Damage identification using inverse methods

BY Michael I. Friswell*

Department of Aerospace Engineering, University of Bristol, Bristol BS8 1TR, UK

This paper gives an overview of the use of inverse methods in damage detection and location, using measured vibration data. Inverse problems require the use of a model and the identification of uncertain parameters of this model. Damage is often local in nature and although the effect of the loss of stiffness may require only a small number of parameters, the lack of knowledge of the location means that a large number of candidate parameters must be included. This paper discusses a number of problems that exist with this approach to health monitoring, including modelling error, environmental effects, damage localization and regularization.

Keywords: damage detection; damage location; inverse methods

1. Introduction

Inverse methods combine an initial model of the structure and measured data to improve the model or test a hypothesis. In practice, the model is based on finite element analysis (FEA) and the measurements are acceleration and force data, often in the form of a modal database, although frequency response function (FRF) data may also be used. The estimation techniques are often based on the methods of model updating, which have had some success in improving models and understanding the underlying dynamics, especially for joints (Mottershead & Friswell 1993; Friswell & Mottershead 1995). Model updating methods may be classified as sensitivity or direct methods. Sensitivity-type methods rely on a parametric model of the structure and the minimization of some penalty function based on the error between the measured data and the predictions from the model. These methods offer a wide range of parameters to update that have physical meaning and allow a degree of control over the optimization process. The alternative is direct updating methods that change complete mass and/or stiffness matrices, although the updated models obtained are often difficult to interpret for health monitoring applications. These methods will be considered in more detail later. However, it should be emphasized that a huge number of papers have been written on the application of inverse methods to damage identification, and this paper aims to give an overview of the approaches rather than a complete literature review.

A different approach is to just use the measured data and identify damage from changes in the mode shapes. Farrar & Jauregui (1998a,b) compared several of these methods, such as the damage index method (Stubbs et al. 1992), the
mode shape curvatures (Pandey et al. 1991), the change in flexibilities (Pandey & Biswas 1994) or the change in stiffness (Zimmerman & Kaouk 1994). The example used was a road bridge with concrete deck and steel supports. Different levels of damage were introduced, but the damage was only clearly located with most methods at the most severe level where the first natural frequency changed by over 7%, and the mode shapes changed significantly. The damage index method was found to be the most promising. Other methods based on pattern recognition, often using neural networks, are also popular (e.g. Sohn et al. 2001, 2002; Trendafilova & Heylen 2003). These methods essentially provide curve fits using interpolation functions and are not based on physical models. Hence, these models are not considered further in this paper. The lack of a physical model also limits the scope for damage prognosis.

The four stages of damage estimation, first given by Rytter (1993), are now well established as detection, location, quantification and prognosis. Detection is readily performed by pattern recognition methods or novelty detection (Worden 1997; Worden et al. 2000). The key issue for inverse methods is location, which is equivalent to error localization in model updating. Once the damage is located, it may be parameterized with a limited set of parameters and quantification, in terms of the local change in stiffness, is readily estimated. Prognosis requires that the underlying damage mechanism is determined, which may be possible using inverse methods using hypothesis testing among several candidate mechanisms. This question is considered in more detail later in the paper. However, once the damage mechanism is determined, the associated model is available for prognosis, and this is a great advantage of model-based inverse methods.

2. The measured data

There are three basic types of data used in the measurement of dynamics: time domain; frequency domain; and modal model. During experimental modal analysis, the sampled time-series data are processed into the FRF data. For ambient excitation, the spectra of the responses are calculated. These frequency data are then further processed by curve fitting to obtain the modal model, which is the natural frequencies, damping ratios and mode shapes. Of course, this curve fitting has already assumed stationarity, linearity and reciprocity. At each stage, the processing involves some compression of the data, and results in a reduction in the volume of data. It may be thought best to use time-series data as the number of data points is very high. For linear and stationary systems, there is a little loss of information when going from the time domain to the frequency domain. Indeed, there is the advantage that the data may be averaged easily and therefore the effect of random noise is reduced.

The modal model represents a further reduction in the number of data points. However, is the quantity and quality of information reduced? The FRF may be reconstructed quite accurately using the modal model, and this is done to check the accuracy of the curve fitting. Theoretically, the FRF data contain some information from the out of range modes. However, in practice, unless the mode is just out of the range, the modes outside the range are easily corrupted by noise and the response is dominated by the in-band modes. Thus, the FRF and the modal model essentially contain the same information. The advantage of the
modal model is that it may be checked to ensure that the data look reasonable. The advantage of the FRF is that the data are closer to its raw state and the curve fitting may eliminate effects that would help in the identification of damage. For algorithms based on a predominantly linear model, this is unlikely to be an advantage. The equivalence of the FRF and modal data means that it should be possible to draw the same conclusions when presented with equivalent data. Demonstrating such equivalence in practice is difficult because the weights required for the data and the initial finite element model to ensure that the identification is equivalent are difficult to determine.

It is commonly acknowledged that natural frequencies may be measured more accurately than mode shapes. Natural frequencies are related to the time response of the structure, and measuring the frequency content of a signal may be performed very accurately. Mode shapes require the spatial distribution of the response to be estimated and in controlled laboratory tests, for example using force appropriation, accurate mode shapes may be obtained. In contrast, obtaining mode shapes from ambient response measurements is particularly difficult because the excitation force is not measured. Furthermore, assessing the error in the mode shapes is very difficult, and comparison with modelled mode shapes requires a consistent normalization. The situation is further complicated for well-separated natural frequencies, where the changes in mode shapes due to damage are often smaller than the error bounds on the corresponding measurements. Owing to these problems, natural frequencies alone are often used in model-based identification exercises. However, mode shapes are still valuable to pair the analytical and experimental modes, and also for symmetrical structures that often have areas which produce similar changes to the natural frequencies (Titurus et al. 2003b). Damage generally affects the flexural rigidity of beam structures, which directly affects the curvature of the mode shapes. Methods to determine damage based on estimating the mode shape curvature have been proposed (for example, Pandey et al. 1991), although obtaining the curvatures from noisy mode shape data is very difficult.

3. Model updating methods

Most of the approaches used for damage identification are based on the methods of model updating. These methods and the specific requirements for health monitoring are now discussed.

(a) Direct methods

Consider first the direct updating methods. The goal is often to reproduce the measured data (usually the modal model) by changing the stiffness matrix as little as possible (in some minimum norm sense). Historically, these methods were among the earliest in model updating (Friswell & Mottershead 1995) and a number of generalizations have been proposed, depending on what is considered to be the reference quantities (e.g. Kenigsbuch & Halevi 1998). A number of problems exist with the direct methods. There is no guarantee that the resulting matrices are positive definite (or semi-definite for structures with free–free modes), and extra modes may be introduced into the frequency range of interest. The standard methods do not enforce the connectivity of the structure,
represented by the bandedness of the matrices and the pattern of zero terms, although Kabé (1985) gave a method that enforced the expected connectivity. More fundamental is that forcing the model to reproduce the data does not allow for the errors that will be present in the measured data. Mode shapes, in particular, can only be measured with a limited accuracy. Further errors are introduced when the mode shapes are expanded, as they need to be, since very few of the analytical degrees of freedom will be measured. The major problem for damage location, and indeed for error location in model updating, is that all elements in the matrices may be changed. If only a small number of sites are modelled incorrectly (or are damaged) then only a small number of the matrix elements will be changed. Generally, owing to the minimum norm optimization in the updating method, all the matrix elements would be changed a little, rather than a small number of elements changed substantially. Thus, the effect of any damage present would be spread over all the degrees of freedom making location difficult.

Kaouk & Zimmerman (1994) and Zimmerman & Kaouk (1994) proposed that the change in the stiffness matrix should be low rank. This does not ensure that the change in stiffness will be local, as the stiffness change could be global but low in rank. The method requires the rank of the stiffness change to be less than or equal to the number of measured modes used in the update. Zimmerman et al. (1995) gave an overview of this approach, and discussed issues such as the number of measured modes to use. Doebling (1996) extended the method by updating the elemental parameter vector rather than the global stiffness matrix. Abdalla et al. (1998, 2000) developed methods by minimizing the change in the stiffness matrix, while enforcing constraints, such as symmetry, sparsity and positive definiteness.

It should be apparent by now that methods updating whole mass and stiffness matrices have significant disadvantages, which are not outweighed by the major advantage of not requiring parametric models of the damage mechanisms. Of course, the original purpose of these methods was not damage detection but often vibration control and stability. However, it is very unlikely that these methods will prove useful in the majority of structural health monitoring situations.

\( (b) \) Sensitivity methods

Sensitivity-based methods allow a wide choice of physically meaningful parameters and these advantages have led to their widespread use in model updating. The approach is very general and relies on minimizing a penalty function, which usually consists of the error between the measured quantities and the corresponding predictions from the model. Parameters are then chosen that are assumed uncertain, and these are usually estimated by approximating the penalty function using a truncated Taylor series and iterating to obtain a converged solution. If there are sufficient measurements and a restricted set of parameters, then the identification may be well-conditioned. Often some form of regularization must be applied, and this is considered in detail later. Other optimization methods may be used, such as quadratic programming, simulated annealing or genetic algorithms, but these are not considered further in this paper. Problems will also arise if an incorrect or incomplete set of parameters is chosen, or even worse, if the structure of the model is wrong.
Friswell & Mottershead (1995) discussed sensitivity-based methods in detail. The approach minimizes the difference between modal quantities (usually natural frequencies and less often mode shapes) of the measured data and model predictions. This problem may be expressed as the minimization of $J$, where

$$J(\theta) = \|z_m - z(\theta)\|^2 = \epsilon^T \epsilon,$$

and

$$\epsilon = z_m - z(\theta).$$

Here, $z_m$ and $z(\theta)$ are the measured and computed modal vectors, $\theta$ is a vector of all unknown parameters and $\epsilon$ is the modal residual vector. The modal vectors may consist of both natural frequencies and mode shapes, although often mode shapes are only used to pair individual modes. If mode shapes are included, then they must be carefully normalized, the sensor locations must be carefully matched to the finite element degrees of freedom and weighting should be applied to (3.1). Frequency response functions may also be used, although a model of damping is required, and the penalty function is often a very complicated function of the parameters with many local minima, making the optimization very difficult. The modal residual in (3.1) is a nonlinear function of the parameters and the minimization is solved using a truncated linear Taylor series and iteration. Thus, the Taylor series is

$$z_m = z_j + S_j \delta \theta_j + \text{higher order terms},$$

where

$$z_j = z(\theta_j), \quad S_j = S(\theta_j), \quad \delta \theta_j = \theta_m - \theta_j.$$  

The matrix $S_j$ consists of the first derivatives of the modal quantities with respect to the model parameters; index $j$ denotes the $j$th iteration and $\theta_m$ is the parameter vector that gives the measured outputs. Standard methods exist to calculate the modal derivatives required (Fox & Kapoor 1968; Nelson 1976; Friswell & Mottershead 1995; Friswell 1996; Adhikari & Friswell 2001). By neglecting higher order terms in (3.3), an iterative scheme may be derived using the linear approximation

$$\delta z_j = S_j \delta \theta_j,$$

where $\delta z_j = z_m - z_j$ and $\delta \theta_j = \theta_{j+1} - \theta_j$. Often, for damage location studies, only the residual and sensitivity matrices for the initial model are used. Avoiding iteration reduces the computation required, particularly where multiple parameter sets have to be estimated. However, particularly if the damage is severe, there is a risk that the wrong location is identified.

A frequent problem that arises in model-based vibration-based damage detection, whether parametric or non-parametric, is the need for a very accurate mathematical model, so that it correctly captures the actual structural dynamic behaviour in some predetermined frequency range. Often, in structural health monitoring, the changes in the measured quantities caused by structural damage are smaller than those observed between the healthy (i.e. undamaged) structure and the mathematical model. Consequently, it becomes almost impossible to discern between inadequate modelling and actual changes due to damage. There are two alternative approaches to this problem. The first is to update the healthy
model so that the correlation between the model and the measured data is improved. This approach requires that the errors that remain after updating are smaller in magnitude than the changes due to the damage. Furthermore, the changes to the model should be physically meaningful, so that the updating process corrects actual model errors, and does not merely reproduce the measured data. The second approach is based on the use of (relative) differences between data measured on healthy and potentially damaged structures. In this case, assuming that the only changes in the structure are due to damage, the problem may be reduced to finding those parameters that reproduce the measured changes.

As indicated above, one of the problems with sensitivity methods is the need for a parametric model of the damage. Mottershead et al. (1999) proposed an approach where the system was constrained so that unknown stiffnesses are replaced with rigid connections. The constraint is not imposed physically but the behaviour inferred from the unconstrained measurements. The best fit between the measured and the predicted data is obtained when the damage is located in the substructure that is made rigid.

4. Parameterization of damage

One of the key aspects of a model-based identification method is the parameterization of the candidate damage. Since inverse approaches rely on a model of the damage, the success of the estimation is dependent on the quality of the model used. The type of model used will depend on the type of structure and the damage mechanism, which leads to an increase either in local or in distributed flexibility. The damage model may be simple or complex. For example, a cracked beam may be modelled as a reduction in stiffness in a large finite element or substructure, or alternatively using a very detailed model from fracture mechanics. Whether such a detailed model is justified will often depend on the requirements of the estimation procedure and the quality of the measured data. Using a measured modal model consisting of the lower natural frequencies and associated mode shapes will mean that only a coarse model of the damage may be identified.

(a) Generic elements

One major difficulty in parametric approaches is that a model, which accurately reflects the effect of damage on the mass and the stiffness matrices, is required. To some extent, the situation is helped when low-frequency vibration measurements are used because any local stiffness reduction will have a very similar effect on the dynamic response. Thus, it is possible to use equivalent parameters, such as element stiffnesses, to model the damage. Generic elements (Gladwell & Ahmadian 1995; Friswell et al. 2001) take this approach further by allowing changes to the eigenvalues and eigenvectors of the stiffness matrices of structural elements or substructures. These changes are usually constrained so that properties such as the rigid body modes and the geometric symmetry are retained.

Generic elements introduce flexibility into the joint in a controlled way. Other equivalent models, such as discrete rotational springs, offset parameters or changing element properties, may also be used, although generic parameters do
have advantages (Friswell et al. 2001). In particular, all models prejudge how the damage will affect the full model of the structure, whereas the generic element approach automatically finds the probable low-frequency motion of the joint. Consider a two-dimensional $T$ joint constructed from three beam elements. Each node has three degrees of freedom and, since the substructure has four nodes, the substructure stiffness matrix has three rigid body eigenvectors and nine flexible eigenvectors (Titurus et al. 2003a). The lower eigenvectors have much simpler deformation shapes that are more likely to represent the motion the substructure would undergo in many of the global modes of the structure. Thus, reducing the eigenvalues corresponding to these eigenvectors makes the joint substructure more flexible in the frequency range of the global dynamics, and may be used to model damage. Higher frequency eigenvectors of the substructure may also be included if the motion of the joint is more complex; however, the lower eigenvectors of the joint are likely to adequately characterize the low-frequency dynamics of the structure. Gladwell & Ahmadian (1995) gave a further explanation of the physical meaning of generic elements.

(b) Crack models

The modelling of cracks in beam structures and rotating shafts has been a significant research topic. The models fall into three main categories: local stiffness reduction; discrete spring models; and complex models in two or three dimensions. Dimarogonas (1996) and Ostachowicz & Krawczuk (2001) gave comprehensive surveys of crack modelling approaches. The simplest methods for finite element models reduce the stiffness locally, for example by reducing a complete element stiffness to simulate a small crack in that element (Mayes & Davies 1984). This approach suffers from problems in matching damage severity to crack depth, and is affected by the mesh density. An improved method introduces local flexibility based on physically based stiffness reductions, where the crack position may be used as a parameter for identification purposes. The second class of methods divides a beam-type structure into two parts that are pinned at the crack location and the crack is simulated by the addition of a rotational spring. These approaches are a gross simplification of the crack dynamics and do not involve the crack size and location directly. The alternative, using beam theory, is to model the dynamics close to the crack more accurately, for example producing a closed form solution giving the natural frequencies and mode shapes of cracked beam directly or using differential equations with compatible boundary conditions satisfying the crack conditions (Christides & Barr 1984; Lee & Chung 2001; Sinha et al. 2002). Friswell & Penny (2002) compared several of the simple crack models that may be used for health monitoring, for both linear and nonlinear responses. Alternatively, two- or three-dimensional finite element meshes for beam-type structures with a crack may be used. Meshless approaches may also be used, but are more suited to crack propagation studies. No element connectivity is required and therefore the task of remeshing as the crack grows is avoided, and a growing crack is modelled by extending the free surfaces corresponding to the crack (Belytschko et al. 1995). However, the computational cost of these meshless methods generally exceeds that of the conventional FEA. Rao & Rahman (2001) avoided this difficulty by coupling a meshless region near the crack with an FEA model in the remainder of

Phil. Trans. R. Soc. A (2007)
the structure. The two- and three-dimensional approaches produce detailed and accurate models but are a complicated and computational-intensive approach to model simple structures like beams, and are unlikely to lead to practical algorithms for damage identification.

(c) Composite structures

Composite structures have an excellent performance, although this deteriorates significantly with damage. Unfortunately damage due to impact events, for example, is difficult to detect visually, and hence some method of non-destructive testing of these structures is required. Zou et al. (2000) reviewed the vibration-based methods that are available to monitor composite structures. Since this paper considers inverse methods for damage estimation, this section will only consider the parameterization of the damage in composite structures, and in particular the modelling of delaminations. Although composite structures have other modes of failure, such as matrix cracking, fibre breakage or fibre–matrix debonding (Ostachowicz & Krawczuk 2001), these damage mechanisms produce similar changes in the vibration in response to that obtained for damage in metallic structures. However, delamination is a serious problem in composite structures, and has no parallel to damage mechanisms in other materials. Once the damage is parameterized, inverse methods, such as sensitivity analysis, may be applied.

Zou et al. (2000) reviewed methods to model delaminations, and here we will concentrate on simple models. For example, if a structure is modelled with beam or plate elements, then only beam or plate elements should be used to model the structure with delaminations. Delamination occurs when adjacent plies in a laminated composite debond. For beam structures, the simplest case of a through width delamination, parallel to the beam surface, was modelled using four beam segments (Majumdar & Suryanarayan 1988; Tracy & Pardoen 1989). Separate beam elements were used above and below the delamination, and the constraints to join these elements to those of the undamaged parts of the beam needed to be applied carefully. Zou et al. (2000) detailed further development of these models. One difficulty with using these models for parameter-based identification is that changing the length and position of a delamination requires the model to be remeshed, and care must be exercised in calculating the associated sensitivity matrices. The techniques detailed by Sinha et al. (2002) for the position of cracks might be extended to this case. Paolozzi & Peroni (1990) highlighted that the most sensitive modes are those whose wavelength is approximately the same size as the delamination. Luo & Hanagud (1995) used a sensitivity-based method to detect delaminations, and they also discovered that some modes split to give two closely spaced natural frequencies.

(d) Distributed damage

Teughels et al. (2002) presented a sensitivity-based finite element updating method for damage assessment that minimized differences between the experimental and the predicted modal data. The parameterization of the damage (both localization and quantification) was represented by a reduction factor of the element bending stiffness. The number of unknown variables was reduced to obtain a physically meaningful result, by using a set of damage functions to determine the
spatial bending stiffness distribution. The updating parameters were then the multiplication factors of the damage functions. The procedure was illustrated on a reinforced concrete beam and on a highway bridge (Teughels & De Roeck 2004).

5. Regularization

The advantages of sensitivity-type model updating methods have been highlighted in this paper. However, there are significant differences in the application of these methods in model updating and damage location, which necessitates different methods of regularization. In both the cases, the number of potential parameters is very large and the estimation process is likely to be ill-conditioned unless the physical understanding can be used to introduce extra information.

In model updating, the number of parameters may be reduced by only including those parameters that are likely to be in error. Thus, if a frame structure is updated, the beams are likely to be modelled accurately but the joints are more difficult to model. It would therefore be sensible to concentrate the uncertain parameters to those associated with the joints. Even so, a large number of potential parameters may be generated, the measurements may still be reproduced and the parameters are unlikely to be identified uniquely. In this situation, all the parameters are changed and regularization must be applied to generate a unique solution (Friswell et al. 2001). Regularization generally applies extra constraints to the parameter estimation problem to ensure a unique solution. Applying the standard Moore–Penrose pseudo-inverse is a type of regularization where the parameter vector with the minimum norm is chosen. The parameter changes may be weighted separately to give a weighted least-squares problem, where the penalty function is a weighted sum of squares of the measurement errors and the parameter changes. Such weighting may also be extended to include minimizing the difference between equivalent parameters that are nominally equal in different substructures such as joints. Although using parametric models can reduce the number of parameters considerably, there will still be a large number of parameters for the damage location. Most regularization techniques rely on minimum norm-type solutions that will tend to spread the identified damage over a large number of parameters. Using subset selection, where only the optimum subset of the parameters are used for the estimation (Friswell et al. 1998), has been used for model updating and also for damage location.

(a) Subset selection and statistical approaches

Parameter subset selection is a technique that selects the best subset of parameters from a candidate set, utilizing some application-dependent cost function that provides a measure of goodness of each subset. Often, these techniques only obtain a suboptimal estimate of the best subsets in some sense due to the excessive computational burden posed by the original problem. These techniques are firmly rooted in statistics and related fields (Millar 1990), although recently applications in structural mechanics have appeared. Friswell et al. (1997) gave an overview of subset selection and also proposed the use of this technique for damage detection. They suggested an approach based on forward parameter subset selection, which is especially suited to local damage, and applied the method to a simulated cantilever beam example with physical
parameters corresponding to either element or node properties. Different selection and iteration strategies were evaluated, and the case where multiple measurement sets are available was handled by computing the principal angles between two vector subspaces. Fritzen et al. (1998) used an orthogonalization scheme for subset selection.

In damage location, statistical methods and performance measures have been used that work on a similar principle (Cawley et al. 1978; Cawley & Adams 1979; Friswell et al. 1994). Only a limited number of sites are assumed to be damaged, and the model updated based on the reduced number of parameters. This process is repeated for all possible combinations of the damage site, and possibly even the damage mechanism. The results from all the updated models are compared and the one that matches the measured data best is chosen.

The major problem with both the subset selection and the statistical-type approaches is that many smaller model updating exercises have to be performed. To optimally derive the best set of parameters, or the best damage location, requires the evaluation of many subsets of parameters. With a large number of parameters evaluating all subsets of even two or three parameters can become daunting. Thus, suboptimal methods must be used to derive good, but not necessarily the best, subsets of the parameters. In the forward approach parameters are chosen one at a time, and the parameters selected previously are retained. However, there is no guarantee that the optimal subset will be found. The number of candidate damage locations may be controlled based on the expected reduction in the residual (Millar 1990). The addition of a parameter to a previously selected subset inevitably reduces the residual terms, and thus there is a trade off between the number of parameters selected and the magnitude of the residual. Often only a single damage location will be required in which case the optimal parameter may be determined. Often a reasonable number of parameter subsets (say between 3 and 20) are selected for a more detailed study (Millar 1990). Friswell et al. (1998) reviewed the relationship between subset selection and matrix decomposition, and also expanded the methods to parameter groups using subspace angles. Titurus et al. (2003b) considered the weighting requirements within the inner product defining the subspace angles, following the work by Knyazev & Argentati (2002).

6. Applications of inverse methods to health monitoring

There are many examples of the practical application of inverse methods to the updating and health monitoring of real structures. Link et al. (2004) gave an overview of model updating with large-scale examples from the aerospace industry. Other papers in this special issue highlight the application to health monitoring in the aerospace and the civil sectors. Doebling et al. (1998) and Sohn et al. (2004) described a wide range of health monitoring applications. The following conferences also give a good range of examples: the SEM IMAC conference; the ISMA conference; the SPIE Nondestructive Evaluation for Health Monitoring and Diagnostics conference; the International Conference on Damage Assessment of Structures (DAMAS); the International Workshop Structural Health Monitoring; and the European Workshop on Structural Health Monitoring.
The European COST Action F3 on Structural Dynamics provided two benchmarks for health monitoring (Golinval & Link 2003). A number of European research groups used the same data to identify damage in the structures. The first was called the Steelquake structure, and consisted of a welded steel frame, and steel and concrete floors. Damage was induced by large static loads, and FRFs were measured before and after damage. Visual inspection determined the actual damage locations. The response of the Z24 highway bridge was measured during an earlier European Brite-Euram project. The bridge was a pre-stressed bridge with three spans and was 60 m long. Damage was induced by cutting one of the piers to simulate settlement. Frequency response functions and ambient response data were available. Teughels & De Roeck (2004) also used data from the Z24 bridge to identify the damage using model updating.

7. Problems and errors in damage identification

The discussion thus far has indicated some of the problems with damage identification. There are always errors in the measured data and the numerical model that affect all of the algorithms. These errors, and the adequacy of the data, are now discussed. Damage identification algorithms should always be tested on realistic experimental examples, as many methods that work well on simulated data often fail due to the problems highlighted in this section. As a first step, methods may be tested using simulated data, but even then realistic systematic errors should be incorporated.

(a) Modelling errors

One of the major problems in damage location is the reliance on the finite element model. This model is also an important strength because the very incomplete set of measured data requires extra information from the model to be able to identify damage location. There will undoubtedly be errors even in the model of the undamaged structure. Thus, if the measurements on the damaged structure are used to identify damage locations, the methods will have great difficulty in distinguishing between the actual damage sites and the location of errors in the original model. If suitable parameters are not included to allow for the undamaged model errors, then the result will be a systematic error between the model and the data. Identification schemes generally have considerable difficulty with systematic errors. It is very likely that the original errors in the model will produce frequency changes that are far greater than those produced by the damage. There are two basic approaches to reducing this problem, although both rely on having measured data from an undamaged structure. The first is to update the finite element model of the undamaged structure to produce a reliable model (Friswell & Mottershead 1995). Obviously, the quality of the damage location assessment is critically dependent upon the updated model being physically meaningful (Friswell et al. 2001; Link & Friswell 2003). Generally, this requires model validation using a control set of data not used for the updating. The second alternative uses differences between the damaged and the undamaged response data in the damage location algorithm (Parloo et al. 2003; Titurus et al. 2003b). To the first order, any error in the undamaged model
of the structure that is also present in the damaged structure will be removed. This does rely on the structure remaining unchanged, except for the damage, between the two sets of measurements.

Another potential source of error is the mismatch between the measurement locations and the model degrees of freedom. Such a mismatch makes the direct comparison of FRFs and mode shapes impossible, and the generation of residuals inaccurate. The magnitude of the errors involved will depend on the mesh density in the sensor region and the complexity of the mode shapes. The best solution is to ensure nodes in the model exist at the sensor locations. Alternatively, interpolation techniques may be used.

(b) Environmental and other non-stationary effects

One very difficult aspect of damage assessment is the change in the measured data due to environmental effects. This is one undesirable non-stationary effect and makes damage location very difficult. Of course, progressive damage is also a non-stationary phenomenon, and damage can be difficult to identify if other non-stationary effects are also present. Typical environmental effects are demonstrated by highway bridges, especially those constructed using concrete, which have been the subject of many studies in damage location. For example, temperature changes can cause the stiffness properties of a bridge to change significantly, and the difficulty is to predict the effects of temperature from readily available measurements. Peeters & De Roeck (2001) report on measurements of the Z24 bridge over a whole year and suggest a ‘black box’ model to predict the temperature variation. Sohn et al. (1999) considered the effect of temperature on the Alamosa Canyon Bridge. Sohn et al. (2002) used a combination of time-series analysis, neural networks and statistical inference to determine damage state for structures affected by the environmental conditions. Mickens et al. (2003) corrected FRF measurements by assuming the temperature affected the global stiffness of the structure. On a highway bridge, the changing traffic conditions cause different mass loading effects that can change the natural frequencies by as much as 1% (Zhang et al. 2002). There are further difficulties with highway bridges because they are highly damped with low natural frequencies. They are in a noisy environment and are difficult to excite. The frequency resolution in the measurements is invariably quite low, leading to considerable difficulties in detecting small frequency changes due to damage.

(c) The effect of frequency range

The range of frequencies employed in damage location has a great influence on the resolution of the results and also the physical range of application. A great advantage in using low-frequency vibration measurements is that the low-frequency modes are generally global and therefore the vibration sensors may be mounted remotely from the damage site. Equally fewer sensors may be used. The problems with low-frequency modes are that the spatial wavelengths of the modes are large and typically far larger than the extent of the damage. The spatial resolution of the damage identification scheme requires that there is a significant change in response between two adjacent potential damage sites. If low-frequency modes are used, then this resolution is closely related to the spatial wavelengths of the modes. Using high-frequency excitation uses highly local
modes which are able to accurately locate damage, but only very close to the sensor and the actuator position. Estimating accurate models at these high-frequency ranges is also very difficult, and often changes in the response are used for damage identification. For example, Park et al. (2000, 2001) used changes in measured impedance to identify damage in civil structures and pipeline systems. Schulz et al. (1999) used high-frequency transmissibility to detect delaminations in composite structures.

Moving to even higher frequencies can also yield good results. Acoustic emission (Rogers 2001) is a transient elastic wave, typically in the region of 50–500 kHz, and is able, for example, to detect the energy released when cracks propagate. One approach to damage assessment is to use physical models to deduce quantitative relationships between the measured acoustic emission signals and the damage mechanism that cause them. Significant research has been undertaken to obtain a physical understanding of various source mechanisms (Scruby & Buttle 1991) and the radiation pattern of bulk shear and longitudinal acoustic waves that they produce (Ono 1991). The difficulty in using these models in inverse estimation procedures is the accuracy of these high-frequency models, and the huge computational requirements. In recent years, a pattern recognition philosophy has dominated, which relies on using large databases of empirical data from which correlations between measured acoustic emission signals and damage mechanisms are inferred. Many advanced signal processing algorithms have been employed to interpret experimental data. Damage location is often determined using time of flight methods, which require the events to be well separated in time, the wave speed to be approximately equal in all parts of the structure, and the effects of reflection and refraction to be insignificant.

(d) Nonlinearity

Many forms of damage cause a change in the stiffness nonlinearity that qualitatively and quantitatively affects the dynamic response of a structure. For example, Nichols et al. (2003a,b) used the features of the chaotic response of a structure to detect changes in a joint. Adams & Nataraju (2002) gave a variety of features based on the nonlinear dynamic response. Kerschen et al. (2003) considered model-based estimation methods and identified the form of nonlinearity that is most likely present in the measured data. Meyer & Link (2003) identified a parametric nonlinear model using harmonic balance, using a model updating approach. A breathing crack, which opens and closes, can produce interesting and complicated nonlinear dynamics. Brandon (1998) and Kisa & Brandon (2000) gave an overview of some of the techniques that may be applied. Many techniques to analyse the resulting nonlinear dynamics are based on approximating the bilinear stiffness when the crack opens and closes. Linear approaches to damage estimation approximate a local reduction in the stiffness matrix of the beam. Since the nonlinearity introduced by a crack is often weak, many of the common testing techniques will tend to linearize the response (Leonard et al. 2001; Worden & Tomlinson 2001; Friswell & Penny 2002). Sinusoidal forcing will tend to emphasize the nonlinearity, and damage detection methods based on detecting harmonics of the forcing frequency have been proposed (Shen 1998). In rotor dynamic applications, these approaches are useful because the forcing is inherently sinusoidal (Dimarogonas 1996). However, in
structural health monitoring applications, this approach requires considerable hardware and software to implement, and also requires a lengthy experiment. Johnson et al. (2004) used a transmissibility approach that was insensitive to boundary condition nonlinearities. Neild et al. (2003) investigated the potential of a time frequency analysis procedure to identify damage in concrete beams.

Although using the nonlinear response has a huge potential in health monitoring, model-based inverse approaches have a number of difficulties owing to the high number of degrees of freedom required, and therefore the computational burden imposed. In practice, any realistic multi degree of freedom nonlinear analysis would have to be based on a reduced order model of the structure. Furthermore, many of the difficulties outlined in this section for linear systems are also a problem for nonlinear systems.

(e) Prognosis

Rytter (1993) gave prognosis as the fourth level of damage estimation and Farrar et al. (2003) gave a summary of the ‘state-of-the-art’. The philosophy of damage detection using measured vibration data is based on the premise that the damage will change the stiffness of the structure. In some instances, there is a significant difference between strength and stiffness. Indeed, estimating the remaining useful life of a component based on conclusions from a dynamic analysis is very difficult. For example, a concrete highway bridge will have steel reinforcement cables running in channels in the concrete. The cables are tensioned, either before or after the concrete has set, to ensure that the concrete remains in compression. One major failure mechanism is by the corrosion of these cables. Once the cables have failed, the concrete has no strength in tension and therefore the bridge is liable to collapse. Unfortunately, the stiffness of the bridge is mainly due to the concrete, and therefore the progressive corrosion of the cables is very difficult to identify from stiffness changes. Essentially, the dynamics of the bridge changes very little until it collapses. Even for metallic structures, the estimation of remaining life requires an estimate of the damage present, an assessment of the probable future loads, and an accurate model of how the damage may develop and the structure might fail. Although this process is very difficult, the use of inverse methods to generate a physically meaningful model offers a route to prognosis.

8. Conclusions

This paper has given a brief overview of the huge literature available on the approaches of damage identification based on inverse methods. The sensitivity-based methods to identify physical parameters using subset selection for error localization have been suggested as a viable approach. However, many difficulties remain to be fully solved, such as the modelling error between the model and the physical structure, and the influence of environmental factors. The most promising route is to include measurements of temperature, humidity and other environmental variables within the model, although this requires more stringent conditions on modelling error. At the very least, these errors give a lower bound on the level of damage that can be detected and localized, and this can be formalized using statistics from the response of the undamaged structure.
in its normal operating environment. The application of robust damage detection and location algorithms based on monitoring the in-service response of a structure remains a challenge, although the availability of a model does open the way to more accurate prognosis and the estimation of the remaining life.

The author gratefully acknowledges the support of the Royal Society through a Royal Society-Wolfson Research Merit Award.

References


Phil. Trans. R. Soc. A (2007)


Damage identification using inverse methods


Phil. Trans. R. Soc. A (2007)

Downloaded from http://rsta.royalsocietypublishing.org/ on July 4, 2017


