Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Computers and Structures 81 (2003) 1149–1176
Computers & Structures
www.elsevier.com/locate/compstruc

Structural system identification: from reality to models

K.F. Alvin [a], A.N. Robertson [b], G.W. Reich [c], K.C. Park [d,*]

[a] Structural Dynamics Research, Sandia National Laboratories, Albuquerque, NM 87185, USA
[b] Los Alamos National Laboratory, ESA-WR, MS P946, Los Alamos, NM 87545, USA
[c] Air Force Research Laboratory, Air Vehicles Directorate, 2210 Eighth St., Room 219, WPAFB, OH 45433, USA
[d] Department of Aerospace Engineering Sciences, Center for Aerospace Structures, University of Colorado, Campus Box 429, Boulder, CO 80309-0429, USA

Abstract

The paper is an expository contribution on the subject of structural system identification, measured signal processing and their applications to model-based structural health detection. The materials covered in this paper are by and large extracted from the three theses of the first three authors. The paper focuses on the state-space oriented system identification theory as specialized to structural dynamics governing equations of motion, a judicious use of wavelet transformation techniques for extracting impulse response functions, various input–output combinations for multi-input and multi-output problems, robust ways of identifying both proportional and non-proportional damping parameters, and the use of localized identification theory for damage detection from measured response data. The authors then offer several outstanding challenges in structural system identification theory and their applications.

1. Introduction: background on structural system identification

1.1. Modal testing for model validation

The experimentally validated modeling of complex structures subjected to dynamic loading and active controls has been a subject of intense research for the last three to four decades. While the variety and efficiency of finite element modeling and analysis tools have undergone enormous growth over the same period, systematic improvements in the accuracy and reliability of finite element analysis for structural dynamics have proven much more difficult to achieve. This is due in part to the governing assumptions of linearity in the analysis, the statistical variance in the properties of structural components due to fabrication tolerance, and the increased heterogeneity of the structures being modeled. It is also true, however, that much of the limitation of model accuracy is due to the ever-growing complexity of lightweight structures, including articulating joint designs and the introduction of new hybrid materials. Even a high degree of modeling precision cannot compensate for a lack of experience with the real measured behavior of complex structures.

Test-validated finite element models have therefore played a key role in the design of high performance and high reliability structures. Even as testing validates current analysis methodologies, it stimulates an enhanced understanding of complex dynamic behavior. Model validation testing is particularly critical in the design qualification of structures. This is because verification of structural strength and dynamic performance, even under extensive environmental testing, remains strongly dependent on the choice of analytical models. For example, the specification of test load levels for design qualification and the determination of appropriate test boundary conditions depend on a reasonably accurate finite element model of the tested subsystem. Correlation of finite element models is therefore a necessary component of progress in the design and optimization of structures subject to dynamic excitation.

* Corresponding author. Tel.: +1-303-492-6330; fax: +1-303-492-4990.
E-mail address: kcpark@titan.colorado.edu (K.C. Park).

0045-7949/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0045-7949(03)00034-8

Modal testing [1] is perhaps the most versatile form of structural model validation testing. Modal testing is also problematic because, unlike mass property measurements or static influence coefficient testing, modal tests capture global response measures which characterize mass, damping and stiffness behavior simultaneously. With the development of digital signal processing (DSP) techniques, and in particular the fast Fourier transform (FFT), frequency domain methods in modal testing became competitive with characterizing the modal properties of structures [2]. With the use of multiple simultaneous random excitations with multiple output readings, closely-spaced modes were able to be identified, shifts caused by the movement of actuators was alleviated, and random inputs enabled all modes in the system to be excited simultaneously. All of these attributes are the reason that multiple input/output (MIMO) testing is the most prominent method used today. Single input methods and sine dwelling are also still in use, but not to as large an extent. Another more recent approach is the incorporation of natural excitation in modal testing. In some cases, it is either impractical or impossible to use artificial inputs to excite the system, so natural excitation must be measured along with the system response to assess the dynamic characteristics of the structure.

1.2. Estimation of response functions from measured data

The identification or extraction of modal parameters from the measured force inputs and sensor outputs is a complex process. In order to determine modal parameters, the frequency response functions (FRFs), or equivalently the impulse response functions in the time domain (sometimes referred to as Markov parameters), must first be estimated from the available vibrations records for the excitations and sensors. In a traditional modal testing laboratory, the preferred approach is to estimate FRFs using very long time histories from random inputs. This allows for the time records to be partitioned into numerous independent test records for ensemble averaging purposes. Alternatively, if hammer excitations are used, many tests will be performed in order to obtain sufficient data for averaging. Auto and cross-spectral densities of the inputs and outputs are then computed from discrete Fourier transforms of each test record, and averaged over all the tests. This frequency domain-based approach is efficient and generally effective in such a laboratory setting, when large amounts of data are available.

There are other drawbacks of the frequency domain approach, apart from the aforementioned need for large amounts of data or tests. The first is that the discrete forward and inverse Fourier transforms engender a phenomenon termed leakage, in which the non-periodicity of the signal leads to corruption of the spectral density magnitudes. Windowing methods are used to offset this problem by smoothing the end of each function to zero, thus forcing the signals to be periodic. The use of windowing, however, has a deleterious effect on the modal parameter estimation, especially in the damping estimation. Second, in order for an FFT-based procedure to be practical, the input signals must have a rich frequency spectrum. Input signals such as simple harmonic oscillations have a sparse representation in the frequency domain causing numerical ill-conditioning in the computation of the FRFs. This necessitates the adoption of a random input, which is most commonly used in experimental testing today. However, this limits the ability of spectral methods to be used for such applications as on-line system identification where the input signal is the actual disturbance to the structure and cannot be specified. The third problem associated with spectral methods is aliasing, which occurs when time data are not sampled at a sufficient rate. In the transformation from the time to the frequency domain and back again, aliasing has a compounding effect. Filtering methods are often used to alleviate aliasing before the Fourier transform is performed, but can also detrimentally affect determination of the modal parameters.

The above-mentioned issues with the discrete forward and inverse Fourier transforms instigated research back into the use of time domain methods. These methods solve the deconvolution problem in the time domain through the creation of a convolution matrix, therefore eliminating the problems of leakage and aliasing. However, depending on the type of input used, the convolution matrix may be strongly ill-conditioned resulting in an unstable impulse response [3]. A number of algorithms [4–6] have been proposed to solve this problem, from which the best results have been obtained with the least-squares solutions [7]. Least-squares methods perform a singular value decomposition of the matrix from which certain values are truncated below a threshold. Other methods involve an observer based approach to by-pass the inversion of the large, possibly ill-conditioned convolution matrix, known as the observer/Kalman filter identification (OKID) method [8].

Other problems associated with time domain methods are noise and computation time. Time domain methods, like spectral methods, have the ability to average data to either help alleviate noise or make the system determinant, but this is not always sufficient. There is no systematic way of actually filtering the noise when averaging does not work. In addition, solving the deconvolution problem in the time domain is computationally expensive, taking several times as long as spectral methods. These weaknesses necessitate the investigation of new methods of extracting Markov parameters.

Recently, engineers and mathematicians have looked at a new method of analysis using a wavelet transform.

A wavelet is a mathematical basis function that breaks up a signal into its frequency components as a function of time [9]. It was developed independently as an analysis tool in several areas including mathematics, quantum physics, electrical engineering, and seismic geology [10]. Its ability to retain the time quality of a signal while also exhibiting frequency content makes it a prime candidate for use in signal analysis, specifically impulse response determination. Recent applications in this area include acoustical analysis, seismology, and structural vibration analysis [11].

1.3. Identification of models from response functions

There are numerous methods which can be applied to discrete FRFs or their associated impulse response functions (Markov parameters) in the time domain [12–16]. These methods have varying performance characteristics depending on the organization and condensation of data, methods of determining model size, ability to use multiple inputs and detect repeated modal frequencies, etc. The earliest approaches to model estimation involved circle-fitting methods [1], which yields the estimated frequencies, mode shapes and damping for a single FRF (i.e. a single input–output pair). More systematic and rigorous approaches for the single-input-multiple-output (SIMO) model estimation problem were eventually developed, starting with the Ibrahim time domain (ITD) [12] and complex exponentials [13], both of which operate in the time domain on the impulse response functions and determine least-squares solutions. It is not possible, however, to distinguish modes with close or identical frequencies using these methods, and so a Polyreference method [14], which determines a solution for the multiple-input-multiple-output (MIMO) model estimation problem, was developed. The damped modal characteristics are then determined from the poles and residues of the estimated model. Polyreference is similar to the least-square methods except for its multiple-input characteristics, and reduces to the complex exponential method in the case of a single input.

A key characteristic of these methods is that the model order, or the number of intrinsic modes, is not determined systematically by the model estimation solution. The estimated model will typically retain a higher order so that the modes of interest are identified accurately. Then the roots of the estimated model must be studied to determine which are structural and which are residual modes due to the high order of the model. Thus, in terms of linear systems theory, the estimated models lack properties such as model order minimization and model uniqueness. By formally addressing these properties through system realization theory, modern methods of model estimation have been developed which are advantageous for determining minimal-order realizations of MIMO systems. Notable amongst these modern model estimation techniques is the eigensystem realization algorithm (ERA) [15]. While most model estimation methods determine approximate models which fit the given measured data, the system realization theory that ERA is based on renders a minimum model order realization in the absence of noise. This characteristic enhances the application of ERA to the systematic identification needs of complex structures. System realization theory has been developed within the framework of linear dynamic systems analysis and control (see [17], for example). Realizations of linear systems are models which accurately express the system dynamics inherent in the transfer functions relating the system inputs and outputs [18–20]. State space realizations are relevant to modal testing because the first-order form encompasses all linear system behavior, including damped structural dynamics. Furthermore, the result of any model estimation method can be cast as a system realization of the measured data. Literature on system realization theory and its recent application to modal testing is extensive (see Bibliography in [21] and [22–24], among others).

1.4. Extraction of modal parameters from identified models

There are two fundamental types of modal parameters of interest in linear structural dynamics: normal modes and damped modes. Normal modes, alternatively referred to as undamped or classical modes, are intrinsic properties of the conservative system, i.e., they are the eigensystem characterized by the mass and stiffness matrices. The undamped eigenvalues Ω^2 and their mode shapes Φ can be predicted through finite element modeling and analysis. Although damping always exists in real systems, at present it is not accurately modeled or predicted. The normal modes, therefore, are of primary interest in model validation and structural analysis because they are the dynamic response components which are direct expressions of the system mass and stiffness without the influence of unmodeled damping. Complex modes, on the other hand, are the intrinsic modes of the damped structure, also referred to as damped or complex damped modes. The term complex modes refers to the complex roots of a first-order system of equations, which can be transformed into the vibrational characteristics of a damped structure [25]. The damped eigenvalues Λ and their damped mode shapes Ψ are complex quantities, and in general the real and imaginary components of the mode shapes are not linearly dependent [26,27].

The solution provided by model estimation methods determines the damped modes and complex mode shapes, but does not determine the normal modal parameters of the associated conservative structure. If the real and imaginary components of the damped mode shapes are collinear (i.e. linearly dependent), the system

is said to be proportionally or classically damped. In this particular case, the complex modes are also collinear with the normal mode shapes, and the complex and normal modal parameters are easily related to one another, mode by mode. Therefore, for proportional damping, the normal modal parameters are obtained directly from the damped modal parameters, thereby providing the necessary modal data for finite element model correlation or other applications. The model correlation process then proceeds by evaluating the normal modes of the finite element model and, possibly, calibrating the model at the element level or global matrix level to improve the correlation between the normal modes of the model and of the test.

Generally, however, the damping is non-proportional or non-classical and the normal modes which uncouple the mass and stiffness matrices do not simultaneously uncouple the physical damping matrix. The equivalence between the damped and normal modes is lost, and the modal parameters resulting from the finite element analysis are no longer directly comparable to the modal parameters obtained from dynamic testing. Therefore, there is a missing link between the intrinsic damped modes Λ and Ψ of the tested structure as obtained through modal testing and model estimation, and the intrinsic undamped modes Ω^2 and Φ of the finite element model. In a physical sense, the missing link is the damping, which is not accurately predicted from the finite element model, nor isolated by the system realization of the measured modal test data. One solution to this dilemma is to transform the state space-based system realizations which are characterized by Λ and Ψ to an equivalent second-order canonical basis. The equivalent second-order realization is significant in that it isolates the influence of damping into a damping matrix which can be used in finite element simulations, and simultaneously yields the undamped modal parameters Ω^2 and Φ of the structure for use in model correlation, controls design and damage detection.

One family of methods, which can be classified as mode shape estimation, uses each complex mode individually to determine an associated normal mode estimate. Thus, there is a one-to-one assumed relationship between the decoupled damped modes and the normal modes, and the resulting modal parameters are in effect a proportionally-damped estimate of the realized system. Mode shape estimation is characteristically straightforward, but the modal parameters cannot maintain system dynamics exactly equivalent to the complex modes model. A recent survey [28] summarizes the common practice in mode shape estimation of using just the modulus of the complex modes and dismissing the non-normal phase information, which was referred to as the standard technique (ST). A new iterative approach is introduced to maximize the modal assurance criterion (MAC) between the damped modes and normal modes estimates. Another recent approach [29] refers back to the measured peaks of the FRFs to extract real mode shape estimates using the eigenvalues (i.e. frequencies and damping rates) from Polyreference or ERA. In [30], an equivalent realization was shown for which the continuous-time ERA state space transition matrix is rotated to a real form representative of second-order dynamics, termed the McMillan normal form. This transformation-based approach provides real-valued mode shape estimation but does not account for the sensitivity to non-normal phase components in the complex modes. In other words, when the damping is non-proportional, the mode shape estimates will be dependent on the complex scaling of the damped modes, which is not generally defined in a physically-relevant or optimal sense. Hence, it is only accurate for proportional damping, for which the standard technique is adequate.

An alternative family of methods, which can be classified as mode shape-damping de-coupling, involves global transformations which attempt to fully uncouple the damping matrix in order to obtain the real modes [31–39]. These methods can correctly account for the nature of non-proportional damping, and generally compute both real mode shapes and revised damping and frequency estimates. The revised modal damping matrix is characteristically non-diagonal, and the generalized form of the new modal parameters make it possible to maintain exact system equivalence with the initial complex damped modal parameters. In [40], a solution for the mass, damping and stiffness matrices in terms of the ERA system realization is given when the number of sensors, actuators, and identified modes are all equal. This method also accomplishes a mode shape-damping de-coupling, though it is not a central point of the algorithm. An important limitation of all the existing methods for the mode shape-damping de-coupling problem is that they require as many sensors as identified modes in order to uniquely define the required coordinate transformations.

1.5. Application of structural system identification to health monitoring

An important application of structural system identification is health monitoring, also referred to as damage detection. This is the practice of conducting non-destructive tests or inspection of a structure to determine the existence, location, and extent of damage, and in some cases make a determination as to the remaining life of the system. This practice has taken on increased importance in aerospace and civil applications over the last several decades, due to the increased use of aircraft structures far beyond their original life expectancy. Additionally, as the civil infrastructure ages, the determination of the integrity of bridges, dams, and buildings

for continued safe usage becomes critical. Costs associated with inspection, maintenance, and system downtime also provide motivation for improved inspection and damage identification practices. For these reasons, new methods of structural health monitoring are being explored to better determine the functional safety of structures.

There are any number of methodologies for structural damage detection [41]. Of interest to this paper are methods that have their roots in structural system identification and modal testing. These vibration-based health monitoring algorithms can be broken down into the following two categories: non-model-based schemes and model-based schemes. Non-model-based schemes determine direct changes in the sensor output signal to locate damage in the structure [42]. These can be thought of as signal-processing solutions to the problem. Model-based techniques utilize changes in response functions or modal parameters such as natural frequencies, mode shapes, or their derivatives, to identify damage locations and levels. Analysis of changes seen in parameters between sequential tests over time is used to determine damage characteristics. An excellent review of model-based health monitoring methods has recently been compiled [43]. Some of the algorithmic approaches include mode shape-based metrics [44], flexibility methods [45] and other matrix perturbation approaches [46], structural control-based or filtering methods [47], and transmission zero monitors [48]. The major drawback to these methods is that the structural variability due to changing boundary conditions, environmental conditions, slightly different loadings, etc., can cause response characteristic changes that are larger than those caused by damage.

Model-based damage detection methods can also be categorized as to whether they deal with global or localized response characteristics. For example, in examining the change in a structure's modal parameters, a detection algorithm could either use the global modes directly, or could use a decomposition of the structural response through an auxiliary inverse method, and then examine the estimated modal characteristics of a single component of the structure. This localized approach frees the user from considering the entire system—the only requirement is to have enough measurements to characterize the substructure under consideration. The remainder of the system is only important indirectly. The localization process can be very effective at accentuating or exaggerating small changes in a structure. For example, a small crack in one truss member of a large structure is unlikely to cause a measurable change in the global fundamental bending mode of the system. However, the effect of that crack on the truss element itself can be seen through localization.

Methods that use the invariance properties of system transmission zeros are interesting in that they are one of the least developed approaches to damage detection. Until recently, only a single paper has suggested the use of system transmission zeros of the measured transfer function as an indicator of structural damage [49]. More recently, a methodology that combined structural flexibility decomposition with MIMO transmission zero monitoring has been developed [48]. The system response functions are partitioned according to an assumed connectivity, and the transmission zeros corresponding to collocated block-diagonal partitions are monitored for changes. A unique feature of this algorithm is that an invariance property of transmission zeros is utilized as the indication mechanism. As such, damage is identified by the substructure who's transmission zeros don't change when others do.

It is important to understand the differences between model-based damage detection methods that utilize stiffness and those that utilize flexibility. Theoretically, there is no difference between the two, since the flexibility matrix is simply the inverse of the stiffness matrix. Practically, however, there is a significant difference. Perturbations in the stiffness matrix affect the highest eigenvalues of the system first. However, in a modal test setting the low-frequency eigenvalues of the system are measured, and therefore the dominant perturbations of the stiffness matrix are not captured. On the other hand, perturbations in the flexibility matrix affect the highest flexibility eigenvalues, which correspond to the lowest system eigenvalues. Consequently, the measured flexibility matrix is better able to capture system perturbations than the measured stiffness matrix, where both are rank-deficient due to the limitations inherent in measuring high-frequency dynamics.

Model-based health monitoring methodologies are excellent tools in the model updating process. For this purpose, one set of data from the analysis model is compared to a set of data from the physical structure. The damage detection methodology is used to locate features of the model that do not match the experiment. Localized methods are much better suited for this purpose, as it is better to update individual elemental or substructural properties rather than global properties across the whole design. Analytical sensitivities of response parameters to changes in physical properties are used to update modeling assumptions, physical sizing, elastic moduli, etc.

2. Modeling of linear structural dynamics in the time domain and the frequency domain

This section reviews the theoretical basis of structural system identification. To this end, we introduce the equations of motion for structures and their transformation to state space canonical forms. The state space or first-order equation form is important because the

generally-damped behavior of the structural system can only be uniquely expressed from model realizations of experimental data in terms of the modal parameters of a first-order system of equations. A review of frequency-domain structural modeling is also presented.

2.1. State space formulations of structural dynamics

Typically, the equilibrium conditions for linear time-invariant continuum mechanics are discretized through spatial displacement interpolation to a finite number of variables (e.g. finite element methods), resulting in the familiar n-dimensional set of second-order linear differential equations

$$\begin{aligned} &M\ddot{q}(t) + D\dot{q}(t) + Kq(t) = \widehat{B}u(t) \\ &y(t) = H_{\mathrm{d}}q(t) + H_{\mathrm{v}}\dot{q}(t) + H_{\mathrm{a}}\ddot{q}(t) \end{aligned} \tag{1}$$

where M, D and K are the mass, damping and stiffness matrices, respectively; q is the n-displacement state vector; u is an m-input force vector; y is an l-sensor output vector, either displacement, velocity or acceleration; $\widehat{B}$ is the input-state influence matrix, and H_{d}, H_{v} and H_{a} are state-output influence matrices for displacement, velocity and acceleration, respectively. The undamped portion of this second-order system can be decoupled through an eigenvector change-of-basis $q(t) = \Phi\eta(t)$, resulting in

$$\begin{aligned} &\ddot{\eta}(t) + \Xi\dot{\eta}(t) + \Omega^2\eta(t) = \Phi^{\mathrm{T}}\widehat{B}u(t) \\ &y(t) = H_{\mathrm{d}}\Phi\eta(t) + H_{\mathrm{v}}\Phi\dot{\eta}(t) + H_{\mathrm{a}}\Phi\ddot{\eta}(t) \end{aligned} \tag{2}$$

where Φ is the mass-orthonormalized eigenvector basis for the generalized undamped eigenproblem

$$K\Phi = M\Phi\Omega^2$$

such that

$$\begin{aligned} &\Phi^{\mathrm{T}}M\Phi = I_{n\times n} \\ &\Phi^{\mathrm{T}}K\Phi = \Omega^2 = \mathrm{diag}\{\omega_{\mathrm{n}i}^2, i = 1,\ldots,n\} \\ &\Phi^{\mathrm{T}}D\Phi = \Xi \end{aligned} \tag{3}$$

where $\omega_{\mathrm{n}i}$ is the undamped natural frequency for mode i. In the case of Rayleigh damping ($D = \alpha M + \beta K$ or more generally $D = \sum \alpha_i M^{-1}K^i$ [25]), this transformation will decouple the damped second-order system, and the modal damping matrix Ξ is given as

$$\Xi = \mathrm{diag}\{2\zeta_i\omega_{\mathrm{n}i},\ i = 1,\ldots,n\} \tag{4}$$

Here ζ_i is the modal damping ratio for mode i, which varies from 0% for undamped vibration to 100% for critically-damped vibration, at which point the system response for the mode becomes non-oscillatory.

2.1.1. General state space realizations

It is also possible to retain model equivalence while transforming the equations of motion from second-order to first-order differential form. The general form of a linear, time-invariant state space realization is

$$\begin{aligned} &\dot{x}(t) = Ax(t) + Bu(t) \\ &y(t) = Cx(t) + Du(t) \end{aligned} \tag{5}$$

where x is the $N \times 1$ state vector, A is the $N \times N$ state transition matrix, B is the $N \times m$ input-state influence matrix, C is the $l \times N$ state-output influence matrix, and D is the $l \times m$ matrix corresponding to direct input/output feedthrough. For structural dynamics, $u(t)$ are generally externally applied forces as in (1) and B is an array which maps the physical locations of the input forces to the internal variables of the realization. Similarly, $y(t)$ are physical sensor measurements such as displacement, velocity or acceleration and C is an array which constructs these physical quantities from the internal variables $x(t)$. For example, if $x(t)$ is a vector of physical displacements and velocities which include as a subset those degrees of freedom driven by u or measured by y, then B and C are typically binary arrays (i.e. values of 0 or 1). The state space formulation of structural dynamics requires $N = 2n$ states to equivalently represent the second-order system (1). Specific state space realizations for structural dynamics will be presented later in this section.

2.1.2. State space realizations for structural dynamics: physical variables

A family of state space realizations of (1) can be expressed through a generalized momentum variable v (see [50]). Define v as

$$v(t) = E_1\dot{q}(t) + E_2q(t) \tag{6}$$

and the state space basis as

$$x(t) = \begin{bmatrix} q(t) \\ v(t) \end{bmatrix} \tag{7}$$

Rather than determining the general state space equations for x, one special basis choice will be considered:

$$\mathrm{I}: \quad E_1 = M^{-1}, \quad E_2 = 0 \tag{8}$$

Since $v = \dot{q}$, the state space basis can be termed a physical displacement–velocity (PDV) model. The resultant model is given by

$$\begin{aligned} &\dot{x}_I(t) = A_Ix_I(t) + B_Iu(t) \\ &y(t) = C_Ix_I(t) + Du(t) \end{aligned} \tag{9}$$

where

$$A_I = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix}, \quad B_I = \begin{bmatrix} 0 \\ M^{-1}\widehat{B} \end{bmatrix}$$
$$C_I = [H_d \quad 0] + [H_v \quad 0]A_I + [H_a \quad 0]A_I^2$$
$$= \left[H_d - H_a M^{-1}K \quad H_v - H_a M^{-1}D\right] \tag{10}$$
$$D = H_a M^{-1}\widehat{B}$$

2.1.3. State space damped modal realizations

As noted previously, for systems with Rayleigh damping, or where $D = \sum \alpha_i M^{-1}K^i$, the modal equations (2) are decoupled, as Ξ is diagonal. This form of damping is referred to variously as diagonal, modal, classical, or proportional damping. When Ξ is not diagonal, the damping is termed non-classical or non-proportional, due to the fact that D cannot be expressed as a proportional combination of M and K. One physical interpretation of this type of damping is that the modes which decouple the system equations are now complex, such that there are phase relationships between the various physical displacements of the structure within each mode. Another interpretation is that the classical undamped modes of the structure are energy-coupled through the off-diagonal terms of Ξ.

Because the second-order equations of motion for non-proportionally-damped systems cannot be decoupled by the normal modes of M and K, the governing eigenproblem for non-proportional damping must be formulated from equivalent first-order equations of motion. To this end, the PDV model (9) and (10) with displacement sensing is rewritten in a symmetrical companion form as

$$\begin{bmatrix} D & M \\ M & 0 \end{bmatrix}\begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} -K & 0 \\ 0 & M \end{bmatrix}\begin{bmatrix} q \\ \dot{q} \end{bmatrix} + \begin{bmatrix} \widehat{B} \\ 0 \end{bmatrix}u$$
$$y_d = [H_d \quad 0]\begin{bmatrix} q \\ \dot{q} \end{bmatrix} \tag{11}$$

such that

$$\begin{bmatrix} X \\ X\Lambda \end{bmatrix}^T \begin{bmatrix} D & M \\ M & 0 \end{bmatrix}\begin{bmatrix} X \\ X\Lambda \end{bmatrix} = I$$
$$\begin{bmatrix} X \\ X\Lambda \end{bmatrix}^T \begin{bmatrix} -K & 0 \\ 0 & M \end{bmatrix}\begin{bmatrix} X \\ X\Lambda \end{bmatrix} = \Lambda \tag{12}$$

whose equivalent normalized damped modal realization may be given by

$$\dot{z}_n = \Lambda z_n + X^T\widehat{B}u$$
$$y_d = H_d X z_n \tag{13}$$

Here the columns of X are the complex damped mode shapes normalized with respect to the physical properties of the structure, and possibly possessing phase relationships within each decoupled mode between spatially distinct points. The symmetry of the damping matrix and the resultant generalized eigenproblem is important because they imply reciprocal behavior between spatially-distinct points. This property of self-adjoint structural systems is useful for identification because it allows the sensor mode shapes and the modal participation factors of the input forces to be treated as equivalent. For systems possessing gyroscopic damping, for example, it is necessary to apply input forces at every sensor degree of freedom in order to discern the structural properties. This is because the right and left eigenvectors are no longer equivalent.

In the case of proportional damping, complex mode shapes and eigenvalues are still obtained when posing the first order eigenproblem as above, but the complex and real modal quantities are directly related to one another for each mode i as

$$\sigma_i = -\zeta_i\omega_{ni} \quad \omega_i = \omega_{ni}\sqrt{1-\zeta_i^2}$$
$$\Re(X_i) = \frac{1}{2\sqrt{\omega_i}}\phi_i \quad \Im(X_i) = -\left(\frac{1}{2\sqrt{\omega_i}}\right)\phi_i \tag{14}$$

where ϕ_i and X_i are normalized as in (3) and (12), respectively.

The intrinsic modes for non-proportional damping are different from those of classically-damped systems. The preceding discussion points out both the distinctions between the undamped and damped eigenproblems and their respective modal parameters, and the importance of the first-order (or state space) form in the behavior of damped structural dynamics.

3. Methods for determining frequency response and impulse response functions

The dynamic properties of a linear, time invariant system can be described by the impulse response function, $h(t)$, which is defined as the output of a system to a unit impulse input. The importance of the impulse response function is that for any arbitrary input $u(t)$, it allows the output of a system $y(t)$ to be expressed by the convolution integral:

$$y(t) = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)\,d\tau \tag{15}$$

If the system is physically realizable, no response will occur until after the input is applied:

$$h(\tau) = 0 \quad \text{for } \tau < 0 \tag{16}$$

If the system also operates for only a finite amount of time, $0 \leqslant \tau \leqslant T$, the convolution equation can be rewritten as

$$y(t) = \int_0^{\mathrm{T}} h(\tau)u(t-\tau)\,\mathrm{d}\tau \tag{17}$$

The task of any impulse response extraction method is to deconvolve the known input from the resultant output. This may be done by either leaving the equation in the time domain and solving a set of linear equations, or by transforming the equation into another basis, such as by Fourier or wavelet transform. The following sections will review the more traditional methods of solving the convolution equation and introduce a new wavelet based procedure.

3.1. Conventional time domain extraction

Time domain procedures attempt to solve Eq. (22) by constructing a matrix U, representing the convolution operator applied to the input. In so doing, the convolution equation can be represented in matrix form as

$$Y = hU \tag{18}$$

where the output matrix Y, the time-domain impulse response matrix h and the input matrix U are given, respectively, by

$$\begin{aligned} Y &= [\,y(0) \quad y(1) \quad \cdots \quad y(s-1)\,](m \times s) \\ h &= [\,h(0) \quad h(1) \quad \cdots \quad h(rp)\,](m \times r(p+1)) \\ U &= \begin{bmatrix} u(0) & u(1) & \cdots & u(p) & \cdots & u(s-1) \\ 0 & u(0) & \cdots & u(p-1) & \cdots & u(s-2) \\ 0 & 0 & \cdots & u(p-2) & \cdots & u(s-3) \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & u(0) & \cdots & u(s-p-1) \end{bmatrix} \\ &\quad (r(p+1) \times s) \end{aligned} \tag{19}$$

in which m, s, r and p are the number of measurement vectors, the number of measurement samples, the number of input signals, and the length of the impulse response, respectively. The discrete deconvolution problem is then essentially reduced to solving a set of linear equations given by Eq. (18), or in more traditional form:

$$U^{\mathrm{T}}h^{\mathrm{T}} = Y^{\mathrm{T}} \tag{20}$$

The direct method of solving (18) is to multiply both sides of the equation by U^{-1}. Unfortunately, depending on the kind of input used, the matrix U may be strongly ill-conditioned, leading to an exploding impulse response. Also, for multiple input systems, the additional DOFs are stacked on top of one another forming a matrix U that has more rows than columns, meaning the matrix is underdetermined. Most cases therefore require the use of a pseudo-inverse rather than a direct one to overcome the underdetermined state, performed through the creation of auto and cross correlation functions:

$$\begin{aligned} &\text{cross correlation:} \quad R_{YU} = YU^{\mathrm{T}} \\ &\text{auto correlation:} \quad R_{UU} = UU^{\mathrm{T}} \end{aligned} \tag{21}$$

where h is then found by

$$h = R_{YU}R_{UU}^{-1} \tag{22}$$

which is also a definition of a pseudo-inverse.

Conventional methods attempt to overcome some of these problems by ensemble-averaging and linear least-squares methods. Ensemble-averaging helps to make the problem over-determined, while linear least-squares truncates all singular values below a certain threshold in an attempt to help the conditioning of the matrix. Presumably the truncated values correspond to low level noise present in the system, otherwise truncation may only worsen the results.

Even when the U matrix is invertible, inversion times can become quite large, making this procedure computationally inefficient. For this reason, other procedures attempt to reduce the size of the matrix to be inverted or perform deconvolution in another set of basis where computational efficiency can be greatly improved.

3.2. FFT-based extraction

Spectral methods of impulse response extraction attempt to solve Eq. (17) by transforming the convolution equation to the frequency domain. The continuous Fourier transform is defined as

$$X(f) = \int_{-\infty}^{\infty} x(t)\,\mathrm{e}^{-\mathrm{j}2\pi ft}\,\mathrm{d}t \tag{23}$$

where $x(t)$ is a continuous time function and $X(f)$ its continuous Fourier transform. In the frequency domain, transforms of the input u and output y are related through a transfer function or frequency response function $H(f)$ given by

$$Y(f) = H(f)U(f) \tag{24}$$

The time domain convolution is now simply a multiplication in the frequency domain, making the determination of Markov parameters a much easier task. Instead of matrix inversion, the solution of (24) is found by vector division.

However, due to problems associated with noise, Eq. (24) must be modified slightly. Using the discrete Fourier transform of the input and output, the auto and cross spectral densities, which are the Fourier transforms of the auto and cross correlations found in (21) are given:

$$G_{uu}(f_k,m) = \hat{u}^*(f_k,m)\cdot\hat{u}(f_k,m)$$
$$G_{yu}(f_k,i,m) = \hat{y}^*(f_k,i)\cdot\hat{u}(f_k,m) \quad (25)$$

where $\hat{u}(f_k,m)$ and $\hat{y}(f_k,i)$ are the Fourier transformed input and output data, and $\hat{u}^*(f_k,m)$ and $\hat{y}^*(f_k,i)$ are their complex conjugates, with m referring to the input number and i referring to the output number. The frequency response function can then be given by

$$H(f_k,i,m) = G_{yu}(f_k,i,m)/G_{uu}(f_k,m) \quad (26)$$

After the FRFs are found, the Markov parameters of the system can be determined by transforming back to the time domain via the inverse discrete Fourier transform:

$$h(t) = \mathrm{IDFT}(H(f_k)) \quad (27)$$

As mentioned previously, spectral methods offer a much more computationally efficient procedure for finding the Markov parameters of the system due to the transformation of time convolution to multiplication in the frequency domain. In addition, the use of the FFT [51], a computationally efficient algorithm, can help to improve time efficiency even more.

There are also some drawbacks, however, to using spectral methods. Foremost is the loss of accuracy engendered with FFTs and IFFTs. In addition, there is the critical shortcoming of the division by G_{uu} to obtain $H_{yu}(f_k,i,m)$. For example, if $u(i)$ consists of only a single or a few frequencies, then the resulting G_{uu} is often badly conditioned, leading to an erratic set of temporal impulse response data. Therefore, $\hat{u}(f_k,m)$ should consist of a wide band of frequencies in order to render a rich frequency spectrum in the structural responses and thus avoid the leakage induced ill-conditioning of G_{uu}.

3.3. Impulse response extraction using wavelet transforms

In contrast to the FFT-based extraction procedure which must process the data both in the time and frequency domains, the discrete wavelet transform (DWT)-based extraction procedure handles the experimental data only in the time domain, from the beginning to the end [52]. This involves the forward and inverse DWT plus an inversion operation, which is indeed preferable. In addition, it is able to obtain reasonable Markov parameters for even a single frequency input. The ability to use a variety of inputs is an asset for real-time health monitoring of structures such as operational airplanes, bridges and offshore oil-drilling platforms, where one must use the disturbance sources as the input, and the sensor output corresponding to the on-site disturbances as the response data. In other words, where one must deal with the structural responses as they occur without the use of random excitations which are often used in laboratory structural system identification experiments.

3.3.1. Wavelet transform background

The best way to introduce wavelets is through their comparison to Fourier transforms, a common signal analysis tool. Wavelet and Fourier transforms represent a signal through a linear combination of their basis functions. For Fourier transforms, the basis functions are dilations of cosine and sine signals (each spanning the entire time interval). For wavelet transforms, they are different translations and dilations of one function termed the Mother wavelet, along with a scaling function, (each spanning a logarithmically reduced subinterval). The dilations of both sets of basis functions are possible due to their frequency localization, thus allowing us to obtain frequency information about the signal being analyzed. This leads to the most important difference between the two sets of basis functions, time localization. The wavelet transform basis functions are compact, or finite in time, while the Fourier sine and cosine functions are not. This feature allows the wavelet transform to obtain time information about a signal in addition to frequency information.

Fourier transforms are also capable of obtaining time information about a signal if a windowing procedure is used to create a short time Fourier transform (STFT). The window is a square wave which truncates the sine or cosine function to fit a window of a particular width. Since the same window is used for all frequencies, the resolution is the same at all positions in the STFT time-frequency plane as seen in Fig. 1. The discrete wavelet transform (DWT), on the other hand, has a window size that varies with frequency scale. This is advantageous for the analysis of signals containing both discontinuities and smooth components. Short, high frequency basis functions are needed for the discontinuities, while at the same time, long low frequency ones are needed for the smooth components. This is exactly the type of time-frequency tiling obtainable from wavelet transforms [10]. Fig. 1 depicts this relationship by showing how the time resolution gets finer as the scale (or frequency) increases. Each basis function is represented by a tile, where the shading corresponds to the value of the expansion coefficient.

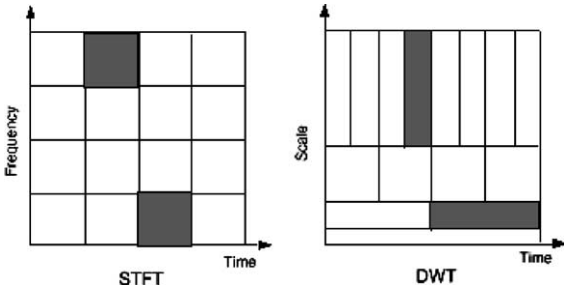

Fig. 1. STFT and DWT frequency/time tiling.

Unlike the Fourier basis functions, there are an infinite number of possible sets of wavelet basis functions. A wavelet is formed from a set of filter coefficients that must satisfy a given set of conditions [53]. Any set of filter coefficients which satisfy the given conditions can be used to create a wavelet function.

We can now see clearly an important comparison between the Fourier and the wavelet expansions of $f(t)$. The Fourier expansion may be advantageous in capturing frequency characteristics [54] in $f(t)$ whereas the wavelet expansion directly captures the temporal properties of $f(t)$. Hence, in a typical signal processing of vibration data, the discrete Fourier expansion of $f(t)$ involves first the convolution integral in the frequency domain, usually via FFT, and then an inverse FFT. In other words, in the Fourier expansion the data must be transformed from the time domain into the corresponding frequency domain and then converted again back into time domain. On the other hand, the wavelet expansion preserves the temporal nature of the data while also showing frequency content during both the forward and inverse wavelet transforms.

3.3.2. Basic wavelet algorithm for extraction of the impulse response function

The DWT method for impulse response function extraction starts with the same convolution integral as the time domain approach with some alterations:

$$y(t) = \int_0^{\mathrm{T}} h(\theta)u(t-\theta)\,\mathrm{d}\theta \tag{28}$$

where $h(\theta)$ is the temporal impulse response function. Note that the impulse response function is an intrinsic function of a specific system, $h(t) = \mathrm{e}^{At}$. Therefore, the output at the discrete time t_{d} is influenced only by the nature of the input (the forcing function). This observation enables us to expand $h(\theta)$ in terms of the wavelet basis functions for the entire response time interval, $\{0 \leqslant t \leqslant T \Rightarrow 0 \leqslant \theta \leqslant 1\}$:

$$h(\theta) = h_0^{\mathrm{DWT}} + \sum_j \sum_k h_{(2^j+k)}^{\mathrm{DWT}} \varphi(2^j\theta - k) \tag{29}$$

where h^{DWT} is the discrete wavelet transform of $h(\theta)$.

For the DWT characterization of $u(t-\theta)$, first $u(\theta)$ is reversed in time to obtain $u(-\theta)$, then it is shifted toward the positive time axis by the amount t with $\{u(\theta) = 0, \text{for } (\theta > t)\}$. With this convention, the DWT series of $u(t-\theta)$ is expressed as

$$u(t-\theta) = u_0^{\mathrm{DWT}} + \sum_j \sum_k u_{(2^j+k)}^{\mathrm{DWT}} \varphi(2^j\theta - k) \tag{30}$$

where u^{DWT} is the discrete wavelet transform of $u(t-\theta)$.

After the substitution, the following formula is obtained for one point in time, $t = t_0$ [11]:

$$y(t_0) = h_0^{\mathrm{DWT}} u_0^{\mathrm{DWT}} + \sum_j \sum_k \left(\frac{1}{2^j}\right) h_{(2^j+k)}^{\mathrm{DWT}} u_{(2^j+k)}^{\mathrm{DWT}} \tag{31}$$

The preceding expression can be represented as

$$\begin{aligned} y(t_0) &= h^{\mathrm{DWT}} u^{\mathrm{DWT}} \\ h^{\mathrm{DWT}} &= [\,h_0 \quad h_1 \quad h_2 \quad h_3 \quad \cdots \quad h_{n-1}\,] \\ (u^{\mathrm{DWT}})^{\mathrm{T}} &= [\,u_0 \quad u_1 \quad u_2/2 \quad u_3/2 \quad \cdots \quad u_{n-1}/2^j\,] \end{aligned} \tag{32}$$

where h^{DWT} is the wavelet transform of $h(\theta)$ and $u^{\mathrm{DWT}}(\theta)$ is the wavelet transform of $\{u(t-\theta), 0 \leqslant \theta \leqslant t\}$ with $\{j = \log_2(n)\}$. It should be noted that the preceding convolution integral does not require the restriction $\{h(\theta) \approx 0, \text{for } \theta > 1\}$ as implied in [11].

Eq. (32) gives a representation of the output at one point in time. For the entire response data, viz. $\{y(t_0), y(t_1), y(t_2), \ldots\}$, one can arrange the input and output relation in the following matrix equation:

$$\begin{aligned} \underset{(m\times s)}{Y} &= \underset{(m\times rl)}{h^{\mathrm{DWT}}}\,\underset{(rl\times s)}{U^{\mathrm{DWT}}} \\ Y &= \{y(t_0), y(t_1), y(t_2), \ldots\} \\ U^{\mathrm{DWT}} &= \{u^{\mathrm{DWT}}(0), u^{\mathrm{DWT}}(1), u^{\mathrm{DWT}}(2), \ldots\} \end{aligned} \tag{33}$$

where m, s, r and l are the number of measurement vectors, the number of measurement samples, the number of input signals, and the depth of the wavelet transform levels, respectively.

Solving for h^{DWT} from the above relation, one obtains

$$h^{\mathrm{DWT}} = Y^{\mathrm{DWT}}(U^{\mathrm{DWT}})^{\mathrm{T}}(U^{\mathrm{DWT}}(U^{\mathrm{DWT}})^{\mathrm{T}})^{-1} \tag{34}$$

and finally the inverse DWT of h^{DWT} yields the desired temporal impulse response data:

$$\{h(t),\ t = t_0, t_1, \ldots, t_{(s-1)}\} = \mathrm{IDWT}\{h^{\mathrm{DWT}}\} \tag{35}$$

Remark 1. The wavelet-transformed input coefficients $\{u^{\mathrm{DWT}}(k),\ k = 0, \ldots, (l-1)\}$ consist of a set of orthogonal local basis functions. Hence, when $rl \leqslant s$, the rank of U^{DWT} would be (rl) provided the multiple input vectors $u(t)$ of size $(r \times 1)$ are linearly independent. Therefore, $(U^{\mathrm{DWT}}(U^{\mathrm{DWT}})^{\mathrm{T}})$ is invertible. When $rl \geqslant s$, which would be the case for a full-depth wavelet resolution, one must seek a least-squares solution using $(U^{\mathrm{DWT}}(U^{\mathrm{DWT}})^{\mathrm{T}})$ as is done in Eq. (34).

The foregoing properties of the present wavelet method imply that the wavelet-based extraction of impulse response functions would not have to be subjected to the requirement $\{h(t_p) \approx \Longleftrightarrow CA^{(p-1)}B \approx 0\}$ that must be imposed on existing time-domain techniques. Consequently, $\{h^{\mathrm{DWT}}(k),\ k = 0, \ldots, r(l-1)\}$ consists of linearly independent basis functions.

Remark 2. On the other hand, one must resort in general to a pseudo-inverse of the input matrix U^+ for time-domain techniques. Experience shows that the rank deficiency in U^+ in the time-domain method or ill-conditioning of:

$$H_{yu}(\omega,i,j)=\frac{\hat{y}^*(\omega,i)\cdot\hat{u}(\omega,j)}{\hat{u}^*(\omega,j)\cdot\hat{u}(\omega,j)} \tag{36}$$

in the FFT-based method adversely influences the low-mode or static deformation amplitudes and phases. On the other hand, the wavelet-based method does not infuse any rank-deficiency, a key attribute of the proposed method.

Remark 3. The one major drawback of the DWT extraction method is the computation time required. Each row of the U matrix must be wavelet transformed, for a total of (rl) transformations. It can be argued, however, that the Markov parameter determination as a part of the overall system identification process is but a minute part of the over-all time required. Therefore, a time increase for this section would not significantly affect the time needed for the total process.

3.3.3. Wavelet algorithm improvements

The procedure presented above will be termed the basic wavelet algorithm for determining the Markov parameters of a system. It is basic in the sense that the data are not conditioned in any way. More traditional methods of system identification, such as spectral methods, rely heavily on data conditioning from both filtering and ensemble-averaging [1]. With this in mind, improvements to the basic wavelet algorithm were made in an attempt to achieve similar results.

Ensemble-averaging is a method of taking multiple input and output time series and averaging them together. Each time series ensemble may either come from extracts of one long input/output time history or from multiple tests. Most averaging is then done on the cross and auto correlations of the input/output data, such as the FFT method that averages the signal's spectral densities. With the wavelet method there are actually a variety of ways to determine the Markov parameters and subsequently a variety of ways to average the cross and auto correlation matrices.

If the basic method, as explained above, is being used, one would start out with a description of the wavelet transformed matrices:

$$\begin{aligned} h &= h^{\mathrm{DWT}}\cdot W^{\mathrm{T}} \\ U &= W\cdot U^{\mathrm{DWT}} \end{aligned} \tag{37}$$

where W is a wavelet transformation matrix that is applied to the columns of a matrix. W^{T} is the transpose of this matrix and represents the application of the wavelet transform to the rows of the preceding matrix. The rows or the columns of these two matrices consist of the basis functions of the type of wavelet being used. The previous equations are then substituted into Eq. (18) to obtain a representation of the output:

$$Y=h^{\mathrm{DWT}}\cdot W^{\mathrm{T}}\cdot W\cdot U^{\mathrm{DWT}}=h^{\mathrm{DWT}}\cdot D\cdot U^{\mathrm{DWT}} \tag{38}$$

This equation is the same as (33) with the D being a diagonal matrix of scaling values resulting from the multiplication of $W^{\mathrm{T}}\cdot W$, as shown in the description of $(u^{\mathrm{DWT}})^{\mathrm{T}}$ (32). This results when the wavelet matrix is orthogonal, but not orthonormal. Ensemble-averaging is then applied to this equation by first multiplying both sides of the equation by the transpose of the input matrix, to create the auto, $\{U^{\mathrm{DWT}}\cdot(U^{\mathrm{DWT}})^{\mathrm{T}}\}$, and cross, $\{Y\cdot(U^{\mathrm{DWT}})^{\mathrm{T}}\}$, correlation matrices, and then summing them together:

$$\sum_{i=1}^{\mathrm{nens}}[Y_i(U_i^{\mathrm{DWT}})^{\mathrm{T}}]=h^{\mathrm{DWT}}\cdot D\cdot\sum_{i=1}^{\mathrm{nens}}[U_i^{\mathrm{DWT}}(U_i^{\mathrm{DWT}})^{\mathrm{T}}] \tag{39}$$

h is then found by

$$\begin{aligned} h^{\mathrm{DWT}} &= \left\{\sum_{i=1}^{\mathrm{nens}}[Y_i(U_i^{\mathrm{DWT}})^{\mathrm{T}}]\right\}\cdot\left\{\sum_{i=1}^{\mathrm{nens}}[U_i^{\mathrm{DWT}}(U_i^{\mathrm{DWT}})^{\mathrm{T}}]\right\}^{-1} \\ &\quad\cdot D^{-1} \\ h &= \mathrm{IDWT}(h^{\mathrm{DWT}}) \end{aligned} \tag{40}$$

To achieve a representation of the time correlation function in the wavelet domain, transforming h and U is not the only possible method. Other methods include the wavelet transformation of Y in place of either U or h, two dimensional transformations, and wavelet transformations of the auto and cross correlations themselves. The last method mentioned is implemented by first finding the auto and cross correlations in the time domain:

$$\begin{aligned} Y &= h\cdot U \\ Y\cdot U^{\mathrm{T}} &= h\cdot U\cdot U^{\mathrm{T}} \end{aligned} \tag{41}$$

and then substituting the following representation for the auto and cross correlations in the wavelet domain:

$$\begin{aligned} Y\cdot U^{\mathrm{T}} &= [Y\cdot U^{\mathrm{T}}]^{\mathrm{DWT}}\cdot W^{\mathrm{T}} \\ U\cdot U^{\mathrm{T}} &= [U\cdot U^{\mathrm{T}}]^{\mathrm{DWT}}\cdot W^{\mathrm{T}} \end{aligned} \tag{42}$$

back into Eq. (41) to obtain:

$$\begin{aligned} [Y\cdot U^{\mathrm{T}}]^{\mathrm{DWT}}\cdot W^{\mathrm{T}} &= h\cdot[U\cdot U^{\mathrm{T}}]^{\mathrm{DWT}}\cdot W^{\mathrm{T}} \\ [Y\cdot U^{\mathrm{T}}]^{\mathrm{DWT}} &= h\cdot[U\cdot U^{\mathrm{T}}]^{\mathrm{DWT}} \end{aligned} \tag{43}$$

The ensembling process and then averaging is done by

$$\sum_{i=1}^{\mathrm{nens}}[Y\cdot U^{\mathrm{T}}]_i^{\mathrm{DWT}}=h\cdot\sum_{i=1}^{\mathrm{nens}}[U\cdot U^{\mathrm{T}}]_i^{\mathrm{DWT}} \tag{44}$$

Note that in this case there is no scaling matrix, D, since there is a multiplication of $W^{\mathrm{T}} \cdot (W^{\mathrm{T}})^{-1}$ and not $W^{\mathrm{T}} \cdot W$. These two ensembling methods presented produce slightly different results when only a few ensembles are used, but converge to the same solution as nens $\to \infty$. The differences of these two methods become important in the filtering process, a task that will not be addressed in the present paper. It should be noted that ensembling of the Markov parameters, h, themselves is possible, but will not improve results.

3.3.4. Example: four DOF spring–mass system

A four DOF spring–mass system will now be examined to show the ability of the wavelet procedure for impulse response extraction. The system as shown in Fig. 2 has four inputs and four outputs, one at each of the masses. The masses are all valued at 1 and the spring stiffnesses at 10 000 each.

For this example, the system was excited by a random input at all four nodes. Fig. 3 shows that the Markov parameters determined via the WT method are slightly better than the FFT-based ones. When a harmonic input is used, the superiority of the WT method is more apparent.

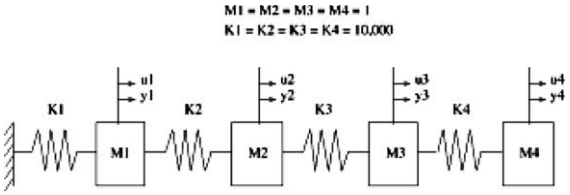

Fig. 2. Four DOF spring–mass system.

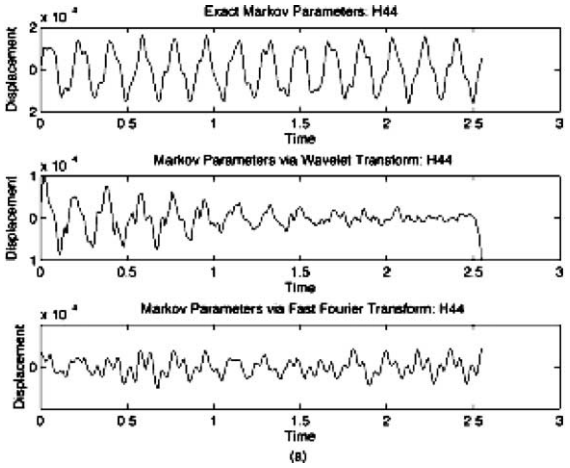

Fig. 3. Markov parameters of a four DOF system with 1 ensemble.

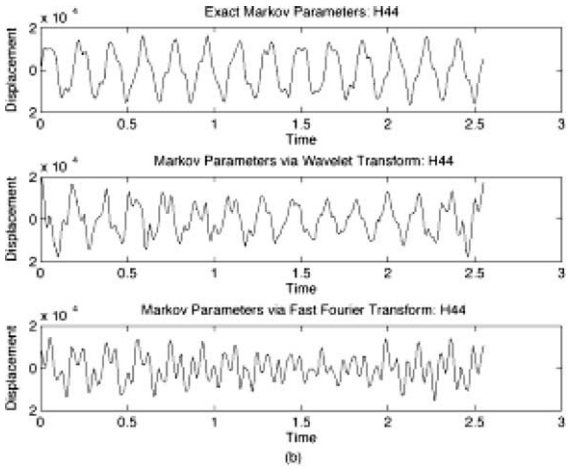

Fig. 4. Markov parameters of a four DOF system with 5 ensembles.

When using random inputs, ensemble-averaging is a necessity to alleviate the noise found in the resultant Markov parameters. Fig. 4 shows how when just five ensembles are averaged together, the results improve significantly. More ensemble-averaging of the random data will improve results further. On the other hand, when harmonic excitations are used to determine Markov parameters, ensemble-averaging does not improve accuracy.

Once the Markov parameters of a system have been determined, modal and physical models of the system can be formed. The next step in the construction of these models is termed system realization and is detailed in the next section.

4. System realization methods for fitting models to response functions

4.1. The system realization problem and ERA

Generally the measured data obtained from modal testing, typically in the form of discrete FRFs or impulse response functions for each input–output pair, is not utilized directly in model correlation analyses, vibration control design or damage detection algorithms. It is necessary instead to determine a small number of intrinsic modal (hopefully normal mode) parameters which equally represent the large quantity of redundant response samples in the FRFs. The various state space forms discussed previously for a linear time-invariant system are recast in a general form as

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{45}$$

The general state solution of (45) is

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau)\,d\tau \tag{46}$$

The corresponding discrete-time (sampled data) model with a zero-order hold and a sampling rate of Δt is given by

$$\begin{aligned} x(k+1) &= \bar{A}x(k) + \bar{B}u(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{47}$$

where $\bar{A} = e^{A\Delta t}$ and $\bar{B} = \int_0^{\Delta t} e^{A(\Delta t-\tau)}B\,d\tau$. The pulse response of (47) is given as

$$y(k) = \sum_{i=1}^{k} Y(k-i)u(i) \tag{48}$$

where $Y(k)$ are the Markov parameters. The Markov parameters are given in terms of the state space model matrices as

$$Y(0) = D \quad Y(k) = C\bar{A}^{k-1}\bar{B}, \quad k > 0 \tag{49}$$

The system realization problem is stated briefly as follows: Find the state-space model, described by the matrices $\bar{A}$, $\bar{B}$, C and D, which best approximates (i.e. minimizes some error norm) the given pulse response of the system

$$Y(k) \quad 0 \leqslant k < \infty \tag{50}$$

where the pulse response is described by the Markov parameters Y. The approach first developed by Ho and Kalman [18] uses the discrete time shift of the Hankel matrix, defined as

$$H(k) = \begin{bmatrix} Y(k+1) & Y(k+2) & \cdots & Y(k+d) \\ Y(k+2) & Y(k+3) & \cdots & Y(k+d+1) \\ \vdots & \vdots & \cdot & \vdots \\ Y(k+q) & Y(k+q+1) & \cdots & Y(k+q+d-1) \end{bmatrix} \tag{51}$$

The eigensystem realization algorithm [15] provides a numerical procedure for identifying the state space model in discrete form from an estimate of the Markov parameters. Using the resultant system realization provided by ERA, the damped modal parameters are obtained as follows. Using the eigenvectors Ψ of $\bar{A}$, defined as

$$\bar{A}_0\Psi = \Psi\bar{\Lambda} \tag{52}$$

the damped mode shapes C_z and continuous-time complex eigenvalues are

$$\bar{\Lambda} = \Psi^{-1}\bar{A}_0\Psi \quad \Lambda = \frac{\ln\bar{\Lambda}}{\Delta t} = \operatorname{diag}\{\sigma_i \pm j\omega_i\}$$

$$\begin{aligned} C_z &= \begin{cases} C_0\Psi & \text{for displacement sensing} \\ C_0\Psi\Lambda^{-1} & \text{for velocity sensing} \\ C_0\Psi\Lambda^{-2} & \text{for acceleration sensing} \end{cases} \\ &= \lfloor \ldots, \Re(C_{zi}) \pm j\Im(C_{zi}), \ldots \rfloor \end{aligned} \tag{53}$$

The damped modal realization for displacement output, in continuous-time, is given by

$$\begin{aligned} \dot{z} &= \Lambda z + B_z u \\ y_d &= C_z z \end{aligned} \tag{54}$$

where

$$B_z = \Psi^{-1}B_0 \tag{55}$$

and B_0 is the continuous-time transform of $\bar{B}_0$.

The damped modal realization (54) is one of the primary model realizations sought in structural system identification, as it is equivalent to the normalized damped modal realization (13) to within an arbitrary complex scaling of the state basis. As noted previously, when driving-point measurements (from sensors collocated with the forcing inputs) are available, the symmetry of the input and output matrix coefficients in (13) can be exploited to determine the complex scaling relating C_z and X, and thus determine the "mass-normalized" damped mode shapes X.

Note that, for a collocated actuator J and sensor K, $H_k = \widehat{B}_J$. Thus, in the displacement output realization, the force-state input influence array and the state-output influence arrays are transposes of one another. This provides a criterion for physically-based normalization of the damped mode shapes such that they can be interpreted as the normalized damped mode shapes X without a priori knowledge of M, D and K. The correct physically-based normalization of the damped modal realization is as follows. The damped modal states z and their normalized counterparts z_n are related by a complex scaling transformation F_n, viz.

$$z_n = F_n z \tag{56}$$

where f_n is diagonal. Applying the similarity transformation F_n to the damped modal realization (54) with displacement output, the normalized damped modal realization is given by

$$\begin{aligned} \dot{z}_n &= \Lambda z_n + F_n B_z u \\ y_d &= C_z F_n^{-1} z_n \end{aligned} \tag{57}$$

The normalization of the damped mode shapes X implies that, for a collocated force input J and displacement output K, the normalized force-state influence matrix $F_n B_z$ is the transpose of the state-output influence matrix $C_z F_n^{-1}$. Therefore, for each state z_i,

$$f_i B_{ziJ} = \frac{C_{ziK}}{s_i} \Rightarrow f_i = \sqrt{\frac{C_{ziK}}{B_{ziJ}}} \qquad F_{\mathrm{n}} = \mathrm{diag}\{f_i, i = 1, \ldots, n\} \tag{58}$$

Thus, the normalized mode shapes $HX = C_z F_{\mathrm{n}}^{-1}$ can be determined from the unscaled damped modal realization (54) using driving point measurements. If the driving point sensor measures acceleration, then the output influence matrix in the z basis $C_0\Psi$ can be effectively integrated by postmultiplying with Λ^{-2} and then treated as the damped displacement mode shapes C_z thereafter.

In summary, the ERA realization can be applied to the discrete time impulse response parameters estimated from empirical data, and the resultant first-order model can be transformed into the damped modal realization which yields the damped modal parameters Λ and C_z. The damped modal parameters are derivable in a general sense from broadband modal testing and system realization methods. However, these modal parameters still implicitly contain the influence of system damping, and are thus only directly applicable to physical parameter derivation and model correction when the damping can be accurately characterized independently of the structure's mass and stiffness. The normal modes, which express the undamped system behavior, are usually the desired modal quantities for model updating and physical parameter determination. In the next section, methods for estimating normal modal parameters from Λ and C_z will be presented for approximating the normal modal parameters from the damped modal parameters of individual damped first-order modes.

5. Relating system realization models to modal parameters

This section is an extension of the CBSI method presented previously in [55]. That work has been extended to correctly determine objective bases in general symmetrically-damped multiple-input-multiple-output (MIMO) systems. This is accomplished by minimizing partitions of either the B or C arrays through an optimization-based formulation, which leads effectively to a least-square solution for the transformation parameters. In addition, an alternate symmetrical formulation of CBSI, which is alluded to in [55], is formally developed, leading to a symmetrical companion form realization. The symmetrical CBSI requires driving point measurements in order to correctly rotate the damped mode shapes, as opposed to a rotation of the real mode estimates as in [55]. The least-square CBSI does not specifically require this normalization, although a mass normalization of the mode shapes can be effectively found using driving point measurements with any of the CBSI algorithm variants. We begin by reviewing the classical problem of normal mode estimation from the damped modes of the system realization.

5.1. Normal mode estimation from damped modes

The traditional approach to normal mode estimation is based on the expression of proportionally-damped behavior in the first-order damped or complex modes. Using (14), the undamped frequency and damping ratio are found as

$$\omega_{n_i}^2 = \sigma_i^2 + \omega_i^2 \quad \zeta_i = -\frac{\sigma_i}{\omega_{n_i}} \tag{59}$$

The above result is unambiguous and common to most normal mode estimation methods, including CBSI, because it is the only simple result which holds correctly for the exact case, i.e. when the damping is proportional. Note, however, that (59) does not hold for the non-proportional damping case. Hence, the existence of non-proportional damping affects not only the accuracy of estimated normal mode shapes, but also the estimated natural frequency and modal damping ratio.

The problem of normal mode shape estimation, however, is not as well determined. Ignoring the scaling of the vector, the correct normal mode shape can be determined by any combination of the real and imaginary components of the damped mode shape in the case of proportional damping. The standard technique (ST) (e.g. see Imregun and Ewins [28]) is evolved from interpreting the complex mode shape for mode i as vectors of magnitudes $\overline{X}_{ij}$ and phase angles α_{ij} at spatial points j such that

$$\left.\begin{aligned} \Re(X_{ij}) &= \overline{X}_{ij}\cos\alpha_{ij} \\ \Im(X_{ij}) &= \overline{X}_{ij}\sin\alpha_{ij} \end{aligned}\right\} \Rightarrow \begin{cases} \overline{X}_{ij} = \sqrt{\Re(X_{ij})^2 + \Im(X_{ij})^2} \\ \alpha_{ij} = \mathrm{atan}\left(\frac{\Im(X_{ij})}{\Re(X_{ij})}\right) \end{cases} \tag{60}$$

For proportional damping, the phase angles α_{ij} are given as

$$\alpha_{ij} - \alpha_{0_i} = (\mathrm{sgn}(j))\frac{\pi}{2} \tag{61}$$

where α_0 is an arbitrary angle dependent on the complex scaling of X_i and $(\mathrm{sgn}(j))$ is the sign function which is equal to ± 1 and varies with spatial location j. The components X_{ij} are said to be purely in-phase or out-of-phase if (61) holds. In the case of modes which are close to proportionally damped, the phase angles are clustered about $\alpha_{0_i} \pm \pi/2$, viz.

$$\alpha_{ij} - \alpha_{0_i} = (\mathrm{sgn}(j))\frac{\pi}{2} + \varepsilon_j \tag{62}$$

The standard technique for normal mode shape estimation is then to neglect the variation ε_j in non-normal phase components, i.e. let $\varepsilon_j \approx 0$. Thus, the normal mode estimates are given as

$$\phi_{ij} = \overline{X}_{ij}(\mathrm{sgn}(j)) \tag{63}$$

While the standard technique is straightforward, it lacks a basis in the context of model equivalence. That is, the relationship between the damped modal coordinates, and thus back to the original states of the estimated model, is lost. An alternative to the standard technique is to utilize similarity transformations such that a new realization of the dynamics is obtained which is fully equivalent to the original realization but which directly expresses the modal parameters sought by the normal mode estimation problem.

5.2. Transformations to canonical variables

The CBSI method is based on a consistent transformation which correctly satisfies the conditions for a second-order canonical state basis. In this section, the criteria for transformations to the second-order canonical form are developed. A second-order canonical realization of the system dynamics is one for which the basis is defined as

$$x(t) = \begin{bmatrix} p_{\mathrm{d}}(t) \\ \dot{p}_{\mathrm{d}}(t) \end{bmatrix} \tag{64}$$

where $p_{\mathrm{d}}(t)$ is a second-order state variable such that the first-order basis definition $x(t)$ spans the state space of the original realization. If a given realization of arbitrary basis definition is given by

$$\begin{aligned} \dot{x}_0(t) &= A_0 x_0(t) + B_0 u(t) \\ y(t) &= C_0 x_0(t) + D u(t) \end{aligned} \tag{65}$$

then an order-n state variable $p(t)$ can be defined in terms of the order-$2n$ state vector $x_0(t)$, viz.

$$p_{\mathrm{d}}(t) = p x_0(t) \tag{66}$$

Differentiating (66), the time derivative of the second-order state $p_{\mathrm{d}}(t)$ is given as

$$\dot{p}_{\mathrm{d}}(t) = P\dot{x}_0(t) = PA_0 x_0(t) + PB_0 u(t) \tag{67}$$

Now, define a new state basis and transformation as

$$\begin{bmatrix} p_{\mathrm{d}}(t) \\ p_{\mathrm{v}}(t) \end{bmatrix} = V^{-1} x_0(t) = \begin{bmatrix} P \\ PA_0 \end{bmatrix} x_0(t) \tag{68}$$

Combining, $p_{\mathrm{v}}(t)$ and $\dot{p}_{\mathrm{d}}(t)$ are related as

$$\dot{p}_{\mathrm{d}}(t) = p_{\mathrm{v}}(t) + PB_0 u(t) \tag{69}$$

Therefore, for the new state basis to be equivalent to a second-order canonical basis, the constraint

$$PB_0 = 0 \tag{70}$$

must be satisfied. The transformation V which satisfies this constraint leads to an new state space realization which is equivalent to n second-order equilibrium equations plus n identity equations

$$\dot{p}_{\mathrm{d}}(t) = p_{\mathrm{v}}(t) \tag{71}$$

Note that the transformation P can be arbitrarily scaled without violating the constraint condition. Thus, satisfying the constraint $PB_0 = 0$ does not uniquely define the transformation P, so that some scaling for P must also be chosen in order to objectively define the resultant second-order basis x. It will be seen that in applying this transformation theory to the problem of normal mode estimation, various scaling choices affect only the real scaling of the normal mode shapes.

To apply the general transformation theory to the problem of normal mode estimation, the initial basis chosen is the individual damped modal displacements $z_i(t)$ and $\bar{z}_i(t)$ from (54). Therefore, a second-order variable $\eta(t)$ can be defined as

$$\eta(t) = P_i \begin{Bmatrix} z_i(t) \\ \bar{z}_i(t) \end{Bmatrix} = d_i \begin{bmatrix} 1 & e_i \end{bmatrix} \begin{Bmatrix} z_i(t) \\ \bar{z}_i(t) \end{Bmatrix} \tag{72}$$

Using (68) and (69), the definition for the required transformation is

$$\begin{aligned} \begin{bmatrix} \eta(t) \\ \dot{\eta}(t) \end{bmatrix} &= V_i^{-1} \begin{Bmatrix} z_i(t) \\ \bar{z}_i(t) \end{Bmatrix} \\ &= d_i \begin{bmatrix} 1 & e_i \\ \sigma_i + \mathrm{j}\omega_i & e_i(\sigma_i - \mathrm{j}\omega_i) \end{bmatrix} \begin{Bmatrix} z_i(t) \\ \bar{z}_i(t) \end{Bmatrix} \end{aligned} \tag{73}$$

with the constraint from (71)

$$P_i B_{zi} = b_{z_i}^{\mathrm{T}} + e_i \bar{b}_{z_i}^{\mathrm{T}} = 0 \tag{74}$$

Thus e_i is specifically defined by the constraint, while d_i is an arbitrary real scalar quantity.

Letting $d_i = 1$ and $e_i = 1$ leads to the McMillan normal form solution from [30] for V_i^{-1}, viz.

$$\left\{ V_i^{-1} \right\}_{\mathrm{MNF}} = \begin{bmatrix} 1 & 1 \\ \sigma_i + \mathrm{j}\omega_i & \sigma_i - \mathrm{j}\omega_i \end{bmatrix} \tag{75}$$

Evaluating the constraint, however, results in

$$\Re(b_{z_i}^{\mathrm{T}}) = 0 \tag{76}$$

Thus, in order for the McMillan normal form transformation to yield a canonical basis, the damped modal coordinates $z_i(t)$ must be rotated using an arbitrary complex scalar such that (76) holds. Correcting the complex normalization of the damped modes is one approach for determining a correct canonical basis. Another approach, which is used by the basic CBSI algorithm, simply applies a second transformation to the McMillan realization obtained by letting $d_i = 1$ and $e_i = 1$. In this second transformation, the same constraint (70) is applied to the new B matrix, resulting in the transformation

$$V_{i2} = \begin{bmatrix} 1 + 2\sigma_i e_i & -e_i \\ e_i(\sigma_i^2 + \omega_i^2) & 1 \end{bmatrix} \tag{77}$$

where

$$b_{i1} + b_{i2}e_i = 0 \tag{78}$$

and b_{i1} and b_{i2} are the upper and lower part of the B matrix in McMillan normal form. Note that if the McMillan normal form is already the correct canonical basis, then $b_{i1} = 0$ and thus $e_i = 0$ and the CBSI second transformation reduces to the identify matrix.

5.3. The CBSI algorithm

Although it is possible to correctly determine a transformation which yields a second-order canonical basis from the damped modal realization equation (7), it is also possible to use the McMillan normal form realization equation (6) as a starting point since it is a fully equivalent realization. This is the approach for the basic CBSI algorithm. Thus, applying V_i from (75) to Eq. (7), the McMillan normal form realization is found as

$$\begin{aligned} \dot{z}_{r_i}(t) &= \begin{bmatrix} 0 & 1 \\ -\sigma_i^2 - \omega_i^2 & 2\sigma_i \end{bmatrix} z_{r_i}(t) + \begin{bmatrix} b_{i1}^{\mathrm{T}} \\ b_{i2}^{\mathrm{T}} \end{bmatrix} u(t) \\ y_{\mathrm{d}}(t) &= \sum_{i=1}^{n} \begin{bmatrix} c_{i1} & c_{i2} \end{bmatrix} z_{r_i}(t) \end{aligned} \tag{79}$$

where, b_{i1}, b_{i2}, c_{i1} and c_{i2} are real-valued partitions of the transformed input and output arrays, respectively, and $z_{r_i}(t)$ is the real-valued basis resulting from the McMillan transformation and corresponding to the complex modal basis $z_i(t)$ and its complex conjugate.

Therefore, in order to transform (79) to a correct second-order basis, define the normal modal displacement $\eta(t)$ as

$$\eta(t) = d_i \begin{bmatrix} 1 & e_i \end{bmatrix} z_{r_i}(t) \tag{80}$$

and the transformation is given as

$$\begin{aligned} V_{i2}^{-1} &= d_i \begin{bmatrix} 1 & e_i \\ -e_i(\sigma_i^2 + \omega_i^2) & 1 + 2\sigma_i e_i \end{bmatrix} \\ V_{i2} &= \frac{1}{\bar{d}_i} \begin{bmatrix} 1 + 2\sigma_i e_i & -e_i \\ e_i(\sigma_i^2 + \omega_i^2) & 1 \end{bmatrix} \\ \bar{d}_i &= d_i(1 + 2\sigma_i e_i + (\sigma_i^2 + \omega_i^2)e_i^2) \end{aligned} \tag{81}$$

with the constraint

$$b_{i1} + b_{i2}e_i = 0 \tag{82}$$

This is the fundamental theoretical development for the basic CBSI algorithm. Two variants of the method focus on the solution for e_i in the cases of proportional and non-proportional damping.

5.4. Basic CBSI for proportional damping

In determining a solution to the transformation constraint equation (82), it is first necessary to consider whether such a solution exists. In the case of proportional damping, it must be true that the vectors b_{i1} and b_{i2} are collinear, since both are linear combinations of the real and imaginary parts of the damped mode shape which are collinear. Therefore, the constraint is satisfied exactly for a single value e_i. From (82), e_i is determined by

$$e_i = -\frac{b_{i1}^{\mathrm{T}} b_{i2}}{b_{i2}^{\mathrm{T}} b_{i2}} \tag{83}$$

Furthermore, the vectors c_{i1} and c_{i2} are also linear combinations of the real and imaginary parts of the damped mode shape, and hence must also be collinear. In fact, they can be shown to satisfy

$$e_i c_{i1} = c_{i2} \tag{84}$$

Thus, applying (81) to (79), the CBSI realization is given as

$$\begin{aligned} \begin{bmatrix} \dot{\eta} \\ \ddot{\eta} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -\sigma_i^2 - \omega_i^2 & 2\sigma_i \end{bmatrix} \begin{bmatrix} \eta \\ \dot{\eta} \end{bmatrix} + \begin{bmatrix} 0 \\ \phi_{ui}^{\mathrm{T}} \end{bmatrix} u \\ y_{\mathrm{d}} &= \sum_{i=1}^{n} \begin{bmatrix} \phi_{yi} & 0 \end{bmatrix} \begin{bmatrix} \eta \\ \dot{\eta} \end{bmatrix} \end{aligned} \tag{85}$$

where

$$\begin{aligned} \phi_{ui} &= d_i b_{i2} + d_i e_i (2\sigma_i b_{i2} - (\sigma_i^2 + \omega_i^2) b_{i1}) \\ \phi_{yi} &= \frac{1}{\bar{d}_i} c_{i1} + \frac{e_i}{\bar{d}_i} (2\sigma_i c_{i1} + (\sigma_i^2 + \omega_i^2) c_{i2}) \end{aligned} \tag{86}$$

By inspection, (85) is a correct second-order canonical realization for mode i. In order to provide a common basis normalization for ϕ_{ui} and ϕ_{yi}, some scaling definition independent of the particular model realization is needed. For example, the modal participation factor ϕ_{uij} of some input u_j could be normalized to 1.0, viz.

$$d_i = \frac{1}{b_{ij2}(1 + 2\sigma_i e_i - (\sigma_i^2 + \omega_i^2)e_i^2)} \tag{87}$$

If the system damping is proportional, it is also possible in some cases to normalize such that the mode shape data extracted from ϕ_{yi} is mass-normalized, provided there exists at least one collocated actuator and sensor pair. This allows ϕ_{ui} and ϕ_{yi} to be scaled as follows. For displacement sensing, collocation requires that

$$H_{\mathrm{d}} = \widehat{B}^{\mathrm{T}} \tag{88}$$

Therefore, for each mode i, $\phi_{ui} = \phi_{yi}$ for collocated input–output pairs, and d_i is found as

$$d_i^2 = \frac{b_{i2}^{\mathrm{T}} c_{i1}}{b_{i2}^{\mathrm{T}} b_{i2}(1 + 2\sigma_i e_i - (\sigma_i^2 + \omega_i^2)e_i^2)} \tag{89}$$

5.5. Least-squared CBSI for general viscous damping

In the case of non-proportional damping, the vectors b_{i1} and b_{i2} are no longer collinear, and hence (83) does not satisfy (82) for the multiple inputs u. This implies that the CBSI transformation (81) cannot exactly determine the desired second-order canonical basis variables η_i using the uncoupled complex modal variables z_i individually. In this case, it is possible to determine a basis using *quasi-normal* modal displacements $\tilde{\eta}_{\mathrm{d}i}$ and velocities $\tilde{\eta}_{\mathrm{v}i}$ through the CBSI algorithm. The resultant CBSI transformation and basis is written as

$$\begin{bmatrix} \tilde{\eta}_{\mathrm{d}i} \\ \tilde{\eta}_{\mathrm{v}i} \end{bmatrix} = \frac{1}{d_i} \begin{bmatrix} 1 + 2\sigma_i e_i & -e_i \\ e_i(\sigma_i^2 + \omega_i^2) & 1 \end{bmatrix} z_{r_i} \tag{90}$$

and from (69) the error between the quasi-normal modal velocities and the time derivative of the quasi-normal modal displacements is

$$\dot{\tilde{\eta}}_{\mathrm{d}i} - \tilde{\eta}_{\mathrm{v}i} = d_i(b_{i1}^{\mathrm{T}} + e_i b_{i2}^{\mathrm{T}})u \tag{91}$$

Thus, in order to minimize this error for arbitrary inputs u, it is necessary to minimize the quantity, where

$$J_u = d_i^2(b_{i1} + e_i b_{i2})^{\mathrm{T}}(b_{i1} + e_i b_{i2}) \tag{92}$$

Since this is effectively a least-square criterion for e_i, the method is termed the least-square CSBI (CBSI-LS) algorithm.

At this point it might appear that (83) would yield the correct least-square solution for (92). However, in order to determine an optimal solution for e_i independent of the scaling parameter d_i, it is necessary to impose an additional scaling condition to constrain d_i. Otherwise, it would be necessary to minimize J_u with respect to both e_i and d_i. The optimal solution in that case, however, is trivial $(d_i = 0)$, since J_u is quadratic. Therefore, if the CBSI-transformed input influence array is written as

$$V_{i2}^{-1} \begin{bmatrix} b_{i1}^{\mathrm{T}} \\ b_{i2}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \tilde{\varepsilon}_{ui}^{\mathrm{T}} \\ \tilde{\phi}_{ui}^{\mathrm{T}} \end{bmatrix} \tag{93}$$

then the transformation scaling can be controlled by requiring that

$$\tilde{\phi}_{ui}^{\mathrm{T}} \tilde{\phi}_{ui} = b_{i2}^{\mathrm{T}} b_{i2} \tag{94}$$

This implies that the norm of $\tilde{\phi}_{ui}$, which is an estimate of the mode shape at the input locations, is constrained to the norm of b_{i2}. From (92), d_i can be eliminated and J_u is a function strictly of e_i, viz.

$$J_u = \frac{S_{22} e_i^2 + 2S_{12} e_i + S_{11}}{a_{\mathrm{d}} e_i^2 + 2b_{\mathrm{d}} e_i + S_{22}} \tag{95}$$

where

$$\begin{aligned} &S_{11} = b_{i1}^{\mathrm{T}} b_{i1} \quad S_{12} = b_{i1}^{\mathrm{T}} b_{i2} \quad S_{22} = b_{i2}^{\mathrm{T}} b_{i2} \\ &a_{\mathrm{d}} = 4\sigma_i^2 S_{22} - 4\sigma_i(\sigma_i^2 + \omega_i^2) S_{12} + (\sigma_i^2 + \omega_i^2)^2 S_{11} \\ &b_{\mathrm{d}} = 2\sigma_i S_{22} - (\sigma_i^2 + \omega_i^2) S_{12} \end{aligned} \tag{96}$$

To minimize J_u, the necessary condition $\delta J_u = 0$ requires that

$$\frac{\mathrm{d}J_u}{\mathrm{d}e_i} = 0 \tag{97}$$

leading to the solution for e_i

$$e_i = \frac{-b_{\mathrm{e}} \pm \sqrt{b_{\mathrm{e}}^2 - 4a_{\mathrm{e}}c_{\mathrm{e}}}}{2a_{\mathrm{e}}} \tag{98}$$

where

$$\begin{aligned} &a_{\mathrm{e}} = b_{\mathrm{d}} S_{22} - a_{\mathrm{d}} S_{12} \\ &b_{\mathrm{e}} = S_{22}^2 - a_{\mathrm{d}} S_{11} \\ &c_{\mathrm{e}} = S_{12} S_{22} - b_{\mathrm{d}} S_{11} \end{aligned} \tag{99}$$

The two solutions for e_i from (98) correspond to a minimum and maximum value of J_u. Applying the CBSI transformation (90) to (79), the CBSI realization for the non-proportional damping condition is

$$\begin{aligned} &\begin{Bmatrix} \dot{\tilde{\eta}}_{\mathrm{d}i} \\ \dot{\tilde{\eta}}_{\mathrm{v}i} \end{Bmatrix} = \begin{bmatrix} 0 & 1 \\ -\tilde{\omega}_{\mathrm{n}i} & 2\tilde{\zeta}_i \tilde{\omega}_{\mathrm{n}i} \end{bmatrix} \begin{Bmatrix} \tilde{\eta}_{\mathrm{d}i} \\ \tilde{\eta}_{\mathrm{v}i} \end{Bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_{ui}^{\mathrm{T}} \\ \tilde{\phi}_{ui}^{\mathrm{T}} \end{bmatrix} u \\ &y_{\mathrm{d}} = \sum_{i=1}^{n} \begin{bmatrix} \tilde{\phi}_{yi} & \tilde{\varepsilon}_{yi} \end{bmatrix} \begin{Bmatrix} \tilde{\eta}_{\mathrm{d}i} \\ \tilde{\eta}_{\mathrm{v}i} \end{Bmatrix} \end{aligned} \tag{100}$$

where $\|\tilde{\varepsilon}_{ui}\|_2$ has been minimized such that $\|\tilde{\phi}_{ui}\|_2 = \|b_{i2}\|_2$. Note that the undamped natural frequency $\tilde{\omega}_{\mathrm{n}i}$ and modal damping ratio $\tilde{\zeta}_i$ are now approximations because of the effects of non-proportional damping.

An alternate approach for determining the least-square CBSI transformation is through use of the damped mode shapes at the sensor locations, i.e. by use of c_{i1} and c_{i2} in (79). Recall the constraint $C_{\mathrm{d}}B = 0$ relating the force input influence array B and the displacement output influence array C_{d}, expressed in terms of the partitions of (79)

$$C_{\mathrm{d}}B = c_{i1} b_{i1}^{\mathrm{T}} + c_{i2} b_{i2}^{\mathrm{T}} = 0 \tag{101}$$

This implies that the first-order delay term in the impulse response of force input-displacement output transfer function is zero, which in turn is a necessary condition for the impulse response behavior of a second-order system. Thus, an equivalent expression for e_i in terms of c_{i1} and c_{i2} is found by substituting the transformation constraint (101), viz.

$$(e_i c_{i1} - c_{i2}) b_{i2}^{\mathrm{T}} = 0 \tag{102}$$

This relationship must hold for inputs at any physical degree of freedom spanned by the modal state space, i.e. for all b_{i2} in R^n. Hence, an equivalent transformation constraint condition is given as

$$e_i c_{i1} - c_{i2} = 0 \tag{103}$$

This is the same condition claimed in (82), which is now proven to hold for both proportional and non-proportional damping. The remainder of the problem is basically equivalent to the optimal solution for e_i in terms of the input matrix coefficients b_{i1} and b_{i2}.

A key motivation for the preceding derivation is that, for normal mode estimation in the presence of non-proportional damping, it is advantageous to use the maximum quantity of damped modal information available. Determination of the optimal e_i using c_{i1} and c_{i2} generally leads to a more balanced estimate of the normal modes. This comes from the observation that the solution for e_i can be heavily biased towards the phase quantities of a small number of measured degrees of freedom. So, although both the force input matrix coefficients b_{i1} and b_{i2} and the displacement output matrix coefficients c_{i1} and c_{i2} both express the values of the complex mode shapes at a number of physical degrees of freedom, the reality in modal testing is that there are a significantly larger number of measured sensors than force actuators. Furthermore, the test instrumentation may include driving point measurements, which are sensor collocated with the force inputs. In this case, the output matrix coefficients possess not only the mode shapes at the numerous sensor locations, but also the mode shapes at the actuator locations.

5.6. A symmetrical CBSI method

The final variant of the CBSI procedure to be presented determines a real-valued symmetrical companion form realization directly from the damped modal realization. In place of the second CBSI rotation, which transforms the McMillan normal form realization, the symmetrical CBSI (CBSI-SYM) method determines an objective normalization of the damped modes leading to a symmetrical form of the first-order realization, then applies a single transformation which preserves the symmetry of the matrix coefficients. The resultant CBSI realization is in a symmetrical companion form of the normal modal variables, viz.

$$\begin{bmatrix} \Xi & I \\ I & 0 \end{bmatrix} \begin{Bmatrix} \dot{\eta} \\ \ddot{\eta} \end{Bmatrix} = \begin{bmatrix} -\Omega^2 & 0 \\ 0 & I \end{bmatrix} \begin{Bmatrix} \eta \\ \dot{\eta} \end{Bmatrix} + \begin{bmatrix} \Phi^{\mathrm{T}} \widehat{B} \\ 0 \end{bmatrix} u$$
$$y_{\mathrm{d}} = [H\Phi \quad 0] \begin{Bmatrix} \eta \\ \dot{\eta} \end{Bmatrix} \tag{104}$$

A particular restriction on this method is that it requires the physically-based normalization of the damped modal coordinates as detailed in (57) and (58). This in turn requires driving point measurements (i.e. collocated sensors at the force inputs).

The equivalent normalized damped modal form of (104), as obtained from the driving point scaling (58), is given as

$$\begin{aligned} \dot{z}_{\mathrm{n}} &= \Lambda z_{\mathrm{n}} + X^{\mathrm{T}} \widehat{B} u \\ y_{\mathrm{d}} &= H X z_{\mathrm{n}} \end{aligned} \tag{105}$$

where

$$\begin{aligned} &\Lambda = \{\mathrm{diag}\, \sigma_i \pm \mathrm{j}\omega_i, \ i = 1, \ldots, n\} \\ &X = \lfloor \ldots, \Re(X_i) \pm \mathrm{j}\Im(X_i), \ldots \rfloor \\ &\sigma_i = -\zeta_i \omega_i \quad \omega_i = \omega_{\mathrm{n}i} \sqrt{1 - \zeta_i^2} \\ &\Re(X_i) = \frac{1}{2\sqrt{\omega_i}} \phi_i \quad \Im(X_i) = \frac{-1}{2\sqrt{\omega_i}} \phi_i \end{aligned} \tag{106}$$

Therefore, we can define the second-order transformation as

$$\hat{\eta}_{\mathrm{d}i} = P_i z_{\mathrm{n}i} = \frac{1}{\sqrt{2\mathrm{j}\omega_i}} [1 \quad \mathrm{j}] \begin{Bmatrix} z_{\mathrm{n}i} \\ \bar{z}_{\mathrm{n}i} \end{Bmatrix} \tag{107}$$

such that the symmetrical CBSI transformation is

$$V_i = \sqrt{\frac{\mathrm{j}}{2\omega_i}} \begin{bmatrix} \mathrm{j}(\sigma_i - \mathrm{j}\omega_i) & -\mathrm{j} \\ -(\sigma_i + \mathrm{j}\omega_i) & 1 \end{bmatrix} \tag{108}$$

and

$$V_i^{\mathrm{T}} V_i = \begin{bmatrix} -2\sigma_i & 1 \\ 1 & 0 \end{bmatrix} \tag{109}$$

Then, applying the basis definition (107) to (105) and premultiplying by V_i^{T} to preserve symmetry, the resultant symmetrical CBSI realization for non-proportional damping is given by

$$\begin{bmatrix} -2\tilde{\zeta} & 1 \\ 1 & 0 \end{bmatrix} \begin{Bmatrix} \dot{\hat{\eta}}_{\mathrm{d}i} \\ \dot{\hat{\eta}}_{\mathrm{v}i} \end{Bmatrix} = \begin{bmatrix} -\tilde{\omega}_{\mathrm{n}i}^2 & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} \hat{\eta}_{\mathrm{d}i} \\ \hat{\eta}_{\mathrm{v}i} \end{Bmatrix} + \begin{bmatrix} \hat{\phi}_{ui}^{\mathrm{T}} \\ \hat{\varepsilon}_{ui}^{\mathrm{T}} \end{bmatrix} u$$
$$y_{\mathrm{d}} = \sum_{i=1}^{n} \begin{bmatrix} \underline{\hat{\phi}}_{yi} & \hat{\varepsilon}_{yi} \end{bmatrix} \begin{Bmatrix} \hat{\eta}_{\mathrm{d}i} \\ \hat{\eta}_{\mathrm{v}i} \end{Bmatrix} \tag{110}$$

In (110), because of the symmetry of the companion form realization, the relevant mode shapes at the inputs are given by $\hat{\phi}_{ui}$ in the upper partition of the input influence coefficient matrix, rather than the lower partition characteristic of the realization in (100). The particular feature and potential advantage of the symmetrical CBSI method is that the solution is not biased towards minimizing the residual quantities at particular input or output locations. Instead, the basis definition is "balanced" by the normalization of the damped modes, which leads to a realization with equivalent mode shape

estimates and modal participation factors for the collocated input–output pairs independent of whether the damping is proportional or non-proportional. In fact, the symmetrical CBSI realization follows as the only realization possessing the given symmetrical state coefficient matrices which allows the symmetry of the input and output influence matrices in (105) to be preserved. And, as demonstrated, the transformation does not consider the residual model quantities as in the least-square CBSI, although the solution obtained in practice yields similar mode shape estimates, especially as compared to the CBSI-LS solution using the sensor matrix partitions c_{i1} and c_{i2}.

For structures with non-proportional damping, the CBSI-LS or CBSI-SYM methods lead to generally accurate and reliable estimates of the normal modes and their mode shapes by transforming the given system realization to a ''pseudo-normal'' state variable basis. This pseudo-normal modes basis is defined by minimizing the difference between the estimated modal velocities and the time derivative of the estimated modal displacements. As the CBSI transformations are applied to the individual damped modes, however, its ability to determine the true normal modes is necessarily limited when the system damping is non-proportional. Therefore, in some cases where structures exhibit a high degree of modal energy coupling due to repeated frequencies and the presence of non-proportional damping, it is necessary to utilize a mode shape-damping decoupling method, which allows for the coupling of normal modes. Such a procedure, based on the second-order transformation theory detailed in this section, has been developed in [56].

5.7. Numerical example: 36-DOF planar truss with light non-proportional damping

We now consider a 36-DOF two-dimensional truss structure with 3 actuators and 18 sensors with fixed–fixed conditions, shown in Fig. 5. The structure exhibits relatively light damping averaging around 1% of critical, but possesses repeated frequencies which introduce the potential for modal coupling.

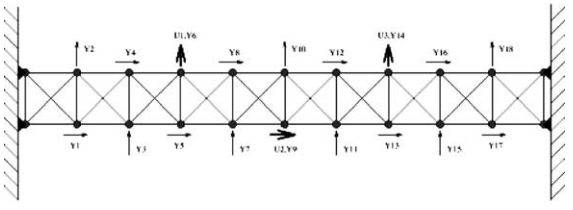

Fig. 5. Planar truss example model showing input force and output displacement locations.

In order to evaluate the mode shape estimates given by various algorithms, it is helpful to define particular measures for modal parameter comparisons. The first of these, modal phase collinearity (MPC) [57] is an indicator of the inherent degree of collinearity between the real and imaginary parts of the damped mode shapes. Modes with MPC values equal to 1.00 are effectively equivalent to normal modes, and the various methods of normal mode shape estimation should work equally well. Modes with MPC values of less than 0.90, on the other hand, do not possess unambiguous corresponding normal modes, and so the normal mode shape estimates will be sensitive to the methods used for their estimation. The MPC for damped mode shape C_{zi} is given as

$$\mathrm{MPC}_i = \frac{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}{(S_{xx} + S_{yy})^2} \tag{111}$$

where

$$S_{xx} = \Re(C_{zi})^{\mathrm{T}}\Re(C_{zi}) \quad S_{xy} = \Re(C_{zi})^{\mathrm{T}}\Im(C_{zi})$$
$$S_{yy} = \Im(C_{zi})^{\mathrm{T}}\Im(C_{zi}) \tag{112}$$

Thus, the MPC is effectively a measure of the degree to which a damped mode is directly related to an undamped mode.

A second important indicator is the modal assurance criterion (MAC), which is used to compare the relative collinearity of different normal mode shapes. The MAC is used, for example, to compare test-derived normal mode shapes to those predicted by finite element models. In the case of studying different normal mode estimation methods, the MAC can be used to evaluate the relative shape agreement between the exact normal mode shapes, given by known mass and stiffness matrices, and the estimated mode shapes. The MAC is generally defined as

$$\mathrm{MAC}_{jk} = \frac{(\phi_j^{\mathrm{T}}\phi_k)^2}{(\phi_j^{\mathrm{T}}\phi_j)(\phi_k^{\mathrm{T}}\phi_k)} \tag{113}$$

where ϕ_j and ϕ_k are corresponding mode shapes from models j and k. Note that both the MAC and the MPC measure the collinearity between two real vectors apart from the scaling, and so they are related quantities. They have been traditionally used for assessing different system properties, however. This paper follows that convention by using the MPC to measure the inherent phase angle scatter of the damped modes, and using the MAC to compare the estimated normal mode shapes from different models or estimation methods.

In Table 1, the accuracy indicators for the normal mode estimates are given for 13 modes possessing MPC values less than 0.99. Note in particular modes 27 and 28, which are more highly coupled. Overall, the performance of CBSI-LS and CBSI-SYM, in terms of the MAC between the correct mode shapes and the CBSI

Table 1
Comparison of normal mode shape estimates for a 36 DOF non-proportionally damped truss

Mode	Modal phase collinearity	Standard technique	McMillian normal form	CBSI	CBSI-LS	CBSI-SYM
16	0.9887	0.9993	0.1187	0.9995	0.9995	0.9995
17	0.9866	0.9993	0.2163	0.9994	0.9994	0.9994
19	0.9481	0.9983	0.0567	0.9982	0.9986	0.9986
20	0.9743	0.9991	0.0394	0.9995	0.9994	0.9994
22	0.9771	0.9617	0.0028	0.9446	0.9448	0.9444
23	0.9748	0.9617	0.0153	0.9402	0.9414	0.9410
25	0.9755	0.9998	0.0397	0.9999	0.9999	0.9999
26	0.8468	0.9934	0.0030	0.9977	0.9981	0.9980
27	0.5722	0.8755	0.0709	0.9058	0.9257	0.9241
28	0.5990	0.8896	0.0841	0.9236	0.9302	0.9299
29	0.9404	0.9964	0.0290	0.9985	0.9986	0.9986
30	0.9853	0.9996	0.1171	1.0000	1.0000	1.0000
31	0.9893	0.9998	0.1860	1.0000	1.0000	1.0000
Mean 1–36	0.9639	0.9909	0.3618	0.9918	0.9926	0.9926

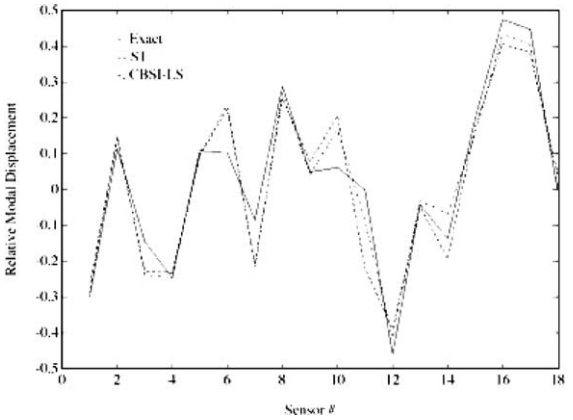

Fig. 6. Comparison of estimated normal mode shapes for planar truss, mode 27.

estimates, are comparable. This is true for both the most complex of the modes, and also in terms of the mean indicator values over all 36 modes. Only in modes 22 and 23 does the ST approach identify a better estimated normal mode shape. Note also that in this general numerical example, the mode shapes obtained from the McMillan form realization, without the improvement of the second CBSI transformation, are quite poor as estimates of the normal modes, even for a lightly damped structure. Finally, Fig. 6 plots the mode shape estimates from ST and CBSI-LS as compared to the exact mode shapes from the known mass and stiffness for mode 27. Thus, as quantified through the MAC indicators with respect to the exact mode shapes, and as illustrated in the previous figures, application of the CBSI algorithm leads to a unique definition of the estimated normal mode shapes and shows a consistent improvement in accuracy over the standard technique for non-proportionally-damped systems.

5.8. Concluding remarks

The present CBSI procedure offers a consistent link between the various system identification techniques and the second-order based equations of motion for structures. Utilizing the general transformation to second-order form, CBSI can determine an optimal estimate of second-order model states which correspond individually to the damped modes of the given system realization. In the presence of strictly proportional viscous damping, the CBSI algorithm systematically determines the accurate normal modes in the form of the well-known second-order structural dynamics equations of motion. For structures with non-proportional damping, the CBSI-LS or CBSI-SYM methods lead to generally accurate and reliable estimates of the normal modes and their mode shapes by transforming the given system realization to a "pseudo-normal" state variable basis. This pseudo-normal modes basis is defined by minimizing the difference between the estimated modal velocities and the time derivative of the estimated modal displacements. As the CBSI transformations are applied to the individual damped modes, however, its ability to determine the true normal modes is necessarily limited when the system damping is non-proportional. Therefore, in some cases where structures exhibit a high degree of modal energy coupling due to repeated frequencies and the presence of non-proportional damping, it is necessary to utilize a mode shape-damping decoupling method, which allows for the coupling of normal modes. This problem is addressed in [56].

6. Damage detection using structural system identification

In this section, the application of system identification to the detection of damage in a structure is ad-

dressed. To this end, a brief treatment of the ILTZ method [48] is presented, along with a companion method based on test-realized flexibility matrices [58]. Both methods require a reference finite element model only for an assumed connectivity and model element type, but not for any physical or analytical characteristics. Likewise, both methods compare realization results from two sets of tests, using the first set as the baseline or nominal condition, and the second set as the current state or damaged condition. These two methods are useful in highlighting many of the important issues in vibration-based health monitoring.

6.1. Dynamic substructural flexibility partitioning theory

As discussed in Section 1, damage detection methods which focus on localized properties have an advantage over so-called global methods because the decomposition process allows small localized changes to have greater impact on the observed system dynamics. The decomposition process can result in any number of equivalent localized variables. In order to make this process more applicable to vibration testing, localized variables that can be directly measured during testing should be utilized if possible [59]. Strains are one such localized variable, and can be easily related to traditional displacements. To understand how partitioning can help the damage detection process, we present a brief review of a dynamic flexibility partitioning theory [60] that reveals the relation between the global equations of motion and the equations of motion in a strain-based form.

6.1.1. Global dynamic systems

To begin, the energy of a spatially discrete global dynamic system can be written in the following functional form:

$$\Pi(q_g) = q_g^T\left(\widehat{B}u_g - \tfrac{1}{2}K_g q_g - D_g\dot{q}_g - M_g\ddot{q}_g\right) \tag{114}$$

where the subscript 'g' is added to indicate values relating to the "assembled" physical structure rather than partitioned substructures, which will be introduced shortly. The equations of motion are found by computing the stationary value of the functional ($\delta\Pi = 0$) which yields (along with the accompanying output equation):

$$\begin{aligned} &M_g\ddot{q}_g + D_g\dot{q}_g + K_g q_g = \widehat{B}u_g \\ &y_g = H_d q_g + H_v\dot{q}_g + H_a\ddot{q}_g \end{aligned} \tag{115}$$

Note here that in a typical vibration test setting it is not practical to excite all the modes, nor place sensors sufficient to cover every discrete degree of freedom desired. Nevertheless, for the clarity of our discussion, we will assume from this point on that the entire displacement vector is measured, and $\widehat{B}$ is an identity matrix whose size is the same as q_g, so that $y_g = q_g$. From this, the global FRF relating the displacement and its conjugate force input are determined as

$$q_g(\omega) = H_g(\omega)u_g(\omega) = (K_g + j\omega D_g - \omega^2 M_g)^{-1}u_g(\omega) \tag{116}$$

where $\omega = 2\pi f$.

6.1.2. Partitioning

Partitioning is the decomposition process given by

$$q = Lq_g \quad L^T u = u_g \tag{117}$$

where u is the elemental force vector conjugate to q and L is a Boolean disassembly matrix that relates the global and substructural displacements. L is typically created based on an underlying finite element model connectivity. In addition, the global stiffness matrix K_g is formed by the assembly of individual substructural stiffnesses via the assembly operator L:

$$K_g = L^T K L \quad K = \begin{bmatrix} K^{(1)} & & & \\ & K^{(2)} & & \\ & & \ddots & \\ & & & K^{(n_s)} \end{bmatrix} \tag{118}$$

where K is the block-diagonal collection of unassembled substructural stiffness matrices $K^{(s)}$ which correspond to each individual substructure that make up the global system.

In order to maintain the kinematic compatibility along the boundary, the substructural displacement vector q must satisfy the following relation:

$$B_\lambda^T(q - Lq_g) = 0 \tag{119}$$

A vector of Lagrange multipliers λ_b is introduced to enforce this constraint. By combining (118) and (119) with (114), the energy functional is expressed in partitioned form

$$\begin{aligned} \Pi(q, \lambda_b, q_g) = {} & q^T\left(u - \tfrac{1}{2}Kq - D\dot{q} - M\ddot{q}\right) \\ & - \lambda_b^T B_\lambda^T(q - Lq_g) \end{aligned} \tag{120}$$

where M and D are the substructural mass and damping matrices, respectively. The matrix B_λ can be chosen in a variety of ways. One method is to define it as a Boolean matrix such that its non-zero entries correspond to the measured output nodes. Alternatively, it can be defined as a null space of L such that $B_\lambda^T L = 0$. If the latter definition is chosen, Π takes its final form of

$$\Pi(q, \lambda_b) = q^T\left(u - \tfrac{1}{2}Kq - D\dot{q} - M\ddot{q}\right) - \lambda_b^T B_\lambda^T q \tag{121}$$

This choice of B_λ is significant because it removes the contribution of the global displacement to the partitioned energy functional.

6.1.3. Decomposition into deformation modes and rigid-body modes

A fundamental problem in determining strain-based frequency response functions is the expression of displacement in terms of strain, rather than the converse. First, note that the displacement q of a free–free partitioned substructure can be decomposed into a deformation d and a rigid-body motion r:

$$q = d + r = d + R\alpha \tag{122}$$

where R are the rigid-body mode shapes of the substructure, and α are their amplitudes. R is determined for each type of element used to model the structure and is dependent only on the geometry of the substructure.

For a discrete element, strain is typically expressed in terms of the elemental displacements as the strain–displacement relation

$$\varepsilon = Sq \tag{123}$$

where S can be determined in a variety of ways, for example from the finite element shape functions of the corresponding discrete elements. This equation can be combined with (122), which results in the expression of strain in terms of deformation:

$$\varepsilon = S(d + R\alpha) = Sd \tag{124}$$

since the rigid-body modes R are orthogonal to the strain modes. The inverse relation is the displacement–strain relation which is an equation for the absolute substructural deformation in terms of the elemental strain, as in

$$d = \Phi_s\varepsilon \quad \Phi_s = S^{+} \tag{125}$$

This is combined with (122) to arrive at the desired relation, which is an expression of the elemental displacements in terms of the elemental strain and rigid-body motion:

$$q = \Phi_s\varepsilon + R\alpha \tag{126}$$

Eq. (126) is now utilized in the energy functional of (121). The resulting energy functional is expressed in terms of the elemental strain, rigid-body motion, and the interface forces

$$\begin{aligned}\Pi(\varepsilon,\alpha,\lambda_b) = (\Phi_s\varepsilon + R\alpha)^T\Big\{u - \tfrac{1}{2}K(\Phi_s\varepsilon + R\alpha)\\ - D(\Phi_s\dot{\varepsilon} + R\dot{\alpha}) - M(\Phi_s\ddot{\varepsilon} + R\ddot{\alpha})\Big\}\\ - \lambda_b^T B_\lambda^T(\Phi_s\varepsilon + R\alpha)\end{aligned} \tag{127}$$

The stationary value of this functional yields the partitioned, strain-based equations of motion, which can be written in the frequency domain as

$$\begin{bmatrix} K_\phi + j\omega D_\phi - \omega^2 M_\phi & B_\phi \\ B_\phi^T & 0 \end{bmatrix}\begin{Bmatrix}\phi \\ \lambda_b\end{Bmatrix} = \begin{Bmatrix} u_\phi \\ 0\end{Bmatrix} \tag{128}$$

where

$$\begin{aligned}&\Phi_\phi = [\,\Phi_s \quad R\,] \quad K_\phi = \Phi_\phi^T K\Phi_\phi = \begin{bmatrix} K_\varepsilon & 0\\ 0 & 0\end{bmatrix}\\ &M_\phi = \Phi_\phi^T M\Phi_\phi = \begin{bmatrix} M_\varepsilon & M_{\varepsilon\alpha}\\ M_{\varepsilon\alpha}^T & M_\alpha\end{bmatrix}\\ &D_\phi = \Phi_\phi^T D\Phi_\phi = \begin{bmatrix} D_\varepsilon & D_{\varepsilon\alpha}\\ D_{\varepsilon\alpha}^T & D_\alpha\end{bmatrix}\\ &\phi = \begin{Bmatrix}\varepsilon\\ \alpha\end{Bmatrix} \quad u_\phi = \Phi_\phi^T u\end{aligned} \tag{129}$$

and $B_\phi = \Phi_\phi^T B_\lambda$. The corresponding strain-based FRF is determined from this equation by eliminating λ_b.

$$\begin{aligned}&\phi = H_\phi(\omega)u_\phi\\ &H_\phi(\omega) = Z_\phi - Z_\phi B_\phi[B_\phi^T Z_\phi B_\phi]^{-1}B_\phi^T Z_\phi\\ &Z_\phi = (K_\phi + j\omega D_\phi - \omega^2 M_\phi)^{-1}\end{aligned} \tag{130}$$

Note that the linear, partitioned dynamics matrix Z_ϕ is a block-diagonal collection of dynamic flexibility matrices corresponding to each substructure. However, within each substructure, coupling exists between the rigid-body motion and the strain energy through the off-diagonal terms in M_ϕ and D_ϕ.

Recalling (126) and (129), the relationship between the substructural variables and the global displacements are determined as

$$\phi = \Phi_\phi^{-1}Lq_g \tag{131}$$

Additionally, invariance of the external work term of the energy functional gives

$$q_g^T u_g = q^T u = \phi^T u_\phi \Rightarrow u_g = L^T\Phi_\phi^{-T}u_\phi \tag{132}$$

Eqs. (131) and (132) are then used in conjunction with the global FRF expression (116) to determine an alternate expression for H_ϕ:

$$H_\phi(\omega) = \Phi_\phi^{-1}LH_g(\omega)L^T\Phi_\phi^{-T} \tag{133}$$

The frequency response functions in (130) and (133) are equivalent because of input–output system invariance. Eq. (130) highlights the dependence of the localized FRF on the local dynamics of the partitioned form, while (133) highlights the dependence on the global dynamics.

Finally, note that from (129) the dynamics of the system in strain-based form are partitioned into the internal strain energy corresponding to the strain DOF and the inertial energy corresponding to the rigid-body motion of the system. In particular, if one is to assume that damage occurs only as a change in the stiffness of the structure, then the only terms of interest are those containing K_ε. This is not to say that no change occurs in

the cross-coupled terms of $H_\phi(\omega)$, but that more insight can be gained by examining the strain partition on its own. From the definition of Φ_ϕ in (129), we can determine the partition of the substructural FRFs corresponding to strain energy as

$$H_\varepsilon(\omega) = SLH_g(\omega)L^T S^T \tag{134}$$

This form of the strain-based FRFs are used by the two damage detection methods that follow.

6.2. Damage detection by flexibility method

The flexibility method [58] of structural health monitoring is based on a comparison of experimentally determined flexibilities. From (134), the quasi-static asymptote $\omega \to 0$ leads to

$$F_\varepsilon = SLF_g L^T S^T \quad F_g = K_g^{-1} = \Phi_g \Omega^{-1} \Phi_g^T \tag{135}$$

where Φ_g are the global undamped modeshapes. If the number of measured modes is less than the number of sensors, a common occurrence in structural realization, then the measured flexibility matrix F_{g_m}, where m represents the identified modes, is found by

$$F_{g_m} = \Phi_{g_m} \Omega_m^{-1} \Phi_{g_m}^T \tag{136}$$

in global variables, or

$$F_{\varepsilon_m} = \Phi_{\varepsilon_m} \Omega_m^{-1} \Phi_{\varepsilon_m}^T \tag{137}$$

if the strain-based modeshapes are measured directly. Note that the eigenvalues are independent of the form used—they represent the poles of the system, which are invariant to the transformation.

The flexibility method then compares F_{ε_m} for nominal and damaged realizations, and a quantitative measure, the relative damage indicator, or RDI, is computed. The RDI is defined as

$$\text{RDI} = \left|\text{diag}(F_{\varepsilon_m}^d - F_{\varepsilon_m}^n)\right| \tag{138}$$

The individual values of RDI correspond to strain DOF, which are unique to individual substructures. Non-zero values of the RDI indicate flexibility changes, and the amount of change can to some degree indicate the relative degree of damage at that location. This will be demonstrated later with a simple example.

6.3. Transmission zero theory

The transmission zeros of a transfer function can be defined in a variety of ways [61], typically depending on how the transfer function itself is defined. In general, transmission zeros are the frequencies at which the value of the transfer function is zero. Physically, this means that no energy is transmitted to the sensor locations when a harmonic input at the transmission zero frequency is applied at the actuator locations. This section reviews the mathematics of linear system theory pertinent to the definition and computation of transmission zeros.

One method of describing the transmission zeros of a transfer function matrix (TFM), equivalent to the frequency response function, is in terms of the system's dynamic flexibility matrix, which is the inverse of the dynamic stiffness matrix. In the Laplace domain, this can be written as

$$H_g(s) = H_d(K_g + sD_g + s^2 M_g)^{-1}\widehat{B} = H_d K_g(s)^{-1}\widehat{B} \tag{139}$$

where

$$K_g(s) = K_g + sD_g + s^2 M_g \tag{140}$$

is defined as the dynamic stiffness matrix. Note that the matrix inverse K_g^{-1} can be expressed [62] in terms of a *cofactor matrix*, namely

$$K_g^{-1} = \frac{\text{cof}(K_g)}{\det(K_g)} \tag{141}$$

where element (i,j) of the cofactor matrix is defined by

$$\text{cof}(K_g)(i,j) = (-1)^{i+j}\det(K_g^{j,i}) \tag{142}$$

The submatrix $K_g^{j,i}$ is obtained by deleting row j and column i from K_g. As before, assume that $H_d = I$ and $\widehat{B} = I$. Thus, the zeros of a scalar matrix element $h_{i,j}$ of $H_g(s)$ are given by

$$\text{zeros of } h_{i,j} \Rightarrow \text{the roots of cof } K_g(j,i) = 0 \tag{143}$$

A generalization of this statement is that the transmission zeros of a block partition of H_g corresponding to multiple degrees of freedom given by (l,m) are computed by finding the roots of $\text{cof}(K_g)(l,m) = 0$; in other words, by finding the determinant of the submatrix of K_g without rows m and columns l.

It is important to note that the transmission zeros of a multi-input, multi-output transfer function do not correspond in general to the zeros, or anti-resonances, of an individual scalar element $h_{i,j}(s)$ of the transfer function matrix $H_g(s)$. However, if the transfer function has only a single input and a single output (SISO), then the transfer function is a scalar function, and the transmission zeros of that transfer function are indeed equal to the anti-resonances of the scalar function. The partition of H_g which has collocated inputs and outputs corresponding to all of the degrees of freedom at a node is said to be the "transfer function of the node". If there is

only a single degree of freedom at the node, then the transfer function is SISO, and the transmission zeros corresponding to the node are the same as the anti-resonances of the scalar partition of the transfer function. If there are multiple degrees of freedom at the node, which occur in structures utilizing beam or shell elements, then the transfer function of the node is MIMO, and the transmission zeros of the node do not correspond to any of the anti-resonances of the scalar TFM elements which make up the matrix partition.

6.4. Damage detection based on localized transmission zeros

Consider a strain-based transfer function matrix H_ε determined from localized inputs and outputs or transformed from the global TFM as in (134). The partition of H_ε which consists of all of the localized inputs and outputs corresponding to a single element is called the "transfer function of the element", and its corresponding transmission zeros are referred to as the transmission zeros of the element. In order to determine properties of localized transmission zeros of element e, the behavior of the block-diagonal partition of H_ε corresponding to e, H_ε^e, must be investigated. From (141), note that the global FRFs can be expressed as the inverse of the global dynamic stiffness matrix, K_g^{-1}, as

$$H_g(\omega)=\frac{1}{P}Z_g,\quad Z_g=\begin{bmatrix} z_{g_{1,1}} & z_{g_{1,2}} & \cdots & z_{g_{1,n}} \\ z_{g_{2,1}} & z_{g_{2,2}} & \cdots & z_{g_{2,n}} \\ \vdots & \vdots & \cdots & \vdots \\ z_{g_{n,1}} & z_{g_{n,2}} & \cdots & z_{g_{n,n}} \end{bmatrix} \tag{144}$$
$$P=\det(K_g),\quad z_{g_{i,j}}=\operatorname{cof}(K_g)(i,j)$$

The global TFM H_g can then be localized according to

$$H(\omega)=LH_g(\omega)L^T=\frac{1}{P}Z,\quad Z=LZ_gL^T \tag{145}$$

since, by (117),

$$q=Lq_g=LH_gu_g=LH_gL^Tu=Hu \tag{146}$$

Specifically, a matrix partition $H(k,k)$ corresponding to element e can be expressed as

$$H(k,k)=\frac{1}{P}Z(k,k),\quad Z(k,k)=L(k,l)Z(l,l)L(k,l)^T \tag{147}$$

where k indicate the elemental displacements of e, and the matrix indices l are the non-zero row entries in L corresponding to columns k. Note that the poles of the transfer function matrix $H=LH_gL^T$ remain unchanged while the zeros of a partition of H are localized due to the disassembly matrix L, which acts only on the elements $Z_g(l,l)$. This is an important point. The poles of a system are not affected by the partitioning transformation. The poles are determined by the global structure alone, and the form and location of the input and output have no bearing on them whatsoever.

Through the process represented by (146), Z_g is successfully localized on the elemental displacements; however, this form does not produce the specific invariance property desired. The block-diagonal partition of Z corresponding to element e retains stiffness attributes of e. In order to remove them, further localization into the strain-based form is required. Using (126) and (129), this process is

$$\begin{aligned} &H_\varepsilon=SHS^T\Rightarrow Z_\varepsilon=SZS^T \\ &Z_\varepsilon(m,m)=S(m,k)Z(k,k)S(m,k)^T \end{aligned} \tag{148}$$

This second transformation effectively annihilates the attributes corresponding to the strain-based block-diagonal rows and columns m of element e. This means that the transmission zeros of $H_\varepsilon(m,m)=H_\varepsilon^e$ do not contain attributes from e, except perhaps its mass and damping. The ILTZ method utilizes this invariance property by comparing the transmission zeros from several localized substructures (that is, the transmission zeros of H_ε^e for several substructures e) for two or more tests. If there is some observed change in the system, then the substructure who's transmission zeros do not change is the location of damage.

At first glance, this invariance property may seem counter-intuitive. Since damage creates changes in mechanical and/or physical properties, it seems logical that any dynamic response characteristics corresponding to the damaged substructure would change, while the response characteristics corresponding to all of the undamaged substructures would remain the same. However, the opposite is true for transmission zeros, due to the effect of the matrix cofactor of the dynamic flexibility in combination with the strain-based localization scheme. Note that a major limitation of ILTZ is that the method is unable to handle the situation where multiple damage sites occur simultaneously. In that case, the transmission zeros of every substructure would change, including those which contain damage. While this is an important limitation, it does not in general rule out the use of ILTZ.

The process of determining the transmission zero variance depends on the particular problem and the form of the data which is available. The localized transfer function matrix H_ε can either be determined directly from localized time history data, or it can be found as a transformation from the global TFM H_g as in (134). H_ε is then partitioned into elemental blocks H_ε^e along the diagonal, and the transmission zeros of each

block are computed by a standard numerical method [63]. A numerical measure is used to quantitatively determine the relative variance or invariance of a set of transmission zeros. Because a set of transmission zeros vary over a range of frequencies, the quantitative measure is defined as a cumulative error over the range of transmission zeros determined for the particular problem. This error factor is called the cumulative transmission zero deviation, represented by D_{TZ}. For the jth zero computed for a given transfer function, D_{TZ} is given as

$$D_{TZ} = \text{mean}\left(\left|\frac{z_d(1:j) - z_n(1:j)}{z_n(1:j)}\right|\right) \tag{149}$$

where z_n indicates the transmission zero values from a nominal healthy or reference test case and z_d represent those from a follow-up or damaged test case. D_{TZ} values are computed separately for each block-diagonal partition of the full transfer function matrix corresponding to separate elements or substructures. The element which has the lowest values of D_{TZ} over the range of transmission zeros is defined as the element where damage occurs.

There are a number of similarities between the localized flexibility method and the ILTZ method. The localized flexibility method directly compares the diagonal values of the partitioned flexibility matrix to determine locations of flexibility changes, while the ILTZ method compares parameters determined from the partitioned form of the dynamic flexibility. Their differences are perhaps more illuminating, because their differences highlight the separate motivations for development of the two systems. The flexibility method can determine multiple damage locations simultaneously, while ILTZ can determine only a single damaged location at a time. Additionally, the flexibility comparison method requires a large number of sensors which enable partitioning to be done in an intelligent manner. This reflects the fact that this method is designed to determine damage locations over an entire structure. This method has been shown to work for a very small number of measured modes of the system, as few as three modes [58]. In contrast, the ILTZ method requires a larger amount of spectral information in order to adequately realize the system dynamics. However, ILTZ can be applied to a single focus "hot spot" of the structure, where damage is likely to occur. In this manner, fewer sensors are required in a smaller region in order to determine damage in the specified area of the structure.

6.5. A simple example

To illustrate the two damage identification methods, consider a spring–mass system as shown in Fig. 7. The global system properties are

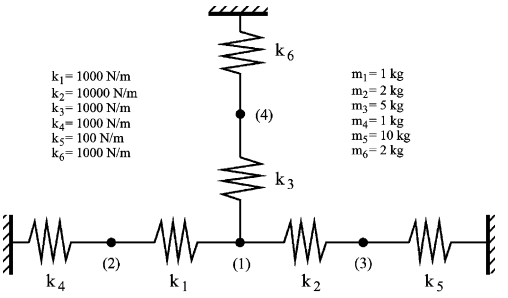

Fig. 7. A simple spring–mass system.

$$K_g = \begin{bmatrix} k_1+k_2+k_3 & -k_1 & -k_2 & -k_3 \\ -k_1 & k_1+k_4 & 0 & 0 \\ -k_2 & 0 & k_2+k_5 & 0 \\ -k_3 & 0 & 0 & k_3+k_6 \end{bmatrix}$$

$$SL = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_g = \begin{bmatrix} \frac{1}{2}(m_1+m_2+m_3) & 0 & 0 & 0 \\ 0 & \frac{1}{2}(m_1+m_4) & 0 & 0 \\ 0 & 0 & \frac{1}{2}(m_2+m_5) & 0 \\ 0 & 0 & 0 & \frac{1}{2}(m_3+m_6) \end{bmatrix} \tag{150}$$

where the bar lengths are taken to be unity, and the masses are lumped at the two elemental nodes. Carrying out the necessary localization, the matrix element of the strain-based numerator polynomial $z_{\varepsilon_{1,1}}$ corresponding to the first element is given by

$$\begin{aligned} z_{\varepsilon_{1,1}} &= (\bar{k}_2+\bar{k}_5)(\bar{k}_3+\bar{k}_6)(\bar{k}_2+\bar{k}_3+\bar{k}_4+2\omega^2 m_1) \\ \bar{k}_i &= k_i - \tfrac{1}{2}\omega^2 m_i \end{aligned} \tag{151}$$

The zeros of the strain-based FRF for element 1 are given by the roots of this polynomial:

$$\bar{k}_2+\bar{k}_5 = 0 \quad \bar{k}_3+\bar{k}_6 = 0 \quad \bar{k}_2+\bar{k}_3+\bar{k}_4+2\omega^2 m_1 = 0 \tag{152}$$

Observe that the stiffness of element 1 is not present in the expression for $z_{\varepsilon_{1,1}}$. The only presence of attributes from element 1 comes from the mass m_1. Hence, we conclude that unless the mass is changed from the reference model, the zeros of strain-based FRFs corresponding to element 1 are *invariant* with respect to its corresponding element stiffness. For other elements, a similar claim can be established.

The transmission zero variation measure can be computed for the example above if numerical values are

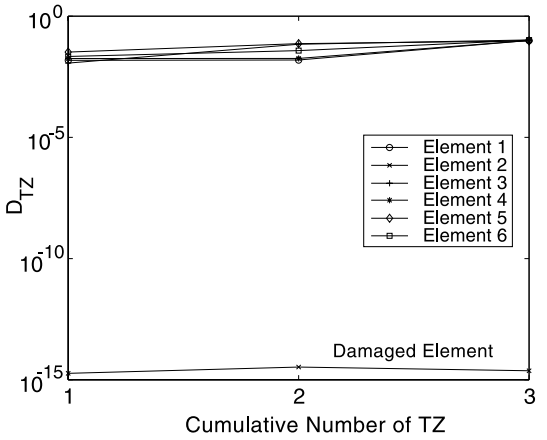

Fig. 8. Strain-based cumulative TZ deviation.

assigned to the mass and stiffness. For the values given in Fig. 7, with damage modeled as a 50% reduction in stiffness in k_2, the plot of D_{TZ} for the localized collocated transfer function partitions is shown in Fig. 8. The strain-based transmission zeros give a clear indication of the damaged element. The variation in the transmission zero set corresponding to element 2 is very small, basically to the order of the machine accuracy. This is many orders of magnitude smaller than the variation in the other transmission zero sets. Therefore, it can be correctly concluded that element 2 is the location of damage.

In comparison to ILTZ, the flexibility method can also be used to determine the damage location. Following (135) and (136), values of RDI for all 4 modes, and for the lowest 2 modes, are computed. Fig. 9 shows that in both cases, the RDI clearly indicates the correct location of damage. For the 4 mode case, there is a small spillover into element 5, which is the neighboring element between element 2 and the boundary condition. In this case, the relative value of the RDI for the damaged element does not reflect the level of damage. Theoretically, the increase in flexibility should be 2.0, because the reduction in stiffness was 50%. Neither of the two RDI values approach that.

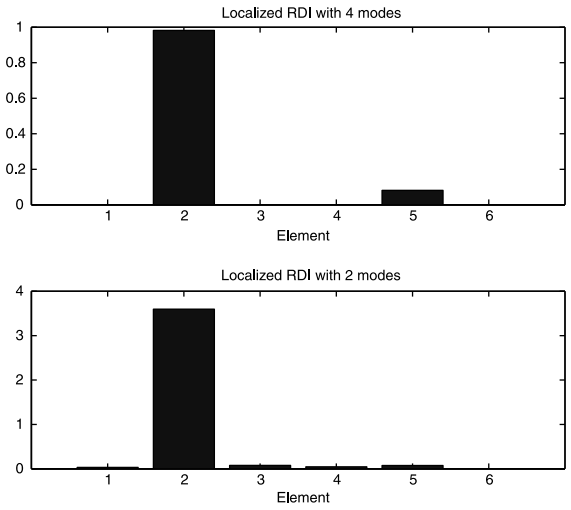

Fig. 9. Strain-based RDI by flexibility method.

6.6. Summary

This section presents two related damage detection methods that are based on a flexibility decomposition theory. The ILTZ method is based on an invariance property of transmission zeros corresponding to substructural transfer functions. The flexibility method is a simple comparison of diagonal values of the localized measured flexibility matrix. The two methods are applied to the health monitoring problem of a simplistic spring–mass system in order to demonstrate their utility. Refs. [64] and [65] present experimental tests, and address issues involved in the successful application of these methods to experimental test problems.

7. Conclusions

The paper has presented a state space-based structural identification theory, its implementation and applications. Hence, from the theoretical viewpoint it has attempted to offer a complete coverage of the underlying theoretical details sufficient for implementing the present theory. There are several topics that are important in carrying out modal testing, which have not been covered in the present paper. These include sensor selection and instrumentation, signal conditioning and data acquisition process, denoising of the measured data, and correlation of the constructed FRFs with the measured FRFs, among others.

From the theoretical aspects, much work still needs to be advanced, especially for the treatment of non-proportional damping, ways to improve the accuracy of identified mode shapes, rigorous ways to extract residual flexibility in order to improve identified model accuracy, and improved data ensembling methods, among others.

Recent advances in sensor technology in terms of the sensitivity improvement of sensors (i.e., gyroscopes to measure accurately the rotational motions) augur well for the construction of high-fidelity structural models from measured response data in routine manners in a near future.

Challenge remains to develop applicable non-linear system identification theory so that much of the present limits on the understanding of non-linear behavior can be alleviated and computer modeling of non-linear behavior can be validated via system-identified non-linear models.

References

[1] Ewins DJ. Modal testing: theory and practice. Research Studies Press Ltd; 1984.

[2] Klosterman A, Zimmerman R. Modal survey activity via frequency response function, SAE Paper 751063, 1975.

[3] Piombo B, Giorcelli E, Garibaldi L, Fasana A. The deconvolution technique applied to mdofs systems. In: Proceedings of the 11th International Modal Analysis Conference, 1993.

[4] Commenges D. The deconvolution problem: fast algorithms including the preconditioned conjugate-gradient to compute a MAP estimator. IEEE Trans Automatic Control 1984;AC-29(3):229–43.

[5] Phillips DJ. A technique for the numerical solution of certain integral equations of the first kind. J Assoc Comput Machinery 1962;9:84–97.

[6] Hanson RJ. A numerical algorithm for solving Fredholm integral equations of the first kind using singular values. SIAM J Numer Anal 1971;8(3):616–22.

[7] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes. Cambridge, MA: Cambridge University Press; 1989.

[8] Juang JN, Phan M, Horta LG, Longman RW. Identification of observer/Kalman filter Markov parameters: theory and experiments. J Guidance, Control Dyn 1993;16(2):320–9.

[9] Daubechies I. Ten lectures on wavelets. Philadelphia, PA: SIAM; 1992.

[10] Graps A. An introduction to wavelets. IEEE Computat Sci Eng 1995;2(Summer):50–61.

[11] Newland DE. Random vibrations, spectral and wavelet analysis. Essex, England: Longman Scientific and Technical; 1993.

[12] Ibrahim SR, Mikulcik EC. A method for the direct identification of vibration parameters from the free response. Shock Vib Bull 1977;47(4):183–98.

[13] Brown D, Allemang R, Zimmerman R, Mergeay, M. Parameter estimation techniques for modal analysis, SAE Paper 790221, SAE Transactions, Vol. 88/1; 1979. p. 828–46.

[14] Vold H, Kundrat J, Rocklin GT, Russell R. A multiple-input modal estimation algorithm for mini-computers, SAE Paper 820194, SAE Transactions, Vol. 91/1; 1982. p. 815–21.

[15] Juang JN, Pappa RS. An eigensystem realization algorithm for modal parameter identification and model reduction. J Guidance, Control Dyn 1985;8(5):620–7.

[16] Liu KT, Skelton RE. The Q-Markov COVER algorithm for flexible structure identification. J Guidance, Control Dyn 1993;16(2):308–19.

[17] Skelton RE. Dynamic systems control. New York: John Wiley and Sons; 1988.

[18] Ho BL, Kalman RE. Effective construction of linear state-variable models from input/output data. In: Proceedings of the Third Annual Allerton Conference on Circuit and System Theory. p. 449–59; also Regelungstechnik, Vol. 14; 1966. p. 545–48.

[19] Kung S. A new identification and model reduction algorithm via singular value decomposition. In: Proceedings of the 12th Asilomar Conference on Circuits, Systems, and Computers, November 1978. p. 705–14.

[20] Wagie DA, Skelton RE. A projection approach to covariance equivalent realizations of discrete systems. IEEE Trans Automatic Control 1986;AC-31:1114.

[21] Juang J-N. Mathematical correlation of modal parameter identification methods via system realization theory. Int J Anal Expt Modal Anal 1987;2(1):1–18.

[22] Juang J-N, Pappa RS. Effects of noise on modal parameters identified by the eigensystem realization algorithm. J Guidance, Control Dyn 1986;9(3):294–303.

[23] Juang J-N, Cooper JE, Wright JR. An eigensystem realization algorithm using data correlations (ERA/DC) for modal parameter identification. Control Theor Adv Technol 1986;4(1):294–303.

[24] Peterson LD. Efficient computation of the eigensystem realization algorithm. In: Proceedings of the 10th International Modal Analysis Conference, February 1992, J Guidance, Control Dyn 1995;18(3):395–403.

[25] Meirovitch L. Analytical methods in vibrations. New York: McGraw-Hill; 1960.

[26] Caughey TK, O'Kelly MMJ. Classical normal modes in damped linear dynamical systems. J Appl Mech 1965;32:583–8.

[27] Lang GF. Demystifying complex modes. J Sound Vib 1989;23:36–40.

[28] Imregun M, Ewins DJ. Realisation of complex mode shapes. In: Proceedings of the 11th Annual International Modal Analysis Conference, February 1993. p. 1303–9.

[29] Mayes RL. A multi-degree-of-freedom mode shape estimation algorithm using quadrature response. In: Proceedings of the 11th International Modal Analysis Conference, February 1993. p. 1026–34.

[30] Longman RW, Juang JN. A variance based confidence criterion for ERA-identified modal parameters, Paper AAS 87-454, AAS/AIAA Astrodynamics Specialist Conference, Kalispell, MT, August, 1987.

[31] Ibrahim SR. Computation of normal modes from identified complex modes. AIAA J 1983;21(3):446–51.

[32] Niedbal N. Analytical determination of real normal modes from measured complex responses. In: Proceedings of the 25th Structures, Structural Dynamics and Materials Conference, Palm Springs, CA, May 1984. p. 292–5.

[33] Zhang Q, Lallement GL. New method of determining the eigensolutions of the associated conservative structure from the identified eigensolutions. In: Proceedings of the Third International Modal Analysis Conference; 1985. p. 322–8.

[34] Zhang Q, Lallement GL. Comparison of normal eigenmodes calculation method based on identified complex eigenmodes. J Spacecraft Rockets 1987;24(1):69–73.

[35] Wei ML, Allemang RJ, Brown DL. Real-normalization of measured complex modes. In: Proceedings of the Fifth International Modal Analysis Conference; 1987. p. 708–12.

[36] Placidi F, Poggi F, Sestieri A. Real modes computation from identified modal parameters with estimate of generalized damping. In: Proceedings of the Ninth International Modal Analysis Conference; 1991. p. 572–9.

[37] Minas DJ, Inman DJ. Identification of a non-proportional damping matrix from incomplete model information. J Vib Acoust 1991;113(2):219–24.

[38] Sestieri A, Ibrahim SR. Analysis of errors and approximations in the use of modal coordinates. In: Proceedings of the 11th International Modal Analysis Conference; 1993. p. 425–33.
[39] Hsu C, Ju MS, Tsuei YG. Extraction of normal modes from experimental data. In: Proceedings of the 11th International Modal Analysis Conference; 1993. p. 1612–17.
[40] Yang C-D, Yeh F-B. Identification, reduction and refinement of model parameters by the eigensystem realization algorithm. J Guidance, Control Dyn 1990;13(6):1051–9.
[41] Witherell CE. Mechanical failure avoidance: strategies and techniques. New York: McGraw-Hill; 1994.
[42] Hickman GA, Gerardi JJ, Feng Y. Application of smart structure to aircraft health monitoring. J Intelligent Mater Syst Struct 1991;2:411–30.
[43] Doebling SW et al., 1996. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review, Los Alamos National Laboratory, Paper No. LA-13070-MS.
[44] Pandey AK, Biswas M, Samman MM. Damage detection from changes in curvature mode shapes. J Sound Vib 1991;145(2):321–32.
[45] Toksoy T, Atkan AE. Bridge-condition assessment by modal flexibility. Exp Mech 1994;34:271–8.
[46] Zimmerman DC, Kaouk M. Structural damage detection using a minimum rank update theory. J Vib Acoust 1994;116:222–30.
[47] Kranock SJ, Peterson LD. Real-time structural health monitoring using model-based observers, Paper no. 3323-83. In: Proceedings of the Fifth SPIE International Symposium on Smart Structures and Materials, San Diego, CA; 1998.
[48] Reich GW, Park KC. Use of substructural transmission zeros for structural health monitoring. AIAA J 2000;38(6):1040–6.
[49] Afolabi D. An anti-resonance technique for detecting structural damage. In: Proceedings of the Fifth International Modal Analysis Conference; 1987. p. 491–5.
[50] Felippa CA, Park KC. Computational aspects of time integration procedures in structural dynamics. Part 1. Implementation. J Appl Mech 1978;45(3):595–602.
[51] Bendat JS, Piersol AG. Engineering applications of correlation and spectral analysis. New York: John Wiley and Sons; 1980.
[52] Robertson A, Park KC, Alvin KF. Extraction of impulse response data via wavelet transforms for structural system identification. J Vib Acoust 1998;120(1):252–60.
[53] Williams JR, Amaratunga K. Introduction to wavelets in engineering. Int J Numer Meth Eng 1994;37:2365–88.
[54] Strang G. Wavelet transforms vs. Fourier transforms. Bull (New Series) AMS 1993;28(2):288–305.
[55] Alvin KF, Park KC. Second-order structural identification procedure via state space-based system identification. AIAA J 1994;32(2):397–406.
[56] Alvin KF, Peterson LD, Park KC. Extraction of normal modes and full modal damping from complex modal parameters. AIAA J 1997;35(7):1187–94.
[57] Pappa RS, Elliot KB, Schenk A. A consistent-mode indicator for the eigensystem realization algorithm. In: Proceedings of the 1992 AIAA Structures, Structural Dynamics, and Materials Conference, AIAA Paper 92-2136-CP, April 1992. p. 556–65.
[58] Park KC, Reich GW, Alvin KF. Structural damage detection using localized flexibilities. J Intelligent Mater Syst Struct 1998;9(11):911–9.
[59] Reich GW, Park KC. A theory for strain-based structural system identification. J Appl Mech 2001;68(4):521–7.
[60] Park KC, Felippa CA. A variational framework for solution method developments in structural mechanics. J Appl Mech 1998;65(1):242–9.
[61] MacFarlane AGJ, Karcanias N. Poles and zeros of linear multivariable systems: a survey of the algebraic, geometric, and complex-variable theory. Int J Control 1976;24(1):33–74.
[62] Horn RA, Johnson CR. Matrix analysis. Cambridge, MA: Cambridge University Press; 1985.
[63] Emami-Naeini A, Van Dooren P. Computation of zeros of linear multivariable systems. Automatica 1982;18(4):415–30.
[64] Reich GW, Park KC. Experimental damage detection using substructural transmission zeros, AIAA Paper No. 2000-1699. In: Proceedings of the 41st AIAA Structures, Structural Dynamics, and Materials Conference, Atlanta, GA.
[65] Reich GW, Park KC. Experimental application of a structural health monitoring methodology, Paper No. 3323-83. In: Proceedings of the SPIE International Symposium on Smart Structures and Materials 2000, Vol. 3988, Newport Beach, CA.