Application of Support Vector Machines for Damage Detection in Structures

By

Siddharth Sharma

A thesis submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

In

Mechanical Engineering

December 2008

Approved:

_____
Professor Zhikun Hou (Thesis Advisor)

_____
Professor Mikhail Dimentberg (Committee Member)

_____
Professor John Hall (Committee Member)

_____
Professor John Sullivan (Graduate Committee Representative)

Abstract

Support vector machines (SVMs) are a set of supervised learning methods that have recently been applied for structural damage detection due to their ability to form an accurate boundary from a small amount of training data. During training, they require data from the undamaged and damaged structure. The unavailability of data from the damaged structure is a major challenge in such methods due to the irreversibility of damage. Recent methods create data for the damaged structure from finite element models. In this thesis we propose a new method to derive the dataset representing the damage structure from the dataset measured on the undamaged structure without using a detailed structural finite element model. The basic idea is to reduce the values of a copy of the data from the undamaged structure to create the data representing the damaged structure. The performance of the method in the presence of measurement noise, ambient base excitation, wind loading is investigated. We find that SVMs can be used to detect small amounts of damage in the structure in the presence of noise. The ability of the method to detect damage at different locations in a structure and the effect of measurement location on the sensitivity of the method has been investigated. An online structural health monitoring method has also been proposed to use the SVM boundary, trained on data measured from the damaged structure, as an indicator of the structural health condition.

Acknowledgements

I would like to thank Prof Zhikun Hou for his guidance and encouragement through the course of this research. His critical analysis and deep understanding of the underlying issues in the field made me reach the goals that we had set. Thank you for always being able to find the time to answer all my questions during the long discussions we had at various stages of this research.

I thank Prof Dimentberg, Prof Hall and Prof Sullivan for agreeing to be on my thesis committee.

I am grateful to Prof Sullivan for his words of constant encouragement and for always having an open door to any questions that I had.

I am thankful to Prof Hall for all the deeply inspiring conversations during our work together. It provided me with a different perspective of looking at engineering issues.

I am grateful to the ladies in the ME Department, Barbara Edilberti, Barbara Furhman and Pam Kay Loius for their help and kindness.

I am indebted to my family for the constant encouragement and support without which I would not have been able to complete this project.

I would also like to thank my friends at WPI for being very understanding and making my stay at WPI a very enjoyable and memorable one.

Table of Contents

List of Figures

1 Introduction

Millions of dollars are spent every year in the maintenance of large structures like buildings, bridges etc. As these structures become older and new structures are built, their maintenance and health monitoring becomes more and more critical to ensure the safety and performance for their users. Damage detection is an important task for Structural Health Monitoring (SHM) which can lead to considerable benefits by indicating which structures require retrofitting and which do not. Damage detection has been an active area of research for decades and a large number of techniques and methods have been developed. A brief overview of the more common techniques is given below.

1.1 Classification of SHM techniques

SHM involves detecting the presence of damage in a structure. To be able to effectively repair a damaged structure, the exact location of the damage in the structure needs to be identified. The knowledge of the nature (rusting, breaking, etc.) and extent of damage provides a better estimate of the usability of the structure and can help predict the remaining life of the structure. Information about the extent of damage can help derive the effectiveness of the repair measures once they are applied. With these ideas, structural damage detection tasks were broadly classified into the following levels by Rytter (1993):

1) Level 1: Determination of the presence of damage
2) Level 2: Determination of the geometric location of damage

3) Level 3: Determination of the extent of damage

4) Level 4: Prediction of the remaining life of the structure

To be able to monitor the health of a structure some structural property is used to indicate damage in it. Damage in the structure causes a change in its physical properties (mass, stiffness, damping). A change in these physical properties in turn changes the dynamic characteristics (modal characteristics) or response of the structure. SHM methods that do not use the dynamic characteristics of the structure like the visual test, tap tests, acoustic or ultrasonic methods, magnetic field methods, radiographs, eddy-current methods and thermal field methods require an approximate location of the damage to be already known. Hence they are known as “local” damage detection methods as opposed to dynamic (vibration) response based methods which are “global” methods. Global methods are indicators of the overall health of a system and have been investigated in great detail for SHM. The amount of change in modal properties (natural frequencies, mode shape and the modal damping) can be used to detect, locate and determine the extent of damage.

The basic principle of vibration based methods is that a change in the physical properties modifies the modal properties (frequencies, mode shapes and modal damping) of the structure. This change can be detected by changes in the vibration measurements from the structure. A limitation of global methods is that while they indicate the global health of the structure, they are not sensitive to damage in redundant or structurally less important members. Existing methods try to extract the change in natural frequencies and mode shapes of the structure. This change pattern of

natural frequencies or mode shapes is then used to estimate the extent and location of damage. As these methods extract modal parameters from the vibration signals, the measurement locations have to be decided based on the mode shapes of the structure to capture required modes.

Local structural health monitoring methods can be used to locate the damage and provide an estimate of the extent of damage. However, there are a number of challenges while using local SHM methods. They are typically very expensive, time consuming and access to all locations might not be feasible to apply these methods. As mentioned earlier, some prior knowledge of damaged locations or the critical sections of the structure is required for these methods to perform effectively. The best solution for applying SHM methods is using a combination of both global and local structural health monitoring methods.

In the last decade, a lot of methods have been developed to evaluate damage in the structure based on either a change in the modal properties or from the physical properties estimated using the measured signal. The former set of methods depends on evaluating the change in the natural frequencies and mode shapes extracted from the vibration signal. The latter method depends on estimating the mass, stiffness and damping matrices derived from the structural vibration data. A change in the matrices provides an estimate of the presence, location and extent of damage.

The use of each parameter and methods has its own advantages and disadvantages (Doebling, 1996a). An overview of health monitoring methods of civil infrastructure has been provided in Chang et al. (2003). Detection of damage and its location can be done

based on change in structural parameters, estimation of the extent of damage requires a model of the structure (Doebling, 1996a). Recently, damage detection has been seen as a pattern recognition problem where a mapping is derived from some parameter affected by a change in the physical properties of a system to the health condition of a structure. Pattern recognition methods like Neural Networks and Support Vector Machines require data from the undamaged and the damaged structure to be able to successfully train and then later be able to classify the structure into damaged and undamaged classes. The main challenge with these methods is that if the structure is considered undamaged in its current state, data from the damaged class is not available unless some detailed structural models such as a finite element model can be used to generate data for the damaged class.

Support Vector Machines (SVM) have gained significance recently due to their superior ability of generate an accurate mapping between the input and output from a small amount of training data. This ability to create an accurate mapping between the input and output classes from a small amount of data is called the generalization ability of the method. SVM have better generalization performance than other existing methods for problems of damage detection and location (Samanta, 2003). A benchmark study comparing the performance of SVMs with other classification techniques for natural and artificial datasets has been presented in Meyer (2003).

In the following few sections we provide an overview of some major techniques being used for SHM.

1.1.1 SHM methods based on modal properties

Damage in a structure causes a reduction in the stiffness of the structure and in turn modifies the modal properties of the structure. These changes in the modal properties can be used to detect damage in the structure. The amount of change can be used to estimate the damage extent.

1.1.1.1 Damage detection methods based on shift in the natural frequency

The natural frequency of a structure is one of the fundamental dynamic characteristics of the structure at which it resonates with an external disturbance of different frequency contents. The natural frequency is a function of the physical properties (mass and stiffness) of a structure and can be extracted from the vibration signal measured from the structure. Damage in the structure, whether local or distributed, reduces the stiffness and hence the natural frequencies of the structure. The relative shift of natural frequencies of the structure can be used as an indicator of the presence and location of damaged elements of the structure. The advantage of using natural frequencies as a parameter is that they have much less statistical variation from random sources as compared to other modal parameters (Doebling, 1996a). The identification of a damage location and the severity of damage based on the change in the global properties derived from measurements at a limited number of sensor locations is a problem that has a non-unique solution (Humar et al., 2006). A large number of frequencies need to be measured for precise damage extent and location prediction.

Measurement of multiple frequencies is generally not possible with good accuracy due to the presence of noise, coupling of the higher modes and the limitations of the measuring equipment etc. Global properties like the natural frequency and mode shapes are not very sensitive to local damage in the structure (Chang et al., 2003). This is because the stress distribution in a structure is non-uniform and different for each natural frequency (mode). Resultantly, the relative effect that a change in stiffness of each member will have on the modal properties will be different (Cawley et al., 1979). It is easier to detect damage using frequency shifts but more difficult to locate it because damage in more than one location can cause similar frequency shifts.

Cawley et al. (1979) used natural frequencies for damage detection in structures. In the paper they compared the sensitivity matrix of each natural frequency for different damage locations with the actual vibration signal and selected the one with the least error. As the method compares model with the actual signal, for accurate damage detection and location with this method, a detailed mathematical model of the structure is required.

A method to improve the accuracy of damage detection using frequency shifts by applying a preload to the structure before dynamic and static tests was presented by Chen (1999). They demonstrated that with preloads the structure behaves linearly because the load keeps the cracks open. Brincker et al., (1995a) use the difference between the natural frequencies of the vibration data measured from a concrete offshore oil platform structure and the auto regressive moving average (ARMA) model of the structure to evaluate the goodness of fit of the model.

1.1.1.2 Damage detection methods based on a change in the mode shapes

Damage in a structure also changes its mode shapes. This change can be used to detect and locate damage in the structure. The mode shapes can be derived from the vibration signal. The standard test in the 1970s for the accuracy of measured modal vectors was the orthogonality test (Allemang, 2003). The essential condition to validate the experimental modal model was that each experimentally measured mode of a structure should be normal to other mode vectors (experimental or analytically derived from FEM model) when weighed by the analytical mass, stiffness or damping vector (derived from the FEM model). Unfortunately experimental modal vectors are not always orthogonal because of inaccuracies in measurement, incorrect mass matrix in the FEM model and errors introduced by the reduction algorithm used to modify the FEM mass matrix to match the dimensions corresponding to the number of measurement locations. Allemang (1980) introduced the modal assurance criterion (MAC) which is a statistical measure of the consistency (degree of linearity) between estimates of a modal vector. MAC and many of its variants have been used for the validation of experimental models, structural fault detection and optimal sensor placement among other applications.

Liu (1995) use the modal shapes and natural frequencies to identify the element parameters like stiffness and density of a structure from vibration measurements. He used the derived element properties to locate the damaged element in a truss.

Cao et al. (1999) introduced a method to extract load dependent Ritz vectors from experimental data using the state space formulation. Ritz vectors, just like modal

vectors, give an alternative form to represent the system (like modal equations) and can be used to reproduce the output vibration signal for a given input excitation signal (Sohn et al., 2001). Ritz vectors (or Lanczos vectors) can be analytically derived from the model of the structure and can also be experimentally measured as explained in this paper.

1.1.2 SHM methods based on change in the dynamically measured flexibility and other matrix update methods

Another set of methods attempts to use the change in the structural properties (mass, stiffness and damping) for detection of damage. Finite element models are developed to numerically model structures for static and dynamic analysis. The mass, stiffness and damping matrices of the structure can be derived from these models. When there is damage in the structure, the measured vibration signal from the damaged structure can be used to update the parameters of the FEM model representing the undamaged structure. The changes in parameters required to update the model can then be used to estimate the damage.

Mottershead et al. (1993) did a survey of the existing finite element model update methods and classified the major forms of model error as:

1. Model structure errors: the model does not reflect the system dynamics
2. Model parameter errors: inaccurate assumptions and boundary conditions in the model
3. Model order errors: discretizing the structure for modeling leads to a model of inaccurate order

Once the model order and structure have been decided upon, the problem is then one of parameter estimation for the mass, stiffness and damping matrices (system identification). To solve the system identification problem the least squares method was used and then later developed into more sophisticated methods like the Kalman, Weiner and Kolmogorov filters.

The higher modes of the structure contribute more to the calculation of the stiffness matrix and so higher frequency modes are needed for the estimation of the stiffness matrix. In practice, only the first few modes can be measured (Pandey et al., 1994). To workaround this challenge, Pandey et al. (1994) used the flexibility matrix in place of the stiffness matrix for damage estimation. The flexibility matrix is defined as the inverse of the stiffness matrix and is affected more by the lower modes hence converging for a fewer number of modes. Each column of the flexibility matrix represents the displacement pattern when a unit force is applied at the associated DOF. The change in the flexibility matrix of a structure can be used as an indicator of damage in a structure.

Pandey et al. (1994) evaluated change in flexibility as an indicator of damage and as a method to locate and estimate damage in a variety of numerical and experimental studies. As the method requires mode shapes for the calculation of the flexibility matrix, vibration data needs to be measured at a number of sensors in the structure for a detailed modal analysis.

Amani et al. (2007) determined the change in the damping and stiffness matrices during damage using perturbation theory. They called the change in damping and stiffness matrices during damage as the “damage” damping and “damage” stiffness matrices

respectively. In their paper they derived the mass, damping and stiffness matrices of the undamaged structure from the FEM model. The “damage” stiffness and damping matrices for the structure (assuming the mass matrix is unchanged) were evaluated from the measured damaged modal matrices. The ability of the method to locate damage in the structure and also get a reasonable estimate of the magnitude of the damage through a number of simulation and experimental tests was illustrated. The “damage” damping matrix is a more sensitive indicator of damage than the “damage” stiffness matrix for some kind of structures. In this method the mass, stiffness and damping matrices for the undamaged structure are required to be available and can either be available from past measurements or can be derived from the output of a FEM model of the structure.

1.1.3 SHM using Statistical Pattern Recognition methods

Recently SHM is being investigated as a pattern recognition problem. The complete pattern recognition process consists of data acquisition, feature extraction, and classification or description. In the field of SHM, physical and modal parameters derived from the vibration response of the structure have been used as features for classification. Schalkoff et al. (1992) broadly divided classification techniques under pattern recognition into statistical and structural methods based on the approach used for classification or description. Statistical pattern recognition methods are based on some statistical measure derived from quantitative properties of the input data. Structural methods, on the other hand, use the interrelationships or interconnections of the input features for classification. Statistical Pattern recognition methods for

classification can also be divided, based on the availability of training data that has been already been classified, into supervised and unsupervised learning methods. Supervised learning methods are machine learning techniques that try to learn a function to map the input class with the output class (labels) in the training dataset. The training dataset consists of input vectors and desired output vectors. Unsupervised methods try to determine how the data is organized from the training set without the availability of labeled examples.

In the field of SHM, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) use the supervised learning paradigm for damage detection and location.

1.1.3.1 Artificial Neural Networks (ANN)

ANNs are computational models used to approximate the relation between inputs and outputs and have been applied to a wide variety of problems. The neural network consists of layers of kernel functions where the output of one layer is the input to the next. The output of each layer of the ANN is multiplied by weights, added, shifted by a bias and then fed to the next layer as input. In networks with back-propagation, the error between measured and predicted outputs is minimized by adjusting the weights and biases.

The concept of ANNs has been around for more than 40 years but was first applied for damage detection by Ghaboussi et al. (1991). In this paper an ANN with back-propagation was used to model the complex relationship between stresses and strains in reinforced concrete for the state of plane stress under monotonic biaxial loading and compressive uniaxial cycle loading.

In Kudva (1992), ANNs that were trained on FEM models of the structure were used to deduce the location and extent of damage. Strain patterns were calculated for different damage sizes and locations using an FEM model of the structure. Measured data was compared against these strain patterns to detect damage locations and damage severity. The main limitation of the method is that representative training sets for all damage types and damage percents have to be carefully chosen. The effect of noise on the ability of the ANN to detect presence and location of damage has not been discussed. In the paper it was reported that estimation of damage location was easier than estimation of the damage extent.

In Lim (1996), ANNs were used for online identification of modal parameters (such as natural frequency, damping ratio, and mode shape vectors) at each measurement location in the structure. The filtered signal measured was used to train a neural network which consists of a linear neuron with 3 weights. A limitation of the method is the extremely high sampling rates required for accurate parameter estimation. The ability of the structure of the ANN to adapt with time has not been demonstrated in the experiments.

1.1.3.2 Support Vector Machines (SVM)

These are a set of supervised learning methods for classification and regression. When applied to a binary classification problem, the SVM technique involves finding the separating plane between the two classes of training data that maximizes the distance between the classes (Burges, 1998). The assumption is that the maximum margin classifier plane is the best decision boundary between the two sets of data. This

classifier plane can then be used to classify test data based on which side of the classifier the data point lies. SVMs have been used for SHM by treating the damage detection and location problem as a classification problem. A feature that is derived from the vibration data measured on the structure is used by the SVM technique to create the maximum margin classifier between the datasets derived from the damaged and the undamaged structure. The maximum margin classifier can then be used to classify the structure into either healthy or damaged in the future.

The main difference between ANNs and SVMs is the risk minimization. The risk is defined as the expectation of an error in classification of test data that is drawn from an unknown probability distribution. ANNs try to minimize the traditional empirical risk. The empirical risk is defined as the measured mean error rate on the training set (for a fixed, finite number of observations) (Bishop, 1995). SVMs use the "structural risk" minimization principle which minimizes a cost function based on the empirical risk and the complexity of the model (Burges, 1998). More is explained about structural risk minimization later in the mathematical background of SVMs.

When using SVMs in multiple-class classification problems, 3 main approaches are used (Widodo, 2006):

1) One Against All (OAA): constructs k SVM models where k is the number of output classes.

2) One Against One (OAO): constructs k*(k-1)/2 classifiers where each classifier is trained on data from 2 classes. The final class is predicted using the largest vote.

3) Direct Acyclic Graph (DAG): constructs k*(k-1)/2 classifiers for training. During testing, a rooted binary direct acyclic graph is created which has k*(k-1)/2 internal nodes and k leaves.

Widodo et al. (20007) summarize that so far the following parameters have been used as input features for damage detection in SHM:

1) Couples, moments and other statistical features

2) Frequency domain analysis

3) Other parameters specific to the problem like, temp, depth of cut, feed rate etc.

In Yang et al. (2005), a comparison has been made between the performance of SVMs and two ANN methods (self-organizing feature maps and learning vector quantization) for damage detection. Features are extracted using the wavelet transform from the raw noise and vibration data. It was found that SVMs and the Learning Vector Quantization (LVQ) were most accurate.

Shimada et al. (2005) used SVMs to solve the multiple-class classification problem of damage detection and prediction of damage location. The change pattern of the natural frequencies with damage was used as a feature for training the SVM and later detecting the location of damage in the structure. Abbasion et al. (2007) used SVM with the radial basis function kernel to detect the kind of damage in roller bearings. The vibration data measured from the structure was denoised using wavelet analysis. Comparison of the damage type prediction results of SVM with other methods showed that SVMs required only one training set with 2 input dimensions to train the SVM classifier for many more (7) output classes.

Widodo et al. (2007) summarized the various machine condition monitoring and fault diagnosis applications using Support Vector Machines. They mention that SVMs perform better than ANN when generalizing from a small number of input data points or when there are a fewer number of dimensions in the training data. The use of SVMs for damage detection in structures like buildings or bridges is still missing. For a more detailed discussion on the methods used for cleaning data or for feature extraction the reader is encouraged to refer to Widodo (2007).

Zhang et. al. (2008) tested the applicability of the support vector regression (SVR) method for the estimation of the physical parameters of the structure. The stiffness and damping parameters for a five-floor shear-building shaking table computed by the SVR technique were compared against values computed using the H-infinity method. The SVR method was found to perform better than the Least Squares method for the estimation of physical parameters of the structure from the vibration signal. A limitation of the parameter estimation method used in this paper is that only simple structures can be analyzed and the estimation becomes a challenge as the number of model parameters increases.

Noori et al. (2008) used SVMs for damage detection and reliability assessment. The SVM classifier was used as an approximate limit state function for reliability analysis to calculate the probability of the structure to fail. The paper also presented some representative results in this thesis.

In the training stage SVMs require data from the undamaged and damaged structure. In industrial applications, many samples of data for each class (damaged and undamaged)

are available. However, the SVM technique is not directly suited for use in structural health monitoring of structures (like building, bridges etc.) because in real structures data is not available from both classes (undamaged and damaged) due to the irreversibility of damage. To address this challenge a mathematical model is generally used to simulate data for the damaged structure. Using a model of the structure to generate data has its own challenges because of modeling inaccuracies, non-linearity of the structure and the inability of the model to accurately predict the output of the structure for changing operating conditions and environmental conditions.

1.2 Limitations of current methods

An effective SHM strategy for a structure involves the application of both global and local SHM methods. The disadvantage of global SHM methods is that they are dependent on global properties of the structure which are insensitive to small local damage. The extracted properties are also sensitive to the location of the sensors and the location of the damage in the structure. Measurement noise in the vibration data poses a challenge to extraction of structural and modal parameters of the structure. SVMs applied for damage detection in industrial applications using modal properties (global) have proved very promising in their ability to be able to detect damage in the presence of noise. SVMs have been found to have very good generalization ability due to their underlying technique of maximizing the margin between 2 classes of the data. The main challenge while using SVMs for damage detection in structures is the requirement of data from the damaged structure during the training phase. To address

the issue of unavailability of damaged data from a structure, in this thesis, we propose a method to generate the damaged dataset from the undamaged dataset measured in the structure.

1.3 The proposed methodology and expected results

In this thesis a method is proposed to address the issue of dependence of SVMs on data from the damaged structure during training. In the proposed method, we derive the damaged dataset from the undamaged dataset. The undamaged dataset is acquired from the structure. Due to the unavailability of testing equipment, all the results described in this thesis have been derived from simulations.

To create the training set of data for the SVM:

1. Vibration data is first measured from the undamaged structure.

2. Representative features of the structural health condition are extracted from the vibration signal. These features could be the natural frequencies, the mode shapes or some other numerical parameter extracted from the vibration signal. All the selected features form a so-called input feature space for SVMs and the number of features used is the dimension of the input feature space. In this thesis the first two natural frequencies have been used to represent a point in the input feature space of the SVM. Multiple points extracted from the vibration signal constitute the undamaged “class” of data.

3. The dataset for the damaged class is then derived from the undamaged class of data by creating a copy of the undamaged class and then reducing the values of this

copy by “some” percent. This value of the percent reduction of the undamaged dataset depends on the spread of the data. The intention of the reduction is to create two separable datasets. The reduced values of the natural frequencies can be used to create the damaged class in the training dataset for the structure since damage in the structure would reduce the stiffness of the structure and hence the natural frequencies. It should be noted that the damaged class dataset has been derived from the undamaged dataset without using a numerical/FEM model of the damaged structure.

4. Both the undamaged and damaged classes of data are then used in the SVM technique as training data. Using the training dataset, the best possible hyperplane is then computed between these two classes of data. The evaluation of the best hyperplane involves choosing the parameters of the SVM (discussed later), the optimum percent reduction of the training undamaged class and then solving the SVM optimization problem. The health condition of a structure can then be predicted by finding on what side of the SVM hyperplane does a test point derived from the structure lie.

It is expected that the trained SVM will be able to detect damage present in the structure under different working environments.

The SVM boundary was found to shift when the undamaged class of training data it was trained on was collected from a structure with damage in it. As the damage in the structure increased, the SVM boundary trained on it was found to shift gradually. A method has been proposed for online health monitoring based on this shift of the SVM

boundary. It is expected that the SVM boundary will act as an accurate indicator of damage in the structure.

1.4 Scope of thesis

In this thesis a method has been proposed to address the challenge of unavailability of data from the damaged structure to train the SVM. The issues that were faced during the development of the method are explained in the following sections. The applicability of the method to detect damage was then tested under different operating conditions.

In the next section, the mathematical background of linear and non-linear SVMs has been discussed. This is followed by a description of the method of numerically solving the SVM technique.

Due to the unavailability of measuring equipment, all results have been derived from simulations. After the mathematical background of the method has been developed, we discuss the numerical model that was used to derive all the results. It should be noted that though the method has been demonstrated for use with the 4 degree of freedom (DOF) model, it can be used for any general structure.

Section 3 discusses the results that were obtained by applying the method for damage detection in the model. The effect of different parameters in the SVM algorithm on the final hyperplane selected is explored. The applicability of the method is tested in this section under different working conditions. The ability of the method to detect damage in different floors and the effect of sensor location on the sensitivity of the method is investigated.

Section 4 introduces an online SHM technique using the SVM boundary. This method provides an easy visual indication of the health condition of the structure.

In Section 5, the conclusions from the study have been explained. Section 6 mentions areas where there is scope for improvement of the method to be able to apply the method more effectively for damage detection.

2 Methodology

2.1 Mathematical background for SVMs

Support vector machines (SVMs) are a set of supervised learning methods that have been used for the classification and regression of data. Supervised learning methods are machine learning methods that try to create a map between the inputs and outputs in the training data. The training data consists of points from each class with the expected output specified for each point.

In this thesis damage detection of a structure is treated as a classification problem. During the training phase of the SVM, the algorithm finds the best hyperplane that separates data from the undamaged and the damaged structure in the feature space. The feature space is the space with the dimensions of the input data.

In the SVM technique, the best hyperplane is that plane which divides the target classes with the maximum margin between them. The approach lends itself to a very interesting geometrical interpretation and has scope for further research based on its interpretation.

The best hyperplane can be a “linear” boundary in the input feature space. There can be some cases where a linear boundary might not be able to separate the two classes of data adequately. In such cases a non-linear boundary can be used to separate the target classes better. The mathematical background for finding the best linear and non-linear boundary to separate the target classes has been explained in the next two sections.

2.1.1 Linear SVMs

In this section we discuss the mathematical background for creating a linear maximum margin boundary to separate two classes of data. As an example, the figure below contains points that belong to two linearly separable datasets (Class 1: red and Class 2: blue respectively). There are an infinite number of planes that can be drawn between the 2 sets of points e.g. H1, H2, H3…etc. The best dividing plane is one that will maximize the distance between the closest points on both sides of that plane. For the plane H2, the closest points on both sides are P1 and P2. It should be noted that for a different plane H1 or H3 the points from both classes closest to them can be different from P1 and P2. In the given figure, H2 is the best dividing plane because its distance from the closest points on both sides (P1 and P2) is the maximum possible. The classifier H2 is equidistant from both points P1 and P2.

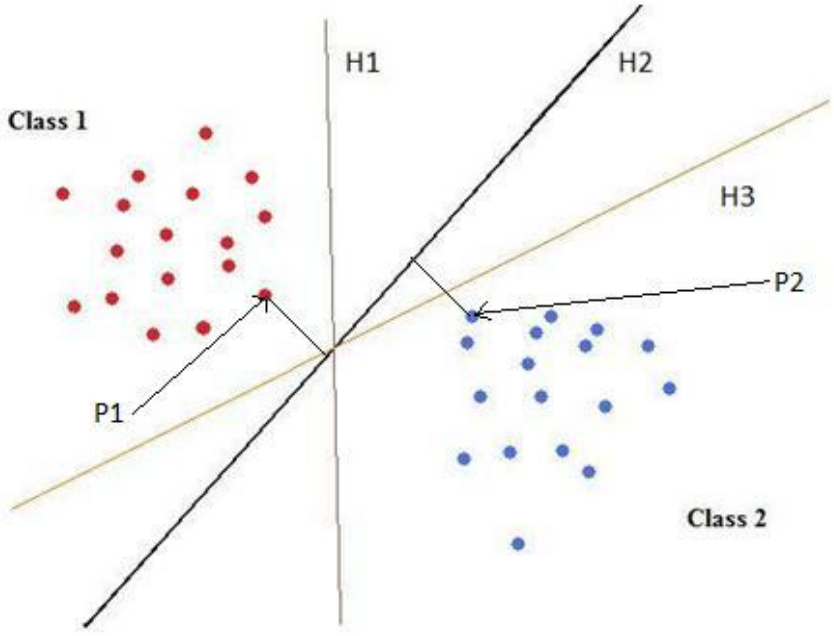

Fig. 1: Best classifier between two classes of data

Applying this approach of finding the best classifier to the problem of damage detection in structures, in Fig. 2 the two classes correspond to two linearly separable datasets extracted from the undamaged and damaged structures. The features 1 and 2 are the parameters extracted from the vibration data of the structure. The best classifier is one that maximizes the distance between the closest points in both classes and the hyperplane (H1). The planes H2 and H3 are called margins and are parallel and equidistant from the plane H1. They pass through the closest points in both classes. The closest points appear to be supporting the margins outwards and hence are called the support vectors. The selected configuration is called the Support Vector Machine. The Equation for H1 can be written as

$$x.w + b = 0 \tag{1}$$

where x is any point that lies on the separating plane (H1) and w is a vector perpendicular to the plane. The distance of the separating hyperplane from the origin is $b/||w||$. The parameter w can be normalized such that the regions away from the two margin planes (H2 and H3) in the two classes are can be represented by:

H2: $x_i.w + b = -1$ for y_i = -1 and (2)

H3: $x_i.w + b = +1$ for y_i = +1, respectively. (3)

The region of Class 2 is: $x_i.w + b \leq -1$ and the region of Class 1 is: $x_i.w + b \geq +1$.

The labels for the target classes, y_i , have been chosen to be ±1 for convenience. The equations (2) and (3) can be combined into one equation:

$$y_i(x_i.w + b) \geq +1 \tag{4}$$

The distance between the margin planes is: $2/||w||$. The best separating plane H1 is that which maximizes the margin: $2/||w||$, which is equivalent to minimizing $||w||$. Hence finding the best plane equates to minimizing $||w||^2/2$. It should be noted that changing of the objective function from $||w||$ to $||w||^2/2$ does not change the solution of the minimization and the factor ½ is used for convenience. This minimization problem is now a quadratic programming problem with the objective function $||w||^2/2$ under the constraints defined by equation (4).

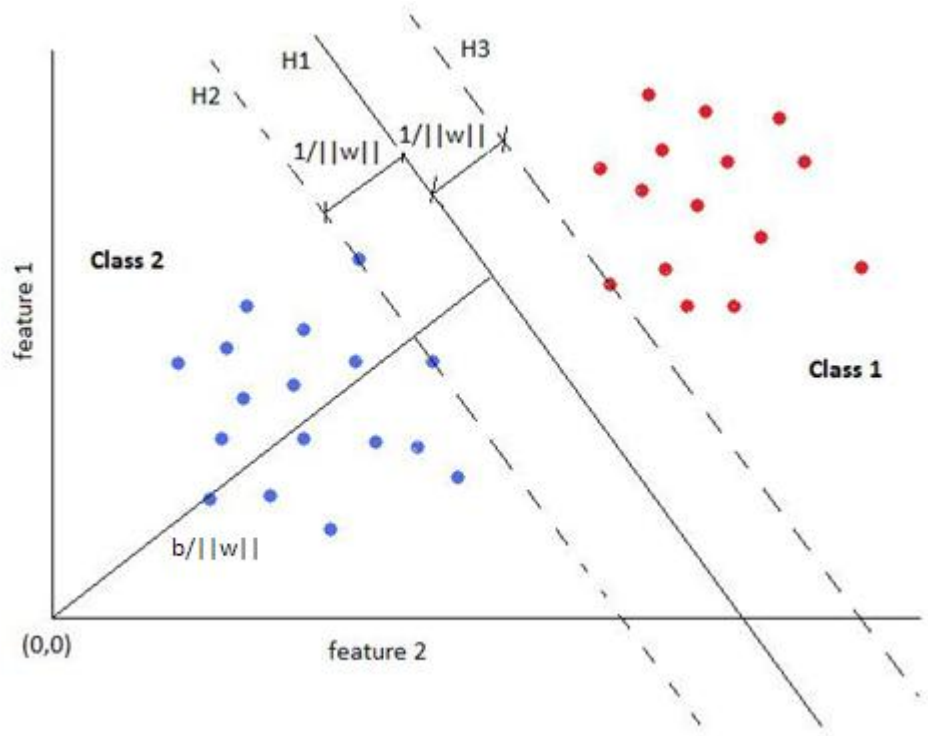

Fig. 2: Best classifier to separate Class 1 (data from undamaged structure) and Class 2 (data from damaged structure)

The optimization problem can be solved by using Lagrangian multipliers. The advantage of using the Lagrangian formulation is that the constraint (4) is changed onto a constraint on the Lagrangian multipliers. Later during the training stage of the SVM, the

training data will appear only as dot products between vectors. The final objective function can be written as:

$$Lp \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i . w + b) + \sum_{i=1}^{l} \alpha_i \quad (5)$$

where α_i are positive Lagrangian multipliers. This is the primal form of the objective function. The aim is to minimize Lp with respect to w and b while requiring that the derivative of Lp with respect to all α_i be zero for $\alpha_i \geq 0$. The objective function of this optimization problem and the linear constraints are convex. The optimization problem with a convex objective function and convex constraints is convex. For a convex quadratic optimization problem, as above in (5), any local solution will be a global solution. The Karush-Kuhn Tucker (KKT) conditions are a generalization of the method of Lagrangian multipliers to inequality constraints. For the SVM convex optimization problem, KKT conditions are necessary and sufficient for $w, b \; and \; \alpha$ to be a unique solution (Fletcher, 1987). Finding the solution to the SVM problem is equivalent to finding the solution to the Karush-Kuhn Tucker (KKT) conditions. The KKT conditions for the above optimization problem can be defined as:

$$\frac{\partial}{\partial w_v} Lp = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad v = 1,2 \ldots d \quad (6)$$

$$\frac{\partial}{\partial b} Lp = -\sum_i \alpha_i y_i = 0 \quad (7)$$

$$y_i . (x_i . w + b) - 1 \geq 0 \quad i = 1, \ldots l \quad (8)$$

$$\alpha_i \geq 0 \text{ for all } i \quad (8)$$

$$\alpha_i (y_i . (x_i . w + b) - 1) = 0 \text{ for all } i \quad (9)$$

where d is the dimension of the training data. The parameter i goes from 1 to the number of training points. The equation (8) is called the Karush-Kuch Tucker complementarity condition. It states that for strict inequality constraints, $(y_i.(x_i.w+b)-1)>0$, α_i=0. For equality constraints, $(y_i.(x_i.w+b)-1)=0$, α_i≥0.

Due to its convexity, this minimization problem in the primal form can be changed into an equivalent maximization problem in the Dual form (Wolfe, 1961). This dual objective function can be calculated by taking the derivative of the primal form, (5), with respect to w and b while requiring $\alpha_i \geq 0$. The Dual form changes the minimization problem of (5) with constraints, say C1, to a maximization problem with constraints, say C2, without changing the point at which the solution occurs. In the Dual form, the problem can be written as:

$$Ld \equiv \sum_i \alpha_i - \tfrac{1}{2}\sum_{i,j} \alpha_i y_i \alpha_j y_j (x_i.x_j) \tag{10}$$

Under the condition:

$$\sum_i \alpha_i y_i = 0 \tag{11}$$

In the Dual form of the optimization problem, all the constraints are placed only on the Lagrangian variables. This form of the equation has easier constraints than the inequality constraints in the primal form to work with. The Dual form also allows the application of a number of algorithmic techniques derived from optimization theory for computational convenience (Cristianini, 2000).

The normal to the separating plane is calculated as:

$$w = \sum_i \alpha_i y_i x_i \tag{12}$$

The training of the Support Vector Machine involves maximizing Ld under the condition (11) and calculating the normal to the final classifier plane by eq. (12). The variable "b" was evaluated implicitly in the optimization and its calculation would require substituting the values of support vectors x_i, the corresponding class labels (y_i) and the value of w in equation (4) for the equality constraint. For numerical precision reasons, it is better to take an average over all points for which α_i are nonzero.

The values of α_i are nonzero only for the points that lie on the margin. These points are called "support vectors" as they can be viewed as supporting the margin boundaries outwards. For all other points in the training dataset $\alpha_i = 0$. Support Vectors are the only points needed in the calculation of the normal to the separating plane using eq. (12). Once the SVM has been trained, all data points in the training set that are not support vectors can be moved in the feature space and as long as none of these points crosses over the margins, training on this changed dataset will create the same SVM boundary as earlier.

The above classification problem only considers problems when the classes are separable. In case there are misclassifications (possibly due to errors in measurement), the dual problem would become very large. The current method can be adapted to suit the situation by introducing positive slack parameters (ξ_i) into the equations of the margins. This in effect makes the margins act as "soft" boundaries capable of absorbing some misclassification. The modified equation of the margins is:

$$\text{H2: } x_i.w + b \leq -1 + \xi_i \text{ for } y_i = -1 \text{ and} \tag{13}$$

$$\text{H3: } x_i.w + b \geq +1 - \xi_i \text{ for } y_i = +1 \text{ respectively.} \tag{14}$$

These functions can be combined into one function in the form:

$$y_i(x_i.w + b) \geq +1 - \xi_i \tag{15}$$

For an incorrect classification, the parameter ξ_i has to be greater than one. $\sum \xi_i$ can provide a measure of the misclassification. The objective function of the minimization problem changes from minimizing $\|w\|^2/2$ to minimizing ($\|w\|^2/2 + C(\sum \xi_i)^p$). Here the parameter C is a user selected cost parameter for incorrect classifications (nonzero ξ_i). The effect of this parameter on the final SVM boundary has been investigated in section 3.2.1.2. The optimization problem stays convex for any value of parameter p. For p=2 and p=1 the problem stays a quadratic programming problem and for k=1 the slack parameters disappear from the dual form of the optimization problem. The Dual form of the optimization problem for non-separable data is:

$$Ld \equiv \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i y_i \alpha_j y_j . \left(x_i . x_j\right) \tag{16}$$

It should be noted that the slack variables (ξ_i) are missing from the dual formulation of the optimization problem.

Under the condition:

$$\sum_i \alpha_i y_i = 0; \tag{17}$$

$$0 \leq \alpha_i \leq C \tag{18}$$

And the SVM is calculated with

$$w = \sum_i \alpha_i y_i x_i \tag{19}$$

The only difference in the above formulation from the separable case is the upper bound on the Lagrangian parameters. There number of Lagrangian parameters is equal to the number of training points and the magnitude of the Lagrangian parameter indicates the effect that the corresponding training point can have on the final separating boundary being calculated. In the separable data case the upper bound of α_i was infinity, implying an infinite effect of each point in the calculation of the classifier. For non-separable data, the

condition $\alpha_i \leq C$ can be interpreted as the maximum effect that any one point can have on the nature of the final decision boundary is limited.

The primal form of the Lagrangian for the non-separable case is:

$$Lp \equiv \frac{1}{2}\|w\|^2 + C\sum_i \xi_i - \sum_{i=1}^{l} \alpha_i\{y_i(x_i.w + b) - 1 + \xi_i\} - \sum_i \mu_i\xi_i \quad (20)$$

Where μ_i are Lagrangian parameters to enforce that $\xi_i \geq 0$.

The KKT conditions for the non-separable data are:

$$\frac{\partial}{\partial w_v} Lp = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad v = 1,2 \ldots d \quad (21)$$

$$\frac{\partial}{\partial b} Lp = -\sum_i \alpha_i y_i = 0 \quad (22)$$

$$\frac{\partial Lp}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \quad (23)$$

$$y_i.(x_i.w + b) - 1 + \xi_i \geq 0 \quad i = 1, \ldots l \quad (24)$$

$$\xi_i \geq 0 \quad (25)$$

$$\alpha_i \geq 0 \text{ for all } i \quad (26)$$

$$\mu_i \geq 0 \text{ for all } i \quad (27)$$

$$\alpha_i(y_i.(x_i.w + b) - 1 + \xi_i) = 0 \text{ for all } i \quad (28)$$

$$\mu_i\xi_i = 0 \text{ for all } i \quad (29)$$

For all points where $\xi_i > 0$, from (29), μ_i = 0. This result when used along with equation (23), means $\alpha_i = C$. Hence, all wrongly classified points are always support vectors and affect the calculation of the decision boundary to the maximum possible. The other support vectors have varying level of effect on the calculation of the calculation of the SVM. It should also be noted that $0 < \alpha_i < C$ when $\xi_i = 0$. To calculate b, an average should be taken of the value calculated from equation (28) for all support vectors where $0 < \alpha_i < C$.

From equation (29) it can be noted that when the inequality condition holds in equation (15), $\alpha_i = 0$. This would mean that all points that are not on the SVM margins and which have not been wrongly classified do not affect the calculation of the SVM boundary. This reduces the amount of computation and the memory required to store matrices during the computation of the SVM boundary. As a comparison, if the normal to a simple regression line created for the training data was used for classification, all the data point would be required for the calculation and, therefore, the computation effort is expensive for a large amount of training data.

Till now we have discussed situations where a linear boundary suffices to act as a separator. There might be situations where a non-linear boundary or classifier is more suitable. There is very minor modification required to the derivation done so far. Note that in equation (16), the linear non-separable dual objective function, the training data x_i exist only as dot products with other data points. This particular form lends itself to a very simple mathematical transformation called the kernel trick when modifying the linear classifier to a non-linear one, as described in the following section.

2.1.2 Non-Linear SVMs

Suppose we use a mapping function vector ϕ to project the data points x from the lower dimensional feature space, say *L,* to a higher dimensional space, say *H.*

$$\phi: L \rightarrow H \tag{30}$$

After the mapping, the dot products between the points would change from $(x_i . x_j)$ to $\varphi(x_i) . \varphi(x_j)$. However, as the dimension of *H* can be very high, possibly infinite,

computations in it would be very expensive. One way to address the challenge would be to find a function K($x_i . x_j$) such that:

$$K(x_i, x_j) = \varphi(x_i) . \varphi(x_j) \tag{31}$$

then that would lead to multiple benefits while creating the SVM in the high dimensional space *H*. Firstly, we do not need to calculate the mapping function φ as we can just use this function K. Secondly, as the computations are done in the data feature space L, the computations take as much time as it would for unmapped data. The function K is called the kernel function. To modify our algorithm to create this SVM in the high dimensional space, we can now replace all dot products between points by their kernel functions. The optimization problem can now be written as:

$$Ld \equiv \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i y_i \alpha_j y_j . K(x_i, x_j) \tag{32}$$

Under the condition:

$$\sum_i \alpha_i y_i = 0; \tag{33}$$

$$0 \leq \alpha_i \leq C \tag{34}$$

And the SVM is calculated from

$$w = \sum_i \alpha_i y_i . \varphi(x_i) \tag{35}$$

It is important to note that the vector *w* is calculated in the higher dimensional space *H* and there might be no vector in the lower dimensional space *L* that can map to it. Also, even though the nonlinearities introduced by the kernel alter the quadratic form of the primal Lagrangian, the dual form is still quadratic in α. However, to use the SVM classifier for classification we do not need to calculate the value of *w* explicitly. To classify a test point we need to calculate to which side of the classifier plane does the test point lie. For a point z, the class that it lies in can be calculated by computing:

$$\text{sgn}(\varphi(z).w + b) \equiv \text{sgn}(\sum_i \alpha_i y_i . \varphi(x_i).\varphi(z) + b) \quad (36)$$

where ∀i = 1… number of support vectors;

x_i : support vectors.

By applying the kernel trick, this can be further simplified to:

$$\text{Class of point z} = \text{sgn}(\sum_{i=1}^{number\ of\ support\ vectors} \alpha_i y_i K(x_i, z) + b) \quad (37)$$

The point "z" for a damage detection problem will be a feature vector, derived from the vibration signal of the structure, which will be used to classify the structure into the undamaged and the damaged classes.

An SVM always finds a global solution in comparison to an ANN where local minima might exist. For SVMs, since the objective function is convex, all local solutions are global. Strict convexity of an objective function means a unique solution. For a quadratic objective function, the Hessian of the objective function is positive definite if and only if the objective function is strictly convex. Hence for a quadratic objective function, if the Hessian is positive definite, the solution will be unique. For non-quadratic objective functions, a positive definite Hessian implies a strictly convex function but not vice-versa. Hence, the objective function could be strictly convex (and resultingly the solution of the optimization unique) even for a positive semi-definite Hessian. If the solution calculated is not unique (which can happen only for a positive semi-definite Hessian), then there is a path between the 2 points that are solutions such that all points on this path are also solutions (Burges, 1998).

In this thesis, we demonstrate the use, and later test, the proposed method for damage detection in structures using simulation data from a 4 degree of freedom spring-mass-

damper model. It should be noted that though the use of the method has been illustrated for a 4-DOF system, it can be applied to detect damage in any structure. In the following section, the major steps involved with using SVMs for damage detection by the proposed method are explained.

2.2 Method for using SVMs for damage detection

In this section, the procedure to use the SVM technique for damage detection in structures is explained. The proposed method for creating the damaged class of training data from the undamaged class has been demonstrated using data from simulation studies. All results have been derived for a 4-DOF spring mass damper system. The same procedure can followed to apply the method for damage detection in any structure. The main steps in the damage detection method are the measurement of data from the undamaged structure, creation of the damaged training class and training the SVM from this dataset. The methodology to test the generated SVM for damage detection in a structure has also been explained.

2.2.1 Creation of "Undamaged" class of simulation data

A 4 degree of freedom spring-mass-damper system was used to create the vibration data. The structure is assumed to be undamaged initially. The Fourier transform of the vibration data collected from the structure is calculated using FFT to extract the natural frequencies of the structure. The first two natural frequencies have been extracted and form the two features of the input data. These features (2 natural frequencies) are the 2

dimensions of the input feature space. In the following sub-sections we describe the model used in the simulations and the procedure to derive damaged class of data.

2.2.1.1 4- DOF Spring Mass Damper Model used

The 4 DOF system is used to represent a lumped mass model of a 4 storey building.

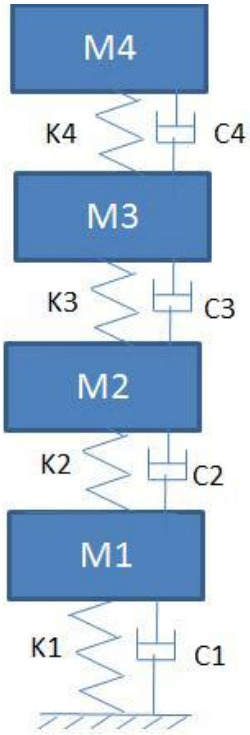

Fig. 3: 4-DOF spring-mass-damper model

The equation of the model is:

$$[M]\ddot{\vec{x}}(t) + [C]\dot{\vec{x}}(t) + [K]\vec{x}(t) = F(t) \quad (38)$$

The vector $\vec{x}(t)$ is a 4 by 1 vector representing the lateral displacement of the 1st to 4th DOF of the structure with respect to the base. The other model parameters have been chosen accordingly to model lateral motion.

In equation (38), [M] is the mass matrix defined by:

$$\begin{bmatrix} M1 & 0 & 0 & 0 \\ 0 & M2 & 0 & 0 \\ 0 & 0 & M3 & 0 \\ 0 & 0 & 0 & M4 \end{bmatrix} \tag{39}$$

The values for the masses of each DOF are as follows:

$$\mathrm{M1} = \mathrm{M2} = \mathrm{M3} = \mathrm{M4} = 2.25 * 10^5 \,\mathrm{kg} \tag{40}$$

The stiffness matrix [K] in equation (38) is defined by:

$$\begin{bmatrix} K1 + K2 & -K2 & 0 & 0 \\ -K2 & K2 + K3 & -K3 & 0 \\ 0 & -K3 & K3 + K4 & -K4 \\ 0 & 0 & -K4 & K4 \end{bmatrix} \tag{41}$$

The stiffness of the springs is as follows:

$$K1 = K2 = K3 = K4 = 2.8147 * 10^7 N/m \tag{42}$$

The effects of nonlinearities have been ignored in the model. Rayleigh damping is assumed in this study, i.e. $C = \alpha\,[M] + \beta[K]$. The values of parameters α and β are $0.2s^{-1}$ and 0.0015 s respectively. The resulting Modal Damping Ratios of the model were $[0.0287 \quad 0.0173 \quad 0.0187 \quad 0.0205]$. The time step for the simulation was 0.02 seconds and the total time for the simulation was 600 seconds.

The natural frequencies and the corresponding modal shape vectors can be found by solving the associated eigenvalue problem (Rao, 2003):

$$\Delta = |[K] - \omega^2[M]| = 0 \tag{43}$$

where ω are the natural frequencies of the system. Their values were: 3.884 rad/s^2, 11.185 rad/s^2, 17.136 rad/s^2 and 21.021 rad/s^2 respectively for the model.

Fig. 4 shows the mode shapes of the 4-DOF model.

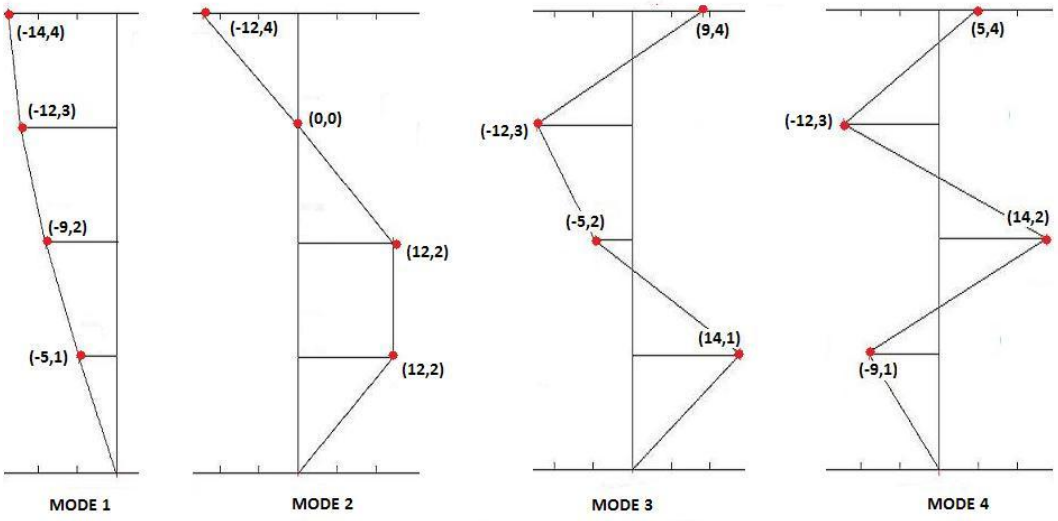

Fig. 4: Mode shapes for the 4-DOF Model

2.2.1.2 Analysis of simulation data without measurement noise:

The vibration data was simulated using the Runge-Kutta method for the 4-DOF numerical model (Rao, 2004). The initial conditions (displacement and velocity) were set to randomly generated values of reasonable order in the simulation. There was no excitation in the model and hence the results in this section are for free vibration of the structure under initial conditions. The structure was assumed to behave linearly with Rayleigh Damping.

Fig. 5a is the free vibration signal simulated from the 4 DOF signal when there is no damage in the structure. It can be noted from Fig. 5(a) that the vibration signal dies out after some time due the presence of damping. The plot contains the displacement measurement of the 1st DOF (the lowest floor of the building). Fig 5b presents a zoomed in plot of Fig. 5a.

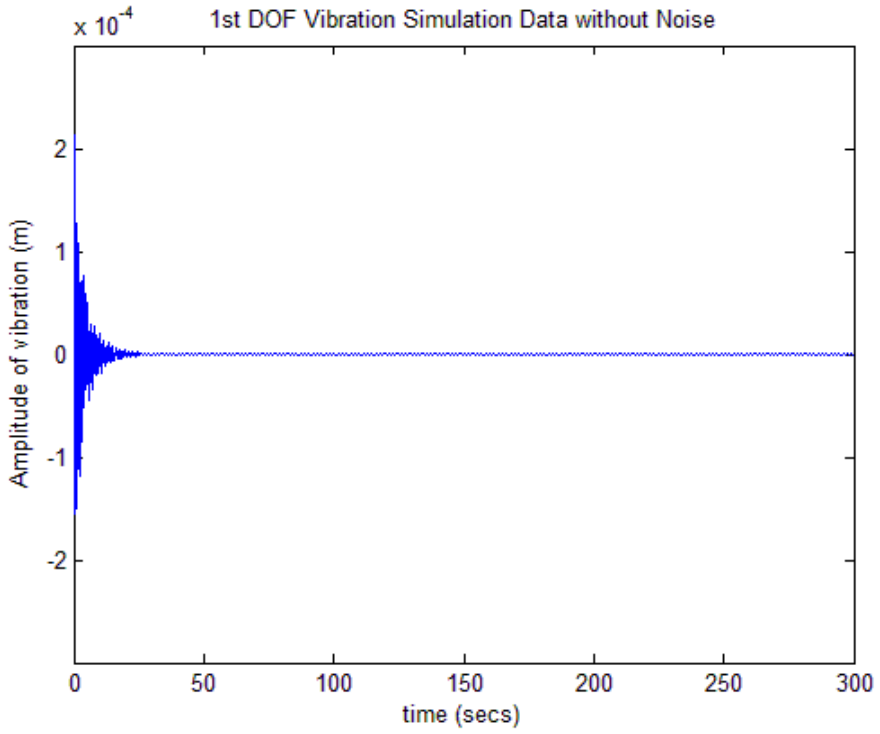

Fig. 5a: Free-Vibration simulation response for the 1st-DOF with random initial conditions

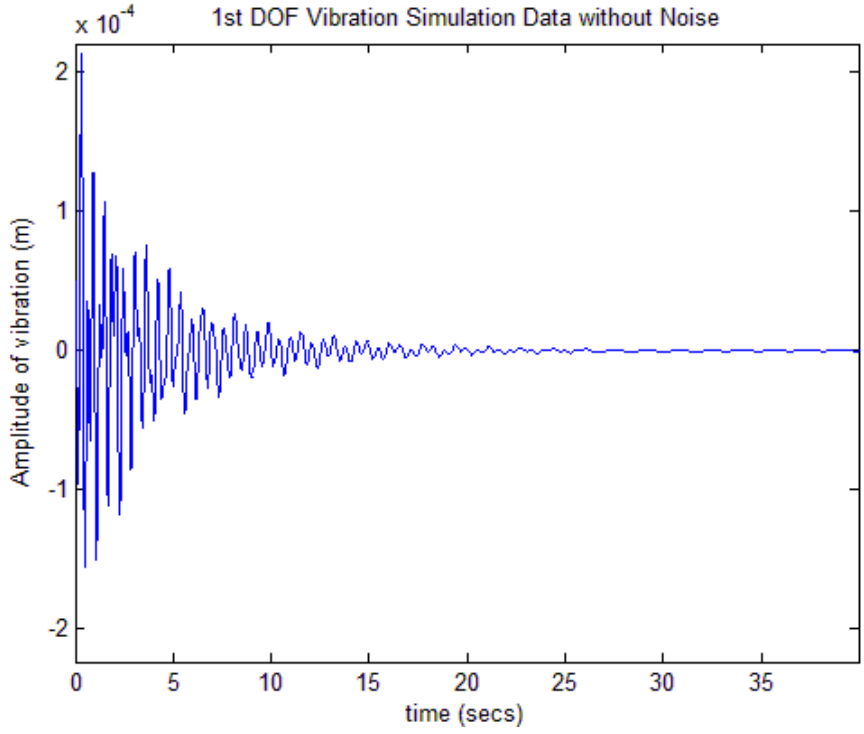

Fig. 5b: Zoomed in plot of Fig. 5a

2.2.1.3 Calculation of Fourier Transform of the vibration signal

For this study the first two natural frequencies of the system have been used as input features for the SVM technique and constitute the feature space for the SVM classifier. To extract the natural frequencies of the system, the Fourier Transform of the simulation signal is computed by using the FFT technique. Fig 6 shows the Fourier amplitude spectrum of the vibration simulation from the 1st DOF of the model. It can be noted that the first two natural frequencies can easily be identified from the plot. A function was written in MATLAB to automate this process.

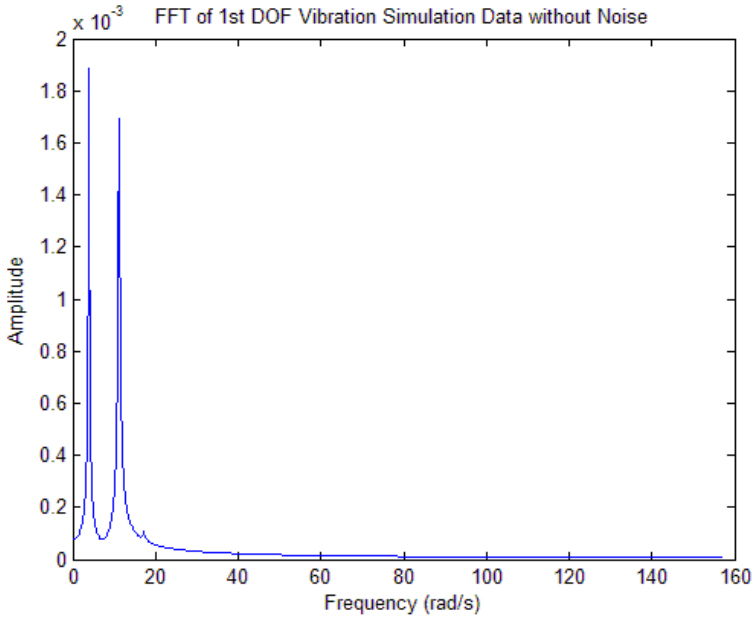

Fig. 6: Fourier amplitude spectrum of the vibration simulation of the 1st-DOF

The length of time signal (seconds) for the simulation was 600 seconds with a time step of 0.02 seconds. The corresponding sampling frequency was 50 samples/second (=1/0.02). The frequency interval was0.00167 Hz (= 1/600). The Nyquist Frequency was 25 Hz (=sampling frequency/2).

When extracting the natural frequencies from the signal, the readings of the first 5 seconds were removed from the time signal to reduce edge effects. The values of the first two natural frequencies form the two dimensions of each training point. Multiple data points are calculated in a similar way from successive vibration signals collected (simulated) from the structure. To account

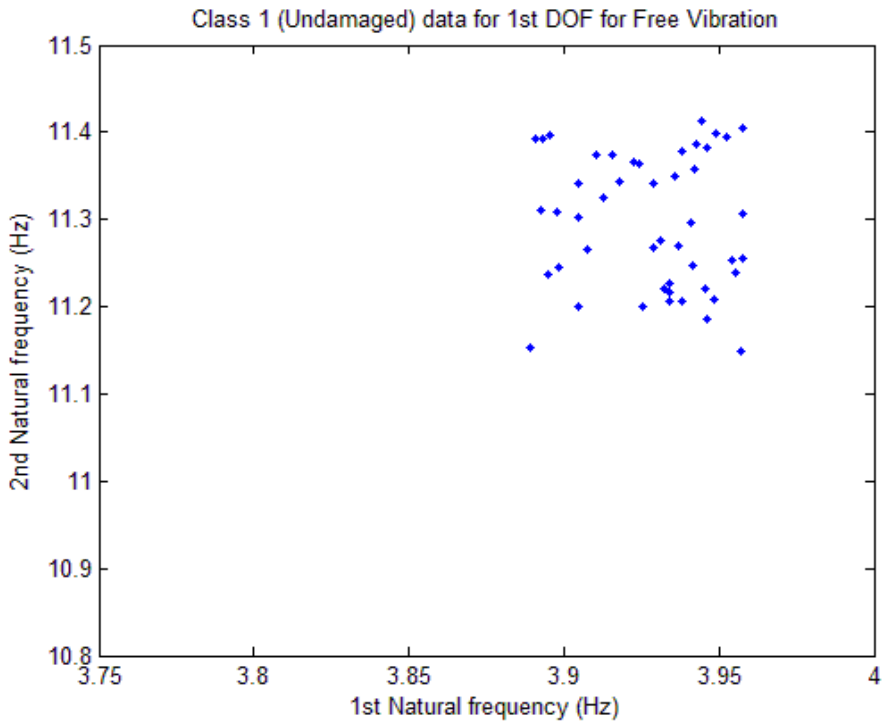

Fig. 7: Undamaged class of data extracted from the vibration simulation of the 1st-DOF of the structure

for modeling error in actual measurements, a random variation equal to 2% of the RMS value of collected natural frequencies was added to the readings. These points form the undamaged class of data. Fig 7 plots the undamaged class of data containing 50 data points. The 1st feature for each data point if the 1st natural frequency extracted from the vibration signal and the 2nd feature is the 2nd natural frequency.

2.2.2 Creation of the "Damaged" class of simulation data

As the SVM method is a supervised learning method, it requires data from both the undamaged structure and the damaged structure to be able to create a classifier during the training phase. In the proposed method we create the second class of training data, representing the damaged structure ("damaged" class), by creating a copy of the undamaged class (generated in the last section) and reducing the numerical values of data points in this copy by "some" percent. Fig 8a plots both the undamaged (blue) and "damaged" (red) classes of data for the 4-DOF model in the same figure.

The motivation behind choosing a particular value of percentage reduction is to be able to create two separable sets of data (damaged and undamaged class). A detailed analysis for this choice is included in section 3.2.3. For now, we will assume that the reduction of the values of the undamaged class to create the damaged class is 2%.

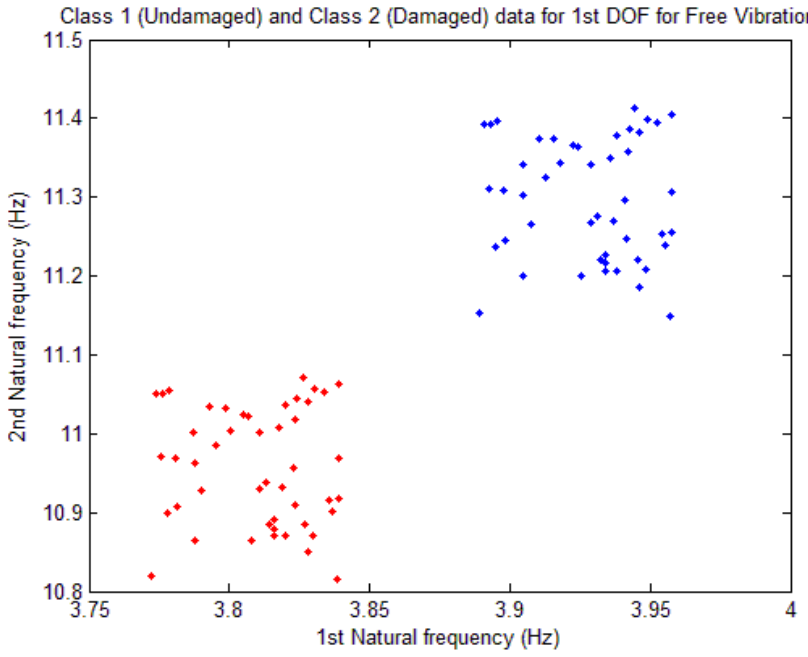

Fig. 8a: "Undamaged" (blue) and "Damaged" (red) classes of data for the training dataset

The new dataset created by reducing the frequencies can be assumed to represent the damaged "class" of the training dataset since it is well known that damage weakens a structure and in turn generally reduces its natural frequencies. As an example, Fig 8b shows the percent reduction of the first two natural frequencies for different levels of damage in a structure model later used in this study.

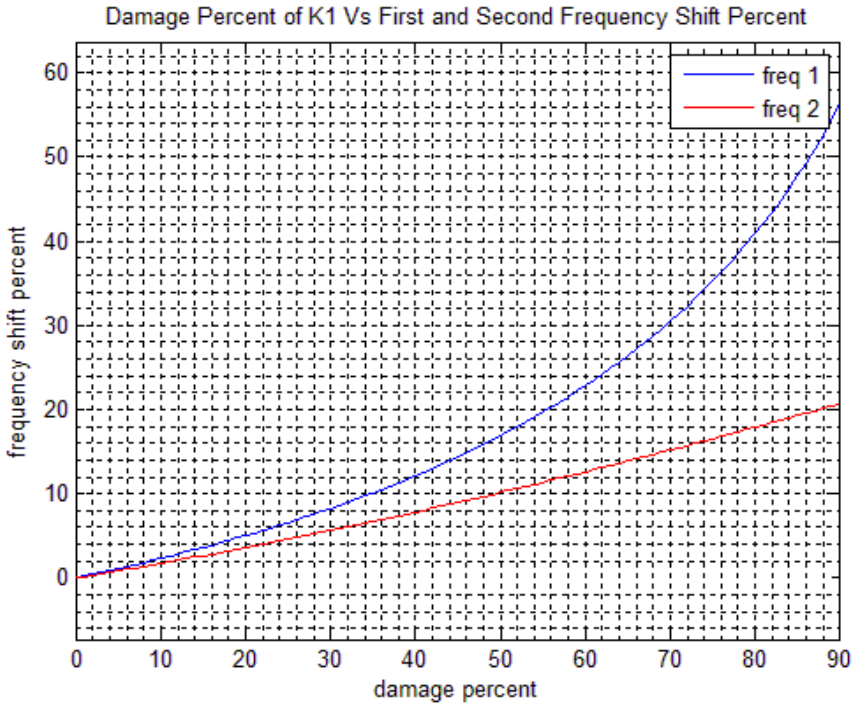

Fig. 8b: Percent reduction of the first two natural frequencies of the structure with an increase in damage in K1.

The reduction of frequency values of the undamaged structure to create the damaged class can be viewed as measuring natural frequencies from a damaged structure.

2.2.3 Development of the SVM classifier

Once the training dataset has been created with data from the undamaged and the damaged class, we need to find the classifier that can separate them. Before we can calculate a solution to the optimization problem an important question that still remains to be answered is what functions can be chosen as kernel functions. The Mercer's condition (Courant, 1953) states that:

There exists a mapping φ and a function

$$K(x, y) = \varphi(x).\varphi(y) \tag{44}$$

if and only if, for any f(x) which has a positive L_2 norm, or $\int f(x)^2 dx$ is finite, then

$$\int K(x, y) f(x) f(y) dx dy \geq 0 \tag{45}$$

As it can be seen, checking the Mercer's condition is not always a trivial task. For the dual form of the optimization problem to be positive definite the kernel function should satisfy the Mercer's condition. Any function that satisfies the MERCER's condition represents a dot product in some higher dimensional space *H* through a function φ and can be used as a kernel function.

The types of kernel functions that were used for the simulations in this thesis are:

a) Linear Kernel:

$$K(x, y) = x.y \tag{46}$$

The only parameter required to be supplied to the SVM for training was the regularization parameter "C" in equations (18) and (34). The regularization parameter is a user selected cost associated with incorrect classifications.

b) Polynomial Kernel:

$$K(x, y) = (x.y + a)^p \quad (47)$$

When using this kernel in our simulations, we chose the parameter a = 0 for convenience. Both the order of the polynomial, p, and the regularization parameter of the SVM "C" needed to be selected while training the SVM using the polynomial kernel.

c) Radial Basis Function Kernel (RBF):

$$K(x, y) = exp(-0.5(x - y)/\sigma^2) \quad (48)$$

The parameters to be selected while training were the "width", σ, and the regularization parameter of the SVM "C". It was found the training process was very sensitive to the parameter "σ". Choosing a very small σ always lead to perfect classification (zero classification error). However a very small "width", σ, created a model that was over fitted to the training dataset. As a result, the trained SVM was able to learn the training dataset correctly but had poor generalization performance. In other words, if the dataset to be classified was different from the training dataset, the error of classification was high. If the value of σ was too high, the curvature for the SVM trained was very low and the SVM trained was similar to the linear SVM. Such a model is called to be over-generalized and leads to high classification errors both while training and testing.

d) Sigmoid Kernel:

$$K(x, y) = tanh(kx.y - \delta) \quad (49)$$

We chose δ =0 to for convenience (reduce iterations for parameter selection). The parameters to be selected during training were the parameter k and the regularization parameter of the SVM "C". The hyperbolic tangent kernel only satisfies the Mercer's condition for a small subset of values of the parameters k and δ.

Once the SVM parameters have been selected, calculating the SVM boundary amounts to solving the optimization problem in equation (32) - (35). To efficiently solve the optimization problem, any implementation of the SVM algorithm should proceed in the following manner to converge to a solution:

- Break data into manageable sizes
- Find a method to continuously and uniformly increase the dual objective function subject to the constraints
- Checking that the KKT conditions are satisfied at the solution

For this study I used the MATLAB toolbox "STPRTool" developed by Vojtech Franc at the Czech Technical University in Prague. This toolbox uses the Sequential Minimal Optimization (SMO) technique (Platt, 1998) to break a large training dataset for the SVM into smaller chunks that can be solved analytically and proceed towards a solution to the optimization problem defined in (33)-(38). There are two parts to this algorithm, one choosing which two Lagrangian multipliers to optimize in each step and another performing the analytical optimization. The two Lagrangian multipliers to optimize are selected by different heuristics. The first point selected is one that violates the KKT conditions the worst. After the selection of the first point, the second point is selected so that an updating of the two points would lead to a large increase in the dual objective. It should be noted that the method optimizes a subset of only two points at each step. The solution to the optimization problem of 2 points can be obtained analytically (Cristianini, 2000). The strength of this method lies in using the analytical solution and hence avoiding a resource intensive numerical optimization problem in the

inner loop of the algorithm. As this algorithm does not require the complete training set to be present in memory for matrix computation during the optimization stage, the memory required for the SMO scales between linear and quadratic in the training set size for test problems. For a more detailed analysis of the SVM implementation using the SMO algorithm the reader is referred to Platt (1998), Cristianini (2000) and Franc (2004).

2.2.4 Creation of the testing dataset

We choose the "best" SVM classifier for the damage detection problem by testing the performance of classifiers trained of different SVM parameters on various training sets. Each classifier is tested by evaluating the accuracy of prediction of the SVM when used to classify the test dataset. The test data set is used at two stages in the methodology proposed. First, when the best SVM kernel and its parameters are being chosen and later when we recheck the value of the percentage shift that was assumed initially for creating the damaged class in the training stage. The procedure for creating the testing dataset is the same in both cases. The points in the testing dataset, like the training dataset, have the values of their two dimensions equal to the first two natural frequencies of the structure. The testing dataset consists of data points from the undamaged structure (which can be measured) and data points from the damaged structure. The data from the damaged structure needs to be simulated from a model of a structure by reducing the stiffness of elements in the structure. From the simulated vibration signal, the natural frequencies are extracted as earlier, using the FFT, to create the damaged testing class of data. Datasets are created for different levels of damage in

the structure. The accuracy of the SVM classifier to detect different levels of damage is then evaluated by classifying these datasets.

As the class of each data point in the testing dataset is already known, the target labels (y_i) can be assigned to the points accordingly. Once the SVM is trained on the training dataset, it is used to classify the test dataset. The known labels for a each point in the test set are compared against the predicted labels for each point to evaluate the performance of the trained SVM on the test set.

3 Results

3.1 Application of SVMs for damage detection

For the selection of the best SVM kernel and its parameters, the performance of a range of parameters for each kernel was evaluated on the test dataset. A classification by the trained SVM is called an error if the target class it predicts differs from the known label of the point. For a given number of test points, the Error Rate is evaluated as the ratio of the number of points incorrectly classified in the testing set to the total number of input test points.

The error rate indicates the ability of the SVM to predict damage in the structure. For example, consider a test set consisting of 50 points with “known” output Class 1 (undamaged signal) and 50 points with “known” output Class 2 (damaged structure signal). Note that the testing set also includes the “known” target classification (“Class 1” or “Class 2”) specified along with the input data. When the trained SVM was used to classify this test dataset, if the SVM predicted the output classes of 3 out of the 100

total points differently from the “known” classes, the SVM error rate on the dataset would be would be 3/100 = 0.03.

In the damage detection problem without any measurement noise, the best SVM kernel was found to be the linear kernel and the regularization parameter “C” parameter was selected as 3000.

Once the SVM parameters have been chosen, these parameters can be used to optimize the shift percent values that were used in the training stage of the SVM to create the damage class. With the best SVM parameters and the best shift percent, the best SVM classifier can be determined for the classification of structures into damaged and undamaged. Fig. 9 shows the SVM that was calculated given the training dataset created. Each class of training data consisted of 50 training points. It can be seen that the SVM calculated is able to clearly separate the training datasets into the undamaged (blue) and the damaged (red) classes. After training, prediction of the health of the structure only involves finding which side of the decision boundary the test point lies. It should be noted that as in this study only 2 natural frequencies of the structure were used, the analysis was performed in a 2-D feature space and the visualization of the linear SVM boundary was not difficult. For a higher dimensional data feature space, visualization of the boundary could be more complicated.

The trained SVM was tested by classifying a test dataset in which the damaged class was created from a model of the structure with 10% reduction in stiffness of K1.

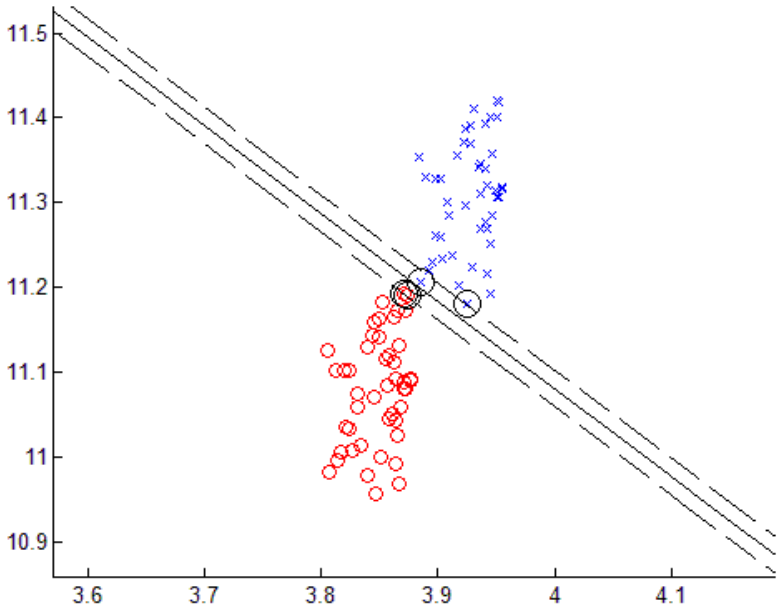

Fig. 9: Best classifier that was determined using the linear kernel and 3% shift for creating the damage class during training. The data shown above is the training dataset.

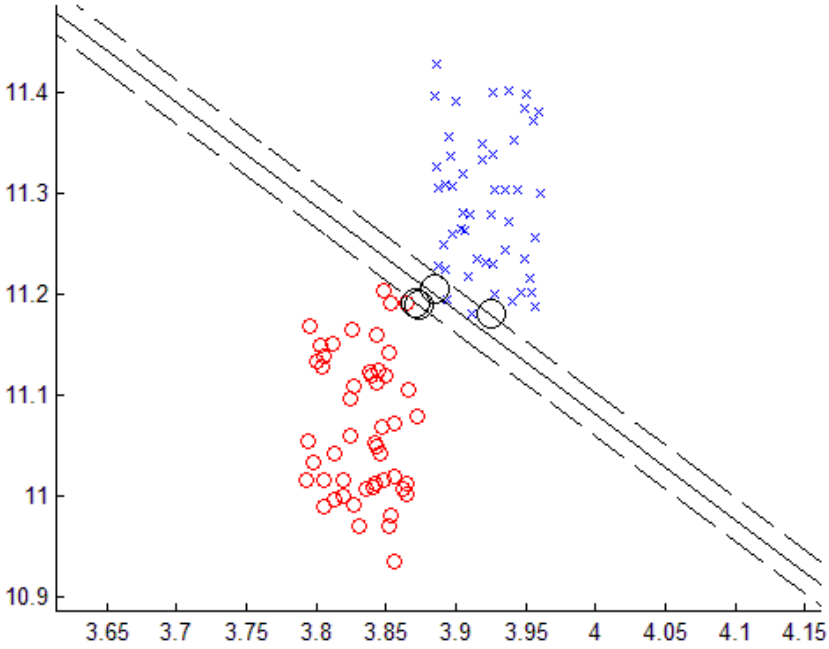

Fig. 10: Classification results when the SVM boundary shown in Fig 9 was used to classify the testing dataset. The damage class of the testing dataset had 10% reduction in stiffness of K1

It can be seen that the trained SVM is able to classify the testing dataset for 10% damage in K1 with 100% accuracy.

3.2 Main issues faced:

During the evaluation of the SVM, there are three variables that need to be checked to find the best SVM:

I. The percent reduction of the undamaged dataset to create the damaged data

II. The kernel function to be used for the SVM and

III. The parameters of the kernel function chosen

When choosing the best kernel function during the training phase, a value of percent reduction of the undamaged class is assumed. After the kernel function and its parameters are selected, the choice of percentage shift for creating the “damaged” class is optimized. The effect of shifting percentages on the accuracy of the SVM has been explored later in this section.

As there are multiple kernel functions available, choosing the best kernel functions and it parameters is a challenge. The behavior of each kernel varies vastly based on its parameters chosen and in turn affects the SVM created using those parameters. An analysis has been done in the following section to illustrate the effect of kernel parameters on the SVM classifier.

The effect of measurement noise and different working conditions on the performance of the method is also investigated in a later section.

3.2.1 Effect of parameters γ and C on SVM created

After the two datasets (training and testing) are available, we decide the best kernel and its parameters by evaluating the performance of the SVMs for a range of parameters with each kernel function when used to classify the testing set.

In the following section we investigate the effect of the kernel parameter (γ) and regularization parameter (C) on the SVM boundary finally created. Though in the cases that follow, the effect of kernel parameters (γ and C) has been demonstrated on the SVM boundary when using the RBF kernel, a similar effect is observed for the polynomial and sigmoid kernels also. The dataset that has been used in this section was artificially created to illustrate the concept.

3.2.1.1 The kernel argument (γ)

This parameter has different interpretations for the polynomial, RBF and the sigmoid kernels but affects the final SVM created using each in a similar manner. This parameter has no effect on the linear classifier. The exact formula for each kernel function is mentioned in section 2.3.3. For the polynomial, RBF and the sigmoid kernel "γ" affects the curvature of the SVM calculated. To demonstrate the effect of the change in "γ" on the SVM, we compute SVMs for the RBF kernel with different values of "γ" and a constant reasonable value of the regularization parameter "C" (=100). The effect of the parameter "C" on the SVM has been explored in a later section.

a. Case 1: Very low "γ" (0.01) , Reasonable C (100)

In this case the RBF kernel parameter was 0.01. The number of kernel evaluations to compute the SVM was 294220, which is quite high. There were 179 support vectors (=number of training points) created during the training. As the number of support vectors was equal to the number of training points, each point in the result is a support vector (circled points in Fig. 11). The width of the margin was 0.0752units. There was no training error (=0.00%). The training error is calculated as the error of classification of the training set with the trained SVM. A 0% training error means that the SVM was able to learn the training set.

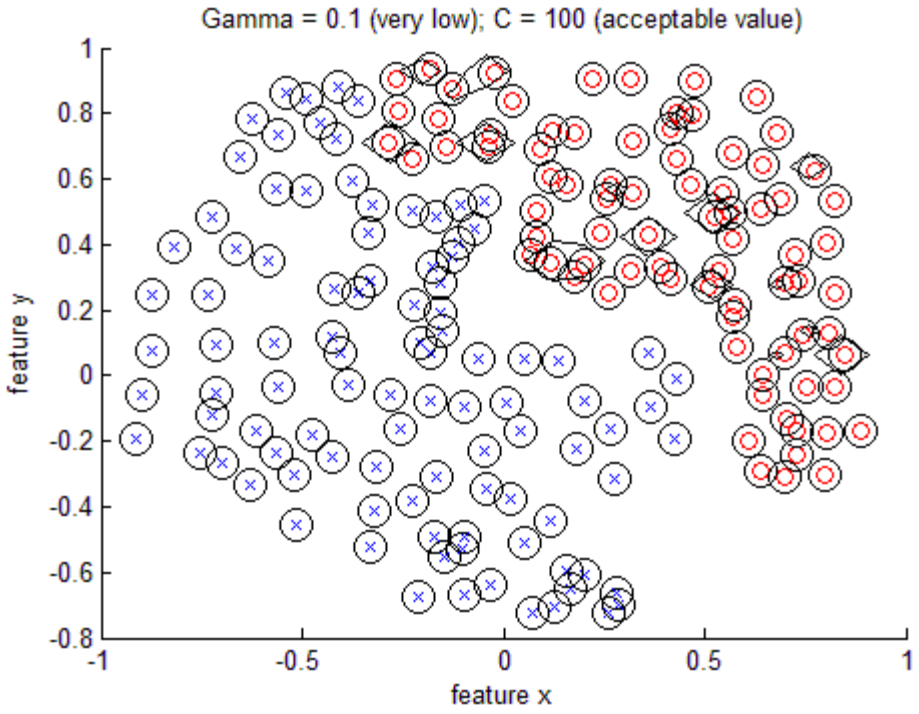

Fig. 11: Trained SVM with reasonable value of "C" (=100) and low value of "γ" (=0.1)

Here the value of "γ" is very low. For the RBF kernel this parameters is the "width" of the radial basis function centered at each training point. The low width of the kernel leads to each point being learned by the algorithm. This would allow the SVM to learn

the training set completely and hence return 0% error when the trained SVM is used to classify the training set again. However, this is a highly overfitted SVM and would result in bad performance during the test phase for any set of data different form the training set. The time taken for the training was 0.2417 seconds.

b. Case 2: Reasonable "γ" (2), Reasonable C (100)

The value of the kernel parameter was reasonable (=2) in this case. The number of kernel evaluations to compute the SVM was 39382 with a computation time of 0.0349 seconds which are both lower than case (a). There were 22 support vectors created during the training phase. As can be noticed this was much lower than the number of training points (circled points in Fig. 12). The width of the margin was 0.0403units, which is lesser than the case with a very low value of "γ". The training error was 3.35%, which is more than case (a) as expected. However it should be noted from Fig. 12 that the training leads to a boundary that is able to classify the two datasets.

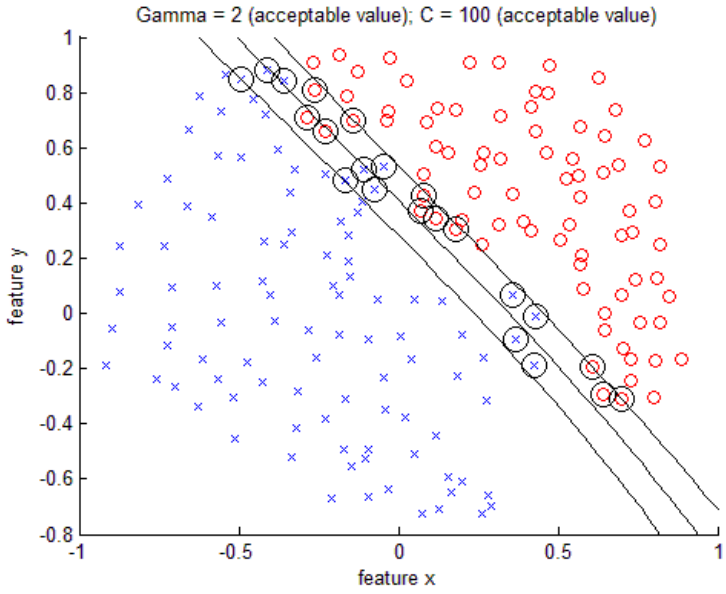

Fig. 12: Trained SVM with reasonable value of "C" (=100) and reasonable value of "γ" (=2)

c. Case 3: High "γ" (10), Reasonable C (100)

In this case we assume a "high" value of the parameter "γ".

The value of the kernel parameter was high (=10) in this case. The training required 97784 kernel evaluations with a computation time of 0.1530 seconds.

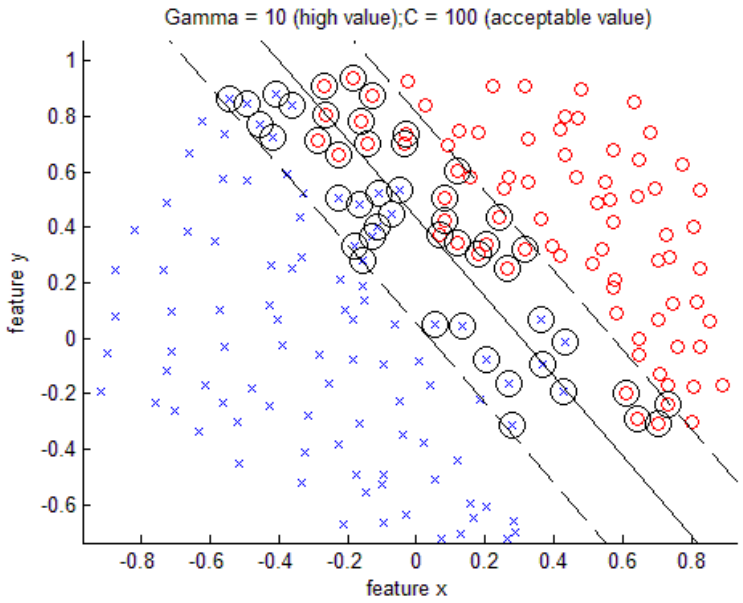

Fig. 13: Trained SVM with reasonable value of "C" (=100) and high value of "γ" (=10)

The number of kernel evaluations and the time taken for the training phase is more than for case (b). There were 48 support vectors created during the training phase. This is expected as the width of the margin was larger than the margin in case (b) and more points lie within the boundary (Fig. 13). The training error was 3.91%, which is more than case (b). This is because the boundary is very straight and hence is over-generalized. This will increase the error at the margins when this SVM boundary is used for classification of the testing dataset.

From the above study it can be seen that too low a value of "γ" or too high a value would lead to poor generalization performance (high error rate) of the SVM classifier.

3.2.1.2 The regularization parameter (C)

The regularization parameter "C" is introduced into the SVM methodology to associate a cost for every incorrect classification during the optimization and to limit the effect that an incorrect classification can have on the final boundary. A high value of "C" would reduce the number of misclassifications in the final SVM. The parameter "γ" majorly affects the curvature of the SVM calculated; the parameter "C" controls the thickness of the margins on both sides. A high value of "C" would lead to a thinner margin on both sides and a slightly more curved classifier. A very low value of "C" would lead to a very broad margin on both sides of the SVM.

We show the effect of the parameter "C" on the SVM calculated in the following section.

a. Case a: Reasonable "γ" (2), Low C (10)

In this case the regularization parameter was 10.

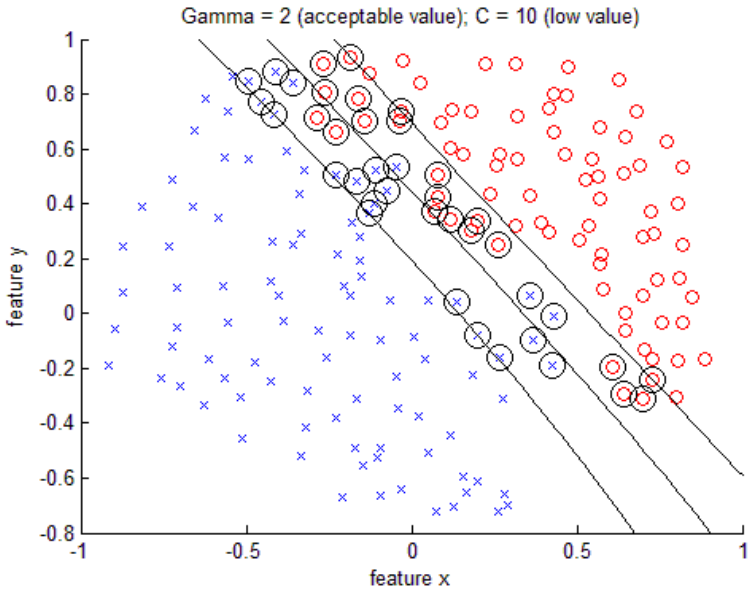

Fig. 14: Trained SVM with low value of "C" (=10) and reasonable value of "γ" (=2)

The number of kernel evaluations to compute the SVM was 469837, which is quite high. There were 39 support vectors created during the training and the margin was 0.0758 units. The training error was 2.79%. The time taken for the training was 0.1148 seconds, which is comparatively high due to the large number of kernel evaluations. The classifier boundary has low curvature.

b. Case b: Reasonable "γ" (2.2), Reasonable C (100)

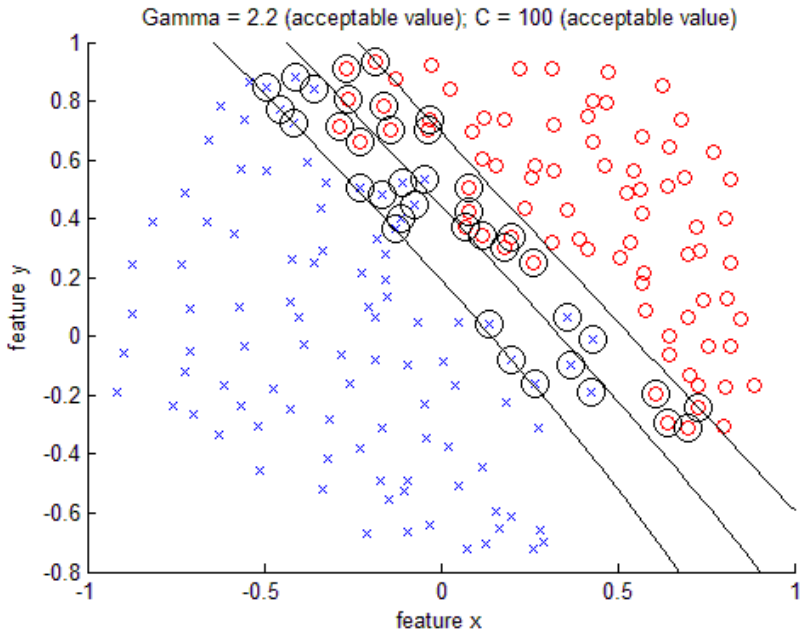

Fig. 15: Trained SVM with reasonable value of "C" (=100) and reasonable value of "γ" (=2.2)

In this case the regularization parameter was 100. The number of kernel evaluations to compute the SVM was 41590, which is lower than the case (a). There were 23 support vectors created during the training and the margin was 0428 units. Both these numbers have reduced as an effect of increasing the value of "C". The training error was still 2.79% because the curvature of the SVM has not changed and then number of

classifications at the boundaries remains the same. The time taken for the training was 0.0284 seconds, which is lower than case (a) since the number of kernel evaluations has decreased. The classifier boundary still has low curvature.

c. Case c: Reasonable "γ" (2), High C (1000)

The regularization parameter was selected as 1000 in this case. The number of kernel evaluations to compute the SVM was 152317, which is lower than the case (b). Increasing the value of "C" reduced the margins and resultingly the number of support vectors. The training error was still 2.79% though there is a slight increase in the curvature of the SVM. The time taken for the training was 0.0417 seconds, which is higher than case (b) since the number of kernel evaluations has increased.

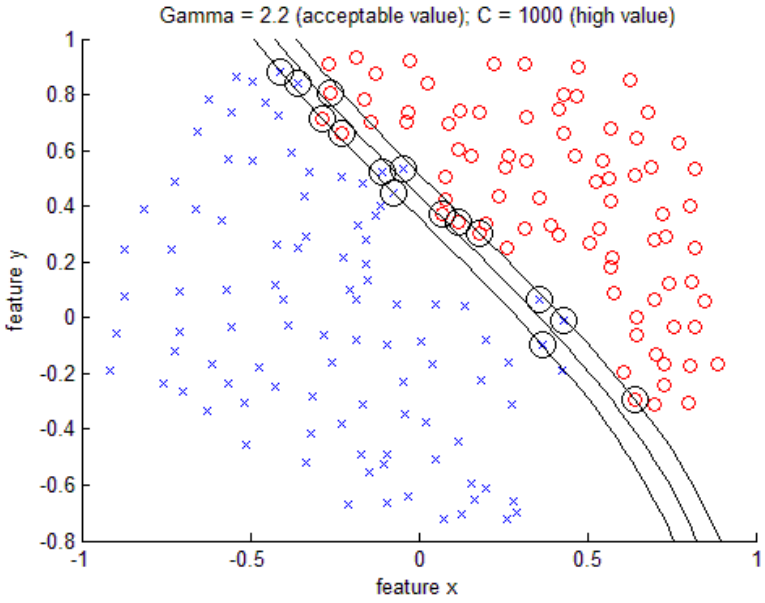

Fig. 16: Trained SVM with high value of "C" (=100) and reasonable value of "γ" (=2.2)

It can be concluded from the above discussion that reasonable values of "C" and "γ" lead to the best performance for the SVM with respect to the times taken for computation and the generalization performance of the classifier.

3.2.2 Selection of the best SVM kernel and its parameters

To select the best kernel and its parameters for the classification problem, SVM boundaries were calculated for different combination of "γ" and "C" values. The error rate was computed for the classification of a test set by each of these SVM classifiers. The test dataset consisted of the undamaged class and damaged class of data. The damaged class was created by extracting the natural frequencies from the simulation response of a model of the structure with reduced stiffness in K1 (10% reduction). Fig. 17 – Fig. 20 plot the error rate for each kernel functions over a range of selected parameters values.

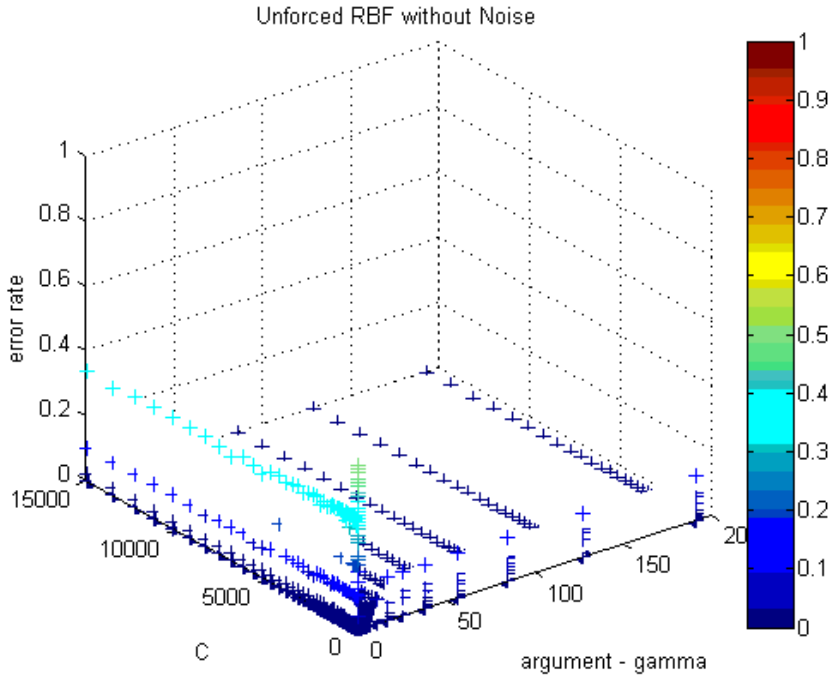

Fig. 17: Error rates when the SVM classifier, calculated with the RBF kernel function on a range of "C" and "γ" values, was used on the testing set.

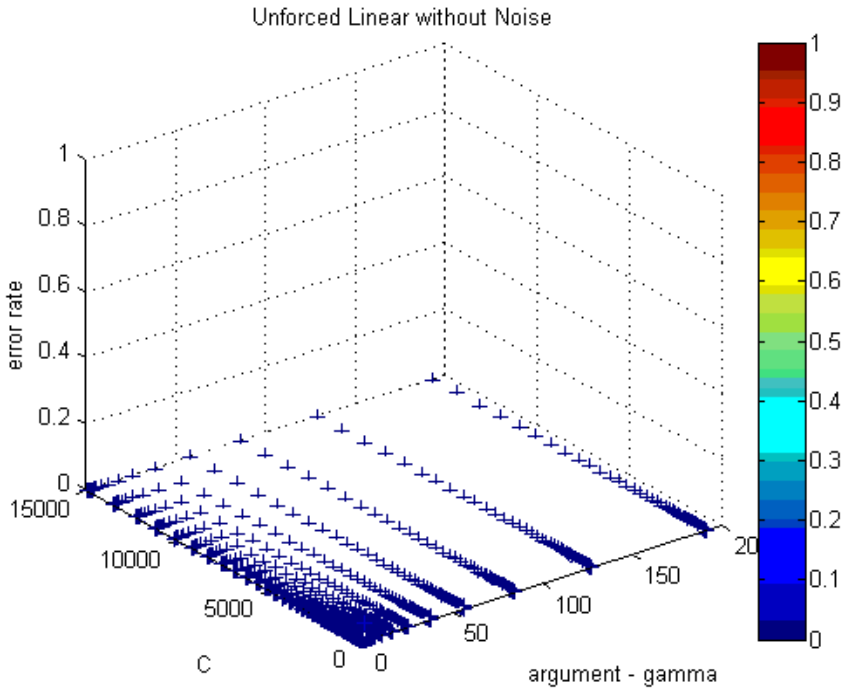

Fig. 18: Error rates when the SVM classifier, calculated with the linear kernel function on a range of "C" and "γ" values, was used on the testing set

From figure 17 it can be seen that the error rates are high for very small values of "γ" and "C". This is expected, from the last section, as a low value of "γ" leads to bad generalized performance and a small value of "C" does not penalize wrong classifications adequately.

Comparing figure 18 with Fig. 17 it can be seen that the linear SVM leads to a low error (=0) rate for all values of C. The parameter "γ" does not affect the linear kernel. This plot also points to the fact that the data in the testing set was separable with a linear boundary.

A look at Fig. 19 shows that the sigmoid kernel has a high error rate for all values of the kernel and regularization parameters and is not able to classify the classes. Fig. 20 shows that polynomial kernels are able to separate the classes in the testing dataset well for

low values of the parameter “γ”. The parameter “γ” in the polynomial kernel was used as the degree of the polynomial and a low value, close to one, would mean a near-linear boundary.

The linear SVM boundary takes the least amount of time to compute and performs well for the classification problem. Hence, the linear kernel with a value of “C” = 3000 was chosen in the simulations.

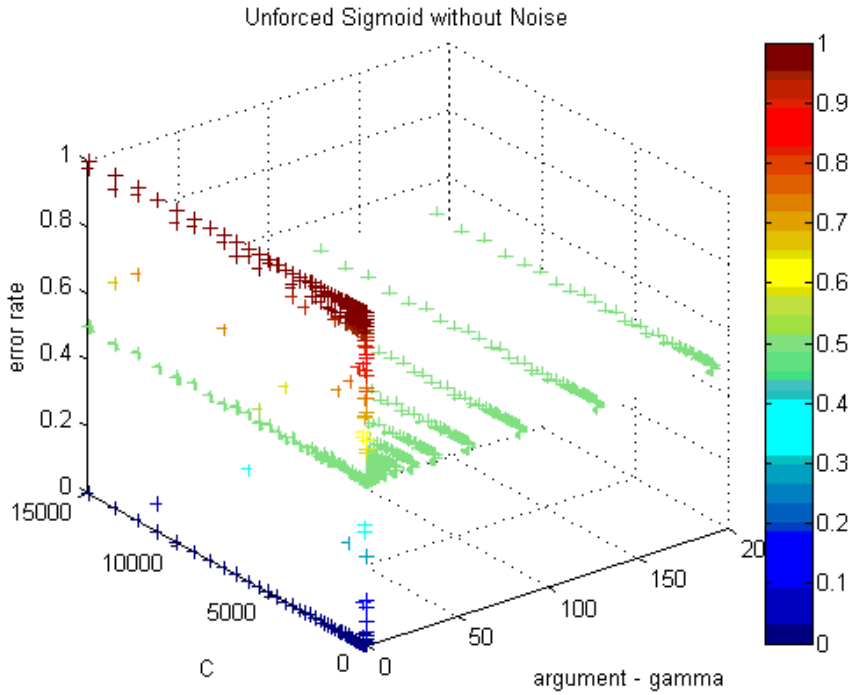

Fig. 19: Error rates when the SVM classifier, calculated with the sigmoid kernel function on a range of “C” and “γ” values, was used on the testing set

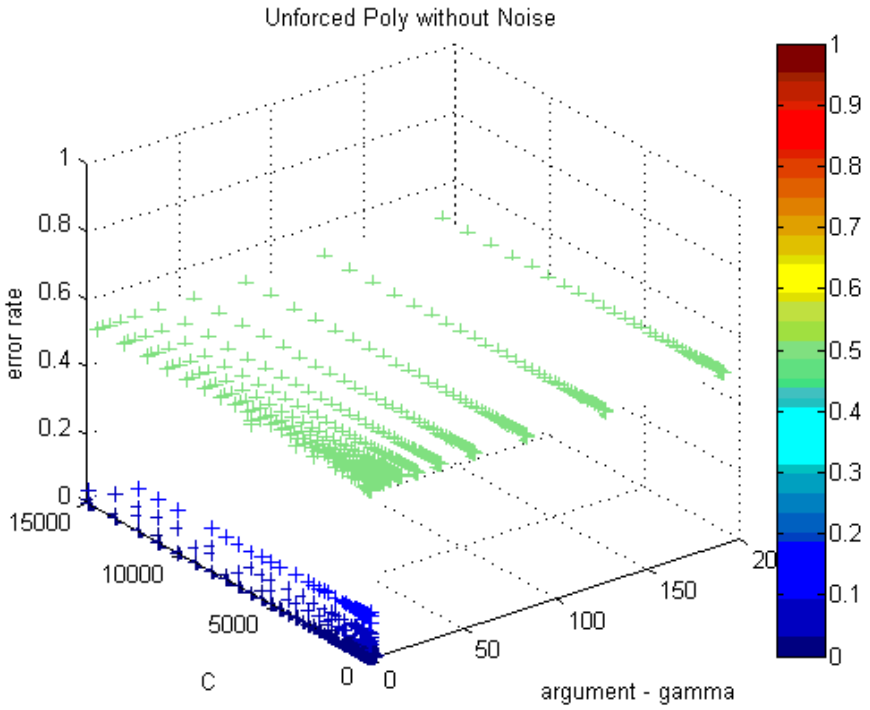

Fig. 20: Error rates when the SVM classifier, calculated with the polynomial kernel function on a range of "C" and "γ" values, was used on the testing set

3.2.3 Effect of percent frequency shifts in the training dataset

After the best SVM parameters have been selected for the given dataset, the percentage reduction of the undamaged training set during training that was used to create the damaged training set needs to be revisited. If the shift of the dataset is not adequate, both the sets of data will overlap and it will not be possible to create a boundary separating the two classes of data. On training, this will create a classifier boundary that passes partially through the undamaged dataset. As in the Fig. 21 below, though the test data point was derived from the undamaged class, it is classified as belonging to the damaged class during testing. Such a misclassification gives a false indication of damage.

Similarly, if the percent reduction of the undamaged training class is too high during training, there might be a huge gap between the two classes of data during training. The classifier boundary created from such a training set would classify data points from structures with small amounts of damage as belonging to the undamaged class. This would reduce the sensitivity of the method to detect damage.

To find the best shift percent we create the damaged class of the training datasets from different percent reduction values of the undamaged training class. SVMs are trained on each of these datasets and tested against the test datasets created using the simulation model. Each test dataset contains data from the undamaged structure and the damaged structure (percentage reduction of stiffness in the model). Multiple testing datasets with different damage levels in the model are created. Each trained SVM is tested on each test dataset to compute the error rate. This error rate indicates the ability of a particular percent shift to detect damage in the structure.

Results compiled in figure 22 below show the ability of the SVM trained on 2% percent shift (black) to detect damage levels as low as 4%. The 1% reduction line in the figure has a high error rate for all damage levels because as described earlier it misclassifies undamaged data as damaged due a very small percent shift. An increase in the shift percent increases the minimum damage that can be detected. From the figure we chose 2 % shift as the best shift for damage detection for this model.

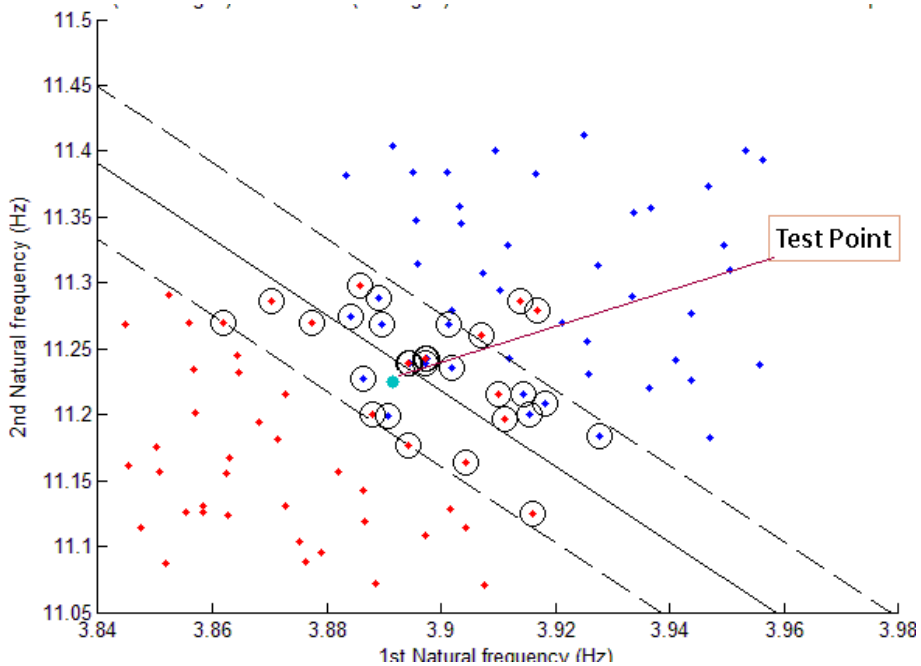

Fig. 21: SVM linear classifier calculated if the percent shift of natural frequencies in the undamaged dataset is too small during training

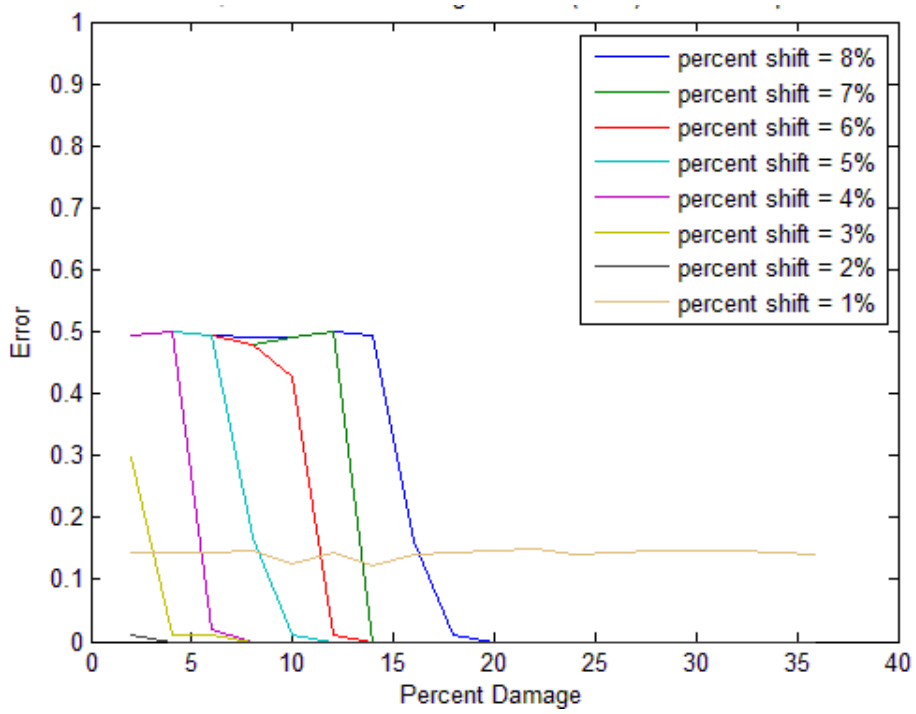

Fig. 22: Error rate versus percent damage for different percent reduction of the undamaged class during training. The error rate is calculated on a test dataset

3.2.4 Effects of measurement noise

Real world measurements are always contaminated by noise from different sources in the working environment. The equipment for vibration measurement also has its limitations on data recording precision. These sources can all add up noise to a significant level and can make the structural modal parameters difficult to extract from the contaminated vibration data.

In this section we introduce a cepstrum-like method to denoise the vibration signal and extract modal parameters from it. Fig. 23 displays the data from the damaged structure contaminated with measurement noise. The Signal to Noise Ratio (SNR) of the signal is 5. The signal to noise ratio is defined as the ratio of the RMS value of the meaningful signal to the RMS value of background noise.

I. Preliminary analysis of simulation data

a. Free vibration simulation data with measurement noise (SNR =5)

The responses of the 1st DOF of the model without and with measurement noise are displayed in Fig 23 (a) and (b) respectively.

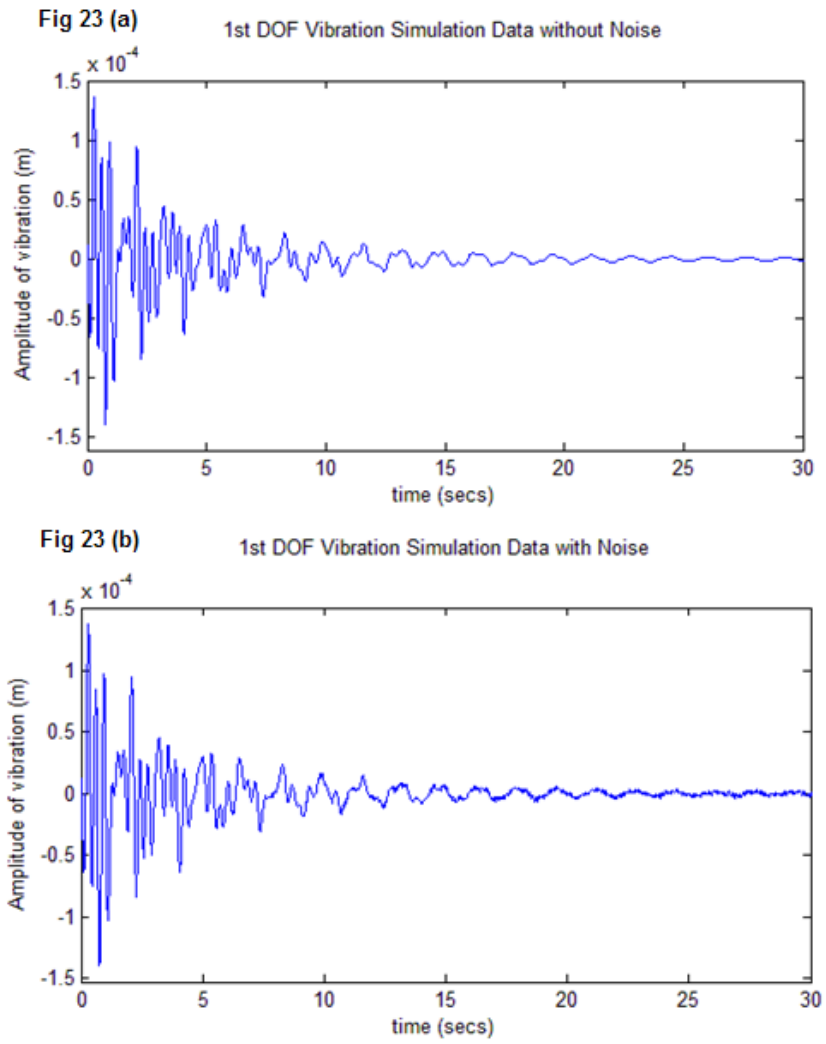

Fig. 23: Simulated free vibration response from 1st-DOF of model (a) without measurement noise (b) with measurement noise

b. Fourier Transform of the noisy signal

Fig.24 is the Fourier amplitude spectrum of the noisy signal in Fig 23. It can be seen that identifying the natural frequency peaks becomes a challenge in the noisy data. Errors in the extraction of the frequencies would affect the sensitivity of the damage detection method. To address this challenge, we require some method to denoise the data.

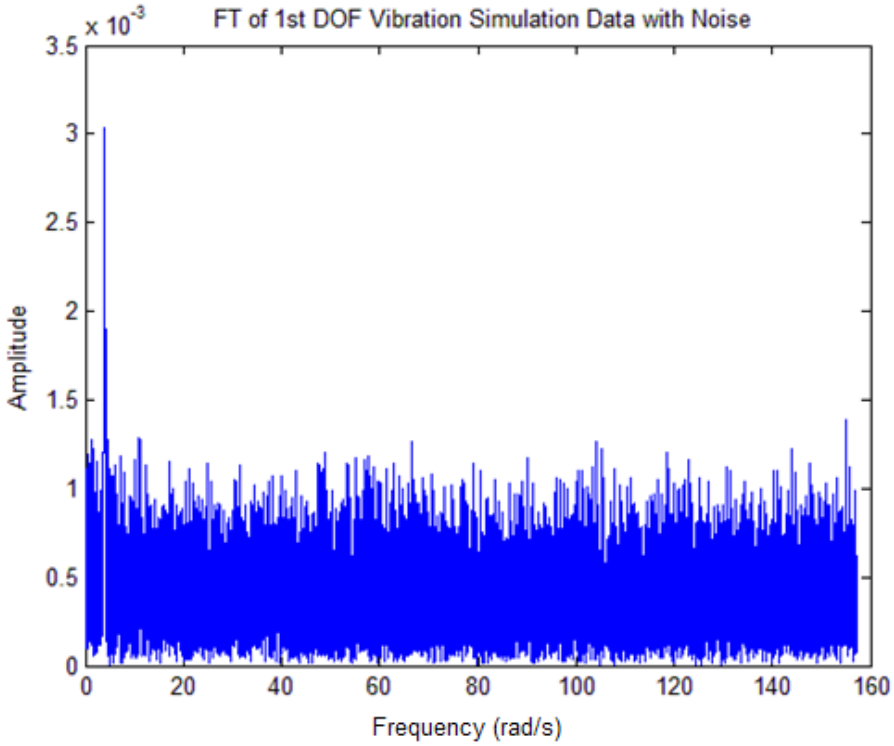

Fig. 24: Fourier amplitude spectrum of the noisy simulated vibration response from the 1st DOF of the model

2. Cepstrum-like technique to reduce the effect of noise from the Fourier amplitude spectrum of the signal

To reduce the effect of noise in the noisy signal and make the extraction of the natural frequencies easier, we remove the higher frequency components from the Fourier Amplitude Spectrum of the signal by the following procedure:

a. Calculation of the Fourier amplitude spectrum (FAS) of the Fourier amplitude spectrum of the signal: First the amplitude of the Fourier transform of the Fourier amplitude spectrum of the vibration signal is calculated (Fig 25).

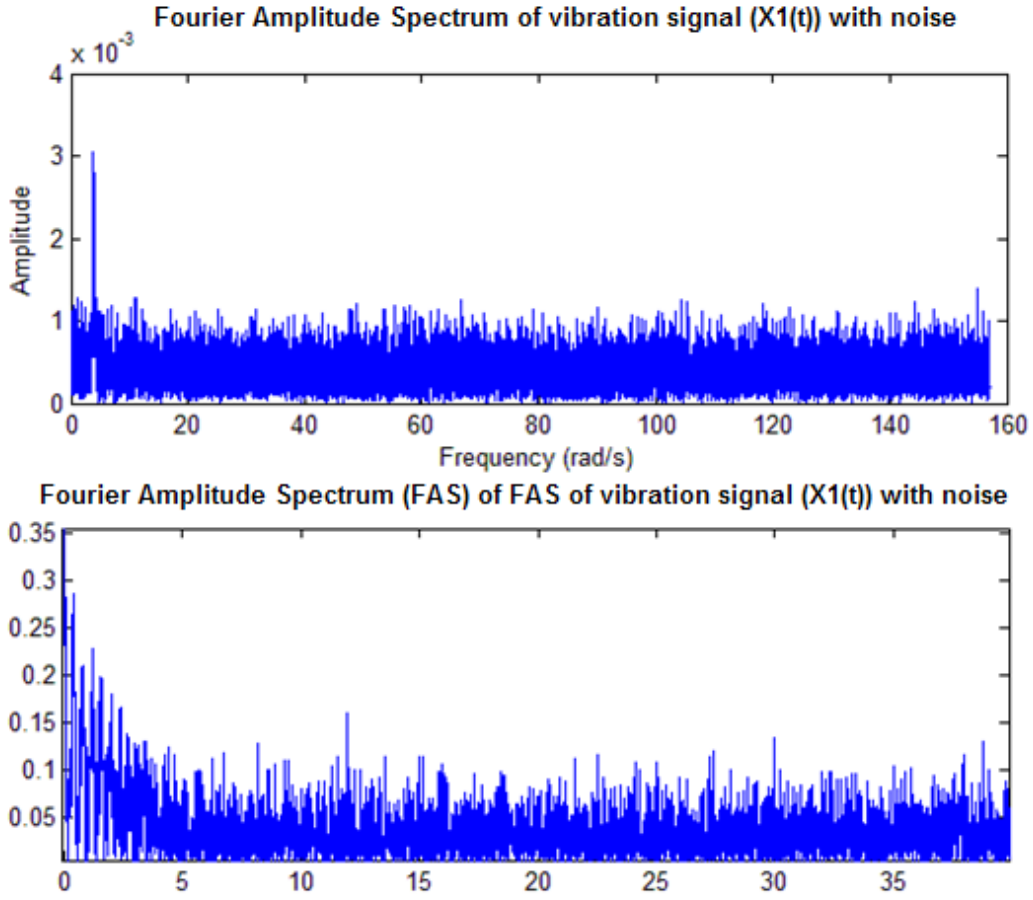

Fig. 25: Fourier amplitude spectrum of the Fourier amplitude spectrum of the vibration response signal of the 1st-DOF of model with measurement noise

b. Filter the Fourier amplitude spectrum of the signal by setting the points after first 500 points to zero:

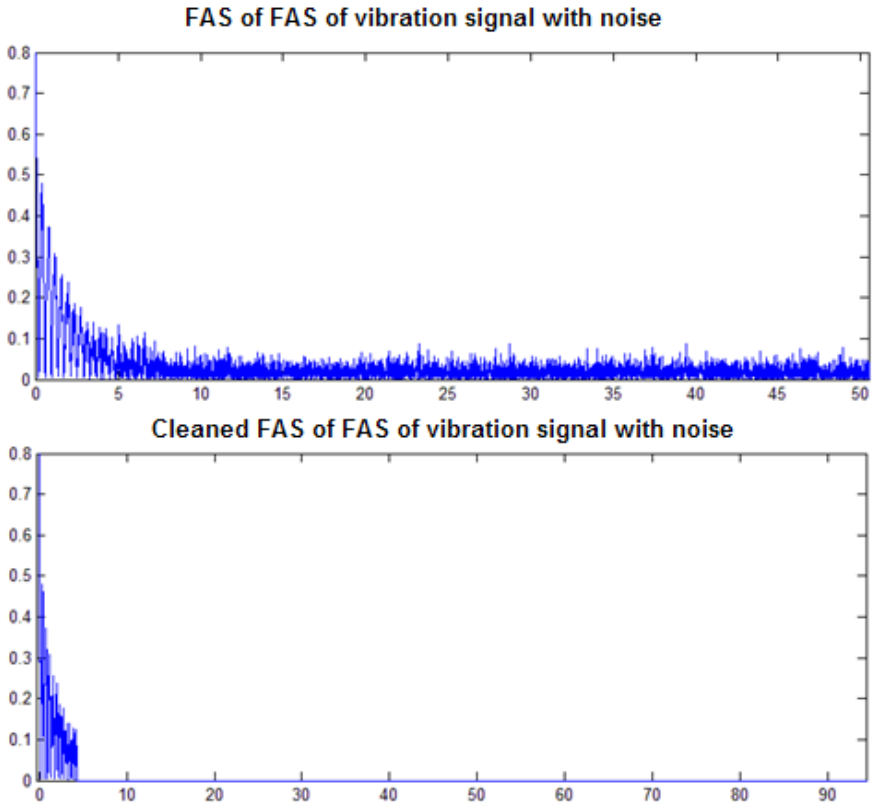

Fig. 26: (a) above: The Fourier amplitude spectrum of the Fourier amplitude spectrum. (b)below: All point after 500 have been set to zero in the Fourier spectrum of the Fourier amplitude spectrum. The plots shows the amplitude of the Fourier spectrum of the FAS.

To "clean" the noise from the Fourier amplitude spectrum, the values of the Fourier transform of the Fourier amplitude spectrum after the first few (=500 in this case) points are set to zero. This has the effect of removing the higher frequency components from the FT of the signal.

c. Take the Inverse FT of the cleaned FAS of FAS of the vibration signal and then used the recovered signal to extract the frequencies.

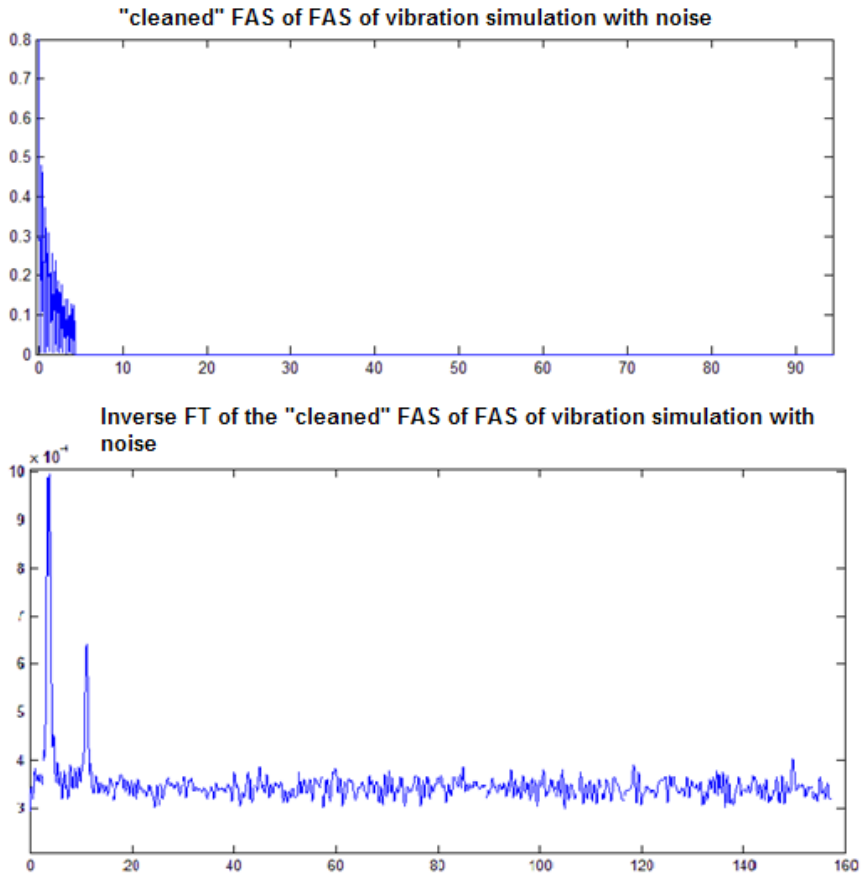

Fig. 27: (a) above: The Fourier amplitude spectrum of the Fourier amplitude spectrum with all points after 500 set to a value of zero (b) below: the inverse Fourier transform of 27 (a)

As can be seen from Fig. 27, the natural frequency peaks are easier to recognize from this cleaned signal as compared to the "unclean" FAS in Fig. 25.

3.2.5 Effects of external excitation

Outdoor structures are exposed to external forcing conditions like winds etc. which can be a major concern for vibration measurements. We introduce forces at the masses proportional to the wind excitation and test the ability of the method to detect damage.

3.2.5.1 Effect on method in the presence of base excitation

Structures like buildings, bridges etc. are exposed to ground based ambient vibration which forces them to vibrate. This ground based excitation can originate from micro tremors, microseisms and various local and periodic sources (like traffic, heavy machinery etc). An analysis of the change of the apparent natural frequencies of a 7-storey building over a period of time was provided in Trifunac et al. (2001a and 2001b). Ivanovic et al. (2007) discuss the ambient vibration survey of the Van-Nuys 7-story building and the ability of the survey to provide conclusive information about local damage. In this section we introduce base excitation into the model and test the ability of the method to detect damage in its presence. Ambient excitations have a very small magnitude (order of 10^-4 m/s^2). The ambient ground acceleration was modeled as white noise with a standard distribution of 0.0005 m/s^2. We first found the best kernel and its parameters for the simulation data with base excited vibrations. This was necessary as the presence of excitation changes the nature of the vibration data collected from the structure. The linear kernel and the parameters used in the earlier case worked satisfactorily for this set of data. Once the best SVM parameters were selected, we recheck the validity of the initially assumed value of percentage reduction of frequencies for the training of the SVM.

Fig. 28 shows that the SVM trained on 3% frequency reduction (shift) was able to detect the lowest damage (~6%). For our simulations we chose to reduce the undamaged class of the training dataset by 3 % as the best shift for detecting damage. The best curve for percent shift found in Fig. 28 has been presented in Fig. 29.

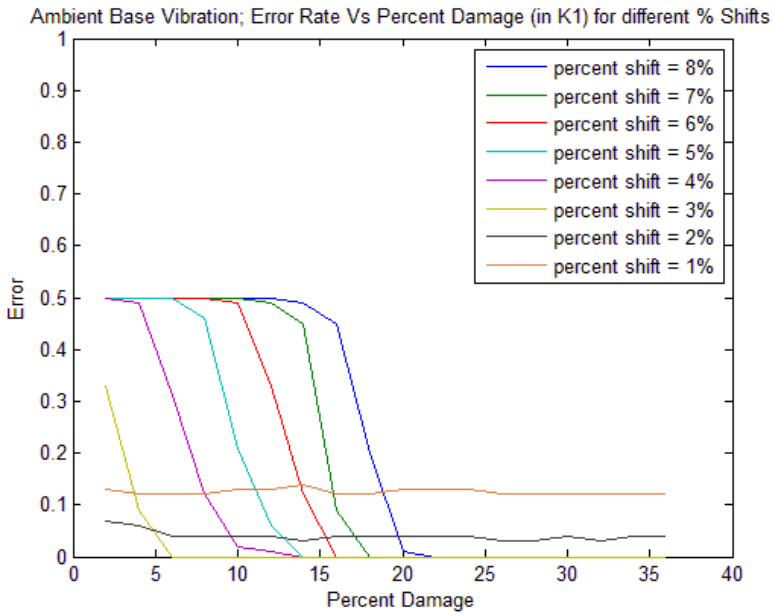

Fig. 28: Error rate versus percent damage for different percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under base excitation

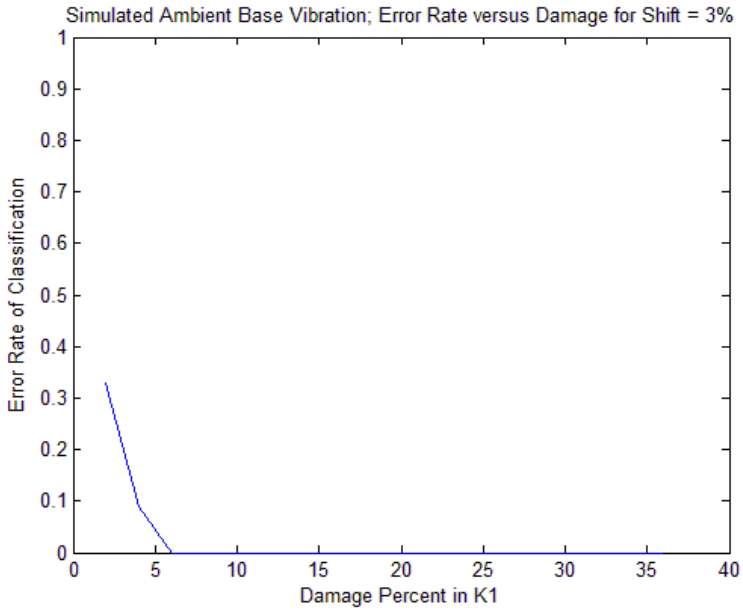

Fig. 29: Selected best percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under base excitation

Fig 30 shows the SVM trained on the chosen kernel parameters and the selected percent shift for damage detection. It can be seen that that the SVM created is able to classify data clearly into two sets.

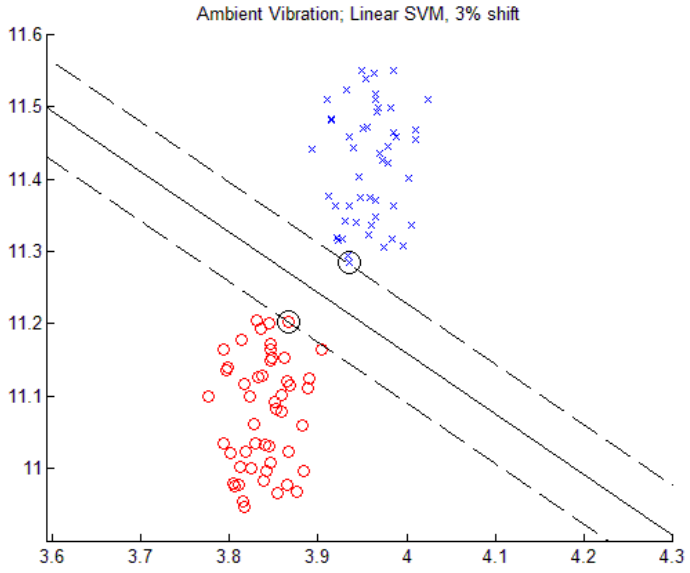

Fig. 30: The training SVM classifier selected using the model under base excitation

3.2.5.2 Effects of wind excitation on the damage

Wind loading can be an important factor to consider when making dynamic measurements in a structure. To test the performance of the damage detection method in the presence of wind loading, we introduce forces at each degree of freedom in the model. The wind load on the building has been modeled as having a static component and a dynamic component. Both components of the load increase from the 1st-DOF to the 4th-DOF. The dynamic component of the load has been modeled as a white noise. To apply the method for detecting damage, the best set of kernel parameters was chosen from the training dataset. The training dataset contained data from the undamaged

class and the damaged class (created with an assumed value of reduction of the training undamaged class equal to 3%). It was found that in the model wind loading introduces larger forces on each DOF of the structure as compared to forces created by base excitation.

An analysis of the SVM kernel and its parameters showed that the linear SVM had the best classification results when applied to the classification of the test dataset. The parameter C was chosen as: C = 6000. The parameter "γ" has no relevance for the SVM trained with a linear kernel.

With the kernel parameters decided, we plot the error rate for different percent shifts to find the best shift percent. In fig 31, and SVM trained with 3% frequency shift can detect damage as low as 10%. The curve for the selected shift percent (3%), decided from Fig 31, is displayed in Fig 32.

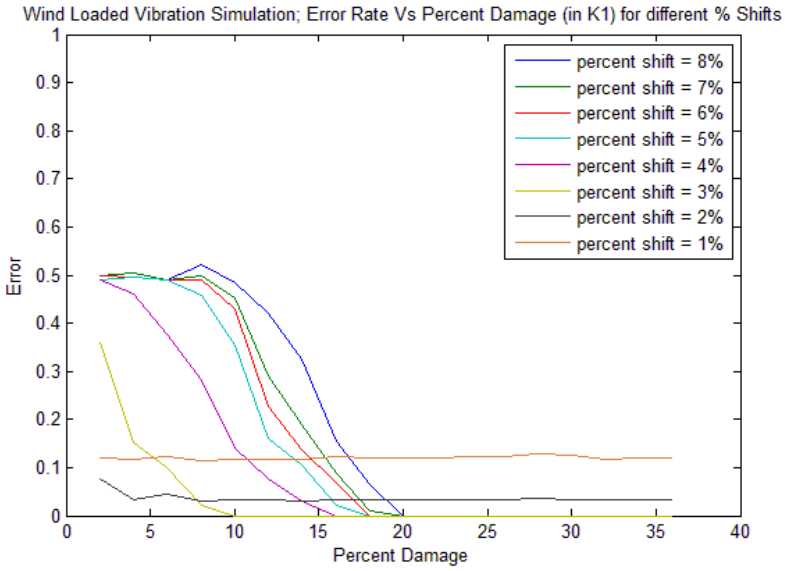

Fig. 31: Error rate versus percent damage for different percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under wind loading

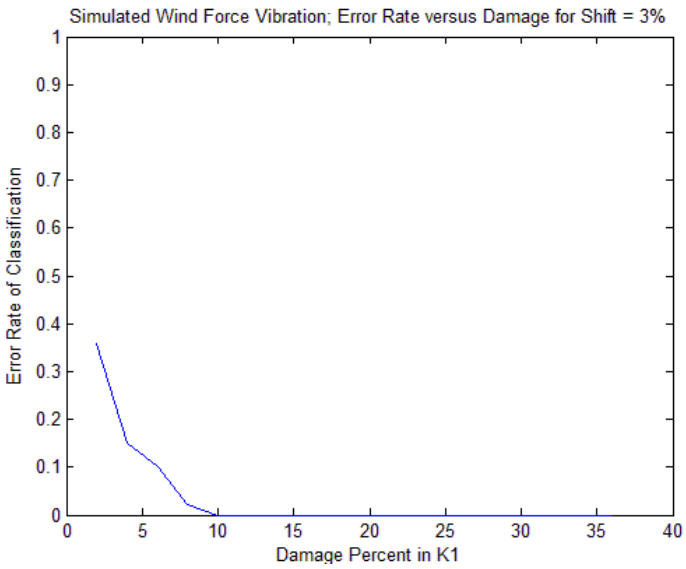

Fig. 32: Selected best percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under wind loading

It can be noticed that the minimum amount of damage that can be measured without any error has increased from 6% (Fig. 29) for the model with only base excitation to 10% (Fig. 32) for the model with base excitation and wind loading. Fig. 33 shows the SVM created by training on the model under wind loading.

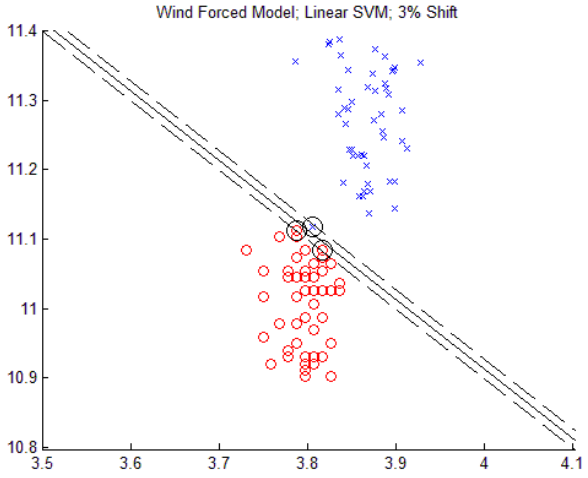

Fig. 33: The trained SVM classifier selected using the model under wind loading

3.3 SVM performance for different damage locations

In the above analysis, damage was introduced in the lowest floor in the structure. This was done in the simulation by reducing the stiffness of K1 in the model. In this section we will check the sensitivity of the method if damage is present in other floors. This test is important when a global property like the natural frequency has been used as a feature for damage detection. Before we proceed it is important to evaluate the effect of damage in different locations of the structure on the natural frequencies. Figures 34 – 37 show the percent change in the 1st and the 2nd natural frequency for increasing amounts of damage in each DOF. It should be noticed that the damage in the 2nd floor does not affect the second natural frequency in this particular model.

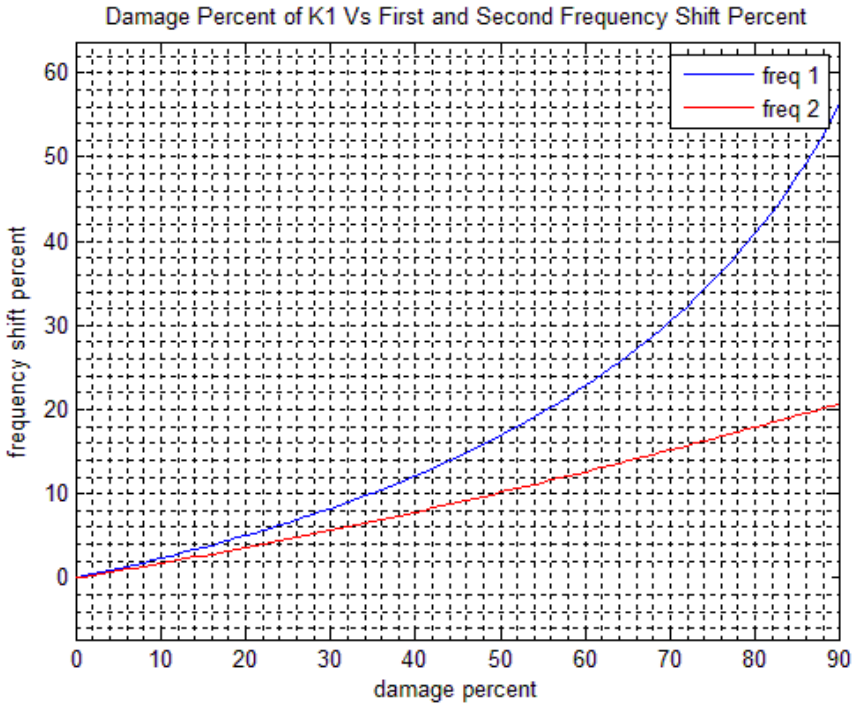

Fig. 34: Percent Change in natural frequency with damage in 1st floor

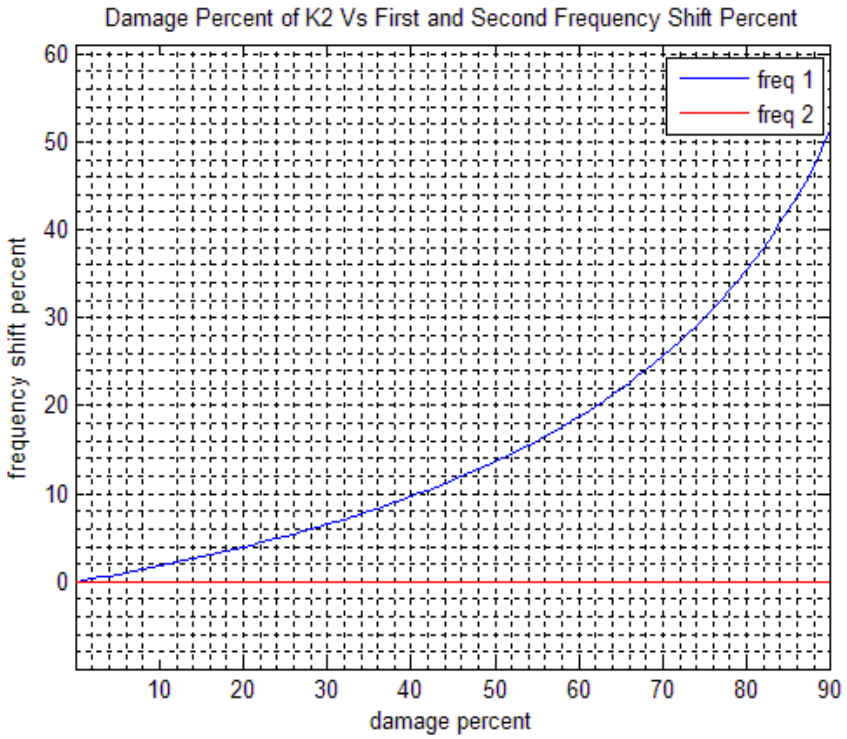

Fig. 35: Percent Change in natural frequency with damage in 2nd floor

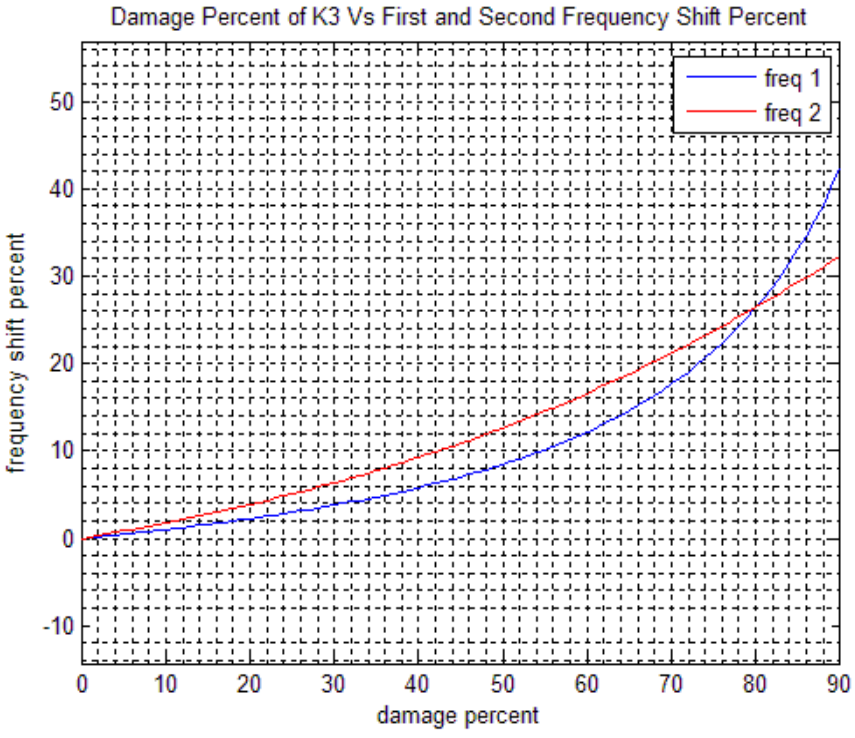

Fig. 36: Percent Change in natural frequency with damage in 3rd floor

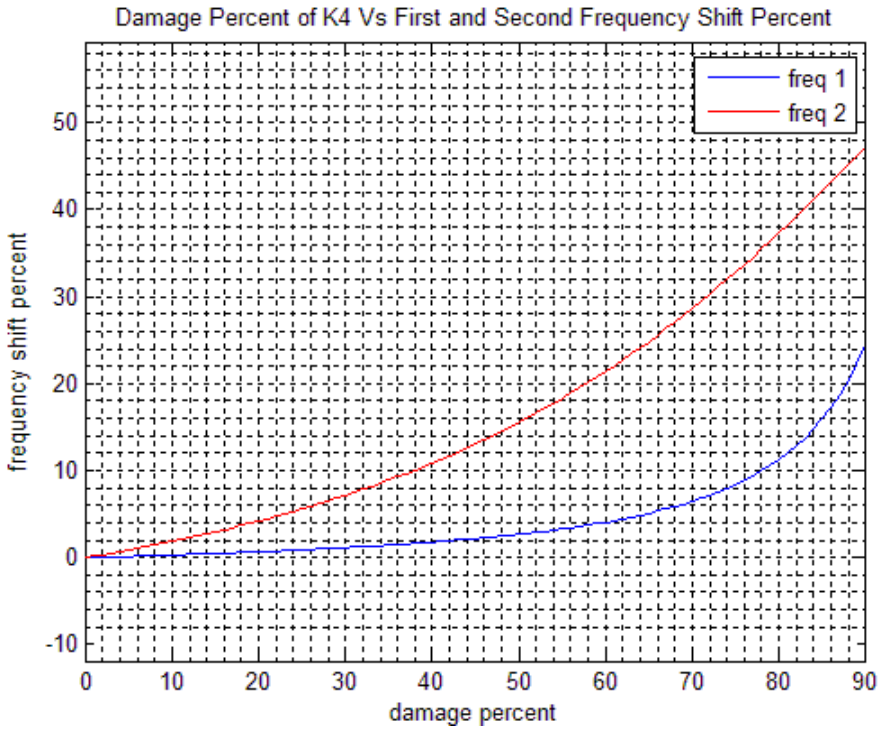

Fig. 37: Percent Change in natural frequency with damage in 4th floor

After an analysis on the performance of the SVM kernels and their parameters we selected the same kernel paramters as for earlier case when the damage was modeled in the 1st floor. We then proceed to evaluate the best percent shift of the undmaged training class for damage in each floor in Fig 38-40. It could again be seen thatand SVM created with 3% frequency shift performed better than other shift percents for detection of damage though the damage detection ability with any shift percent was poor.

The results for Fig. 38 indicate that the best percent shift (3%) can successfully detect damage above 20%. Fig. 39 and Fig. 40 indicate the minimum amount of damage that can be detected is approximately 12% and 15% for damage in K3 and K4 respectively. The results are poor for detecting local damage in the second floor.

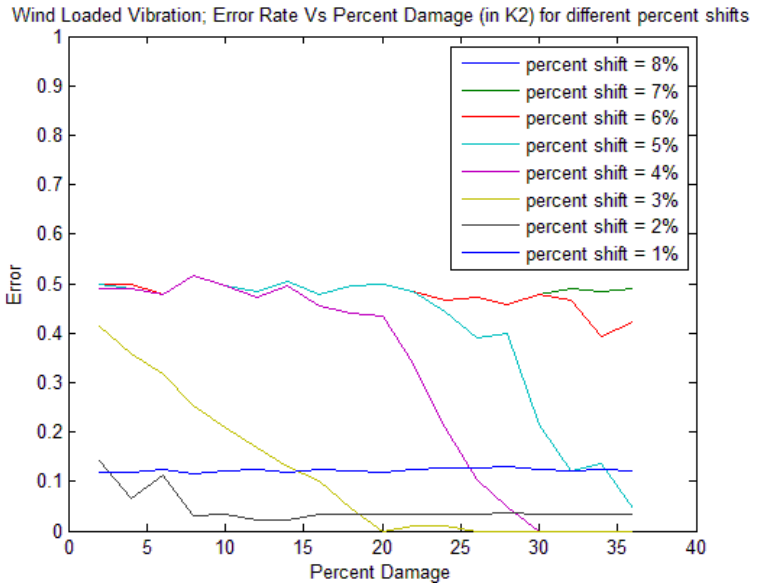

Fig. 38: Error rate versus percent damage (in K2) for different percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under wind loading

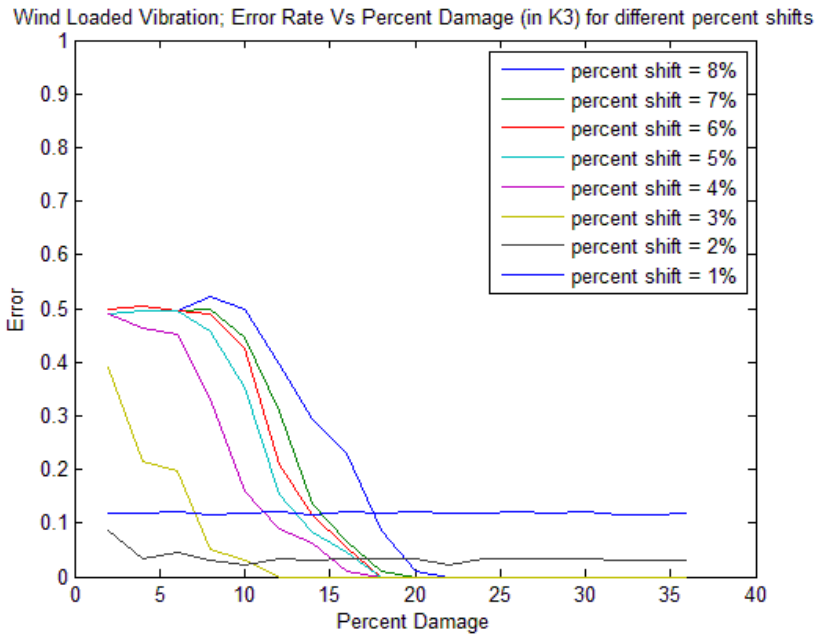

Fig. 39: Error rate versus percent damage (in K3) for different percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under wind loading

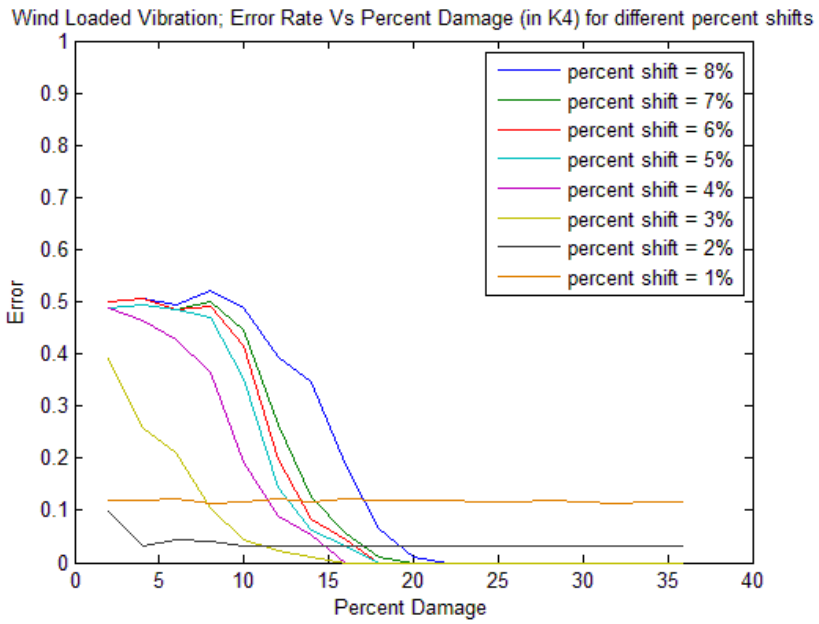

Fig. 40: Error rate versus percent damage (in K4) for different percent reduction of the undamaged class during training. The error rate is calculated on a test dataset. The training and testing datasets were derived from the model under wind loading

An investigation of the poor results for damage detection when the damage was in K2 revealed that the nature of the training and testing dataset were different. This was because in the training set both the natural frequencies were reduced by the same percent to create the damaged class of data. Damage in K2 however does not affect the 2nd natural frequency. This leads to improper training and poor results.

To work around this limitation while keeping the dependence on modal analysis minimal, we modified the data creation method for the damaged training class to account for such situations when only few frequencies were affected during damage. The "new" training data set now consists of:

a. only the first natural frequency reduced by 4% while keeping the 2nd constant

b. only the 2nd natural frequency reduced by 4% while keeping the 1st constant

c. both the frequencies reduced together by (4/√2)%. The parameter √2 was introduced to keep the 3 training datasets of the damaged class equidistant from the undamaged class of the training datset.

We reselected the best SVM paramters and the best shift percent of the undamaged class for this new set of training data. In this case, the RBF kernel performed the best and the parameters chosen were: γ = 1; C= 100.

Fig. 41 shows the best shift percent versus the error rate for the percent damage in K2. From this plot we conclude that the best shift percent is 4% for this training set. The minimum amount of damage that can be detected by this method is ~10% which is an improvement over the damage from the earlier dataset. Fig 42-Fig 44 show the best shift percent versus the error rate for the percent damage in K1, K3 and K4 respectively using the RBF kernel.

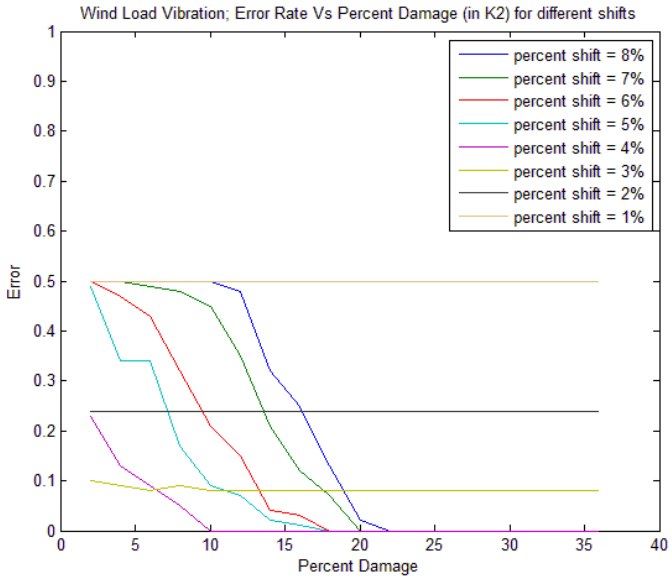

Fig. 41: Error rate versus percent damage (in K2) for different percent reductions of the undamaged class during training. The error rate is calculated on a test dataset. The training ("new") and testing datasets were derived from the model under wind loading

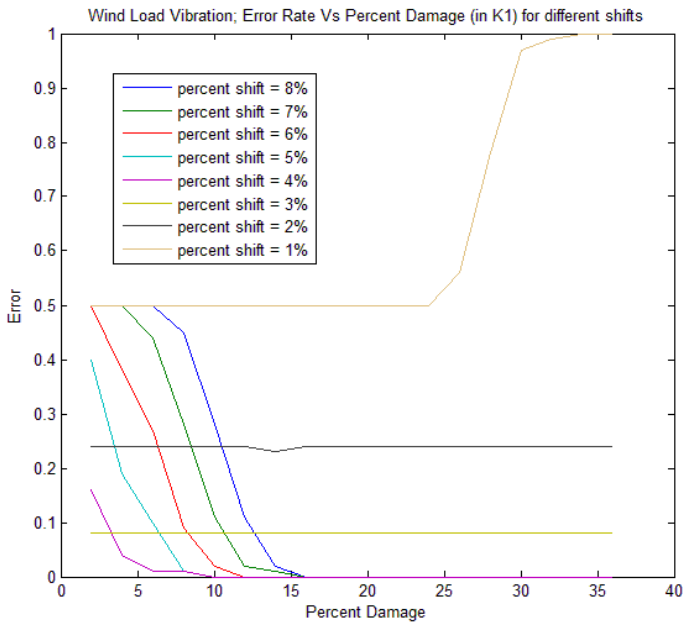

Fig. 42: Error rate versus percent damage (in K1) for different percent reductions of the undamaged class during training. The error rate is calculated on a test dataset. The training ("new") and testing datasets were derived from the model under wind loading

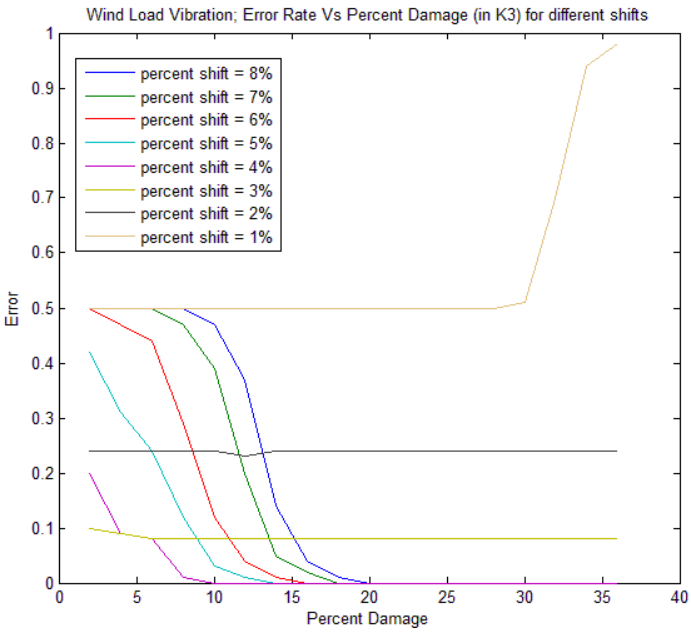

Fig. 43: Error rate versus percent damage (in K3) for different percent reductions of the undamaged class during training. The error rate is calculated on a test dataset. The training ("new") and testing datasets were derived from the model under wind loading

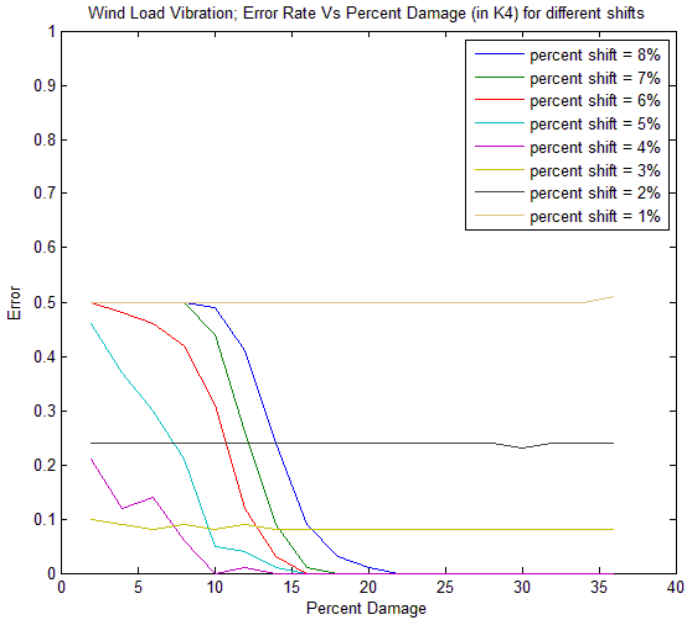

Fig. 44: Error rate versus percent damage (in K4) for different percent reductions of the undamaged class during training. The error rate is calculated on a test dataset. The training ("new") and testing datasets were derived from the model under wind loading

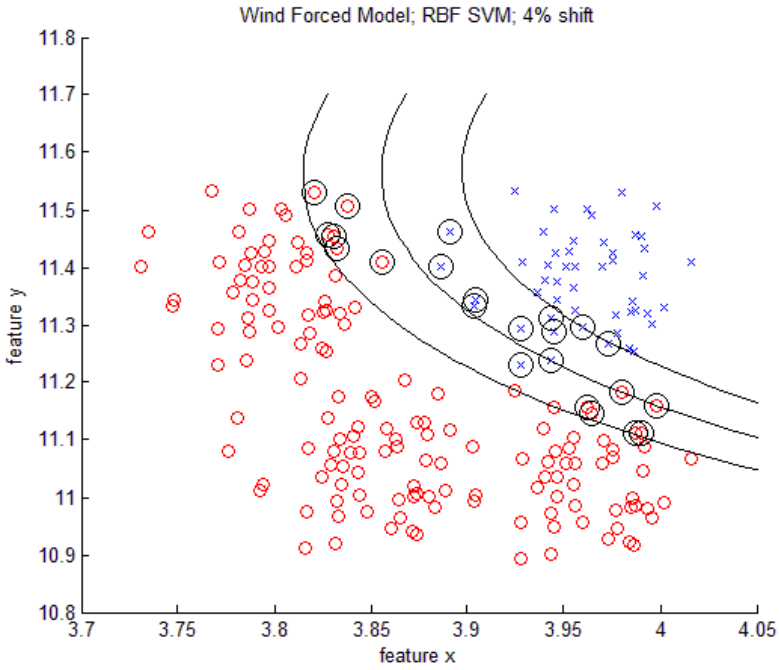

Fig. 45: The SVM classifier, trained on the "new" training dataset, selected using the model under wind loading

The results show that the SVM created using the RBF kernel (Fig. 41, 43, 45) performs better for damage detection that the SVM created using the linear kernel (Fig. 38, 39, 40). The trained SVM for the "new" dataset is present in Fig 45.

3.4 SVM performance for different measurement locations

The above simulation results were calculated for vibration readings only from the 1st DOF. We test the applicability of the method for readings from the 2nd – 4th degrees of freedom when there is damage only in K1. The kernel parameters selected for the SVM were the same as when the measurements were made in the 1st-DOF with damage in K1. Fig. 46-48 show the error rate of classification when SVMs trained with different percent shifts are used to classify the test dataset. In Fig 46-48 the measurements are made in the 2nd, 3rd and the 4th DOF respectively.

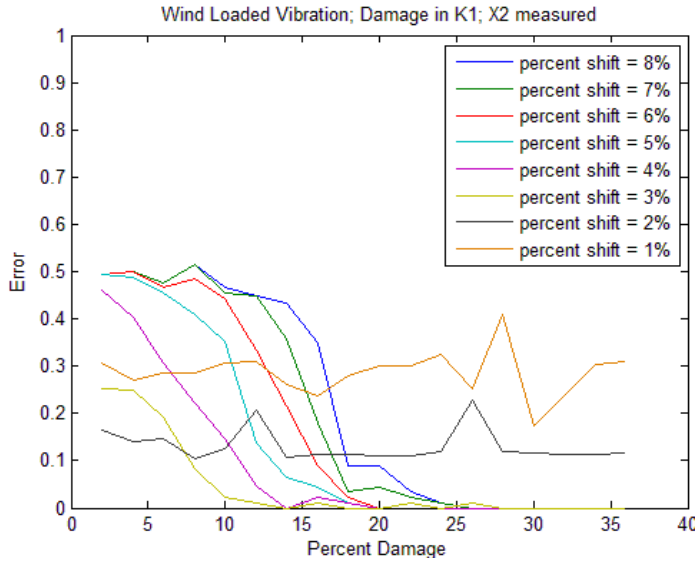

Fig. 46: Error rate versus percent damage (in K1) for different percent reductions of the undamaged class during training when the measurements are made in the 2nd DOF. The training and testing datasets were derived from the model under combined wind loading and ambient base excitation

In Fig. 46, the best curve can be seen to be for a shift of 3%. It can be noticed that the minimum damage in K1 that can be measured with complete accuracy has increased to 15% when the measurements are made in the 2nd DOF.

In Fig. 47, the best curve is for a shift of 3%. It can be seen that the minimum damage in K1 that can be measured is about 13% when the measurements are made in the 3rd DOF. The results are not as clean as when the measurements were made in the 1st DOF because of a combined effect of a high level of excitation and lesser sensitivity to damage.

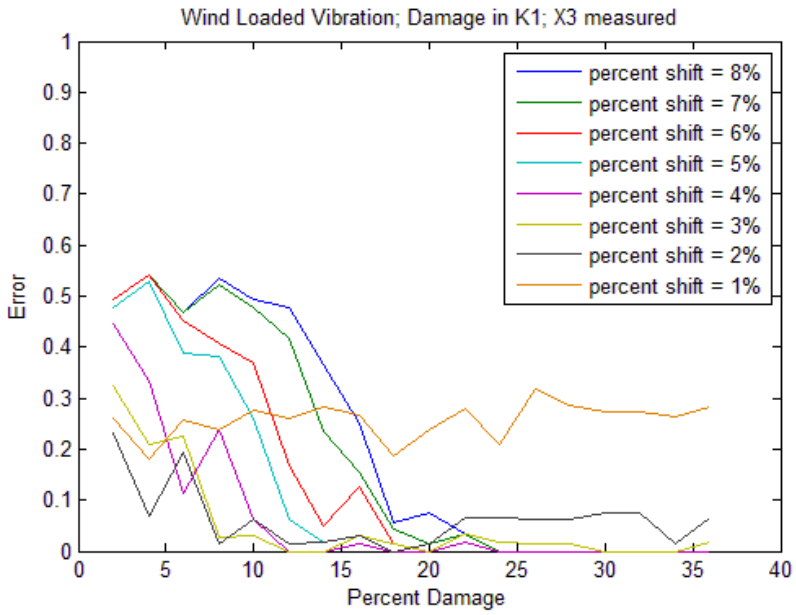

Fig. 47: Error rate versus percent damage (in K1) for different percent reductions of the undamaged class during training when the measurements are made in the 3rd DOF. The training and testing datasets were derived from the model under combined wind loading and ambient base excitation

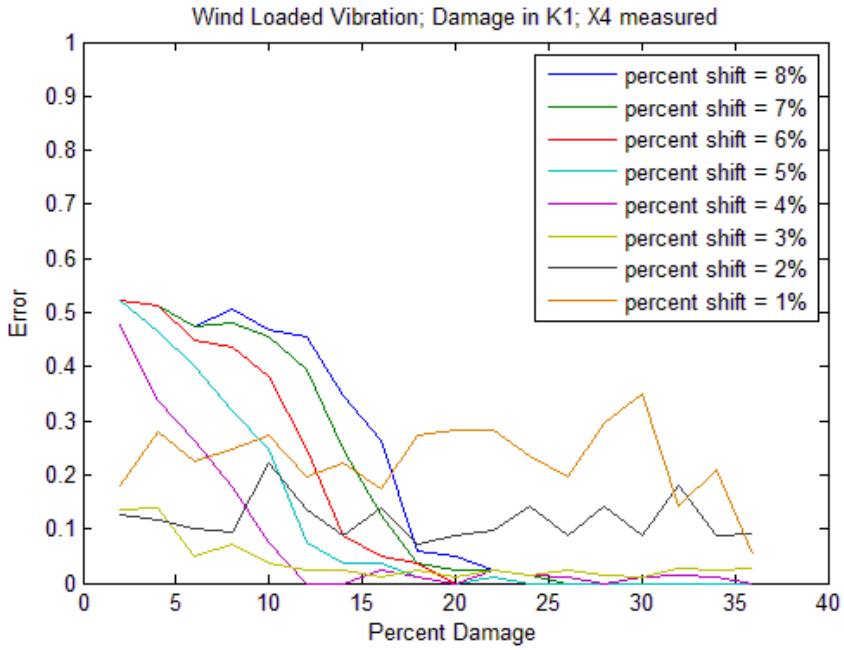

Fig. 48: Error rate versus percent damage (in K1) for different percent reductions of the undamaged class during training when the measurements are made in the 4th DOF. The training and testing datasets were derived from the model under combined wind loading and ambient base excitation

In Fig. 48, the best selected curve is for a shift of 4%. The minimum damage in K1 that can be measured is about 13% for measurements made in the 4th DOF. This is a combination of the presence of a high level of excitation and the fact that readings are made on floors that are not damaged.

3.5 Application of SVMs for Online Health Monitoring

The method developed can be used for online Structural Health Monitoring. A change in the structural properties due to either progressive damage due to fatigue and corrosion or abrupt damage caused by successive stresses would cause a corresponding change in the modal properties of the structure. Due to this change of properties, with time, more and more points tested against the old SVM boundary would be classified as belonging

to the damaged class. More points would be classified to the damaged class even though the structure might seem okay after manual inspection. To update the boundary as per the current health condition of the structure, fresh data can be collected from the structure and used to train a new SVM boundary. We define this boundary that represents the health condition of the structure as the "SVM safety margin". The SVM safety margin would shift from the earlier boundary with time due to a change in the structural properties. The location of this new SVM boundary can be used as an indicator of the health condition of the structure. A limiting condition for the SVM safety margin can be decided by judgment for each structure to represent the limit of damage making the structure unsafe for use.

The amount of shift of the SVM safety margin can be used to estimate the extent of damage caused to the structure. This would be very helpful to indicate the damage caused to the structure by strong motion and the effectiveness of retrofitting methods on a damaged structure.

To test the ability of the SVM safety margin to detect damage in a structure, we trained SVM boundaries on data from increasingly damaged structures. This provides an easy visualization of the health condition of the structure with the gradual shift of the SVM boundary. We test the idea for 3 cases when there is no external excitation on the structure; there is ambient base excitation on the structure; there is combined loading from ambient base excitations and wind loading on the structure. In all cases the damage is modeled in K1 and the measurements are made in the 1^{st} DOF. In the first case we investigate the shift of the SVM boundary when the training data is derived

from a damaged structure in the absence of any external loading. The SVM is trained for the parameters selection in section 3.1.4 when measurement noise was added to the simulation. In Fig. 48 and Fig 49 SVM safety margins are calculated for undamaged, 10% damage, 20% damage and 30% damage in K1.

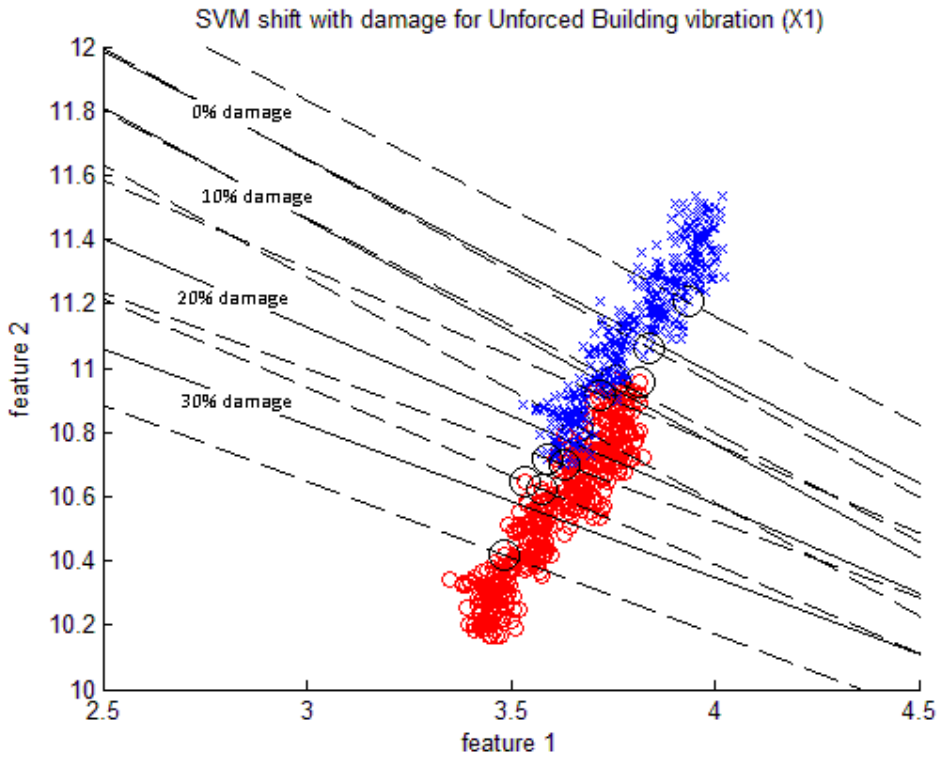

Fig. 49: Shift of the SVM safety margin when trained on data from increasingly damaged structures in the absence of external loading. The SVM safety margins have been created for structures that are 0%, 10%, 20% and 30% damaged.

It can be seen that as the training data is derived from structures with larger levels of damage, the SVM safety margins trained on them shift counter-clockwise. To make the visualization clearer, the boundaries were marked with different colors and the margins on both sides were removed.

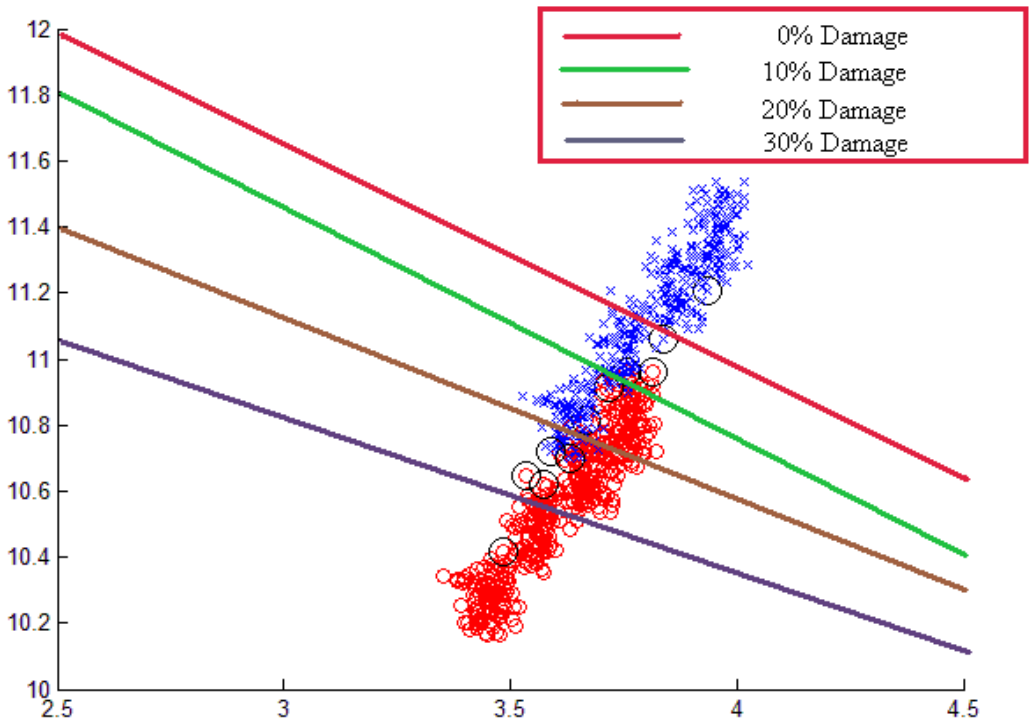

Fig. 50: Shift of the SVM safety margin when trained on data from increasingly damaged structures without any external loading. The safety margins were colored for visualization

In Fig. 51 the shift of the SVM safety margin has been investigated when the structure is affected by ambient base excitations. The best parameters for the training of the SVM have been used from section 3.1.5. It can be seen that the damage in Fig. 51 that there is a gradual progression of the boundary, though not ideal, as the damage in it increases. Fig. 52 illustrates the shift of the SVM safety margin when the structure was affected by combined loading from ambient base excitation and wind loading. The best parameters for the training of the SVM have been used from section 3.1.6. In both the above cases, the SVM has been trained for 0%, 10%, 20% and 30% damage in K1.

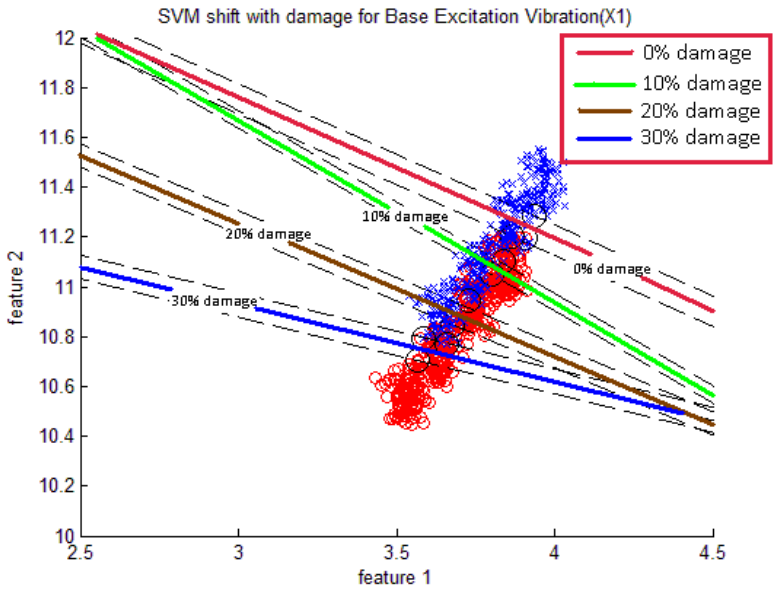

Fig. 51: Shift of the SVM safety margin when trained on data from increasingly damaged structures in the presence of ambient base excitations. The safety margins were colored for visualization

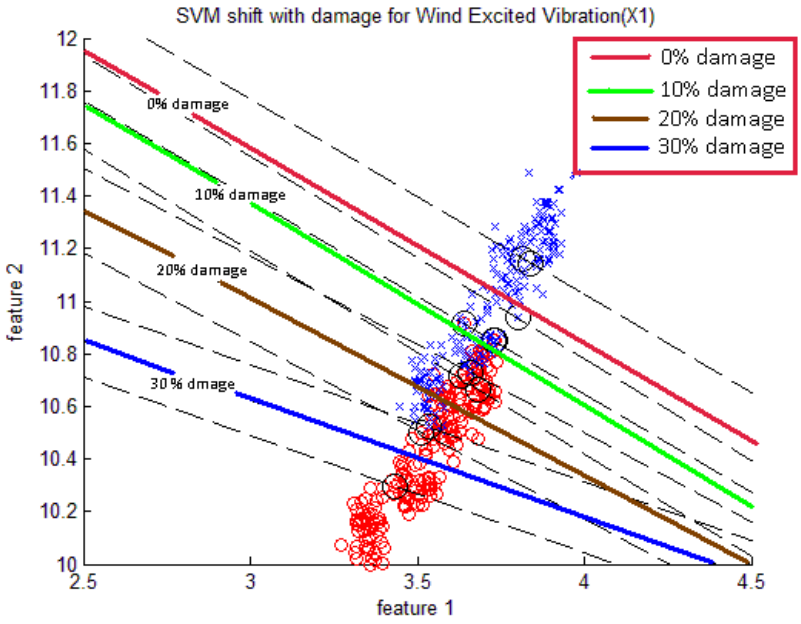

Fig. 52: Shift of the SVM safety margin when trained on data from increasingly damaged structures in the presence of ambient base excitations and wind loading. The safety margins were colored for visualization

From the above analysis that the SVM safety margin shifts gradually when the structure from which its training dataset is measured is increasingly damaged.

4 Conclusion

It this thesis a method has been proposed for using SVMs for structural health monitoring. The method has been illustrated by a simulation study with a four-story building model. The method includes the following essential developments:

1) Creation of the damaged "class" of the training data for SVMs. The unavailability of data from the damaged structure is a challenge for the application of SVMs for damage detection of structures. The proposed method creates the dataset representing the damaged class of data from the undamaged set of data measured on the structure by reducing numerical values of the undamaged dataset by some percent. As the natural frequencies are used as features of the data points, the reduced frequency values of the undamaged structure can be physically interpreted as representing the damaged structure. To account for the situation where only some frequencies are (one frequency in this case) affected by structural damage the damaged class of the training dataset consists of points created by reducing each frequency individually (while keeping the others constant) and also reducing all the frequencies at the same time.

2) Development of a cepstrum-like technique to "denoise" the Fourier spectra of the vibration data for better extraction of the first two natural frequencies. The presence of measurement noise in the structure is a major challenge when extracting

the natural frequencies from the measured vibration signal. To address this challenge, a cepstrum-like method is proposed to “clean” the Fourier amplitude spectrum of the measured vibration signal by taking Fourier transform of the Fourier amplitude spectrum of the vibration data and then filtering out the high frequency part of the signal obtained. An inverse Fourier transform of the truncated cepstrum is implemented to get back a “cleaned” Fourier amplitude spectrum of the vibration signal. The cepstrum-like technique proposed was able to “clean” a reasonably noisy signal successfully.

3) Development of SVM safety margin for structural health monitoring. It was found that the SVM classifier boundary shifted gradually when it was trained on data from a progressively damaged structure. These updated SVM classifiers may be used as good visual indicators of structural damage development and the most recent position for the SVM classifier (referred to as the SVM safety margin) can provide a qualitative assessment of structural damage and help to make an appropriate maintenance decision.

In this thesis, several important issues related to the SVM classifier performance were extensively investigated. The optimal SVM parameters (the kernel and the regularization parameters) were selected based on an extensive comparison. The performance of the damage detection method was proven satisfactory in the presence of measurement noise, ambient base vibrations and wind loading. The damage detection methodology was sensitive to the location of the damage as expected since the natural frequency is a global property of the structure and might not be affected by small local structural

damage. The method was also sensitive to the measurement location in the structure since the ability to detect the natural frequencies (features used in this research) depends on the mode shapes of the structure.

5 Future Work

The results have been successful for detection of damage in all floors though the minimum amount of damage that can be detected varies for different locations. Possible alternatives could be the use of different set of parameters, instead of the natural frequencies, that are more sensitive to local damage. Some kind of averaging of the parameter could also prove to be a more accurate parameter for indicating health of the structure. In the presence of nonlinearities introduced by damage in the structure, such as a beam with a crack, it is expected that use of variation of the natural frequencies of the structure might not be enough for the SVM classifier to perform satisfactorily. A study should be conducted to check which new features are most effective damage indications and how they can be incorporated into the proposed SVM methodology for damage detection.

To be able to successfully detect damage of this nature we modified the method of creation of the damaged class of the training dataset. The new training dataset consisted of a copy of healthy dataset when both the frequencies were reduced and it also included points when only one frequency was reduced and the others left constant. If more than 2 frequencies are measured, the training dataset would consist of a reduction of all frequencies and one frequency at a time. In this situation the Linear

SVM might not prove to be the best and that creates a large set of parameters to choose from when choosing the best SVM parameters. More research needs to be done to be able to associate the nature of the data with the kernels available. A better understanding of the applicability of SVM parameters for different types of data would reduce the time to implement the method in real time.

In the above study environmental factors have not been accounted for. The environmental conditions like moisture, temperature, winds etc. can cause huge variations in the structural response. A method to account for these effects on the structural behavior and correspondingly normalizing the input data to be tested would make the method more robust for real world application.

6 References

Abbasion, S., Rafsanjani, A.,Farshidianfar, A., Irani, N., 2007, " Rolling element bearings multi-fault classification based on the wavelet denoising and Support Vector Machine", *Mechanical Systems and Signal Processing*, 21, pp. 2933-2945

Aizerman, A., Braverman, E., M., Rozoner, L., I., 1964, "L.I. Theoretical foundations of the potential function method in pattern recognition learning", *Automation and Remote Control*, Vol. 25 (1964), pp. 821-837

Allemang R.J., "The modal assurance criterion (MAC): twenty years of use and abuse", *Proceedings of 20th International Modal Analysis Conference*, Los Angeles, CA, USA, pp. 397-405, 2002

Amani, M., G., Riera, J., D., Curadelli, R., O., 2007, "Identification of changes in the stiffness and damping matrices of linear structures through ambient vibrations", *Structural Control and Health Monitoring*, 14, pp. 1155-1169

Boser, B., E., Guyon, I., M., Vapnik, V., N., 1992, "A training algorithm for optimal margin classifiers", *Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, 1992, ACM

Burges Christopher J. C., 1998, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, 2, pp. 121-167

C. M. Bishop, "Neural Networks for Pattern Recognition", Clarendon Press, Oxford, 1995

Cao, T. T., Zimmerman, D. C., 1999, "Procedure to extract Ritz Vectors from dynamic testing data", *Journal of Structural Engineering*, pp 1393- 1400

Cawley, P., Adams, R. D., 1979, " The location of defects in structures from measurements of natural frequencies", *Journal of Strain Analysis*, 14(2), pp. 49-57

Chang, P. C., Flatau, A., Liu, S., C., (2003), "Review Paper: Health Monitoring of Civil Infrastructure", *Structural Health Monitoring*, 2(3), pp. 257-267.

Chen, G., Yang, X., Alkhrdaji, T., Wu, J., Nanni, A., 1999, "Condition assessment of concrete structures by dynamic signature tests", *Proc., 13th Engineering Mechanics Speciality Conference* (CDrom)

Cristianini, N., Taylor, J., W., 2000, "An Introduction to Support Vector Machines and other kernel-based methods", Cambridge University Press, Cambridge

Doebling, S. W., Farrar, C. R., Prime, M. B., and Shevitz, D. W., 1996a, "Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics: A Literature Review", Los Alamos National Laboratory report LA-13070-MS.

Fletcher, R., 1987, "Practical Methods of Optimization", John Wiley and Sons, London

Franc, V., Hlavac, V., 2004, "Statistical Pattern Recognition Toolbox for MATLAB", *Research Reports at Center for Machine Perceptron*, Czech Technical University in Prague, no. 8

Ghaboussi, J., Garrett Jr, J., H., Wu, X., 1991, "Knowledge-based Modeling of Material Behavior with Neural Networks", *Journal of Engineering Mechanics*, 117 (1), pp. 132-153

Kudva, J., N., Munir, N., Tan, P., W., 1992, " Damage Detection in Smart Structures using Neural Networks and Finite Element Analyses", *Smart Materials and Structures*, 1, pp. 108-112

Lim, T., W., Cabell, R., H., Silcox, R., J., 1996, "Online identification of modal parameters using artificial neural networks", *Journal of Vibration and Acoustics*, 118, pp. 649-656

Liu, P. L., 1995, " Identification and damage detection of trusses using modal data", *Journal of Structural Engineering*, 121(4), pp.599-608

Meyer, D., Leisch, F., Hornik, K., 2003, " The support vector machine under test", *Neurocomputing*, 55, pp. 169-186

Mita, A., Fujimoto, A., 2005, "Active detection of loosened bolts using ultrasonic waves and support vector machines", *Proc. of the 5th International Workshop on Structural Health Monitoring*, Stanford University, pp: 1017-1024

Mita, Akira, Hagiwara, Hiromi, 2003, "Qualitative Damage Diagnosis of Shear Structures Using Support Vector Machines", *KSCE Journal of Civil Engineering*, 7(6), pp. 683-689

Mottershead, J., E., Friswell, M., I., 1993, "Model updating in structural dynamics: A survey", *Journal of Sound and Vibration*, 19, pp. 347-375

Noori, M., N., Cao, C., Hou, Z., Sharma, S., 2009, " Application of Support Vector Machine for Reliability Assessment and Structural Health Monitoring", *The 10th International Conference on Structural Safety and Reliability*, (to be published)

Pandey, A., K., Biswas, M., 1994, "Damage detection in structures using changed in flexibility", *Journal of Sound and Vibration*, 169 (1), pp. 3-17

Platt, J., C., 1999, "Fast Training of Support Vector Machines using Sequential Minimal Optimizer", *Advances in kernel methods: support vector learning*, ISBN:0-262-19416-3

Rao, S., S., 2003, "Mechanical Vibrations", Pearson Education, Inc., New Jersey

Rytter, A., 1993, "Vibration based inspection of civil engineering structures," Ph. D. Dissertation, Department of Building Technology and Structural Engineering, Aalborg University, Denmark.

Samanta, B., Al-Balushi, K. R., Al-Araimi,S. A., 2003, " Artificial Neural Networks and Support Vector Machines with genetic algorithm for bearing fault detection", *Engineering Applications of Artificial Intelligence*, 16, pp. 657-665

Schalkoff, R., 1992, "Pattern Recognition: Statistical, Structural and Neural Approaches", John Wiley and Sons, Canada

Shimada, M., Mita, A., 2005, " Damage Detection of bending structures using Support Vector Machines", *Smart Structures and Materials 2005: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, Proceedings of the SPIE*, Volume 5765, pp. 923-930 (2005)

Shimada, M., Mita, A., Feng, M. Q., 2006, " Damage Detection of structures using Support Vector Machines under various boundary conditions*", Proceedings of SPIE, the International Society for Optical Engineering*, vol. 6174 (2), pp. 61742K.1-61742K.9

Sohn, H., Farrar, C. R., Hunter, N., F., Worden, K., 2001, " Structural Health Monitoring using statistical pattern recognition techniques", *Journal of Dynamic Systems, Measurement and Control*, 123 (4), pp. 706-711

Sohn, J., Law, K., H., 2001, "Damage Diagnosis Using Experimental Ritz Vectors", *Journal of Engineering Mechanics*, 127, 1184-1193

Todorovska, M., I., Trifunac, M., D., 2008, "Impulse response analysis of the Van Nuys 7-storey hotel during 11 earthquakes and earthquake damage detection", *Structural Control and Health Monitoring*, 15, pp 90-116

Trifunac, M., D., Ivanovic, M., D., Todorovska, M., I., 2001, "Apparent Periods of a building: 1 fourier analysis", *Journal of Structural Engineering*, 127 (5), pp 517-526

Trifunac, M., D., Ivanovic, M., D., Todorovska, M., I., 2001, "Apparent Periods of a building: 2 time frequency analysis", *Journal of Structural Engineering*, 127 (5), pp 527-537

Widodo, A., Yang, Bo-Suk, 2007, " Support Vector Machines in machine condition monitoring and fault diagnosis", *Mechanical Systems and Signal Processing*, 21, pp. 2560-2574

Wilson, E., L., Yuan, M., W., Dickens, J., M., 2006,"Dynamic analysis by direct superposition of Ritz vectors", *Earthquake Engineering and Structural Dynamics*, 10 (6), PP 813-821

Wolfe, P., 1961, "A duality theorem for nonlinear programming", *Quarterly of Applied Mathematics*, 19, 239-244

Worden, K., Lane, A., J., 2001, " Damage identification using support vector machines", *Smart Materials and Structures*, 10, pp. 540–547

Yang, B., S., Hwang, W., W., Kim, D., J., Tan, A. C., 2005, "Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines", *Mechanical Systems and Signal Processing*, 19, pp. 371-390

Zhang, J., Sato, T., I., 2008, "Experimental verification of the support vector regression based structural identification method by using shaking table test data", *Structural Control and Health Monitoring*, 15, pp. 505-517