essentially a nearest neighbor type of classifier. A radial basis function has the following structure:

\[ f(y, \alpha) = \exp \left( -\sum_{k} \frac{|y_k - \mu|^{\alpha}}{N \theta_{\alpha}} \right) \]  

(A.10)

where the exponent parameter \( \alpha \in (0, \infty) \); and \( \mu \) and \( \theta_{\alpha} \) are the center and \( \alpha^{th} \) central moment of the data set, respectively. For \( \alpha = 2 \), \( f(\bullet) \) becomes Gaussian, which is the typical radial basis function used in the neural network literature. To perform anomaly detection, the first task is to obtain the sampled time series data when the dynamical system is in the nominal condition and then the mean \( \mu \) and the central moment \( \theta_{\alpha} \) are calculated as

\[ \mu = \frac{1}{N} \sum_{k=1}^{N} y_k; \quad \text{and} \quad \theta_{\alpha} = \frac{1}{N} \sum_{k=1}^{N} |y_k - \mu|^{\alpha} \]  

(A.11)

The distance between any vector \( y \) and the center \( \mu \) is obtained as \( d(y, \mu) \equiv (\sum_{n} |y(n) - \mu|^{\alpha})^{\frac{1}{\alpha}} \). Following Eq. (A.10), the radial basis function at the nominal condition is: \( f_0 = f(y) \). Under all conditions including anomalous ones, the parameters \( \mu \) and \( \theta \) are kept fixed. However, at slow time epochs \( \{t_1, t_2, \ldots\} \), the radial basis functions \( \{f_1, f_2, \ldots\} \) are evaluated from the data sets under the (possibly anomalous) conditions. The anomaly measure at the epoch \( t_k \) in the slow time
scale is obtained as a distance from the nominal condition and is given by

\[ \hat{m}_k = d(f_0, f_k) \]

where the \( d(\bullet, \bullet) \) is an appropriately defined path dependant distance function as given by Eq. (3.8).

## A.2 Electronics and instrumentation of the test apparatus

To extract sufficient information from the test apparatus, a variety of sensors have been installed. The sensing system includes: two LVDT’s to measure the position of two of the three moving masses; two load cells for actual force measurement.

The technical specifications of the sensing devices are listed below.

<table>
<thead>
<tr>
<th>Table A.1. LVDT specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vendor</strong></td>
</tr>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td><strong>Range, working</strong></td>
</tr>
<tr>
<td><strong>Max. usable</strong></td>
</tr>
<tr>
<td><strong>Input volts DC</strong></td>
</tr>
<tr>
<td><strong>Frequency Response</strong></td>
</tr>
<tr>
<td><strong>Wiring Code</strong></td>
</tr>
</tbody>
</table>
Table A.2. Load cell specifications

<table>
<thead>
<tr>
<th>Vendor</th>
<th>SENSOTEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>34</td>
</tr>
<tr>
<td>Load Range</td>
<td>100 lbs</td>
</tr>
<tr>
<td>Output</td>
<td>2 mv/V</td>
</tr>
<tr>
<td>Excitation</td>
<td>10 V DC</td>
</tr>
<tr>
<td>Bridge Resistance</td>
<td>350 ohm (foil)</td>
</tr>
<tr>
<td>Overload</td>
<td>50 percent</td>
</tr>
</tbody>
</table>
| Wiring Code | Red +Excitation  
               Black -Excitation  
               Green -Output  
               Blue +Output |

Table A.3. Inline amplifier for load cell

<table>
<thead>
<tr>
<th>Vendor</th>
<th>SENSOTEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Voltage</td>
<td>+/- 15 V DC</td>
</tr>
<tr>
<td>Excitation voltage</td>
<td>10 V DC</td>
</tr>
<tr>
<td>Output Voltage Range</td>
<td>+/- 5 V DC</td>
</tr>
<tr>
<td>Span Adjustment Range</td>
<td>0.5 to 10 mV/V</td>
</tr>
<tr>
<td>Frequency Response</td>
<td>DC 5 kHz</td>
</tr>
</tbody>
</table>

Table A.4. Power amplifier specifications

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Ashley-Audio</th>
</tr>
</thead>
</table>
| Power Output (per channel, both channels driven at 1 kHz, less than 0.1 percent THD) | 8 ohms : 300 W RMS  
4 ohms : 500 W RMS  
2 ohms : 675 W RMS |
| Frequency Response | +/- 0.2 dB 20 Hz - 20 kHz  
 +/- 3 dB 8 Hz- 100 kHz |
| Amplifier Protection | Fuseless short circuit protection, DC, ultrasonic and RF protection |
### Table A.5. Shaker specifications

<table>
<thead>
<tr>
<th>Vendor</th>
<th>VTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rated Force</td>
<td>100 lbf</td>
</tr>
<tr>
<td>Dynamic Stroke</td>
<td>0.75 inch</td>
</tr>
<tr>
<td>Maximum Acceleration</td>
<td>150g</td>
</tr>
<tr>
<td>Suspension Stiffness</td>
<td>40 lb/inch</td>
</tr>
<tr>
<td>Useful Frequency Range</td>
<td>DC-6500 Hz</td>
</tr>
<tr>
<td>Nominal Impedance</td>
<td>6 ohms</td>
</tr>
</tbody>
</table>

### Table A.6. Scaling factors for calibrated sensors

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Scaling factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDT 1</td>
<td>1.135 VDC/in/V</td>
</tr>
<tr>
<td>LVDT 2</td>
<td>1.135 VDC/in/V</td>
</tr>
<tr>
<td>Load Cell 1 SN 592692</td>
<td>After Cal. 5V/20lb (Gain = 0.5 mv/V)</td>
</tr>
<tr>
<td>Load Cell 2 SN 592691</td>
<td>After Cal. 5V/20lb (Gain = 0.5 mv/V)</td>
</tr>
</tbody>
</table>
Bibliography


Vita

Amol M Khatkhate

Amol Khatkhate received a B.E. in Mechanical engineering from Veer Mata Jijabai Technological Institute (VJTI), Mumbai, India in the year 2001. He currently has two M.S degrees, one in mechanical engineering and the other in electrical engineering, both from Penn State University. His general research interests include: control theory, signal processing, and pattern recognition. His specific areas of interest include diagnostics, prognostics, structural health monitoring and various NDE techniques as applied to aerospace/aircraft structures. He is currently working as a graduate research assistant at the Electromechanical Systems Laboratory (EMSL), PennState.