CHAPTER 21
EXPERIMENTAL MODAL ANALYSIS

Randall J. Allemang
David L. Brown

INTRODUCTION

Experimental modal analysis is the process of determining the modal parameters (natural frequencies, damping factors, modal vectors, and modal scaling) of a linear, time-invariant system. The modal parameters are often determined by analytical means, such as finite element analysis. One common reason for experimental modal analysis is the verification or correction of the results of the analytical approach. Often, an analytical model does not exist, and the modal parameters determined experimentally serve as the model for future evaluations, such as structural modifications. Predominantly, experimental modal analysis is used to explain a dynamics problem (vibration or acoustic) whose solution is not obvious from intuition, analytical models, or previous experience.

The process of determining modal parameters from experimental data involves several phases. The success of the experimental modal analysis process depends upon having very specific goals for the test situation. Every phase of the process is affected by the goals which are established, particularly with respect to the errors associated with that phase. One possible delineation of these phases is as follows:

Modal analysis theory refers to that portion of classical vibration theory that explains the existence of natural frequencies, damping factors, and mode shapes for linear systems. This theory includes both lumped-parameter, or discrete, models and continuous models. This theory also includes real normal modes as well as complex modes of vibration as possible solutions for the modal parameters.[1–3]

Experimental modal analysis methods involve the theoretical relationship between measured quantities and classical vibration theory, often represented as matrix differential equations. All commonly used methods trace from the matrix differential equations but yield a final mathematical form in terms of measured raw input and output data in the time or frequency domains or some form of processed data such as impulse-response or frequency response functions.

Modal data acquisition involves the practical aspects of acquiring the data that are required to serve as input to the modal parameter estimation phase. Much care must be taken to assure that the data match the requirements of the theory as well as the requirements of the numerical algorithm involved in the modal parameter estimation. The theoretical requirements involve concerns such as system linearity and time invariance of system parameters. The numerical algorithms are particularly concerned with the bias errors in the data as well as with any overall dynamic range considerations[4–7] (see Chap. 22).

Modal parameter estimation is concerned with the practical problem of estimating the modal parameters, based upon a choice of mathematical model as justified by the experimental modal analysis method, from the measured data.[8–10]

Modal data presentation/validation is the process of providing a physical view or interpretation of the modal parameters. For example, this may simply be the numerical tabulation of the frequency, damping, and modal vectors along with the associated geometry of the measured degrees-of-freedom. More often, modal data presentation involves the plotting and animation of such information.

Figure 21.1 is a representation of all phases of the process. In this example, a continuous beam is being evaluated for the first few modes of vibration. Modal analysis theory explains that this is a linear system and that the modal vectors of this system should be real normal modes. The experimental modal analysis method that has been used is based upon the relationships of the frequency response function to the matrix differential equations of motion. At each measured degree-of-freedom (DOF), the imaginary part of the frequency response function for that measured response degree-of-freedom and a common input degree-of-freedom is superimposed perpendicular to the beam. Naturally, the modal data acquisition in this example involves the estimation of frequency response functions for each degree-of-freedom shown. The frequency response functions are complex-valued functions, and only the imaginary portion of each function is shown. One method of modal parameter estimation suggests that for systems with light damping and widely spaced modes, the imaginary part of the frequency response function at the damped natural frequency may be used as an estimate of the modal coefficient for that response degree-of-freedom. The damped natural frequency can be identified as the frequency of the positive and negative peaks in the imaginary part of the frequency response functions. The damping can be estimated from the sharpness of the peaks. In this abbreviated way, the modal parameters have been estimated. Modal data presentation for this case is shown as the lines connecting the peaks. While animation is possible, a reasonable interpretation of the modal vector can be gained in this case from plotting alone.

MEASUREMENT DEGREES-OF-FREEDOM

The development of any theoretical concept in the area of vibrations, including modal analysis, depends upon an understanding of the concept of the number of degrees-of-freedom n of a system. This concept is extremely important to the area of modal analysis since the number of modes of vibration of a mechanical system is equal to the number of degrees-of-freedom. From a practical point of view, the relationship between this theoretical definition of the number of degrees-of-freedom and the number of *measurement degrees-of-freedom* N_o, N_i is often confusing. For this reason, the concept of degree-of-freedom is reviewed as a preliminary to the following experimental modal analysis material.

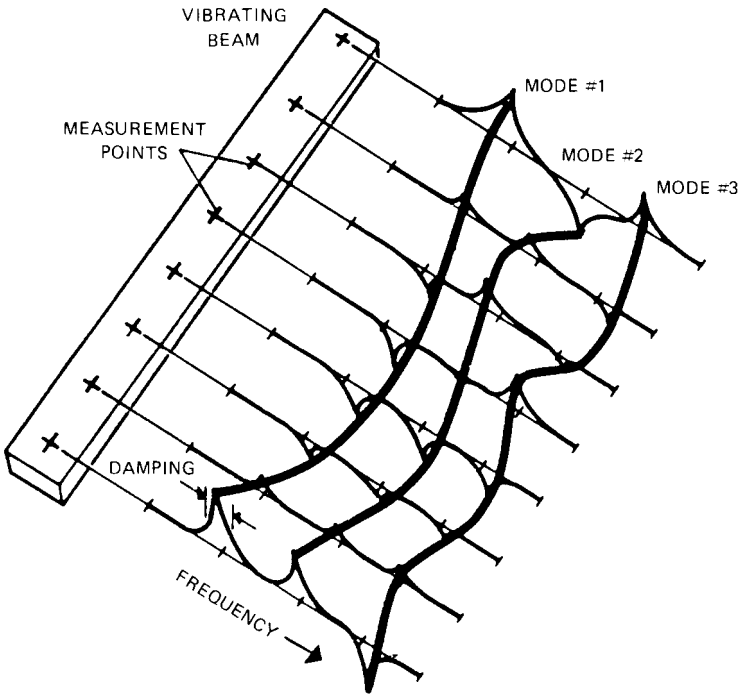

FIGURE 21.1 Experimental modal analysis example using the imaginary part of the frequency response functions.

To begin with, the basic definition that is normally associated with the concept of the number of degrees-of-freedom involves the following statement: *The number of degrees-of-freedom for a mechanical system is equal to the number of independent coordinates (or minimum number of coordinates) that is required to locate and orient each mass in the mechanical system at any instant in time.* As this definition is applied to a point mass, 3 degrees-of-freedom are required since the location of the point mass involves knowing the *x, y,* and *z* translations of the center-of-gravity of the point mass. As this definition is applied to a rigid body mass, 6 degrees-of-freedom are required since θ_x, θ_y, and θ_z rotations are required in addition to the *x, y,* and *z* translations in order to define both the orientation and the location of the rigid body mass at any instant in time. As this definition is extended to any general deformable body, the number of degrees-of-freedom is essentially infinite. However, while this is theoretically true, it is quite common, particularly with respect to finite element methods, to view the general deformable body in terms of a large number of physical points of interest with 6 degrees-of-freedom for each of the physical points. In this way, the infinite number of degrees-of-freedom can be reduced to a large but finite number.

When measurement limitations are imposed upon this theoretical concept of the number of degrees-of-freedom of a mechanical system, the difference between the theoretical number of degrees-of-freedom n and the number of measurement degrees-of-freedom N_o, N_i begins to evolve. Initially, for a general deformable body, the number of degrees-of-freedom n can be considered to be infinite or equal to

some large finite number if a limited set of physical points of interest is considered, as discussed in the previous paragraph. The first measurement limitation that needs to be considered is that there is normally a limited frequency range that is of interest to the analysis. When this limitation is considered, the number of degrees-of-freedom of this system that are of interest is reduced from infinity to a reasonable finite number. The next measurement limitation that needs to be considered involves the physical limitation of the measurement system in terms of amplitude. A common limitation of transducers, signal conditioning and data acquisition systems results in a dynamic range of 80 to 100 dB (10^4 to 10^5) in the measurement. This means that the number of degrees-of-freedom is reduced further because of the dynamic range limitations of the measurement instrumentation. Finally, since few rotational transducers exist at this time, the normal measurements that are made involve only translational quantities (displacement, velocity, acceleration, force) and thus do not include rotational effects, or RDOF. In summary, even for the general deformable body, the theoretical number of degrees-of-freedom that are of interest is limited to a very reasonable finite value ($n = 1$ to 50). Therefore, this number of degrees-of-freedom n is the number of modes of vibration that are of interest.

Finally, then, the number of measurement degrees-of-freedom N_o, N_i can be defined as the number of physical locations at which measurements are made multiplied by the number of measurements made at each physical location. Since the physical locations are chosen somewhat arbitrarily, and certainly without exact knowledge of the modes of vibration that are of interest, there is no specific relationship between the number of degrees-of-freedom n and the number of measurement degrees-of-freedom N_o, N_i. In general, in order to define n modes of vibration of a mechanical system, N_o or N_i must be equal to or larger than n. However, N_o or N_i being larger than n is not a guarantee that n modes of vibration can be found from the measurement degrees-of-freedom. The measurement degrees-of-freedom must include physical locations that allow a unique determination of the n modes of vibration. For example, if none of the measurement degrees-of-freedom are located on a portion of the mechanical system that is active in one of the n modes of vibration, portions of the modal parameters for this mode of vibration cannot be found.

In the development of material in the following text, the assumption is made that a set of measurement degrees-of-freedom exists that allows n modes of vibration to be determined. In reality, either N_o or N_i is always chosen much larger than n since a prior knowledge of the modes of vibration is not available. If the set of N_o or N_i measurement degrees-of-freedom is large enough and if the measurement degrees-of-freedom are distributed uniformly over the general deformable body, the n modes of vibration are normally found.

Throughout this experimental modal analysis reference, the frequency response function notation H_{pq} is used to describe the measurement of the response at measurement degree-of-freedom p resulting from an input applied at measurement degree-of-freedom q. The single subscript p or q refers to a single sensor aligned in a specific direction ($\pm X, Y,$ or Z) at a physical location on or within the structure.

BASIC ASSUMPTIONS

There are four basic assumptions concerning any structure that are made in order to perform an experimental modal analysis:

1. *The structure is assumed to be linear,* i.e., the response of the structure to any combination of forces, simultaneously applied, is the sum of the individual responses

to each of the forces acting alone. For a wide variety of structures this is a very good assumption. When a structure is linear, its behavior can be characterized by a controlled excitation experiment in which the forces applied to the structure have a form that is convenient for measurement and parameter estimation rather than being similar to the forces that are actually applied to the structure in its normal environment. For many important kinds of structures, however, the assumption of linearity is not valid. Where experimental modal analysis is applied in these cases, it is hoped that the linear model that is identified provides a reasonable approximation of the structure's behavior.

2. *The structure is time invariant,* i.e., the parameters that are to be determined are constants. In general, a system which is not time invariant has components whose mass, stiffness, or damping depend on factors that are not measured or are not included in the model. For example, some components may be temperature dependent. In this case, since temperature effects are not measured, the temperature of the component is an unknown time-varying signal. Hence, the component has time-varying characteristics. Therefore, for this case the modal parameters determined by any measurement and estimation process depend on the time (and the associated temperature dependence) when the measurements are made. If the structure that is tested changes with time, then measurements made at the end of the test period determine a different set of modal parameters from measurements made at the beginning of the test period. Thus, the measurements made at the two different times are inconsistent, violating the assumption of time invariance.

3. *The structure obeys Maxwell's reciprocity,* i.e., a force applied at degree-of-freedom p causes a response at degree-of-freedom q that is the same as the response at degree-of-freedom p caused by the same force applied at degree-of-freedom q. With respect to frequency response function measurements, the frequency response function between points p and q determined by exciting at p and measuring the response at q is the same frequency response function found by exciting at q and measuring the response at p ($H_{pq} = H_{qp}$).

4. *The structure is observable,* i.e., the input-output measurements that are made contain enough information to generate an adequate behavioral model of the structure. Structures and machines which have loose components, or, more generally, which have degrees-of-freedom of motion that are not measured, are not completely observable. For example, consider the motion of a partially filled tank of liquid when complicated sloshing of the fluid occurs. Sometimes enough data can be collected so that the system is observable under the form chosen for the model, while at other times an impractical amount of data is required. This assumption is particularly relevant to the fact that the data normally describe an incomplete model of the structure. This occurs in at least two different ways. First, the data are normally limited to a minimum and maximum frequency as well as a limited frequency resolution. Second, no information relative to local rotations is available because of the lack of available transducers in this area.

MODAL ANALYSIS THEORY

While modal analysis theory has not changed over the last century, the application of the theory to experimentally measured data has changed significantly. The advances of recent years with respect to measurement and analysis capabilities have caused a reevaluation of what aspects of the theory relate to the practical world of

testing. Thus, the aspect of transform relationships has taken on renewed importance since digital forms of the integral transforms are in constant use. The theory from the vibrations point of view involves a more thorough understanding of how the structural parameters of mass, damping, and stiffness relate to the impulse-response function (time domain), the frequency response function (Fourier or frequency domain), and the transfer function (Laplace domain) for single and multiple degree-of-freedom systems.

SINGLE DEGREE-OF-FREEDOM SYSTEMS

In order to understand modal analysis, complete comprehension of single degree-of-freedom systems is necessary. In particular, complete familiarity with single degree-of-freedom systems as presented and evaluated in the time, frequency (Fourier), and Laplace domains serves as the basis for many of the models that are used in modal parameter estimation. This single degree-of-freedom approach is trivial from a modal analysis perspective since no modal vectors exist. The true importance of this approach results from the fact that the multiple degree-of-freedom case can be viewed as simply a linear superposition of single degree-of-freedom systems.

The general mathematical representation of a single degree-of-freedom system is expressed by

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) \tag{21.1}$$

where m = mass constant
c = damping constant
k = stiffness constant

This differential equation yields a characteristic equation of the following form:

$$ms^2 + cs + k = 0 \tag{21.2}$$

where s is the complex-valued frequency variable (Laplace variable). This characteristic equation of a single degree-of-freedom system has two roots, λ1 and λ2, which are

$$\lambda_1 = -\sigma_1 + j\omega_1 \qquad \lambda_2 = -\sigma_2 + j\omega_2 \tag{21.3}$$

where σ_1 = damping factor for mode 1
ω_1 = damped natural frequency for mode 1

Thus, the complementary solution of Eq. (21.1) is

$$x(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t} \tag{21.4}$$

A and B are complex-valued constants determined from the initial conditions imposed on the system at time $t = 0$.

For most real structures, unless active damping systems are present, the damping ratio is rarely greater than 10 percent. For this reason, all further discussion is restricted to underdamped systems ($\zeta < 1$). With reference to Eq. (21.2), this means that the two roots λ_1 and λ_2 are always complex conjugates. Also, the two coefficients, A and B, are complex conjugates of each other. For an underdamped system, the roots of the characteristic equation can be written as

$$\lambda_1 = \sigma_1 + j\omega_1 \qquad \lambda_1^* = \sigma_1 - j\omega_1 \tag{21.5}$$

where σ_1 = damping factor
ω_1 = damped natural frequency

The roots of the characteristic equation (21.2) can also be written as

$$\lambda_1 = -\zeta_1\Omega_1 \pm j\Omega_1\sqrt{1-\zeta_1^2} \tag{21.6}$$

The *damping factor* is defined as the real part of a root of the characteristic equation. The damping factor describes the exponential decay or growth of the harmonic. This parameter has the same units as the imaginary part of the root of the characteristic equation, typically radians per second.

Time Domain: Impulse-Response Function. The *impulse-response function* of the single degree-of-freedom system is defined as the time response $x(t)$ of the system, assuming that the initial conditions are zero and that the system excitation $f(t)$ is a unit impulse. The response of the system $x(t)$ to such a unit impulse is known as the impulse-response function $h(t)$ of the system. Therefore

$$h(t) = Ae^{\lambda_1 t} + A^* e^{\lambda_1^* t} = e^{\sigma_1 t}\left[\, Ae^{+j\omega_1 t} + A^* e^{-j\omega_1 t}\,\right] \tag{21.7}$$

Thus, the residue A controls the amplitude of the impulse response, the real part of the pole is the decay rate, and the imaginary part of the pole is the frequency of oscillation. Figure 21.2 illustrates the impulse-response function for a single degree-of-freedom system.

Frequency Domain: Frequency Response Function. An equivalent equation of motion for Eq. (21.1) is determined for the Fourier or frequency (ω) domain. This representation has the advantage of converting a differential equation to an algebraic equation. This is accomplished by taking the Fourier transform of Eq. (21.1). Thus, Eq. (21.1) becomes

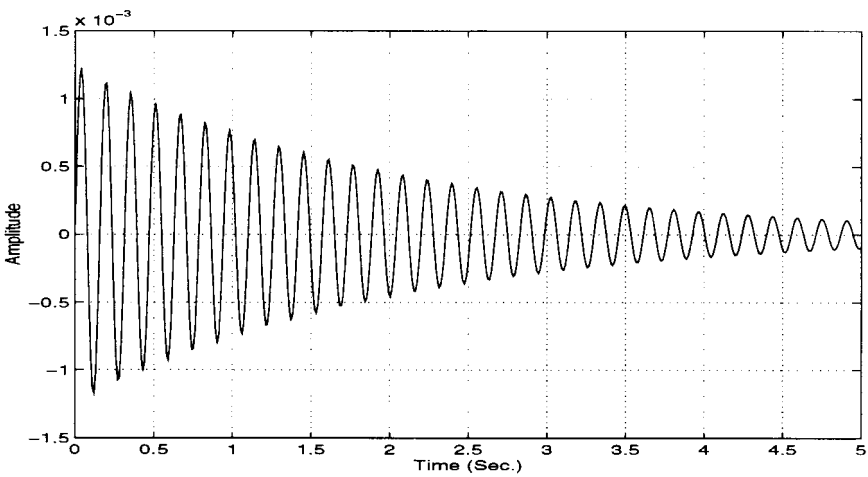

FIGURE 21.2 Single degree-of-freedom impulse-response function.

$$[\, -m\omega^2 + jc\omega + k \,]\, X(\omega) = F(\omega) \tag{21.8}$$

Restating the above equation,

$$X(\omega) = H(\omega)\, F(\omega) \tag{21.9}$$

where

$$H(\omega) = \frac{1}{-m\omega^2 + jc\omega + k}$$

Equation (21.9) states that the system response $X(\omega)$ is directly related to the system forcing function $F(\omega)$ through the quantity $H(\omega)$. If the system forcing function $F(\omega)$ and its response $X(\omega)$ are known, $H(\omega)$ can be calculated. That is,

$$H(\omega) = \frac{X(\omega)}{F(\omega)} \tag{21.10}$$

The quantity $H(\omega)$ is known as the *frequency response function* of the system. The frequency response function relates the Fourier transform of the system input to the Fourier transform of the system response.

The denominator of Eq. (21.9) is known as the *characteristic equation* of the system and is of the same form as Eq. (21.2). Note that the characteristic values of this complex equation are in general complex even though the equation is a function of a real-valued independent variable ω. The characteristic values of this equation are known as the *complex roots* of the characteristic equation or the *complex poles* of the system. In terms of modal parameters, these characteristic values are also called the *modal frequencies*.

The frequency response function $H(\omega)$ can now be rewritten as a function of the complex poles as follows:

$$H(\omega) = \frac{1/m}{(\, j\omega - \lambda_1)(\, j\omega - \lambda_1^*)} \tag{21.11}$$

where $\lambda_1 =$ complex pole $= \sigma + j\omega_1$
$\lambda_1^* = \sigma - j\omega_1$

Since the frequency response function is a complex-valued function of a real-valued independent variable ω, it is represented by a pair of curves, as shown in Fig. 21.3.

Laplace Domain: Transfer Function. Just as in the previous case for the frequency domain, the equivalent information can be presented in the Laplace domain by way of the Laplace transform. The only significant difference in the development concerns the fact that the Fourier transform is defined from negative infinity to positive infinity, while the Laplace transform is defined from zero to positive infinity with initial conditions. The Laplace representation, also, has the advantage of converting a differential equation to an algebraic equation.

The transfer function is defined in the same way that the frequency response function is defined (assuming zero initial conditions):

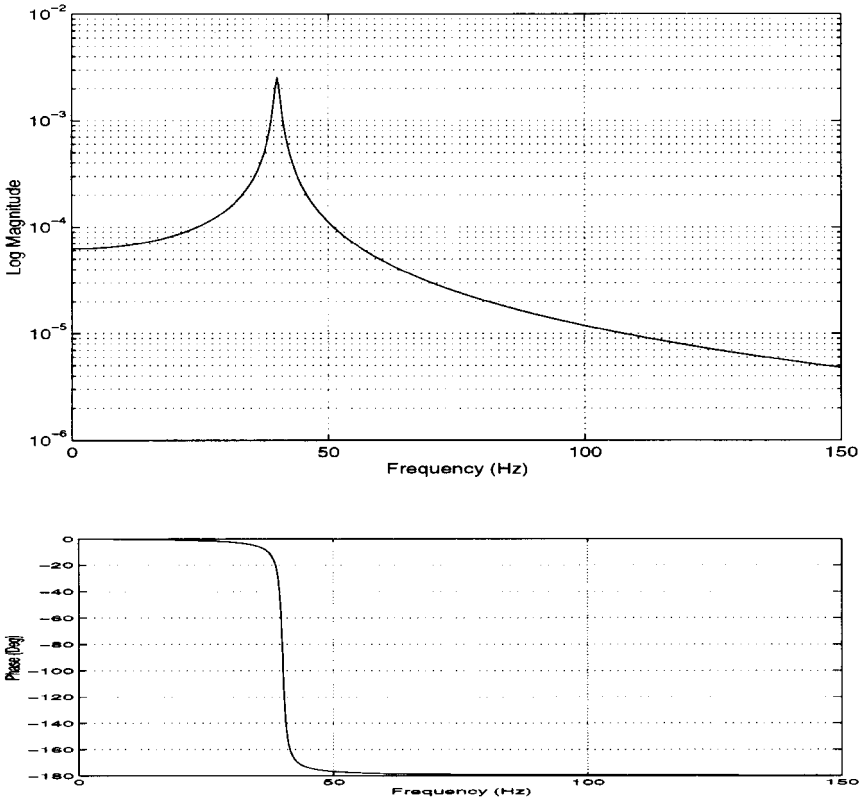

FIGURE 21.3 Single degree-of-freedom frequency response function (log magnitude/phase format).

$$X(s) = H(s)\, F(s) \tag{21.12}$$

where

$$H(s) = \frac{1}{ms^2 + cs + k}$$

The quantity $H(s)$ is defined as the *transfer function* of the system. The transfer function relates the Laplace transform of the system input to the Laplace transform of the system response. From Eq. (21.12), the transfer function is defined as

$$H(s) = \frac{X(s)}{F(s)} \tag{21.13}$$

The denominator term is once again referred to as the characteristic equation of the system. As noted in the previous two cases, the roots of the characteristic equation are given in Eq. (21.5).

The transfer function $H(s)$ is now rewritten, just as in the frequency response function case, as

$$H(s) = \frac{1/m}{(s - \lambda_1)(s - \lambda_1^*)} \tag{21.14}$$

Since the transfer function is a complex-valued function of a complex independent variable *s,* it is represented, as shown in Fig. 21.4, as a pair of surfaces.

The definition of undamped natural frequency, damped natural frequency, damping factor, percent of critical damping, and residue are all relative to the information represented by Fig. 21.4. The projection of this information onto the plane of zero amplitude yields the information shown in Fig. 21.5.

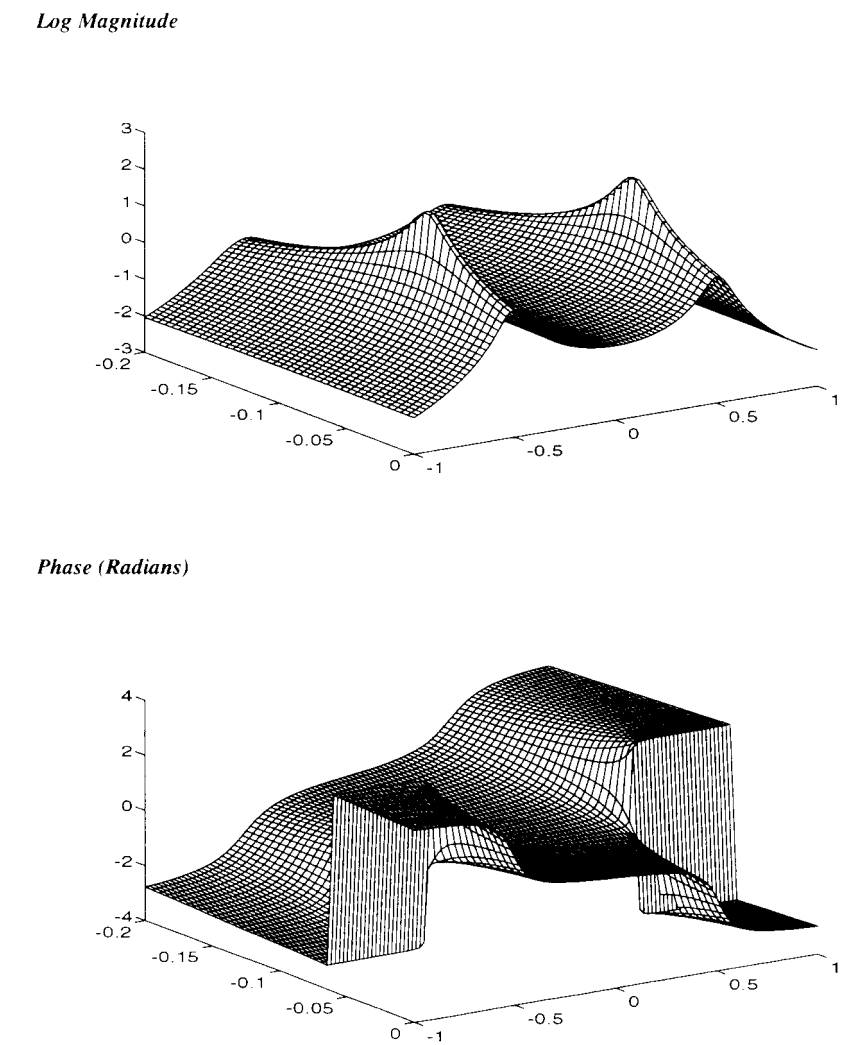

FIGURE 21.4 Single degree-of-freedom transfer function (log magnitude/phase format).

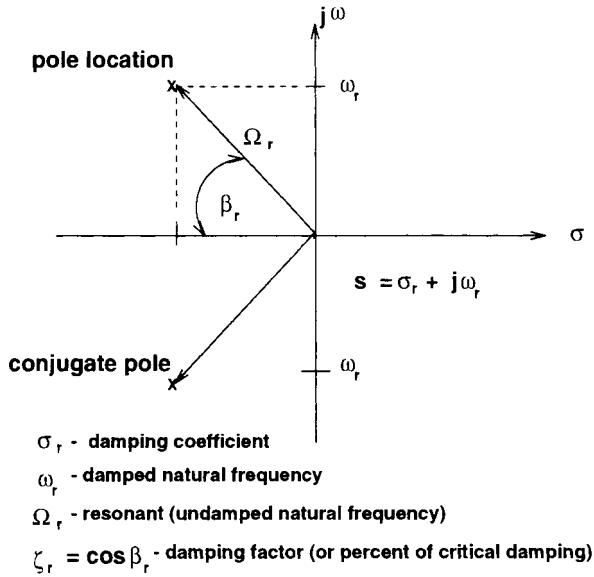

FIGURE 21.5 Transfer function (Laplace domain projection).

The concept of residues is now defined in terms of the partial fraction expansion of the transfer function or frequency response function equation. Equation (21.27) is expressed in terms of partial fractions as follows:

$$H(s) = \frac{1/m}{(s-\lambda_1)(s-\lambda_1^*)} = \frac{A}{s-\lambda_1} + \frac{A^*}{s-\lambda_1^*} \tag{21.15}$$

The *residues* of the transfer function are defined as the constants A and A^*. The terminology and development of residues comes from the evaluation of analytic functions in complex analysis. The residues of the transfer function are directly related to the amplitude of the impulse-response function. In general, the residue A is a complex quantity. As shown for a single degree-of-freedom system, A is purely imaginary. From an experimental point of view, the transfer function is not estimated from measured input-output data. Instead, the frequency response function is actually estimated via the discrete Fourier transform.

MULTIPLE DEGREE-OF-FREEDOM SYSTEMS

Modal analysis concepts are applied when a continuous, nonhomogeneous structure is described as a lumped-mass, multiple degree-of-freedom system. The modal (natural) frequencies, the modal damping, the modal vectors, or relative patterns of motion, and the modal scaling can be found from an estimate of the mass, damping, and stiffness matrices or from the measurement of the associated frequency response functions. From the experimental viewpoint, the relationship of modal parameters with respect to measured frequency response functions is most important.

The development of the frequency response function solution for the multiple degree-of-freedom case parallels that for the single degree-of-freedom case. This development relates the mass, damping, and stiffness matrices to a matrix transfer function model, or matrix frequency response function model, involving multiple degrees-of-freedom. Just as in the analytical case where the ultimate solution can be described in terms of single degree-of-freedom systems, the frequency response functions between any input and response degree-of-freedom can be represented as a linear superposition of the single degree-of-freedom models derived previously.

As a result of the linear superposition concept, the equations for the impulse-response function, the frequency response function, and the transfer function for the multiple degree-of-freedom system are defined as follows:

Impulse-response function:

$$h_{pq}(t) = \sum_{r=1}^{n} A_{pqr} e^{\lambda_r t} + A^*_{par} e^{\lambda_r * t} \tag{21.16}$$

Frequency response function:

$$H_{pq}(\omega) = \sum_{r=1}^{n} \frac{A_{pqr}}{j\omega - \lambda_r} + \frac{A^*_{pqr}}{j\omega - \lambda_r *} \tag{21.17}$$

Transfer function:

$$H_{pq}(s) = \sum_{r=1}^{n} \frac{A_{pqr}}{s - \lambda_r} + \frac{A^*_{pqr}}{s - \lambda_r *} \tag{21.18}$$

where t = time variable
ω = frequency variable
s = Laplace variable
p = measured degree-of-freedom (response)
q = measured degree-of-freedom (input)
r = modal vector number
A_{pqr} = residue
$A = Q_r \psi_{pr} \psi_{qr}$
Q_r = modal scaling factor
ψ_{pr} = modal coefficient
λ_r = system pole
n = number of modal frequencies

It is important to note that the residue, A_{pqr}, in Eqs. (21.16) through (21.18) is the product of the modal deformations at the input q and response p degrees-of-freedom and a modal scaling factor for mode r. Therefore, while the product of these three terms is unique, each of the three terms individually is not unique.

Modal scaling refers to the relationship between the normalized modal vectors and the absolute scaling of the mass matrix (analytical case) and/or the absolute scaling of the residue information (experimental case). Modal scaling is normally presented as *modal mass* or *modal A*. The driving point residue, A_{qqr}, is particularly important in deriving the modal scaling.

$$A_{qqr} = Q_r \psi_{qr} \psi_{qr} = Q_r \psi_{qr}^2 \tag{21.19}$$

For undamped and proportionally damped systems, the *r*th modal mass of a multiple degree-of-freedom system can be defined as

$$M_r = \frac{1}{j2Q_r\omega_r} = \frac{\psi_{pr}\psi_{qr}}{j2A_{pqr}\omega_r} \tag{21.20}$$

where M_r = modal mass
Q_r = modal scaling constant
ω_r = damped natural frequency

If the largest scaled modal coefficient is equal to unity, Eq. (21.20) computes a quantity of modal mass that has physical significance. The physical significance is that the quantity of modal mass computed under these conditions is between zero and the total mass of the system. Therefore, under this scaling condition, the modal mass can be viewed as the amount of mass that is participating in each mode of vibration. For a translational rigid body mode of vibration, the modal mass should be equal to the total mass of the system. The modal mass defined in Eq. (21.20) is developed in terms of displacement over force units. If measurements, and therefore residues, are developed in terms of any other units (velocity over force or acceleration over force), Eq. (21.20) has to be altered accordingly.

Once the modal mass is known, the *modal damping* C_r and *stiffness* K_r can be obtained through the following single degree-of-freedom equations:

$$C_r = 2\sigma_r M_r \tag{21.21}$$

$$K_r = (\sigma_r^2 + \omega_r^2) M_r = \Omega_r^2 M_r \tag{21.22}$$

For systems with nonproportional damping, modal mass cannot be used for modal scaling. For this case, and increasingly for the undamped and proportionally damped cases as well, the modal *A* scaling factor is used as the basis for the relationship between the scaled modal vectors and the residues determined from the measured frequency response functions. This relationship is as follows:

$$M_{A_r} = \frac{\psi_{pr}\psi_{qr}}{A_{pqr}} = \frac{1}{Q_r} \tag{21.23}$$

Note that this definition of modal *A* is also developed in terms of displacement over force units. Once the modal *A* is known, *modal B* (M_{B_r}) can be obtained through the following single degree-of-freedom equation:

$$M_{B_r} = -\lambda_r M_{A_r} \tag{21.24}$$

For undamped and proportionally damped systems, the relationship between the modal mass and the modal *A* scaling factors can be uniquely determined as

$$M_{A_r} = \pm j2M_r\omega_r \tag{21.25}$$

In general, the modal vectors are considered to be dimensionless since they represent relative patterns of motion. Therefore, the modal mass or modal *A* scaling terms carry the units of the respective measurement. For example, the development of the

frequency response is based upon displacement over force units. The residue must have units of length over force-seconds. Since the modal A scaling coefficient is inversely related to the residue, modal A has units of force-seconds over length. This unit combination is the same as mass over seconds. Likewise, since modal mass is related to modal A, for proportionally damped systems, through a direct relationship involving the damped natural frequency, the units of modal mass are mass units, as expected.

DAMPING MECHANISMS

In order to evaluate multiple degree-of-freedom systems that are present in the real world, the effect of damping on the complex frequencies and modal vectors must be considered. Many physical mechanisms are needed to describe all of the possible forms of damping that may be present in a particular structure or system. Some of the classical types are (1) structural damping, (2) viscous damping, and (3) Coulomb damping. It is generally difficult to ascertain which type of damping is present in any particular structure. Indeed most structures exhibit damping characteristics that result from a combination of all the above, plus others that have not been described here. (Damping is described in detail in Chap. 36.)

Rather than consider the many different physical mechanisms, the probable location of each mechanism, and the particular mathematical representation of the mechanism of damping that is needed to describe the dissipative energy of the system, a model is used that is concerned only with the resultant mathematical form. This model represents a hypothetical form of damping that is proportional to the system mass or stiffness matrix. Therefore

$$[C] = \alpha[M] + \beta[K] \tag{21.26}$$

Under this assumption, *proportional damping* is the case where the equivalent damping matrix is equal to a linear combination of the mass and stiffness matrices. For this mathematical form of damping, the coordinate transformation that diagonalizes the system mass and stiffness matrices also diagonalizes the system damping matrix. *Nonproportional damping* is the case where this linear combination does not exist. Therefore when a system with proportional damping exists, that system of coupled equations of motion can be transformed to a system of equations that represent an uncoupled system of single degree-of-freedom systems that are easily solved. With respect to modal parameters, a system with proportional damping has real-valued modal vectors (*real* or *normal modes*), while a system with nonproportional damping has complex-valued modal vectors (*complex modes*).

EXPERIMENTAL MODAL ANALYSIS METHODS

In order to understand the various experimental approaches used to determine the modal parameters of a structure, some sort of outline of the various techniques is helpful in categorizing the different methods that have been developed over the last fifty years. One of several overlapping approaches can be used. One approach is to group the methods according to whether one mode or multiple modes are excited at one time. The terminology that is used for this is

- Phase resonanance (single mode)
- Phase separation (multiple mode)

A slightly more detailed approach is to group the methods according to the type of measured data that is acquired. When this approach is utilized, the relevant terminology is

- Sinusoidal input-output model (forced normal mode)
- Frequency response function model
- Damped complex exponential response model
- General input-output model

A very common approach to comparing and contrasting experimental modal analysis methodologies that is often used in the literature is based upon the type of model that is used in the modal parameter estimation stage. The relevant nomenclature for this approach is

- Parametric model
  - Modal model
  - $[M], [K], [C]$ model
- Nonparametric model

Finally, the different experimental modal analysis approaches may be grouped according to the domain in which the modal parameter estimation model is formulated. The relevant nomenclature for this approach is

- Time domain
- Frequency domain
- Spatial domain

Regardless of the approach used to organize or classify the different approaches to generating modal parameters from experimental data, the fundamental underlying theory is the same. The differences largely are a matter of logistics, user experience requirements, or numerical or computational limitations rather than the fundamental superiority or inferiority of the method. Most methodology is based upon measured frequency response or impulse-response functions. Further discussion of experimental modal analysis is limited to techniques related to the measurement and use of these functions for determining modal parameters. The most widely utilized methods are discussed in detail in a following section on *Modal Parameter Estimation.*

MODAL DATA ACQUISITION

Acquisition of data that are used in the formulation of a modal model involves many important technical concerns. The primary concern is the digital signal processing, or the converting of analog signals into a corresponding sequence of digital values that accurately describe the time-varying characteristics of the inputs to and responses from a system. Once the data are available in digital form, the most common approach is to transform the data from the time domain to the frequency domain by

use of a discrete Fourier transform algorithm. Since this algorithm involves discrete data over a limited time period, there are large potential problems with this approach that must be well understood. (Data acquisition and analysis are discussed in detail in Chap. 27.)

DIGITAL SIGNAL PROCESSING

In order to determine modal parameters, the measured input (excitation) and response data must be processed and put into a form that is compatible with the test and modal parameter estimation methods. As a result, digital signal processing of the data is a very important step in structural testing. This is one of the technology areas where a clear understanding of the time-frequency-Laplace domain relationships is important. The conversion of the data from the time domain into the frequency and Laplace domains is important both in the measurement process and subsequently in the parameter estimation process.

Digital signal processing of the measured input and response data is used for the following reasons:

- *Condensation.* In general, the amount of measured data tremendously exceeds the information present in the desired measurements (frequency response, unit impulse response, coherence function, etc.). Therefore, digital signal processing is used to condense the data.
- *Measurements.* The measurements which are used subsequently in the modal parameter estimation process are estimated. Since there are many excitation, measurement, and modal parameter estimation procedures, there are likewise a large number of digital signal processing options which can be used.
- *Noise reduction.* Signal processing is used to reduce the influences of noise in the measurement process. The types of noise are classified as follows:
  - *Noncoherent noise.* This noise is due to electrical noise on the transducer signals or unmeasured excitation sources, etc., which are noncoherent with respect to the measured input signals or to some other signal which is used in the averaging process. Zero mean noncoherent noise can be eliminated by averaging with respect to a reference signal. This reference signal can be the input signal in terms of a spectrum averaging process, or it can be a synchronization or trigger signal in terms of cyclic averaging or random decrement process.
  - *Signal processing noise.* The signal processing itself may generate noise. For example, *leakage* is a classic source of noise when using fast Fourier transforms (FFT) for computing frequency-domain measurements. This type of noise is reduced or eliminated by using completely observed time signals (periodic or transient), by using various types of windows, or by increasing the frequency resolution.
  - *Nonlinear noise.* If the system is nonlinear, then free decay, frequency response, or unit-impulse function measurements may be distorted, which consequently causes problems when estimating modal parameters. Nonlinear distortion noise is eliminated by linearizing the test structure before testing or by randomizing the input signals to the structure. This causes the nonlinear distortion noise to become noncoherent with respect to the input signal. The nonlinear noise can then be averaged from the data in the same manner as ordinary noncoherent noise.

The process of representing an analog signal as a series of digital values is a basic requirement of digital signal processing analyzers. In practice, the goal of the analog-

to-digital conversion (ADC) process (see Chap. 27) is to obtain the conversion while maintaining sufficient accuracy in terms of frequency, magnitude, and phase. When dealing strictly with analog devices, this concern is satisfied by the performance characteristics of each individual analog device. With the advent of digital signal processing, the performance characteristics of the analog device are only the first criteria considered. The characteristics of the analog-to-digital conversion are also very important.

This process of analog-to-digital conversion involves two separate concepts, each of which is related to the dynamic performance of a digital signal processing analyzer. *Sampling* is the part of the process related to the timing between individual digital pieces of the time-history. *Quantization* is the part of the process related to describing an analog amplitude as a digital value. Primarily, sampling considerations alone affect the frequency accuracy, while both sampling and quantization considerations affect magnitude and phase accuracy. The two constraining relationships that govern the sampling process are known as Shannon's sampling theorem (Fig. 21.6) and Rayleigh's criterion (Fig. 21.7). The selection of the sampling parameters by way of these constraints is discussed in Chaps. 14 and 22.

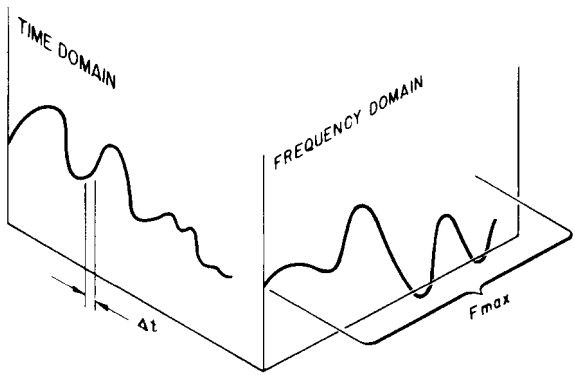

FIGURE 21.6 Shannon's sampling theorem: maximum frequency relationship.

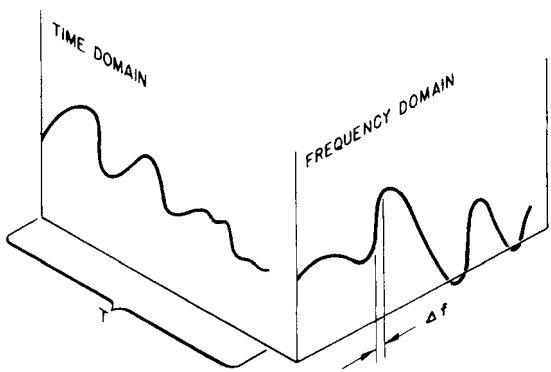

FIGURE 21.7 Rayleigh's criterion: frequency resolution relationship.

DISCRETE FOURIER TRANSFORM

The Fourier series concept explains that any physically realizable signal (signal that satisfies the Dirochlet conditions) can be uniquely separated into a summation of sine and cosine terms at appropriate frequencies. This generates a unique set of sine and cosine terms because of the orthogonal nature of sine functions at different frequencies, the orthogonal nature of cosine functions at different frequencies, and the orthogonal nature of sine functions compared to cosine functions. If the choice of frequencies is limited to a discrete set of frequencies, the discrete Fourier transform describes the amount of each sine and cosine term at each discrete frequency. The real part of the discrete Fourier transform describes the amount of each cosine term; the imaginary part of the discrete Fourier transform describes the amount of each sine term. Figure 21.8 is a graphical representation of this concept for a signal that can be represented by a summation of sinusoids.

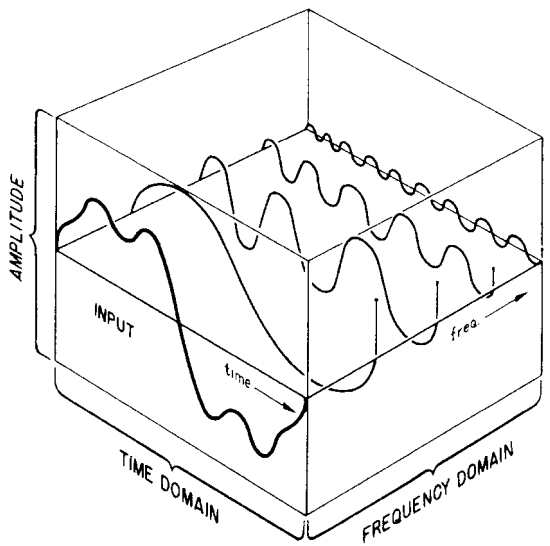

FIGURE 21.8 Discrete Fourier transform concept.

The discrete Fourier transform algorithm is the basis for the formulation of any frequency-domain function in digital data acquisition systems. In terms of an integral Fourier transform, the function must exist for all time in a continuous sense in order to be evaluated. For the realistic measurement situation, data are available in a discrete sense over a limited time period. The discrete Fourier transform, therefore, is based upon a set of assumptions concerning this discrete sequence of values. The assumptions can be reduced to two, of which one must be met by every signal processed by the discrete Fourier transform algorithm. The first assumption is that the signal must be a *totally observed transient* with respect to the time period of observation. If this is not true, then the signal must be composed only of *harmonics of the time period of observation.* If one of these two assumptions is met by any discrete history processed by the discrete Fourier transform algorithm, then the resulting spectrum does not contain bias errors. Much of the data processing effort, with respect to acquisition of data used for the formulation of a modal model, is con-

cerned with the assurance that the input and response histories match one of these two assumptions. A more complete discussion of the discrete Fourier transform algorithm is included in Chap. 14.

ERRORS

The accurate measurement of frequency response functions depends heavily upon the errors involved with the digital signal processing. In order to take full advantage of experimental data in the evaluation of experimental procedures and verification of theoretical approaches, the errors in measurement, generally designated noise, must be reduced to acceptable levels. With the increasing use of personal computer (PC) instrumentation, the user must take great care to be certain that errors are minimized. With respect to the frequency response function measurement, the errors in the estimate are generally grouped into two categories: variance and bias. The *variance* portion of the error is due to random deviations of each sample function from the mean. Statistically, then, if sufficient sample functions are evaluated, the estimate closely approximates the true function with a high degree of confidence. The *bias* portion of the error, on the other hand, is not necessarily reduced by using many samples. The bias error is due to a system characteristic or measurement procedure that consistently results in an incorrect estimate. Therefore, the expected value is not equal to the true value. Examples of this are system nonlinearities or digitization errors such as aliasing or leakage. With this type of error, knowledge of the form of the error is vital in reducing the resultant effect in the frequency response function measurement. See Chap. 22 for details.

TRANSDUCER CONSIDERATIONS

The transducer considerations are often the most overlooked aspect of the experimental modal analysis process. Considerations involving the actual type and specifications of the transducers, mounting of the transducers, and calibration of the transducers are often among the largest sources of error. Chapter 12 discusses transducers and transducer design in significant detail. Calibration of transducers is reviewed in Chap. 18. Chapter 15 discusses measurement techniques, including transducer mounting and alignment. These topics are critical to estimating the accurate frequency response function measurements required for most experimental modal analysis methods.

TEST ENVIRONMENT CONSIDERATIONS

The test environment for any modal analysis test involves several environmental factors as well as appropriate boundary conditions. Primarily, temperature, humidity, vacuum, and gravity effects must be properly considered to match previous analysis models or to allow the experimentally determined model to properly reflect the system. Very few experimental laboratory facilities have the capability to control these factors in other than a rudimentary fashion.

In addition to the environmental concerns, the boundary conditions of the system under test are very important. Traditionally, modal analysis tests have been performed under the assumption that the test boundary conditions can be made to conform to one of four conditions:

- Free-free boundary conditions (impedance is zero).
- Fixed boundary conditions (impedance is infinite).
- Operating boundary conditions (impedance is correct).
- Arbitrary boundary conditions (impedance is known).

Except in very special situations, none of these boundary conditions can be practically achieved. Instead, practical guidelines are normally used to evaluate the appropriateness of the chosen boundary conditions. For example, if a free-free boundary is chosen, the desired frequency of the highest rigid body mode must be at least a factor of 10 below the first deformation mode of the system under test. Likewise, for the fixed-boundary test, the desired interface stiffness must be at least a factor of 10 greater than the local stiffness of the system under test. While either of these practical guidelines can be achieved for small test objects, a large class of systems cannot be acceptably tested in either configuration. Arguments have been made that the impedance of a support system can be defined (via test and/or analysis) and the effects of such a support system eliminated from the measured data. This technique is theoretically sound, but, because of significant dynamics in the support system and limited measurement dynamics, the approach has not been uniformly applicable.

In response to this problem, many alternative structural testing concepts have been proposed. Active suspension systems (see Chap. 32) and combinations of active and passive systems are being evaluated, particularly for application to very flexible space structures. Active inert-gas suspension systems have been used in the past for the testing of smaller commercial and military aircraft, and, in general, such approaches are formulated to better match the requirements of a free-free boundary condition.

Another alternative test procedure is to define a series of relatively conventional tests with various boundary conditions. These various boundary conditions are chosen in such a way that each perturbed boundary condition can be accurately modeled (for example, the addition of a large mass at interface boundaries). Therefore, as the experimental model is acquired for each configuration and used to validate and correct the associated analytical model, the underlying model is validated and corrected accordingly. This procedure has the added benefit of adding the influence of modes of vibration that occur above the maximum frequency of the test into the validation of the model.

MEASUREMENT FORMULATION

For current approaches to experimental modal analysis, the frequency response function is the most important, and most common, measurement to be made. When estimating frequency response functions, a measurement model is needed that allows the frequency response function to be estimated from measured input and output data in the presence of noise (errors). These errors have been discussed in this and other chapters in great detail.

There are at least four different testing configurations that can be considered. These different testing conditions are largely a function of the number of acquisition channels or excitation sources that are available to the test engineer.

- Single input/single output (SISO)
- Single input/multiple output (SIMO)

- Multiple input/single output (MISO)
- Multiple input/multiple output (MIMO)

In general, the best testing situation is the multiple input/multiple output configuration (MIMO), since the data are collected in the shortest possible time with the fewest changes in the test conditions.

FREQUENCY RESPONSE FUNCTION ESTIMATION

The estimation of the frequency response function depends upon the transformation of data from the time to the frequency domain. The Fourier transform is used for this computation. The computation is performed digitally using a fast Fourier transform algorithm. The frequency response functions satisfy the following single and multiple input relationships:

Single input relationship:

$$X_p = H_{pq} F_q \tag{21.27}$$

Multiple input relationship:

$$\begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix}_{N_o \times 1} = \begin{bmatrix} H_{11} & \dots & H_{1q} \\ H_{21} & \dots & H_{2q} \\ \dots & \dots & \dots \\ H_{p1} & \dots & H_{pq} \end{bmatrix}_{N_o \times N_i} \begin{bmatrix} F_1 \\ F_2 \\ \dots \\ F_q \end{bmatrix}_{N_i \times 1} \tag{21.28}$$

The most reasonable, and most common, approach to the estimation of frequency response functions is the use of *least squares* (LS) or *total least squares* (TLS) techniques.[8,9] These are standard techniques for estimating parameters in the presence of noise. Least squares methods minimize the square of the magnitude error and thus compute the *best* estimate of the magnitude of the frequency response function, but they have little effect on the phase of the frequency response function. The primary difference in the algorithms used to estimate frequency response functions is in the assumption of where the noise enters the measurement problem. Three algorithms, referred to as the H_1, H_2, and H_v algorithms, are commonly available for estimating frequency response functions. Table 21.1 summarizes the assumed location of the noise for these three algorithms.

TABLE 21.1 Summary of Frequency Response Function Estimation Models

		Assumed location of noise	
Technique	Solution method	Force inputs	Response
H_1	LS	No noise	Noise
H_2	LS	Noise	No noise
H_v	TLS	Noise	Noise

Consider the case of N_i inputs and N_o outputs measured during a modal test. Based upon the assumed location of the noise entering the estimation process, Eqs. (21.29) through (21.31) represent the corresponding model for the H_1, H_2, and H_v estimation procedures.

H_1 technique:

$$[H]_{N_o \times N_i} \{F\}_{N_i \times 1} = \{X\}_{N_o \times 1} - \{\eta\}_{N_o \times 1} \tag{21.29}$$

H_2 technique:

$$[H]_{N_o \times N_i} \{ \{F\}_{N_i \times 1} - \{\upsilon\}_{N_i \times 1}\} = \{X\}_{N_o \times 1} \tag{21.30}$$

H_v technique:

$$[H]_{N_o \times N_i} \{ \{F\}_{N_i \times 1} - \{\upsilon\}_{N_i \times 1}\} = \{X\}_{N_o \times 1} - \{\eta\}_{N_o \times 1} \tag{21.31}$$

This numerical model can be represented in block diagram form as shown in Fig. 21.9.

Single Input FRF Estimation. With reference to Fig. 21.9 for a case involving only one input and one output (input location q and response location p), the equation that is used to represent the input-output relationship is

$$\hat{X}_p - \eta_p = H_{pq}(\hat{F}_q - \upsilon_q) \tag{21.32}$$

where $F = \hat{F} - \upsilon$ = actual input
$X = \hat{X} - \eta$ = actual output
$\hat{X}$ = spectrum of the pth output, measured
$\hat{F}$ = spectrum of the qth input, measured
H = frequency response function
υ = spectrum of the noise part of the input
η = spectrum of the noise part of the output
X = spectrum of the pth output, theoretical
F = spectrum of the qth input, theoretical

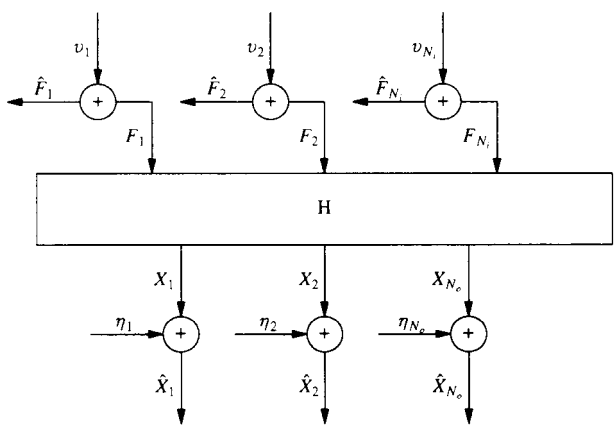

FIGURE 21.9 General system model: multiple inputs.

If $\upsilon = \eta = 0$, the theoretical (expected) frequency response function of the system is estimated. If $\eta \neq 0$ and/or $\upsilon \neq 0$, a least squares method is used to estimate a *best* frequency response function in the presence of noise.

In order to develop an estimation of the frequency response function, a number of averages N_{avg} is used to minimize the random errors (variance). This can be easily accomplished through use of intermediate measurement of the power (auto-) and cross-spectra. The estimate of the auto- and cross-spectra for the model in Fig. 21.9 is defined as follows. Note that each function is a function of frequency.

***Cross-spectra* $\mathbf{WXF_{pq}}$ *and* $\mathbf{WXF_{qp}}$:**

$$WXF_{pq} = \sum_{1}^{N_{avg}} X_p F_q^* \tag{21.33}$$

$$WFX_{qp} = \sum_{1}^{N_{avg}} F_q X_p^* \tag{21.34}$$

***Autospectra* $\mathbf{WFF_{qq}}$ *and* $\mathbf{WXX_{pp}}$:**

$$WFF_{qq} = \sum_{1}^{N_{avg}} F_q F_q^* \tag{21.35}$$

$$WXX_{pp} = \sum_{1}^{N_{avg}} X_p X_p^* \tag{21.36}$$

where F^* = complex conjugate of $F(\omega)$
X^* = complex conjugate of $X(\omega)$

H_1 Algorithm: Minimize Noise on Output (η). The most common formulation of the frequency response function, often referred to as the H_1 algorithm, tends to minimize the noise on the output. This formulation is shown in Eq. (21.37).

$$H_{pq} = \frac{WXF_{pq}}{WFF_{qq}} \tag{21.37}$$

H_2 Algorithm: Minimize Noise on Input (υ). Another formulation of the frequency response function, often referred to as the H_2 algorithm, tends to minimize the noise on the input. This formulation is shown in Eq. (21.38).

$$H_{pq} = \frac{WXX_{pp}}{WFX_{qp}} \tag{21.38}$$

In the H_2 formulation, an autospectrum is divided by a cross-spectrum. This can be a problem since the cross-spectrum can theoretically be zero at one or more frequencies. In both formulations, the phase information is preserved in the cross-spectrum term.

H_v Algorithm: Minimize Noise on Input and Output (η and υ). The solution for H_{pq} using the H_v algorithm is found by the eigenvalue decomposition of a matrix of power spectra. For the single input case, the following matrix involving the auto- and cross-spectra can be defined:

$$[WFFX_p] = \begin{bmatrix} WFF_{qq} & WXF_{pq} \\ WXF_{pq}{}^H & WXX_{pp} \end{bmatrix}_{2 \times 2} \tag{21.39}$$

The solution for H_{pq} is found by the eigenvalue decomposition of the $[WFFX]$ matrix as follows:

$$[WFFX_p] = [V] \lceil \Lambda \rfloor [V]^H \tag{21.40}$$

where $\lceil \Lambda \rfloor$ = diagonal matrix of eigenvalues. The solution for the H_{pq} matrix is found from the eigenvector associated with the smallest (minimum) eigenvalue λ_1. The size of the eigenvalue problem is second-order, resulting in finding the roots of a quadratic equation. This eigenvalue solution must be repeated for each frequency, and the complete solution process must be repeated for each response point X_p.

Alternatively, the solution for H_{pq} is found by the eigenvalue decomposition of the following matrix of auto- and cross-spectra:

$$[WXFF_p] = \begin{bmatrix} WXX_{pp} & WXF_{pq}{}^H \\ WXF_{pq} & WFF_{qq} \end{bmatrix}_{2 \times 2} \tag{21.41}$$

$$[WXFF_p] = [V] \lceil \Lambda \rfloor [V]^H \tag{21.42}$$

where $\lceil \Lambda \rfloor$ = diagonal matrix of eigenvalues. The solution for H_{pq} is again found from the eigenvector associated with the smallest (minimum) eigenvalue λ_1. The frequency response function is found from the normalized eigenvector associated with the smallest eigenvalue. If $[WFFX_p]$ is used, the eigenvector associated with the smallest eigenvalue must be normalized as follows:

$$\{V\}_{\lambda \min} = \begin{Bmatrix} H_{pq} \\ -1 \end{Bmatrix} \tag{21.43}$$

If $[WXFF_p]$ is used, the eigenvector associated with the smallest eigenvalue must be normalized as follows:

$$\{V\}_{\lambda \min} = \begin{Bmatrix} -1 \\ H_{pq} \end{Bmatrix} \tag{21.44}$$

One important consideration in choosing one of the three formulations for frequency response function estimation is the behavior of each formulation in the presence of a bias error such as leakage. In all cases, the estimate differs from the expected value, particularly in the region of a resonance (magnitude maximum) or antiresonance (magnitude minimum). For example, H_1 tends to underestimate the value at resonance, while H_2 tends to overestimate the value at resonance. The H_v algorithm gives an answer that is always bounded by the H_1 and H_2 values. The different approaches are based upon minimizing the magnitude of the error but have no effect on the phase characteristics.

In addition to the attractiveness of H_1, H_2, and H_v in terms of the minimization of the error, the availability of auto- and cross-spectra allows the determination of other

important functions. The quantity γ_{pq}^2 is called the scalar or *ordinary coherence function* and is a frequency-dependent, real value between 0 and 1. The ordinary coherence function indicates the degree of causality in a frequency response function. If the coherence is equal to 1 at any specific frequency, the system is said to have perfect causality at that frequency. In other words, the measured response power is caused totally by the measured input power (or by sources which are coherent with the measured input power). A coherence value less than unity at any frequency indicates that the measured response power is greater than that caused by the measured input. This is due to some extraneous noise also contributing to the output power. It should be emphasized, however, that low coherence does not necessarily imply poor estimates of the frequency response function; it simply means that more averaging is needed for a reliable result. The ordinary coherence function is computed as follows:

$$COH_{pq} = \gamma_{pq}^2 = \frac{|\, WXF_{pq}\,|^2}{WFF_{qq}WXX_{pp}} = \frac{WXF_{pq}WFX_{qp}}{WFF_{qq}WXX_{pp}} \tag{21.45}$$

When the coherence is zero, the output is caused totally by sources other than the measured input. In general, then, the coherence can be a measure of the degree of noise contamination in a measurement. Thus, with more averaging, the estimate of coherence may contain less variance, therefore giving a better estimate of the noise energy in a measured signal. This is not the case, though, if the low coherence is due to bias errors such as nonlinearities, multiple inputs, or leakage. A typical ordinary coherence function is shown in Fig. 21.10 together with the corresponding frequency response function magnitude. In Fig. 21.10, the frequencies where the coherence is lowest are often the same frequencies where the frequency response function is at a maximum or a minimum in magnitude. This is often an indication of leakage since the frequency response function is most sensitive to leakage error at the lightly damped peaks corresponding to the maxima. At the minima, where there is little response from the system, the leakage error, even though it is small, may still be significant.

In all of these cases, the estimated coherence function approaches, in the limit, the expected value of coherence at each frequency, dependent upon the type of noise present in the structure and measurement system. Note that, with more averaging, the estimated value of coherence does not increase; the estimated value of coherence always approaches the expected value from the upper side.

Multiple Input FRF Estimation. Multiple input estimation of frequency response functions is desirable for several reasons. The principal advantage is the increase in the accuracy of estimates of the frequency response functions. During single input excitation of a system, large differences in the amplitudes of vibratory motion at various locations may exist due to the dissipation of the excitation power within the structure. This is especially true when the structure has heavy damping. Small nonlinearities in the structure consequently cause errors in the measurement of the response. With multiple input excitation, the vibratory amplitudes across the structure typically are more uniform, with a consequent decrease in the effect of nonlinearities.

A second reason for improved accuracy is the increase in consistency of the frequency response functions compared to the single input method. When a number of exciter systems are used, the elements from columns of the frequency response function matrix corresponding to those exciter locations are being determined simultaneously. With the single input method, each column is determined independently, and it is possible for small errors of measurement due to nonlinearities and

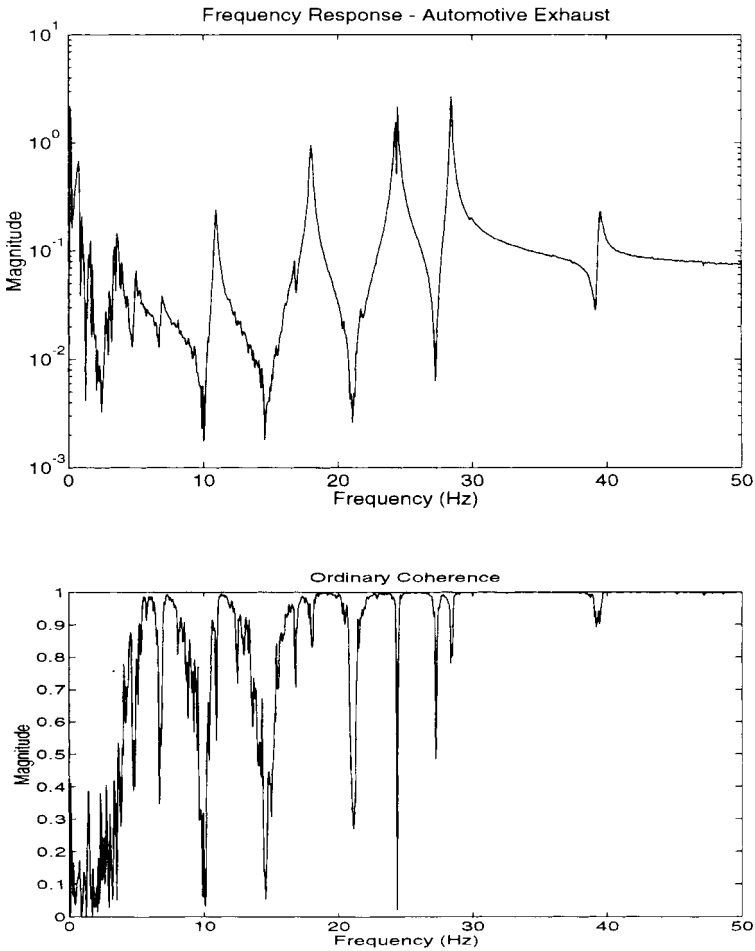

FIGURE 21.10 frequency response function magnitude with associated ordinary coherence function.

time-dependent system characteristics to cause a change in resonance frequencies, damping, or mode shapes among the measurements in the several columns. This is particularly important for the polyreference modal parameter estimation algorithms that use frequency response functions from multiple columns or rows of the frequency response function matrix simultaneously.

An additional, significant advantage of multiple input excitation is a reduction in the test time. In general, when multiple input estimation of frequency response functions is used, frequency response functions are obtained for all input locations in approximately the same time as required for acquiring a set of frequency response functions for one of the input locations using a single input estimation method.[4,11,12]

With reference to Fig. 21.9 for a case involving N_i inputs and N_o outputs, the equation that is used to represent the input-output relationship is

$$\hat{X}_p - \eta_p = \sum_{q=1}^{N_i} H_{pq}(\hat{F}_q - \upsilon_q) \tag{21.46}$$

In order to develop an estimation of the frequency response function for the multiple input case, a number of averages N_{avg} is used to minimize the random errors (variance). This can be easily accomplished through use of intermediate measurement of the auto- and cross-spectra as defined in Eqs. (21.33) through (21.36). Additional matrices, compared to the single input case, need to be defined. These additional matrices are constructed from the auto- and cross-spectra previously defined for the single input case. Each function and, therefore, each resulting matrix is a function of frequency.

Input-output cross-spectra matrix:

$$[WXF] = \{X\}\{F\}^H = \begin{Bmatrix} X_1 \\ X_2 \\ \cdots \\ X_{N_o} \end{Bmatrix} [F_1^* \; F_2^* \ldots F_{N_i}^*] = \begin{bmatrix} WXF_{11} \ldots WXF_{1N_i} \\ \cdots \quad \cdots \quad \cdots \\ WXF_{N_o1} \ldots WXF_{N_oN_i} \end{bmatrix} \tag{21.47}$$

Input cross-spectra matrix:

$$[WFF] = \{F\}\{F\}^H = \begin{Bmatrix} F_1 \\ F_2 \\ \cdots \\ F_{N_i} \end{Bmatrix} [F_1^* \; F_2^* \ldots F_{N_i}^*] = \begin{bmatrix} WFF_{11} \ldots WFF_{1N_i} \\ \cdots \quad \cdots \quad \cdots \\ WFF_{N_i1} \ldots WFF_{N_iN_i} \end{bmatrix} \tag{21.48}$$

The frequency response functions can now be estimated for the three algorithms.

H_1 Algorithm: Minimize Noise on Output (η)

$$[H] = [WXF][WFF]^{-1} \tag{21.49}$$

In the experimental procedure, the input and response signals are measured, and the averaged cross-spectra and autospectra necessary to create the $[WXF]$, $[WFF]$, and $[WXX]$ matrices are computed. The input cross-spectrum matrix must be inverted, at least implicitly, at every frequency in the analysis range. This means that the computational load on the measurement system is greater than for the single input case, in which only the reciprocal of a single input autospectrum is computed.

Equation (21.49) is valid unless the input cross-spectrum matrix $[WFF]$ is singular for specific frequencies or frequency intervals. When this happens, the inverse of $[WFF]$ does not exist and Eq. (21.49) cannot be used to solve for the frequency

response function at those frequencies or in those frequency intervals. A computational procedure that solves Eq. (21.49) for $[H]$ must monitor the rank of the matrix $[WFF]$ that is to be inverted, and provide information on how to alter the input signals or use the available data when a problem exists. The current approach for evaluating whether the inputs are sufficiently uncorrelated at each frequency involves determining the principal/virtual forces using principal component analysis.[10]

H_2 Algorithm: Minimize Noise on Input (υ)

$$[H] = [WXX][WFX]^{-1} \tag{21.50}$$

One problem with using the H_2 algorithm is that the solution for $[H]$ can be found directly using an inverse only when the number of inputs N_i and number of outputs N_o are equal.

H_v Algorithm: Minimize Noise on Input and Output (υ and η). The solution for $[H]$ is found by the eigenvalue decomposition of one of the following two matrices:

$$[WFFX_p] = \begin{bmatrix} [WFF] & [WXF_p] \\ [WXF_p]^H & WXX_p \end{bmatrix}_{(N_i+1)\times(N_i+1)} \tag{21.51}$$

$$[WXFF_p] = \begin{bmatrix} WXX_p & [WXF_p]^H \\ [WXF_p] & [WFF] \end{bmatrix}_{(N_i+1)\times(N_i+1)} \tag{21.52}$$

Therefore, the eigenvalue decomposition is

$$[WFFX_p] = [V]\lceil \Lambda \rfloor [V]^H \tag{21.53}$$

or

$$[WXFF_p] = [V]\lceil \Lambda \rfloor [V]^H \tag{21.54}$$

where $\lceil \Lambda \rfloor$ = diagonal matrix of eigenvalues. The solution for the pth row of the $[H]$ matrix is found from the eigenvector associated with the smallest (minimum) eigenvalue. Note that the size of the eigenvalue problem is $N_i + 1$ and that the eigenvalue solution must be repeated for each frequency. The complete solution process must be repeated for each response point X_p.

The frequency response function associated with a single output p and all inputs is found by normalizing the eigenvector associated with the smallest eigenvalue. If $[WFFX_p]$ is used, the eigenvector associated with the smallest eigenvalue must be normalized as follows:

$$\{V\}_{\lambda_{\min}} = \begin{Bmatrix} H_{p1} \\ H_{p2} \\ \cdots \\ H_{pN_i} \\ -1 \end{Bmatrix} \tag{21.55}$$

If $[WXFF_p]$ is used, the eigenvector associated with the smallest eigenvalue must be normalized as follows:

$$\{V\}_{\lambda_{\min}} = \begin{Bmatrix} -1 \\ H_{p1} \\ H_{p2} \\ \cdots \\ H_{pN_i} \end{Bmatrix} \tag{21.56}$$

The concept of the coherence function, as defined for single input measurement, needs to be expanded to include the variety of relationships that are possible for multiple inputs. *Ordinary coherence* is defined in this general sense as the correlation coefficient describing the linear relationship between any two single spectra. Great care must be taken in the interpretation of ordinary coherence when more than one input is present. The ordinary coherence of an output with respect to an input can be much less than unity even though the linear relationship between inputs and outputs is valid, because of the influence of the other inputs.[4–6]

The ordinary coherence function can be formulated in terms of the elements of the matrices defined previously. The ordinary coherence function between the pth output and the qth input can be computed from the following formula:

Ordinary coherence function:

$$COH_{pq} = \lambda_{pq}^{\ 2} = \frac{|WXF_{pq}|^2}{WFF_{qq}WXX_{pp}} \tag{21.57}$$

where WXX_{pp} = autospectrum of the output p
WFF_{qq} = autospectrum of the input q
WXF_{pq} = cross-spectrum between output p and input q

Partial coherence is defined as the ordinary coherence between a conditioned output and another conditioned output, between a conditioned input and another conditioned input, or between a conditioned input and a conditioned output. The output and input are conditioned by removing contributions from other input(s). The removal of the effects of the other input(s) is formulated on a linear least squares basis. The order of removal of the inputs during "conditioning" has a definite effect upon the partial coherence if some of the input(s) are mutually correlated. There is a partial coherence function for every input-output, input-input, and input-output combination for all permutations of conditioning. The usefulness of partial coherence for experimental modal analysis is limited.

Multiple coherence is defined as the correlation coefficient describing the linear relationship between an output and all known inputs. There is a multiple coherence function for every output. Multiple coherence can be used to evaluate the importance of unknown contributions to each output. These unknown contributions can be measurement noise, nonlinearities, or unknown inputs. In particular, as in the evaluation of ordinary coherence, a low value of multiple coherence near a resonance often means that leakage error is present in the frequency response function.

The formulation of the equations for the multiple coherence functions can be simplified to the following equation:

Multiple coherence function:

$$MCOH_p = \sum_{q=1}^{N_i} \sum_{t=1}^{N_i} \frac{H_{pq} WFF_{qt} H_{pt}^*}{WXX_{pp}} \tag{21.58}$$

where H_{pq} = frequency response function for output p and input q
H_{pt} = frequency response function for output p and input t
WFF_{qt} = cross-spectrum between output q and output t

If the multiple coherence of the pth output is near unity, then the pth output is well predicted from the set of inputs using the least squares frequency response functions.

Multiple Input Force Analysis/Evaluation. Of the variety of situations that can cause difficulties in the computation of the frequency response functions, the one with the highest potential for trouble is the case of coherent inputs. If two of the inputs are fully coherent, then there are no unique frequency response functions associated with those inputs. Unfortunately, there are a number of situations where the input cross-spectrum matrix $[WFF]$ may be singular at specific frequencies or frequency intervals. When this happens, the inverse of $[WFF]$ does not exist, and Eq. (21.49) cannot be used to solve for the frequency response function at those frequencies or in those frequency intervals. First, one of the input autospectra may be zero in amplitude over some frequency interval. Second, two or more of the input signals may be fully coherent over some frequency interval. Third, numerical problems which cause the computation of the inverse to be inexact may be present.

The current approach used to detect correlated inputs involves utilizing principal component analysis to determine the number of forces contributing to the $[WFF]$ matrix. In this approach, a principal component analysis must be conducted on the $[WFF]$ matrix.[10] *Principal component analysis* involves an eigenvalue decomposition of the $[WFF]$ matrix. Since the eigenvectors of such a decomposition are unitary, the eigenvalues should all be of approximately the same size if each of the inputs is contributing. If one of the eigenvalues is much smaller at a particular frequency, one of the inputs is not present or one of the inputs is correlated with the other input(s).

$$[WFF] = [V]\,[\Lambda]\,[V]^H \tag{21.59}$$

where $[\Lambda]$ represents the eigenvalues of the $[WFF]$ matrix. If any of the eigenvalues of the $[WFF]$ matrix are zero or insignificant, then the $[WFF]$ matrix is singular. Therefore, for a three-input test, the $[WFF]$ matrix should have three eigenvalues of approximately the same magnitude. (The number of distinct eigenvalues is equal to the number of uncorrelated inputs.) Figure 21.11 shows the principal force plots for a case with three inputs. At the frequencies where the third principal/virtual force drops (lowest curve), the inputs are mutually correlated.

PRACTICAL MEASUREMENT CONSIDERATIONS

There are several factors that contribute to the quality of actual measured frequency response function estimates. Some of the most common sources of error involve measurement mistakes. With a proper measurement approach, most errors of this type, such as overloading the input, extraneous signal pick-up via ground loops or

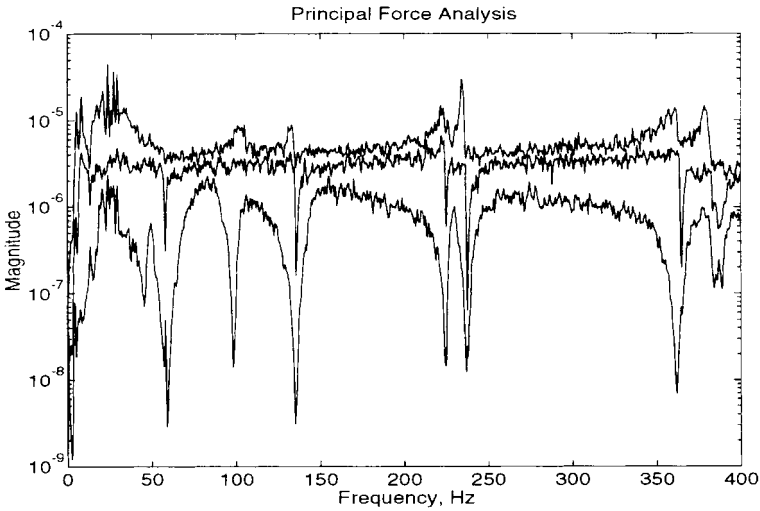

FIGURE 21.11 Principal (virtual) force spectra for three inputs.

strong electric or magnetic fields nearby, etc., can be avoided. Violations of test assumptions are often the source of another inaccuracy and can be viewed as a measurement mistake. For example, frequency response and coherence functions have been defined as parameters of a linear system. Nonlinearities generally shift energy from one frequency to many new frequencies, in a way which may be difficult to recognize. The result is a distortion in the estimates of the system parameters, which may not be apparent unless the excitation is changed. One way to reduce the effect of nonlinearities is to randomize these contributions by choosing a randomly different input signal for each of the contributing averages. Subsequent averaging reduces these contributions in the same way that random noise is reduced. Another example involves control of the system input. One requirement is to excite the system with energy at all frequencies for which measurements are expected. It is important to be sure that the input signal spectrum does not have frequency ranges where little energy exists. Otherwise, coherence is very low, and the variance on the frequency response function is unacceptable.

Assuming that the system is linear, the excitation is proper, and measurement mistakes are avoided, some amount of error and/or noise is still present in the measurement process. Five different approaches can be used to reduce this error involved in frequency response function measurements. First of all, the use of *different frequency response function estimation algorithms* (H_v compared to H_1) reduces the effect of the leakage error on the estimation of the frequency response function computation. The use of *averaging* significantly reduces errors of both variance and bias and is probably the most general technique used to reduce errors in frequency response function measurement. *Selective excitation* is often used to verify nonlinearities or randomize characteristics. In this way, bias errors due to system sources can be reduced or controlled. The *increase of frequency resolution* through the zoom fast Fourier transform improves the frequency response function estimate primarily by reducing the leakage bias error through the use of a longer time sample. The zoom fast Fourier transform by itself is a linear process and does not involve any

specific error reduction characteristics compared to a baseband fast Fourier transform (FFT). Finally, the *use of weighting functions* (*windows*) is widespread, and much has been written about their value.[4–6] Primarily, weighting functions compensate for the bias error (leakage) caused by the analysis procedure.

Signal Averaging. The averaging of signals is normally viewed as a summation or weighted summation process where each sample function has a common abscissa.[6] Normally, the designation of *history* is given to sample functions with the abscissa of absolute time, and the designation of *spectrum* is given to sample functions with the abscissa of absolute frequency. The spectra are normally generated by Fourier transforming the corresponding history. In order to generalize and consolidate the concept of signal averaging as much as possible, the case of relative time is also considered. In this way, *relative history* is discussed with units of the appropriate event rather than seconds, and a *relative spectrum* is the corresponding Fourier transform with units of cycles per event. This concept of signal averaging is used widely in structural signature analysis where the event is a revolution of a rotating shaft. This kind of approach simplifies the application of many other concepts of signal relationships, such as Shannon's sampling theorem and Rayleigh's criterion of frequency resolution.

The process of signal averaging as it applies to frequency response functions is simplified greatly by the intrinsic uniqueness of the frequency response function. Since the frequency response function is expressed in terms of system properties of mass, stiffness, and damping, it is reasonable to conclude that in most realistic structures, the frequency response functions are considered to be constants, just like mass, stiffness, and damping. This concept means that when formulating the frequency response function using H_1, H_2, or H_v algorithms, the estimate of frequency response is intrinsically unique, as long as the system is linear and the noise can be eliminated. In general, the auto- and cross-power spectra are statistically unique only if the input is stationary and sufficient averages are taken. Nevertheless, the estimate of frequency response is valid whether the input is stationary, nonstationary, or deterministic.

The concept of the intrinsic uniqueness of the frequency response function also permits a greater freedom in the testing procedure. Each function is derived as the result of a separate test or as the result of different portions of the same continuous test situation. In either case, the estimate of the frequency response function is the same as long as the time-history data for the auto- and cross-power spectra that are utilized in any computation of the frequency response or coherence function are acquired simultaneously.

The approaches to signal averaging vary only in the relationship between the sample functions used. Since the Fourier transform is a linear function, there is no theoretical difference between the use of histories or spectra. (Practically, though, there are precision considerations.) With this in mind, the signal averaging useful to frequency response function measurements can be divided into three classifications:

- Asynchronous
- Synchronous
- Cyclic

These three classifications refer to the trigger and sampling relationships between sample functions. *Asynchronous averaging* describes the averaging case when each average is acquired without a triggering event; it is sometimes referred to as free-run

averaging. *Synchronous averaging* describes the averaging case when each average is acquired only when an external triggering event occurs. *Cyclic averaging* describes the averaging case when each average is acquired with a specific *absolute* time, or phase, relationship to all previous averages. (Averaging is discussed in detail in Chaps. 13 and 22.)

Excitation. Excitation includes any form of input that is used to create a response in a mechanical system. This can include environmental or operational inputs as well as the controlled force input(s) that are used in a vibration or modal analysis test. In general, the following discussion is limited to force inputs that are measured and/or controlled in some rigorous way.[3,13,14]

Excitation Assumptions. The primary assumption concerning the excitation of a linear structure is that the excitation is observable. Whenever the excitation is measured, this assumption simply implies that the measured characteristic properly describes the actual input characteristics. For the case of multiple inputs, the different inputs must often be uncorrelated for the computational procedures to yield a solution. In most cases this means only that the multiple inputs must not be perfectly correlated at any frequency. As long as the excitation is measured, the validity of these limited assumptions can be evaluated.

There are a number of techniques that can be used to estimate modal characteristics from response measurements with no measurement of the excitation. If this approach is used, the excitation assumptions are much more imposing. If the excitation is not measured, estimates of modal scaling (modal mass, modal A, residues, etc.) cannot be generated. Even when these parameters are not required, all of these techniques have one further restriction: an assumption has to be made concerning the characteristics of the excitation of the system. Usually, the autospectrum of the excitation signal is assumed to be constant over the frequency interval of interest. This is not generally practical.

Classification of Excitation. Inputs which can be used to excite a system in order to determine frequency response functions belong to one of two classifications. The first classification is that of a random signal. Signals of this form can be defined by their statistical properties only over some time period. Any subset of the total time period is unique, and no explicit mathematical relationship can be formulated to describe the signal. Random signals can be further classified as stationary or nonstationary. Stationary random signals are a special case where the statistical properties of the random signals do not vary with respect to translations with time. Finally, stationary random signals can be classified as ergodic or nonergodic. A stationary random signal is ergodic when a time average on any particular subset of the signal is the same for any arbitrary subset of the random signal. All random signals which are commonly used as input signals fall into the category of ergodic, stationary random signals.

The second classification of inputs which can be used to excite a system in order to determine frequency response functions is that of a deterministic signal. Signals of this form can be represented in an explicit mathematical relationship. Deterministic signals are further divided into periodic and nonperiodic classifications. The most common inputs in the periodic deterministic signal designation are sinusoidal in nature, while the most common inputs in the nonperiodic deterministic designation are transient in form.

The choice of input to be used to excite a system in order to determine frequency response functions depends upon the characteristics of the system, the characteristics of the parameter estimation, and the expected utilization of the data. The characterization of the system is primarily concerned with the linearity of the system. As long

as the system is linear, all input forms should give the same expected value. Naturally, though, all real systems have some degree of nonlinearity. Deterministic input signals result in frequency response functions that are dependent upon the signal level and type. A set of frequency response functions for different signal levels can be used to document the nonlinear characteristics of the system. Random input signals, in the presence of nonlinearities, result in a frequency response function that represents the best linear representation of the nonlinear characteristics for a given level of random signal input. For small nonlinearities, use of a random input does not differ greatly from the use of a deterministic input.

The characterization of the parameter estimation is primarily concerned with the type of mathematical model being used to represent the frequency response function. Generally, the model is a linear summation based upon the modal parameters of the system. Unless the mathematical representation of all nonlinearities is known, the parameter estimation process cannot properly weight the frequency response function data to include nonlinear effects. For this reason, random input signals are regularly used to obtain the best linear estimate of the frequency response function when a parameter estimation process using a linear model is to be utilized.

The expected utilization of the data is concerned with the degree of detailed information required by any postprocessing task. For experimental modal analysis, this can range from implicit modal vectors needed for troubleshooting to explicit modal vectors used in an orthogonality check. As more detail is required, input signals, both random and deterministic, need to match the system characteristics and parameter estimation characteristics more closely. In all possible uses of frequency response function data, the conflicting requirements of the need for accuracy, equipment availability, testing time, and testing cost normally reduce the possible choices of input signal.

With respect to the reduction of the variance and bias errors of the frequency response function, random or deterministic signals can be utilized most effectively if the signals are periodic with respect to the sample period or totally observable with respect to the sample period. If either of these criteria is satisfied, regardless of signal type, the predominant bias error, leakage, is eliminated. If these criteria are not satisfied, the leakage error may become significant. In either case, the variance error is a function of the signal-to-noise ratio and the amount of averaging.

Many signals are appropriate for use in experimental modal analysis. Some of the most commonly used signals are described in the following sections. For those excitation signals that require the use of a shaker, Fig. 21.12 shows a typical test configuration; Fig. 21.13 shows a typical test configuration when an impact form of excitation is to be used. The advantages and disadvantages of each excitation signal are summarized in Table 21.2.

Slow swept sine. The slow swept sine signal is a periodic deterministic signal with a frequency that is an integer multiple of the FFT frequency increment. Sufficient time is allowed in the measurement procedure for any transient response to the changes in frequency to decay, so that the resultant input and response histories are periodic with respect to the sample period. Therefore, the total time needed to compute an entire frequency response function is a function of the number of frequency increments required and the system damping.

Periodic chirp. The periodic chirp is a fast swept sine signal that is a periodic deterministic signal and is formulated by sweeping a sine signal up or down within a frequency band of interest during a single sample period. Normally, the fast swept sine signal is made up of only integer multiples of the FFT frequency increment. This signal is repeated without change so that the input and output histories are periodic with respect to the sample period.

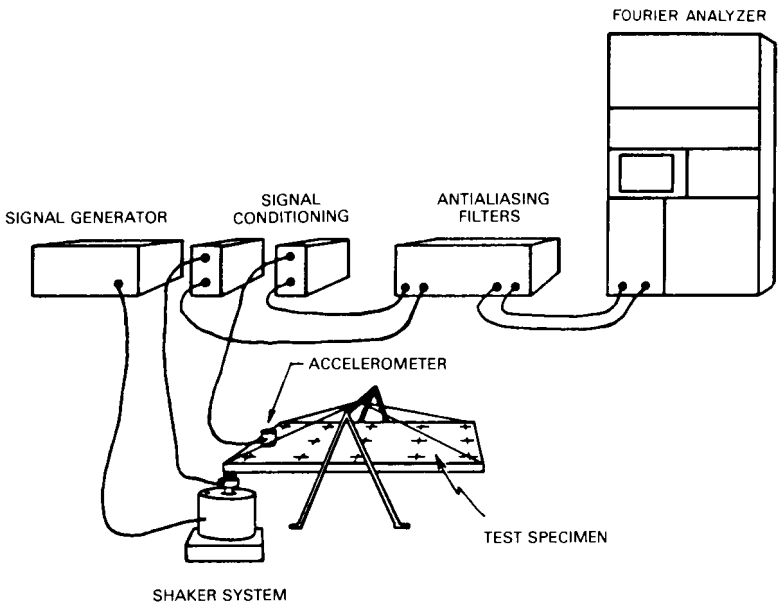

FIGURE 21.12 Typical fixed-input modal test configuration: shaker.

Impact (impulse). The impact signal is a transient deterministic signal which is formed by applying an input pulse lasting only a very small part of the sample period to a system. The width, height, and shape of this pulse determine the usable spectrum of the impact. Briefly, the width of the pulse determines the frequency spectrum, while the height and shape of the pulse control the level of the spec-

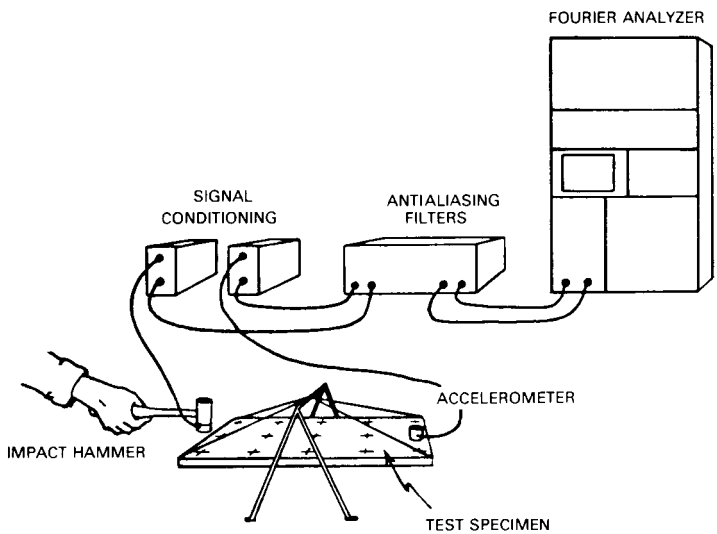

FIGURE 21.13 Typical fixed-response modal test configuration: impact hammer.

TABLE 21.2 Characteristics of Excitation Signals Used in Experimental Modal Analysis

	Slow swept sine	Periodic chirp	Impact	Step relaxation	Pure random	Pseudo-random	Periodic random	Burst random
Minimize leakage	Yes/No	Yes	Yes	Yes	No	Yes	Yes	Yes
Signal-to-noise ratio	Very high	High	Low	Low	Fair	Fair	Fair	Fair
RMS-to-peak ratio	High	High	Low	Low	Fair	Fair	Fair	Fair
Test measurement time	Very long	Very short	Very short	Very short	Good	Very short	Long	Good
Controlled frequency content	Yes*	Yes*	No	No	Yes*	Yes*	Yes*	Yes*
Controlled amplitude content	Yes*	Yes*	No	Yes/No	No	Yes*	Yes*	No
Removes distortion	No	No	No	No	Yes	No	Yes	Yes
Characterize nonlinearity	Yes	Yes	No	No	No	Yes	No	No

* Special hardware required.

trum. Impact signals have proven to be quite popular due to the freedom of applying the input with some form of an instrumented hammer. While the concept is straightforward, the effective utilization of an impact signal is very involved.[14]

Step relaxation. The step relaxation signal is a transient deterministic signal which is formed by releasing a previously applied static input. The sample period begins at the instant that the release occurs. This signal is normally generated by the application of a static force through a cable. The cable is then cut or allowed to release through a shear pin arrangement.

Pure random. The pure random signal is an ergodic, stationary random signal which has a Gaussian probability distribution. In general, the signal contains all frequencies (not just integer multiples of the FFT frequency increment), but it may be filtered to include only information in a frequency band of interest. The measured input spectrum of the pure random signal is altered by any impedance mismatch between the system and the exciter.

Pseudo-random. The pseudo-random signal is an ergodic, stationary random signal consisting only of integer multiples of the FFT frequency increment. The frequency spectrum of this signal has a constant amplitude with random phase. If sufficient time is allowed in the measurement procedure for any transient response to the initiation of the signal to decay, the resultant input and response histories are periodic with respect to the sample period. The number of averages used in the measurement procedure is only a function of the reduction of the variance error. In a noise-free environment, only one average may be necessary.

Periodic random. The periodic random signal is an ergodic, stationary random signal consisting only of integer multiples of the FFT frequency increment. The frequency spectrum of this signal has random amplitude and random phase distribution. Since a single history does not contain information at all frequencies, a number of histories must be involved in the measurement process. For each average, an input history is created with random amplitude and random phase. The system is excited with this input in a repetitive cycle until the transient response to the change in excitation signal decays. The input and response histories should then be periodic with respect to the sample period and are recorded as one average in the total process. With each new average, a new history, uncorrelated with previous input signals, is generated, so that the resulting measurement is completely randomized.

Random transient (burst random). The random transient signal is neither a completely transient deterministic signal nor a completely ergodic, stationary random signal but contains properties of both signal types. The frequency spectrum of this signal has random amplitude and random phase distribution and contains energy throughout the frequency spectrum. The difference between this signal and the periodic random signal is that the random transient history is truncated to zero after some percentage of the sample period (normally 50 to 80 percent). The measurement procedure duplicates the periodic random procedure, but without the need to wait for the transient response to decay. The point at which the input history is truncated is chosen so that the response history decays to zero within the sample period. Even for lightly damped systems, the response history decays to zero very quickly because of the damping provided by the exciter system trying to maintain the input at zero. This damping provided by the exciter system is often overlooked in the analysis of the characteristics of this signal type. Since this measured input, although not part of the generated signal, includes the variation of the input during the decay of the response history, the

input and response histories are totally observable within the sample period and the system damping is unaffected.

Increased Frequency Resolution. An increase in the frequency resolution of a frequency response function affects measurement errors in several ways. Finer frequency resolution allows more exact determination of the damped natural frequency of each modal vector. The increased frequency resolution means that the level of a broad-band signal is reduced. The most important benefit of increased frequency resolution, though, is a reduction of the leakage error. Since the distortion of the frequency response function due to leakage is a function of frequency spacing, not frequency, the increase in frequency resolution reduces the true bandwidth of the leakage error centered at each damped natural frequency. In order to increase the frequency resolution, the total time per history must be increased in direct proportion. The longer data acquisition time increases the variance error problem when transient signals are utilized for input as well as emphasizing any nonstationary problem with the data. The increase of frequency resolution often requires multiple acquisition and/or processing of the histories in order to obtain an equivalent frequency range. This increases the data storage and documentation overhead as well as extending the total test time.

There are two approaches to increasing the frequency resolution of a frequency response function. The first approach involves increasing the number of spectral lines in a baseband measurement. The advantage of this approach is that no additional hardware or software is required. However, FFT analyzers do not always have the capability to alter the number of spectral lines used in the measurement. The second approach involves the reduction of the bandwidth of the measurement while holding the number of spectral lines constant. If the lower frequency limit of the bandwidth is always zero, no additional hardware or software is required. Ideally, though, for an arbitrary bandwidth, hardware and/or software to perform a frequency-shifted, or digitally filtered, FFT is required.

The frequency-shifted FFT process for computing the frequency response function has additional characteristics pertinent to the reduction of errors. Primarily, more accurate information can be obtained on weak spectral components if the bandwidth is chosen to avoid strong spectral components. The out-of-band rejection of the frequency-shifted FFT is better than that of most analog filters that are used in a measurement procedure to attempt to achieve the same results. Additionally, the precision of the resulting frequency response function is improved due to processor gain inherent in the frequency-shifted FFT calculation procedure.[4–6]

Weighting Functions. Weighting functions, or data windows, are probably the most common approach to the reduction of the leakage error in the frequency response function (see Chap. 14). While weighting functions are sometimes desirable and necessary to modify the frequency-domain effects of truncating a signal in the time domain, they are too often utilized when one of the other approaches to error reduction would give superior results. Averaging, selective excitation, and increasing the frequency resolution all act to reduce the leakage error by eliminating the cause of the error. Weighting functions, on the other hand, attempt to compensate for the leakage error after the data have already been digitized.

Windows alter, or compensate for, the frequency-domain characteristic associated with the truncation of data in the time domain. Essentially, again using the narrow bandpass filter analogy, windows alter the characteristics of the bandpass filters that are applied to the data. This compensation for the leakage error causes an attendant distortion of the frequency and phase information of the frequency response

function, particularly in the case of closely spaced, lightly damped system poles. This distortion is a direct function of the width of the main lobe and the size of the side lobes of the spectrum of the weighting function.[4–7]

MODAL PARAMETER ESTIMATION

Modal parameter estimation, or modal identification, is a special case of system identification where the a priori model of the system is known to be in the form of modal parameters. *Modal parameters* include the complex-valued modal frequencies λ_r, modal vectors $\{\psi_r\}$, and modal scaling (modal mass or modal A). Additionally, most algorithms estimate modal participation vectors $\{L_r\}$ and residue vectors $\{A_r\}$ as part of the overall process.

Modal parameter estimation involves estimating the modal parameters of a structural system from measured input-output data. Most modal parameter estimation is based upon the measured data being the frequency response function or the equivalent impulse-response function, typically found by inverse Fourier transforming the frequency response function. Therefore, the form of the model used to represent the experimental data is normally stated in a mathematical frequency response function (FRF) model using temporal (time or frequency) and spatial (input degree-of-freedom and output degree-of-freedom) information.

In general, modal parameters are considered to be global properties of the system. The concept of global modal parameters simply means that there is only one answer for each modal parameter and that the modal parameter estimation solution procedure enforces this constraint. Every frequency response or impulse-response function measurement theoretically contains the information that is represented by the characteristic equation, the modal frequencies, and damping. If individual measurements are treated as independent of one another in the solution procedure, there is nothing to guarantee that a single set of modal frequencies and damping is generated. Likewise, if more than one reference is measured in the data set, redundant estimates of the modal vectors can be made unless the solution procedure utilizes all references in the estimation process simultaneously. Most of the current modal parameter estimation algorithms estimate the modal frequencies and damping in a global sense, but very few estimate the modal vectors in a global sense.

Since the modal parameter estimation process involves a greatly overdetermined problem, the estimates of modal parameters resulting from different algorithms are not the same as a result of differences in the modal model and model domain, differences in how the algorithms use the data, differences in the way the data are weighted or condensed, and differences in user expertise.

MODAL IDENTIFICATION CONCEPTS

The most common approach in modal identification involves using numerical techniques to separate the contributions of individual modes of vibration in measurements such as frequency response functions. The concept involves estimating the individual single degree-of-freedom (SDOF) contributions to the multiple degree-of-freedom (MDOF) measurement.

$$[H(\omega)]_{N_o \times N_i} = \sum_{r=1}^{n} \frac{[A_r]_{No \times Ni}}{j\omega - \lambda_r} + \frac{[A_r^*]_{No \times Ni}}{j\omega - \lambda_r^*} \tag{21.60}$$

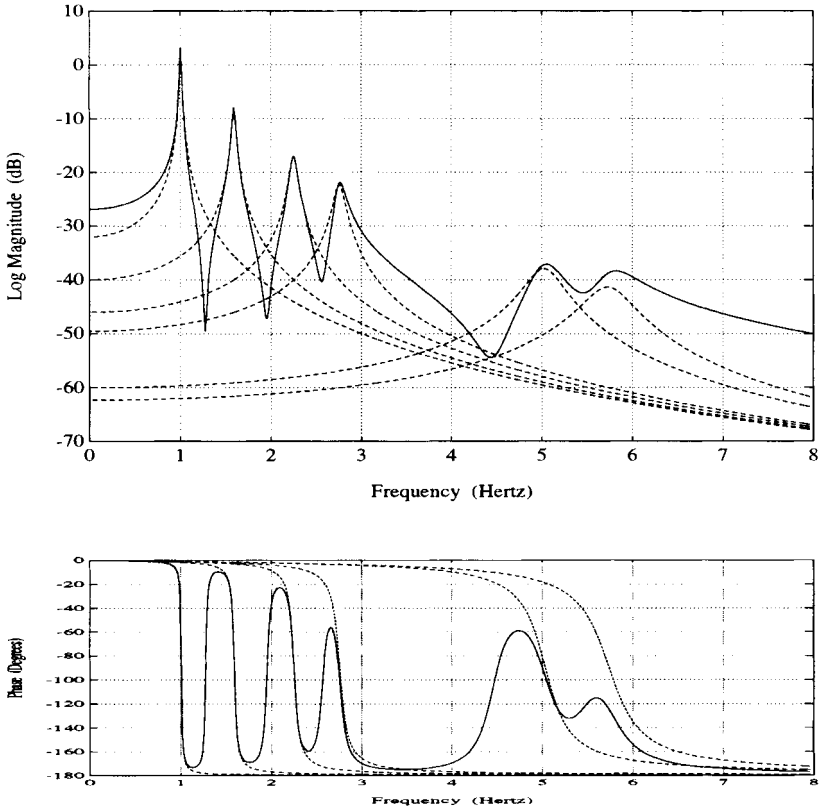

FIGURE 21.14 Modal superposition example (positive frequency poles).

This concept is mathematically represented in Eq. (21.60) and graphically represented in Figs. 21.14 and 21.15.

Equation (21.60) is often formulated in terms of modal vectors $\{\psi_r\}$ and modal participation vectors $\{L_r\}$ instead of residue matrices $[A_r]$. *Modal participation vectors* are a result of multiple reference modal parameter estimation algorithms and relate how well each modal vector is excited from each of the reference locations included in the measured data. The combination of the modal participation vector $\{L_r\}$ and the modal vector $\{\psi_r\}$ for a given mode give the residue matrix $A_{pqr} = L_{qr}\psi_{pr}$ for that mode.

Generally, the modal parameter estimation process involves several stages. Typically, the modal frequencies and modal participation vectors are found in a first stage and residues, modal vectors, and modal scaling are determined in a second stage. Most modal parameter estimation algorithms can be reformulated into a single, consistent mathematical formulation with a corresponding set of definitions and unifying concepts.[15] Particularly, a matrix polynomial approach is used to unify the presentation with respect to current algorithms such as the least squares complex exponential (LSCE), polyreference time domain (PTD), Ibrahim time domain (ITD), eigensystem realization algorithm (ERA), rational fraction polynomial

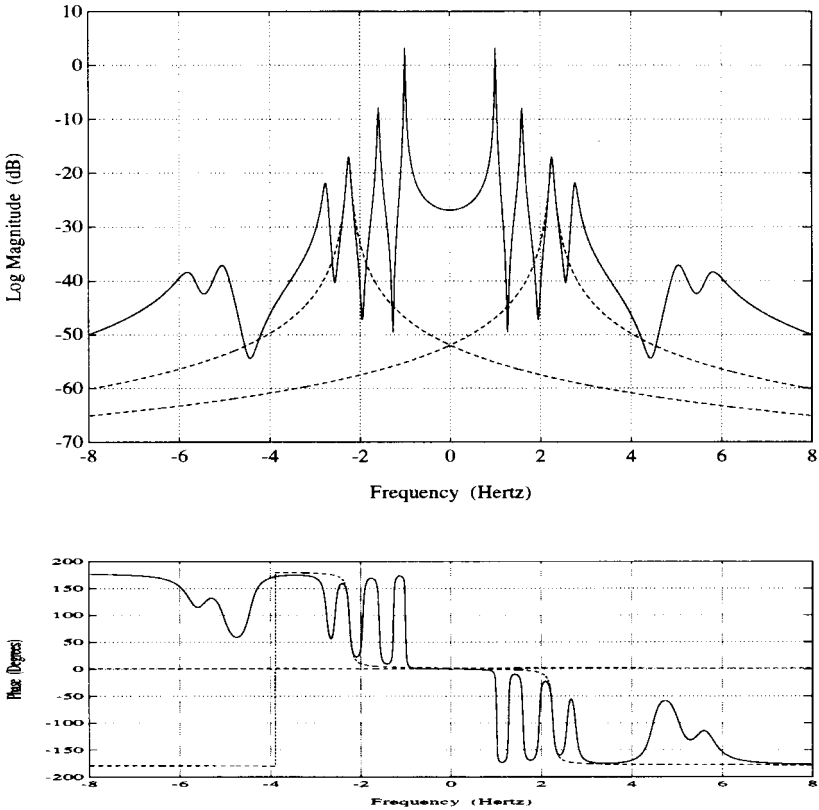

FIGURE 21.15 Modal superposition example (positive and negative frequency poles).

(RFP), polyreference frequency domain (PFD) and complex mode indication function (CMIF) methods. Using this unified matrix polynomial approach (UMPA) allows a discussion of the similarities and differences of the commonly used methods as well as a discussion of the numerical characteristics. Least squares (LS), total least squares (TLS), double least squares (DLS), and singular value decomposition (SVD) methods are used in order to take advantage of redundant measurement data. Eigenvalue and singular value decomposition transformation methods are utilized to reduce the effective size of the resulting eigenvalue-eigenvector problem as well. Many acronyms used in modal parameter estimation are listed in Table 21.3.

Data Domain. Modal parameters can be estimated from a variety of different measurements that exist as discrete data in different data domains (time, frequency, and/or spatial). These measurements can include free decays, forced responses, frequency responses, and unit impulse responses. These measurements can be processed one at a time or in partial or complete sets simultaneously. The measurements can be generated with no measured inputs, a single measured input, or multiple measured inputs. The data can be measured individually or simultaneously. In other words, there is a tremendous variation in the types of measurements and in the

TABLE 21.3 Modal Parameter Estimation Algorithm Acronyms

CEA	Complex exponential algorithm[16]
LSCE	Least squares complex exponential[16]
PTD	Polyreference time domain[17,18]
ITD	Ibrahim time domain[19]
MRITD	Multiple reference Ibrahim time domain[20]
ERA	Eigensystem realization algorithm[21,22]
PFD	Polyreference frequency domain[23–25]
SFD	Simultaneous frequency domain[26]
MRFD	Multireference frequency domain[27]
RFP	Rational fraction polynomial[28]
OP	Orthogonal polynomial[29–31]
CMIF	Complex mode indication function[32]

types of constraints that can be placed upon the testing procedures used to acquire these data. For most measurement situations, frequency response functions are utilized in the frequency domain and impulse-response functions are utilized in the time domain.

Another important concept in experimental modal analysis, and particularly modal parameter estimation, involves understanding the relationships between the temporal (time and/or frequency) information and the spatial (input DOF and output DOF) information. Input-output data measured on a structural system can always be represented as a superposition of the underlying temporal characteristics (modal frequencies) with the underlying spatial characteristics (modal vectors).

Model Order Relationships. The estimation of an appropriate model order is the most important problem encountered in modal parameter estimation. This problem is complicated because of the formulation of the parameter estimation model in the time or frequency domain, a single or multiple reference formulation of the modal parameter estimation model, and the effects of random and bias errors on the modal parameter estimation model. The basis of the formulation of the correct model order can be seen by expanding the theoretical second-order matrix equation of motion to a higher-order model.

$$\left| [m]s^2 + [c]s + [k] \right| = 0 \tag{21.61}$$

The above matrix polynomial is of model order two, has a matrix dimension of $n \times n$, and has a total of $2n$ characteristic roots (modal frequencies). This matrix polynomial equation can be expanded to reduce the size of the matrices to a scalar equation.

$$\alpha_{2N}s^{2N} + \alpha_{2N-1}s^{2N-1} + \alpha_{2N-2}s^{2N-2} + \cdots + \alpha_0 = 0 \tag{21.62}$$

The above matrix polynomial is of model order $2n$, has a matrix dimension of 1×1, and has a total of $2n$ characteristic roots (modal frequencies). The characteristic roots of this matrix polynomial equation are the same as those of the original second-order matrix polynomial equation. Finally, the number of characteristic

roots (modal frequencies) that can be determined depends upon the size of the matrix coefficients involved in the model and the order of the highest polynomial term in the model.

For modal parameter estimation algorithms that utilize experimental data, the matrix polynomial equations that are formed are a function of matrix dimension, from 1×1 to $N_i \times N_i$ or $N_o \times N_o$. There are a significant number of procedures that have been formulated particularly for aiding in these decisions and selecting the appropriate estimation model. Procedures for estimating the appropriate matrix size and model order are another of the differences between various estimation procedures.

Fundamental Measurement Models. Most current modal parameter estimation algorithms utilize frequency- or impulse-response functions as the data, or known information, to solve for modal parameters. The general equation that can be used to represent the relationship between the measured frequency response function matrix and the modal parameters is shown in Eqs. (21.63) and (21.64).

$$[H(\omega)]_{N_o \times N_i} = \left[\psi\right]_{N_o \times 2N} \left[\frac{1}{j\omega - \lambda_r}\right]_{2N \times 2N} [L]^T_{2N \times N_i} \tag{21.63}$$

$$[H(\omega)]^T_{N_i \times N_o} = [L]_{N_i \times 2N} \left[\frac{1}{j\omega - \lambda_r}\right]_{2N \times 2N} \left[\psi\right]^T_{2N \times N_o} \tag{21.64}$$

Impulse-response functions are rarely measured directly but are calculated from associated frequency response functions via the inverse FFT algorithm. The general equation that can be used to represent the relationship between the impulse-response function matrix and the modal parameters is shown in Eqs. (21.9) and (21.10).

$$[h(t)]_{N_o \times N_i} = \left[\psi\right]_{N_o \times 2N} \left[e^{\lambda_r t}\right]_{2N \times 2N} [L]^T_{2N \times N_i} \tag{21.65}$$

$$[h(t)]^T_{N_i \times N_o} = [L]_{N_i \times 2N} \left[e^{\lambda_r t}\right]_{2N \times 2N} \left[\psi\right]^T_{2N \times N_o} \tag{21.66}$$

Many modal parameter estimation algorithms have been originally formulated from Eqs. (21.63) through (21.66). However, a more general development for all algorithms is based upon relating the above equations to a general matrix polynomial approach.

Characteristic Space. From a conceptual viewpoint, the measurement space of a modal identification problem can be visualized as occupying a volume with the coordinate axis defined in terms of three sets of characteristics. Two axes of the conceptual volume correspond to spatial information and the third axis to temporal information. The spatial coordinates are in terms of the input and output degrees-of-freedom (DOF) of the system. The temporal axis is either time or frequency, depending upon the domain of the measurements. These three axis define a 3-D volume which is referred to as the *characteristic space,* as noted in Fig. 21.16. This space or volume rep-

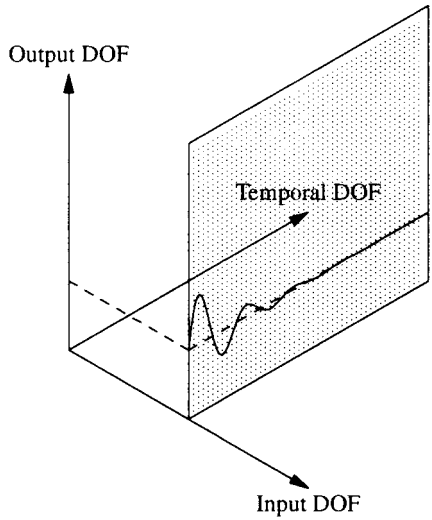

FIGURE 21.16 Conceptualization of modal characteristic space (input DOF axis, output DOF axis, time axis).

resents all possible measurement data as expressed by Eqs. (21.63) through (21.66). This conceptual representation is very useful in understanding what data subspace has been measured. Also, this conceptual representation is very useful in recognizing how the data are organized and utilized with respect to different modal parameter estimation algorithms. Information parallel to one of the axes consists of a solution composed of the superposition of the characteristics defined by that axis. The other two characteristics determine the scaling of each term in the superposition.

In modal parameter estimation algorithms that utilize a single frequency response function, data collection is concentrated on measuring the temporal aspect (time/frequency) at a sufficient resolution to determine the modal parameters. In this approach, the accuracy of the modal parameters, particularly frequency and damping, is essentially limited by Shannon's sampling theorem and Rayleigh's criterion. This focus on the temporal information ignores the added accuracy that use of the spatial information brings to the estimation of modal parameters. Recognizing the characteristic space aspects of the measurement space and using these characteristics (modal vector/participation vector) concepts in the solution procedure leads to the conclusion that the spatial information can compensate for the limitations of temporal information. Therefore, there is a tradeoff between temporal and spatial information for a given accuracy requirement. This is particularly notable in the case of repeated roots. No amount of temporal resolution (accuracy) can theoretically solve repeated roots, but the addition of spatial information in the form of multiple inputs and/or outputs resolves this problem.

Any structural testing procedure measures a subspace of the total possible data available. Modal parameter estimation algorithms may then use all of this subspace or may choose to further limit the data to a more restrictive subspace. It is theoretically possible to estimate the characteristics of the total space by measuring a subspace which samples all three characteristics. However, the selection of the subspace

has a significant influence on the results. In order for all of the modal parameters to be estimated, the subspace must encompass a region which includes contributions of all three characteristics. An important example is the necessity to use multiple reference data (inputs and outputs) in order to estimate *repeated roots.* The particular subspace which is measured and the weighting of the data within the subspace in an algorithm are the main differences among the various modal identification procedures which have been developed.

In general, the amount of information in a measured subspace greatly exceeds the amount necessary to solve for the unknown modal characteristics. Another major difference among the various modal parameter estimation procedures is the type of condensation algorithms that are used to reduce the data to match the number of unknowns [for example, *least squares* (LS), *singular value decomposition* (SVD), etc.]. As is the case with any overspecified solution procedure, there is no unique answer. The answer that is obtained depends upon the data that are selected, the weighting of the data, and the unique algorithm used in the solution process. As a result, the answer is the *best* answer depending upon the objective functions associated with the algorithm being used. Historically, this point has created some confusion since many users expect different methods to give exactly the same answer.

Many modal parameter estimation methods use information (subspace) where only one or two characteristics are included. For example, the simplest (computationally) modal parameter estimation algorithms utilize one impulse-response function or one frequency response function at a time. In this case, only the temporal characteristic is used, and, as might be expected, only temporal characteristics (modal frequencies) can be estimated from the single measurement. The global characteristic of modal frequency cannot be enforced. In practice, when multiple measurements are taken, the modal frequency does not change from one measurement to the next.

Other modal parameter estimation algorithms utilize the data in a plane of the characteristic space. For example, this corresponds to the data taken at a number of response points but from a single excitation point or reference. This representation of a column of measurements is shown in Fig. 21.16 as a plane in the characteristic space. For this case, representing a single input (reference), while it is now possible to enforce the global modal frequency assumption, it is not possible to compute repeated roots and it is difficult to separate closely coupled modes because of the lack of spatial data.

Many modal identification algorithms utilize data taken at a large number of output DOFs due to excitation at a small number of input DOFs. Data taken in this manner are consistent with a multiexciter type of test. Conceptually, this is represented by several planes of data parallel to the plane of data represented in Fig. 21.16. Some modal identification algorithms utilize data taken at a large number of input DOFs and a small number of output DOFs. Data taken in this manner are consistent with a roving hammer type of excitation with several fixed output sensors. These data can also be generated by transposing the data matrix acquired using a multiexciter test. The conceptual representation is several rows of the potential measurement matrix perpendicular to the plane of data represented in Fig. 21.16. Measurement data spaces involving many planes of measured data are the best possible modal identification situations, since the data subspace includes contributions from temporal and spatial characteristics. This allows the best possibility of estimating all the important modal parameters. The data which define the subspace need to be acquired through a consistent measurement process in order for the algorithms to estimate accurate modal parameters. This means that the data must be measured simultaneously and requires that data acquisition, digital signal processing, and instrumentation be designed and operate accordingly.

Fundamental Modal Identification Models. The common characteristics of different modal parameter estimation algorithms can be more readily identified by using a matrix polynomial model rather than using a physically based mathematical model. One way of understanding the basis of this model can be developed from the polynomial model used for the frequency response function.

$$H_{pq}(\omega) = \frac{X_p(\omega)}{F_q(\omega)} = \frac{\beta_n(j\omega)^n + \beta_{n-1}(j\omega)^{n-1} + \cdots + \beta_1(j\omega)^1 + \beta_0(j\omega)^0}{\alpha_m(j\omega)^m + \alpha_{m-1}(j\omega)^{m-1} + \cdots + \alpha_1(j\omega)^1 + \alpha_0(j\omega)^0} \tag{21.67}$$

This can be rewritten as

$$H_{pq}(\omega) = \frac{X_p(\omega)}{F_q(\omega)} = \frac{\sum_{k=0}^{n} \beta_k(j\omega)^k}{\sum_{k=0}^{m} \alpha_k(j\omega)^k} \tag{21.68}$$

Further rearranging yields the following equation, which is linear in the unknown α and β terms:

$$\sum_{k=0}^{m} \alpha_k(j\omega)^k X_p(\omega) = \sum_{k=0}^{n} \beta_k(j\omega)^k F_q(\omega) \tag{21.69}$$

Noting that the response function X_p can be replaced by the frequency response function H_{pq} if the force function F_q is assumed to be unity, the above equation can be restated as

$$\sum_{k=0}^{m} \alpha_k(j\omega)^k H_{pq}(\omega) = \sum_{k=0}^{n} \beta_k(j\omega)^k \tag{21.70}$$

The above formulation is essentially a linear equation in terms of the unknown coefficients α_k and β_k. The equation is valid at each frequency of the measured frequency response function. Since, in the worst case, the number of unknowns is $m + n + 2$, the unknown coefficients can theoretically be determined if the frequency response function has $m + n + 2$ or more discrete frequencies. Practically, this is always the case. Note that the total number of unknown coefficients (or coefficient matrices) is actually $m + n + 1$ since one coefficient (or coefficient matrix) can be assumed to be 1 (or the identity matrix). This is the case because the equation can be divided, or normalized, by one of the unknown coefficients (or coefficient matrices). Note that numerical problems can result if the equation is normalized by a coefficient (or coefficient matrix) that is close to zero. Normally, the coefficient α_0 (or the coefficient matrix $[\alpha_0]$) is chosen as unity (or the identity matrix).

The previous models can be generalized to represent the general multiple input/multiple output case as follows:

$$\sum_{k=0}^{m} \left[[\alpha_k](j\omega)^k \right] \{X(\omega)\} = \sum_{k=0}^{n} \left[[\beta_k](j\omega)^k \right] \{F(\omega)\} \tag{21.71}$$

Note that the size of the coefficient matrices $[\alpha_k]$ and $[\beta_k]$ is normally $N_i \times N_i$ or $N_o \times N_o$ when the equations are developed from experimental data. Rather than the basic model being developed in terms of force and response information, the models can be stated in terms of frequency response information. The response vector $\{X(\omega)\}$ can be replaced by a vector of frequency response functions $\{H(\omega)\}$ where either the input or the output is held fixed. The force vector $\{F(\omega)\}$ is then replaced by an incidence matrix $\{R\}$ of the same size which is composed of all zeros except for unity at the position in the vector consistent with the driving point measurement (common input and output DOF).

$$\sum_{k=0}^{m}\left[(j\omega)^k[\alpha_k]\right]\{H(\omega)\} = \sum_{k=0}^{n}\left[(j\omega)^k[\beta_k]\right]\{R\} \tag{21.72}$$

where

$$\{H(\omega)\} = \begin{Bmatrix} H_{1q}(\omega) \\ H_{2q}(\omega) \\ H_{3q}(\omega) \\ \cdots \\ H_{qq}(\omega) \\ \cdots \\ H_{pq}(\omega) \end{Bmatrix} \qquad \{R\} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{Bmatrix}$$

The above model, in the frequency domain, corresponds to an *autoregressive moving-average* (*ARMA*) model that is developed from a set of finite difference equations in the time domain. The general characteristic matrix polynomial model concept recognizes that both the time- and frequency-domain models generate essentially the same matrix polynomial models. For that reason, the *unified matrix polynomial approach* (*UMPA*) terminology is used to describe both domains since the ARMA terminology has been connected primarily with the time domain.[15]

In parallel with the development of Eq. (21.67), a time-domain model representing the relationship between a single response degree-of-freedom and a single input degree-of-freedom can be stated as follows:

$$\sum_{k=0}^{m} \alpha_k x(t_{i+k}) = \sum_{k=0}^{n} \beta_k f(t_{i+k}) \tag{21.73}$$

For the general multiple input/multiple output case,

$$\sum_{k=0}^{m} [\alpha_k]\,\{x(t_{i+k})\} = \sum_{k=0}^{n} [\beta_k]\,\{f(t_{i+k})\} \tag{21.74}$$

If the discussion is limited to the use of free decay or impulse-response function data, the previous time-domain equations can be greatly simplified by noting that the forcing function can be assumed to be zero for all time greater than zero. If this is the case, the $[\beta_k]$ coefficients can be eliminated from the equations:

$$\sum_{k=0}^{m} [\alpha_k]\left\{h_{pq}(t_{i+k})\right\} = 0 \tag{21.75}$$

In light of the above discussion, it is now apparent that most of the modal parameter estimation processes available can be developed by starting from a general matrix polynomial formulation that is justifiable based upon the underlying matrix differential equation. The general matrix polynomial formulation yields essentially the same characteristic matrix polynomial equation for both time- and frequency-domain data. For the frequency-domain data case, this yields

$$\left| [\alpha_m] s^m + [\alpha_{m-1}] s^{m-1} + [\alpha_{m-2}] s^{m-2} + \cdots + [\alpha_0] \right| = 0 \tag{21.76}$$

For the time-domain data case, this yields

$$\left| [\alpha_m] z^m + [\alpha_{m-1}] z^{m-1} + [\alpha_{m-2}] z^{m-2} + \cdots + [\alpha_0] \right| = 0 \tag{21.77}$$

With respect to the previous discussion of model order, the characteristic matrix polynomial equation, Eq. (21.76) or (21.77), has a model order of *m,* and the number of modal frequencies or roots that are found from this characteristic matrix polynomial equation is *m* times the size of the coefficient matrices [α]. In terms of sampled data, the time-domain matrix polynomial results from a set of finite difference equations and the frequency-domain matrix polynomial results from a set of linear equations, where each equation is formulated at one of the frequencies of the measured data. This distinction is important to note since the roots of the matrix characteristic equation formulated in the time domain are in the *z domain* (z_r) and must be converted to the frequency domain (λ_r), while the roots of the matrix characteristic equation formulated in the frequency domain (λ_r) are already in the desired domain. Note that the roots that are estimated in the time domain are limited to maximum values determined by Shannon's sampling theorem relationship (discrete time steps).

$$z_r = e^{\lambda_r \Delta t} \qquad \lambda_r = \sigma_r + j\omega_r \tag{21.78}$$

$$\sigma_r = \mathrm{Re}\left[\frac{\ln z_r}{\Delta t}\right] \qquad \omega_r = \mathrm{Im}\left[\frac{\ln z_r}{\Delta t}\right]$$

Using this general formulation, the most commonly used modal identification methods can be summarized as shown in Table 21.4.

The high-order model is typically used for those cases where the system is undersampled in the spatial domain. For example, the limiting case is when only one measurement is made on the structure. For this case, the left-hand side of the general linear equation corresponds to a scalar polynomial equation with the order equal to or greater than the number of desired modal frequencies. This type of high-order model may yield significant numerical problems for the frequency-domain case.

The low-order model is used for those cases where the spatial information is complete. In other words, the number of independent physical coordinates is greater than the number of desired modal frequencies. For this case, the order of the left-hand side of the general linear equation, Eq. (21.72) or (21.75), is equal to 1 or 2.

The zero-order model corresponds to a case where the temporal information is neglected and only the spatial information is used. These methods directly estimate the eigenvectors as a first step. In general, these methods are programmed to process data at a single temporal condition or variable. In this case, the method is essentially

TABLE 21.4 Characteristics of Modal Parameter Estimation Algorithms

	Domain		Matrix polynomial order			Coefficients	
Algorithm	Time	Frequency	Zero	Low	High	Scalar	Matrix
CEA	•				•	•	
LSCE	•				•	•	
PTD	•				•		$N_i \times N_i$
ITD	•			•			$N_o \times N_o$
MRITD	•			•			$N_o \times N_o$
ERA	•			•			$N_o \times N_o$
PFD		•		•			$N_o \times N_o$
SFD		•		•			$N_o \times N_o$
MRFD		•		•			$N_o \times N_o$
RFP		•			•	•	Both
OP		•			•	•	Both
CMIF		•	•				$N_o \times N_i$

equivalent to the single degree-of-freedom (SDOF) methods which have been used with frequency response functions. In other words, the comparison between the zeroth-order matrix polynomial model and the higher-order matrix polynomial models is similar to the comparison between the SDOF and MDOF methods used in modal parameter estimation.

Two-Stage Linear Solution Procedure. Almost all modal parameter estimation algorithms in use at this time involve a two-stage linear solution approach. For example, with respect to Eqs. (21.63) through (21.66), if all modal frequencies and modal participation vectors can be found, the estimation of the complex residues can proceed in a linear fashion. This procedure of separating the nonlinear problem into a multistage linear problem is a common technique for most estimation methods today. For the case of structural dynamics, the common technique is to estimate modal frequencies and modal participation vectors in a first stage and then to estimate the modal coefficients plus any residuals in a second stage. Therefore, based upon Eqs. (21.63) through (21.66), most commonly used modal identification algorithms can be outlined as follows:

First stage of modal parameter estimation:

- Load measured data into linear equation form [Eq. (21.72) or (21.75)].
- Find scalar or matrix autoregressive coefficients $[\alpha_k]$.
  - Normalize frequency range (frequency domain only).
  - Utilize orthogonal polynomials (frequency domain only).
- Solve matrix polynomial for modal frequencies.
  - Formulate companion matrix.
  - Obtain eigenvalues of companion matrix λ_r or z_r.
  - Convert eigenvalues from z_r to λ_r (time domain only).
  - Obtain modal participation vectors L_{qr} or modal vectors $\{\psi\}_r$ from eigenvectors of the companion matrix.

Second stage of modal parameter estimation:

- Find modal vectors and modal scaling from Eqs. (21.63) through (21.66).

Equation (21.72) or (21.75) is used to formulate a single, block coefficient linear equation as shown in the graphical analogy of Case 1a, Fig. 21.17. In order to estimate complex conjugate pairs of roots, at least two equations from each piece or block of data in the data space must be used. This situation is shown in Case 1b, Fig. 21.18. In order to develop enough equations to solve for the unknown matrix coefficients, further information is taken from the same block of data or from other blocks of data in the data space until the number of equations equals (Case 2) or exceeds (Case 3) the number of unknowns, as shown in Figs. 21.19 and 21.20. In the frequency domain, this is accomplished by utilizing a different frequency from within each measurement for each equation. In the time domain, this is accomplished by utilizing a different starting time or time shift from within each measurement for each equation.

Once the matrix coefficients $[\alpha]$ have been found, the modal frequencies λ_r or z_r can be found using a number of numerical techniques. While in certain numerical situations, other numerical approaches may be more robust, a companion matrix approach yields a consistent concept for understanding the process. Therefore, the roots of the matrix characteristic equation can be found as the eigenvalues of the associated companion matrix. The companion matrix can be formulated in one of several ways. The most common formulation is as follows:

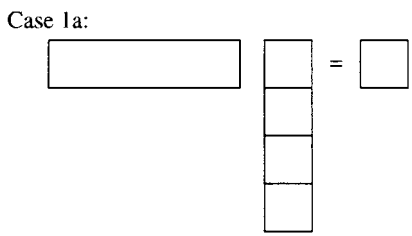

FIGURE 21.17 Underdetermined set of linear equations.

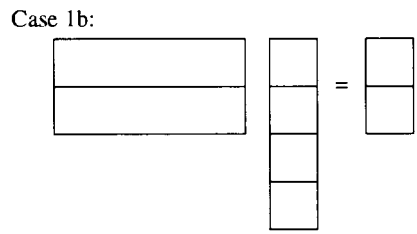

FIGURE 21.18 Underdetermined set of linear equations.

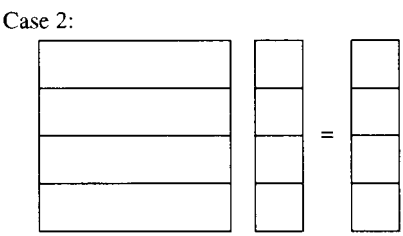

FIGURE 21.19 Determined set of linear equations.

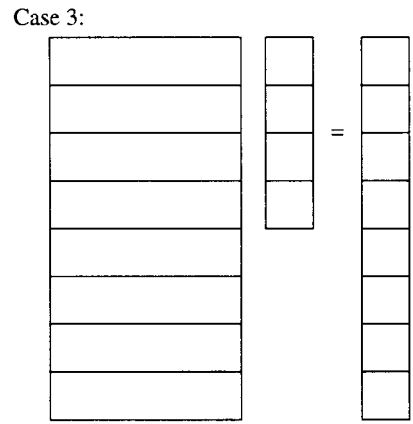

FIGURE 21.20 Overdetermined set of linear equations.

$$[C] = \begin{bmatrix} -[\alpha]_{m-1} & -[\alpha]_{m-2} & \cdots & -[\alpha]_1 & -[\alpha]_0 \\ [I] & [0] & \cdots & [0] & [0] \\ [0] & [I] & \cdots & [0] & [0] \\ [0] & [0] & \cdots & [0] & [0] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ [0] & [0] & \cdots & [0] & [0] \\ [0] & [0] & \cdots & [0] & [0] \\ [0] & [0] & \cdots & [I] & [0] \end{bmatrix} \tag{21.79}$$

Note again that the numerical characteristics of the eigenvalue solution of the companion matrix are different for low-order cases than for high-order cases for a given data set. The companion matrix can be used in the following eigenvalue formulation to determine the modal frequencies for the original matrix coefficient equation:

$$[C]\{X\} = \lambda\,[I]\{X\} \tag{21.80}$$

The eigenvectors that can be found from the eigenvalue-eigenvector solution utilizing the companion matrix may or may not be useful in terms of modal parameters. The eigenvector that is found, associated with each eigenvalue, is of length model order times matrix coefficient size. In fact, the unique (meaningful) portion of the eigenvector is of length equal to the size of the coefficient matrices and is repeated in the eigenvector a model order number of times. Each time the unique portion of the eigenvector is repeated, it is multiplied by a scalar multiple of the associated modal frequency. Therefore, the eigenvectors of the companion matrix have the following form:

$$\{\phi\}_r = \begin{Bmatrix} \lambda_r^{\,m}\{\psi\}_r \\ \cdots \\ \lambda_r^{\,2}\{\psi\}_r \\ \lambda_r^{\,1}\{\psi\}_r \\ \lambda_r^{\,0}\{\psi\}_r \end{Bmatrix}_r \tag{21.81}$$

Note that unless the size of the coefficient matrices is at least as large as the number of measurement degrees-of-freedom, only a partial set of modal coefficients, the modal participation coefficients L_{qr}, are found. For the case involving scalar coefficients, no meaningful modal coefficients are found.

If the size of the coefficient matrices, and therefore the modal participation vector, is less than the largest spatial dimension of the problem, then the modal vectors are typically found in a second-stage solution process using one of Eqs. (21.63) through (21.66). Even if the complete modal vector $\{\psi\}$ of the system is found from the eigenvectors of the companion matrix approach, the modal scaling and modal participation vectors for each modal frequency are normally found in this second-stage formulation.

Data Sieving/Filtering. For almost all cases of modal identification, a large amount of redundancy or overdetermination exists. This means that for Case 3, defined in Fig. 21.20, the number of equations available compared to the number required for the determined Case 2 (defined as the *overdetermination factor*) is quite large. Beyond some value of overdetermination factor, the additional equations contribute little to the result but may add significantly to the solution time. For this reason, the data space is often *filtered* (limited in the temporal sense) or *sieved* (limited in the input DOF or output DOF sense) in order to obtain a reasonable result in the minimum time. For frequency-domain data, the filtering process normally involves limiting the data set to a range of frequencies or a different frequency resolution according to the desired frequency range of interest. For time-domain data, the filtering process normally involves limiting the starting time value as well as the number of sets of time data taken from each measurement. Data sieving involves limiting the data set to certain degrees-of-freedom that are of primary interest. This normally involves restricting the data to specific directions (*X, Y,* and/or *Z* directions) or specific locations or groups of degrees-of-freedom, such as components of a large structural system.

Equation Condensation. Several important concepts should be delineated in the area of equation condensation methods. Equation condensation methods are used to reduce the number of equations based upon measured data to more closely match the number of unknowns in the modal parameter estimation algorithms. There are a large number of condensation algorithms available. Based upon the modal parameter estimation algorithms in use today, the three types of algorithms most often used are

- *Least squares.* Least squares (LS), weighted least squares (WLS), total least squares (TLS), or double least squares (DLS) methods are used to minimize the squared error between the measured data and the estimation model. Historically, this is one of the most popular procedures for finding a pseudo-inverse solution to an overspecified system. The main advantage of this method is computational speed and ease of implementation, while the major disadvantage is numerical precision.
- *Transformation.* There are a large number of transformation that can be used to reduce the data. In the transformation methods, the measured data are reduced by approximating them by the superposition of a set of significant vectors. The number of significant vectors is equal to the amount of independent measured data. This set of vectors is used to approximate the measured data and used as input to the parameter estimation procedures. *Singular value decomposition* (SVD) is one of the more popular transformation methods. The major advantage of such methods is numerical precision, and the disadvantage is computational speed and memory requirements.
- *Coherent averaging.* Coherent averaging is another popular method for reducing the data. In the coherent averaging method, the data are weighted by performing a dot product between the data and a weighting vector (spatial filter). Information in the data which is not coherent with the weighting vectors is averaged out of the data. The method is often referred to as a spatial filtering procedure. This method has both speed and precision but, in order to achieve precision, requires a good set of weighting vectors. In general, the optimum weighting vectors are connected with the solution, which is unknown. It should be noted that least squares is an example of a noncoherent averaging process.

The least squares and the transformation procedures tend to weight those modes of vibration which are well excited. This can be a problem when trying to extract modes which are not well excited. The solution is to use a weighting function for condensation which tends to enhance the mode of interest. This can be accomplished in a number of ways:

- In the time domain, a spatial filter or a coherent averaging process can be used to filter the response to enhance a particular mode or set of modes. For example, by averaging the data from two symmetric exciter locations, the symmetric modes of vibration can be enhanced. A second example is to use only the data in a local area of the system to enhance local modes. The third method is using estimates of the modes' shapes as weighting functions to enhance particular modes.
- In the frequency domain, the data can be enhanced in the same manner as in the time domain, plus the data can be additionally enhanced by weighting them in a frequency band near the natural frequency of the mode of interest.

The type of equation condensation method that is utilized in a modal identification algorithm has a significant influence on the results of the parameter estimation process.

Coefficient Condensation. For the low-order modal identification algorithms, the number of physical coordinates (typically N_o) is often much larger than the number of desired modal frequencies ($2n$). For this situation, the numerical solution procedure is constrained to solve for N_o or $2N_o$ modal frequencies. This can be very time consuming and is unnecessary. The number of physical coordinates N_o can be reduced to a more reasonable size ($N_e \approx N_o$ or $N_e \approx 2N_o$) by using a decomposition transformation from physical coordinates N_o to the approximate number of effective modal frequencies N_e. Currently, SVD or eigenvalue decompositions (ED) are used to preserve the principal modal information prior to formulating the linear equation solution for unknown matrix coefficients.[33,34] In most cases, even when the spatial information must be condensed, it is necessary to use a model order greater than 2 to compensate for distortion errors or noise in the data and to compensate for the case where the location of the transducers is not sufficient to totally define the structure.

$$[H'] = [T][H] \tag{21.82}$$

where $[H']$ = transformed (condensed) frequency response function matrix
$[T]$ = transformation matrix
$[H]$ = original FRF matrix

The difference between the two techniques lies in the method of finding the transformation matrix $[T]$. Once $[H]$ has been condensed, however, the parameter estimation procedure is the same as for the full data set. Because the data eliminated from the parameter estimation process ideally correspond to the noise in the data, the modal frequencies of the condensed data are the same as the modal frequencies of the full data set. However, the modal vectors calculated from the condensed data may need to be expanded back into the full space:

$$[\Psi] = [T]^T [\Psi'] \tag{21.83}$$

where $[\Psi]$ = full-space modal matrix
$[\Psi']$ = condensed-space modal matrix

Model Order Determination. Much of the work on modal parameter estimation since 1975 has involved methodology for determining the correct model order for the modal parameter model. Technically, model order refers to the highest power in the matrix polynomial equation. The number of modal frequencies found is equal to the model order times the size of the matrix coefficients, normally N_o or N_i. For a given algorithm, the size of the matrix coefficients is normally fixed; therefore, determining the model order is directly linked to estimating n, the number of modal frequencies in the measured data that are of interest. As has always been the case, an estimate for the minimum number of modal frequencies can be easily found by counting the number of peaks in the frequency response function in the frequency band of analysis. This is a minimum estimate of n since the frequency response function measurement may be at a node of one or more modes of the system, repeated roots may exist, and/or the frequency resolution of the measurement may be too coarse to observe modes that are closely spaced in frequency. Several measurements can be observed and a tabulation of peaks existing in any or all measurements can be used as a more accurate minimum estimate of n. A more automated procedure for including the peaks that are present in several frequency response functions is to observe the summation of frequency response function power. This function represents the autopower or automoment of the frequency response functions summed over a number of response measurements and is normally formulated as follows:

$$H_{\text{power}}(\omega) = \sum_{p=1}^{N_O} \sum_{q=1}^{N_i} H_{pq}(\omega)\, H_{pq}{}^*(\omega) \tag{21.84}$$

These techniques are extremely useful but do not provide an accurate estimate of model order when repeated roots exist or when modes are closely spaced in frequency. For these reasons, an appropriate estimate of the order of the model is of prime concern and is the single most important problem in modal parameter estimation.

In order to determine a reasonable estimate of the model order for a set of representative data, a number of techniques have been developed as guides or aids to the user. Much of the user interaction involved in modal parameter estimation involves the use of these tools. Most of the techniques that have been developed allow the user to establish a maximum model order to be evaluated (in many cases, this is set by the memory limits of the computer algorithm). Information is utilized from the measured data based upon an assumption that the model order is equal to this maximum. This information is evaluated in a sequential fashion to determine if a model order less than the maximum is sufficient to describe the data sufficiently. This is the point at which the user's judgment and the use of various evaluation aids becomes important. Some of the commonly used techniques are:

- Measurement synthesis and comparison (curve-fit)
- Error chart
- Stability diagram
- Mode indication functions
- Rank estimation

One of the most common techniques is to synthesize an impulse-response function or a frequency response function and compare it to the measured function to see if modes have been missed. This curve-fitting procedure is also used as a measure of the overall success of the modal parameter estimation procedure. The differ-

ence between the two functions can be quantified and normalized to give an indicator of the degree of fit. There can be many reasons for a poor comparison; incorrect model order is one of the possibilities.

Error Chart. Another method that has been used to indicate the correct model order more directly is the error chart. Essentially, the error chart is a plot of the error in the model as a function of increasing model order. The error in the model is a normalized quantity that represents the ability of the model to predict data that are not involved in the estimate of the model parameters. For example, when measured data in the form of an impulse-response function are used, only a small percentage of the total number of data values are involved in the estimate of modal parameters. If the model is estimated based upon 10 modes, only 4×10 data points are required, at a minimum, to estimate the modal parameters if no additional spatial information is used. The error in the model can then be estimated by the ability of the model to predict the next several data points in the impulse-response function compared to the measured data points. For the case of 10 modes and 40 data points, the error in the model is calculated from the predicted and measured data points 41 through 50. When the model order is insufficient, this error is large, but when the model order reaches the correct value, further increase in the model order does not result in a further decrease in the error. Figure 21.21 is an example of an error chart.

Stability Diagram. A further enhancement of the error chart is the stability diagram. The stability diagram is developed in the same fashion as the error chart and involves tracking the estimates of frequency, damping, and possibly modal participation factors as a function of model order. As the model order is increased, more and more modal frequencies are estimated, but, hopefully, the estimates of the physical modal parameters stabilize as the correct model order is found. For modes that are very active in the measured data, the modal parameters stabilize at a very low model order. For modes that are poorly excited in the measured data, the modal parameters may not stabilize until a very high model order is chosen. Nevertheless, the nonphysical (computational) modes do not stabilize at all during this process and can be sorted out of the modal parameter data set more easily. Note that inconsistencies (frequency shifts, leakage errors, etc.) in the measured data set obscure the stability and make the stability diagram difficult to use. Normally, a tolerance, in percentage, is given for the stability of each of the modal parameters that are being evaluated. Figure 21.22 is an example of a stability diagram. In Fig. 21.22, a summation of

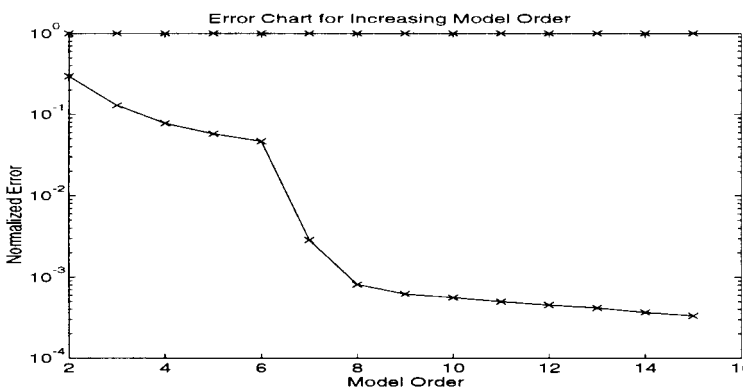

FIGURE 21.21 Model order determination: error chart.

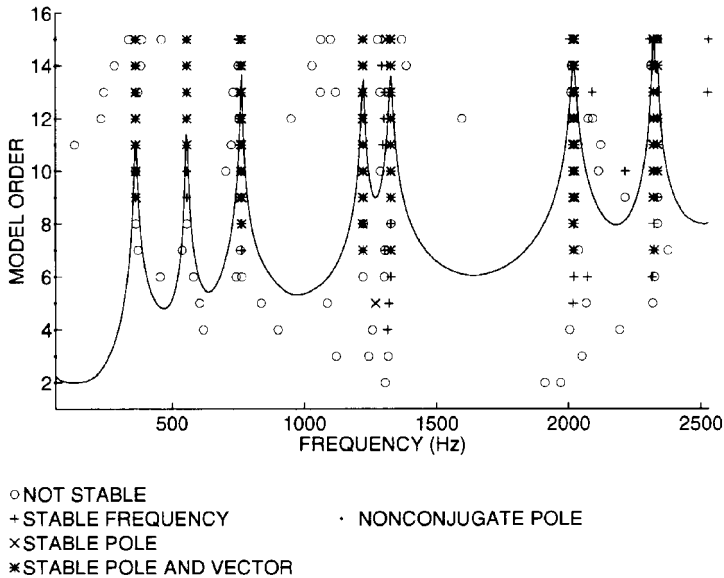

FIGURE 21.22 Model order determination: stability diagram.

the frequency response function power is plotted on the stability diagram for reference. Other mode indication functions can also be plotted against the stability diagram for reference.

Mode Indication Functions. Mode indication functions (MIF) are normally real-valued, frequency-domain functions that exhibit local minima or maxima at the modal frequencies of the system. One mode indication function can be plotted for each reference available in the measured data. The primary mode indication function exhibits a local minimum or maximum at each of the natural frequencies of the system under test. The secondary mode indication function exhibits a local minimum or maximum at repeated or pseudo-repeated roots of order 2 or more. Further mode indication functions yield local minima or maxima for successively higher orders of repeated or pseudo-repeated roots of the system under test.

MULTIVARIATE MODE INDICATION FUNCTION (MvMIF): The development of the multivariate mode indication function is based upon finding a force vector $\{F\}$ that excites a normal mode at each frequency in the frequency range of interest.[35] If a normal mode can be excited at a particular frequency, the response to such a force vector exhibits the 90° phase lag characteristic. Therefore, the real part of the response is as small as possible, particularly when compared to the imaginary part or the total response. In order to evaluate this possibility, a minimization problem can be formulated as follows:

$$\min_{\|F\|=1} \frac{\{F\}^T [H_{\text{Real}}]^T [H_{\text{Real}}] \{F\}}{\{F\}^T ([H_{\text{Real}}]^T [H_{\text{Real}}] + [H_{\text{Imag}}]^T [H_{\text{Imag}}]) \{F\}} = \lambda \tag{21.85}$$

This minimization problem is similar to a Rayleigh quotient, and it can be shown that the solution to the problem is found by finding the smallest eigenvalue $\lambda_{\min}$ and the corresponding eigenvector $\{F\}_{\min}$ of the following problem:

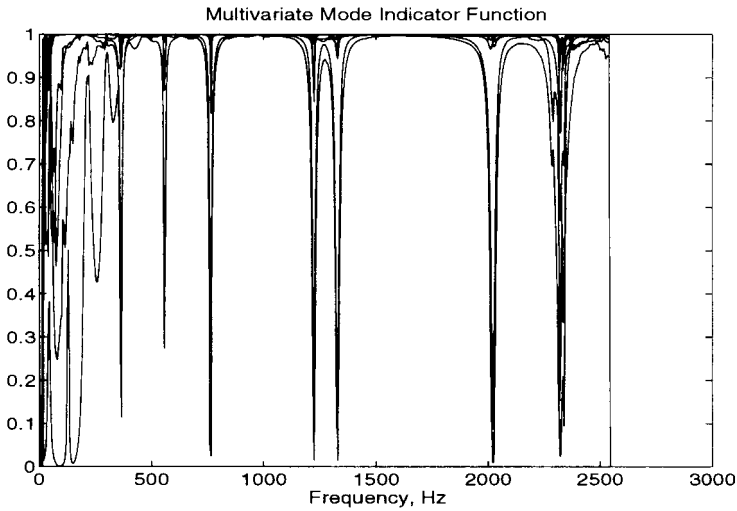

FIGURE 21.23 Multivariate mode indication function: seven-input example.

$$[H_{\text{Real}}]^T [H_{\text{Real}}] \{F\} = \lambda ([H_{\text{Real}}]^T [H_{\text{Real}}] + [H_{\text{Imag}}]^T [H_{\text{Imag}}]) \{F\} \qquad (21.86)$$

The above eigenvalue problem is formulated at each frequency in the frequency range of interest. Note that the result of the matrix product $[H_{\text{Real}}]^T [H_{\text{Real}}]$ and $[H_{\text{Imag}}]^T [H_{\text{Imag}}]$ in each case is a square, real-valued matrix of size equal to the number of references in the measured data $N_i \times N_i$. The resulting plot of a multivariate mode indication function for a seven-reference case can be seen in Fig. 21.23. The frequencies where more than one curve approaches the same minimum are likely to be repeated root frequencies (repeated modal frequencies).

COMPLEX MODE INDICATION FUNCTION (CMIF): An algorithm based on singular value decomposition methods applied to multiple reference FRF measurements, identified as the complex mode indication function (CMIF), is utilized in order to identify the proper number of modal frequencies, particularly when there are closely spaced or repeated modal frequencies.[35] Unlike MvMIF, which indicates the existence of real normal modes, CMIF indicates the existence of real normal or complex modes and the relative magnitude of each mode. Furthermore, MvMIF yields a set of force patterns that can best excite the real normal mode, while CMIF yields the corresponding mode shape and modal participation vector.

The CMIF, in the original formulation, is defined as the eigenvalues, solved from the normal matrix formed from the frequency response function matrix, at each spectral line. The normal matrix is obtained by premultiplying the FRF matrix by its Hermitian matrix as $[H(\omega)]^H [H(\omega)]$. The CMIF is the plot of these eigenvalues on a log magnitude scale as a function of frequency. The peaks detected in the CMIF plot indicate the existence of modes, and the corresponding frequencies of these peaks give the damped natural frequencies for each mode. In the application of CMIF to traditional modal parameter estimation algorithms, the number of modes detected in CMIF determines the minimum number of degrees-of-freedom of the system equation for the algorithm. A number of additional degrees-of-freedom may be needed to take care of residual effects and noise contamination.

$$[H(\omega)]^H [H(\omega)] = [V(\omega)] [\Lambda(\omega)] [V(\omega)]^H \tag{21.87}$$

By taking the singular value decomposition of the FRF matrix at each spectral line, an expression similar to Eq. (21.87) is obtained:

$$[H(\omega)] = [U(\omega)] [\Sigma(\omega)] [V(\omega)]^H \tag{21.88}$$

where N_e = number of effective modes. The effective modes are the modes that contribute to the response of the structure at this particular frequency ω
$[U(\omega)]$ = left singular matrix of size $N_o \times N_e$, which is a unitary matrix
$[\Lambda(\omega)]$ = eigenvalue matrix of size $N_d \times N_e$, which is a diagonal matrix
$[\Sigma(\omega)]$ = singular value matrix of size $N_e \times N_e$, which is a diagonal matrix
$[V(\omega)]$ = right singular matrix of size $N_d \times N_i$, which is also a unitary matrix

Most often, the number of input points (reference points) N_i is less than the number of response points N_o. In Eq. (21.88), if the number of effective modes is less than or equal to the smaller dimension of the FRF matrix, i.e., $N_e \leq N_i$, the singular value decomposition leads to approximate mode shapes (left singular vectors) and approximate modal participation factors (right singular vectors). The singular value is then equivalent to the scaling factor Q_r divided by the difference between the discrete frequency and the modal frequency $j\omega - \lambda_r$. For a given mode, since the scaling factor is a constant, the closer the modal frequency is to the discrete frequency, the larger the singular value is. Therefore, the damped natural frequency is the frequency at which the maximum magnitude of the singular value occurs. If different modes are compared, the stronger the mode contribution (larger residue value), the larger the singular value is.

$$\text{CMIF}_k(\omega) \equiv \Lambda_k(\omega) = \Sigma_k(\omega)^2 \qquad k = 1, 2, \ldots, N_e \tag{21.89}$$

where $\text{CMIF}_k(\omega)$ = kth CMIF as a function of frequency ω
$\Lambda_k(\omega)$ = kth eigenvalue of the normal matrix of FRF matrix as a function of frequency ω
$\Sigma_k(\omega)$ = kth singular value of the FRF matrix as a function of frequency ω

In practical calculations, the normal matrix formed from the FRF matrix, $[H(\omega)]^H [H(\omega)]$, is calculated at each spectral line. The eigenvalues of this matrix are obtained. The CMIF plot is the plot of these eigenvalues on a log magnitude scale as a function of frequency. The peak in the CMIF indicates the location on the frequency axis that is nearest to the pole. The frequency is the estimated damped natural frequency, to within the accuracy of the frequency resolution. The magnitude of the eigenvalue indicates the relative magnitude of the modes, residue over damping factor.

Since the mode shapes that contribute to each peak do not change much around each peak, several adjacent spectral lines from the FRF matrix can be used simultaneously for a better estimation of mode shapes. By including several spectral lines of data in the singular value decomposition calculation, the effect of the leakage error can be minimized. The resulting plot of a complex mode indication function for a seven-reference case can be seen in Fig. 21.24. The frequencies where more than one curve approaches the same maximum are repeated root frequencies (repeated modal frequencies).

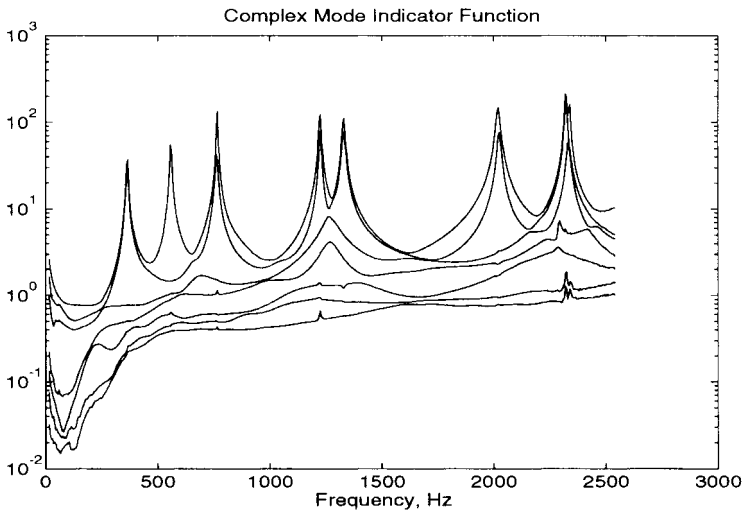

FIGURE 21.24 Complex mode indication function: seven-input example.

Rank Estimation. A more recent model order evaluation technique involves the estimate of the rank of the matrix of measured data. An estimate of the rank of the matrix of measured data gives a good estimate of the model order of the system. Essentially, the rank is an indicator of the number of independent characteristics contributing to the data. While the rank cannot be calculated in an absolute sense, it can be estimated from the singular value decomposition (SVD) of the matrix of measured data. For each mode of the system, one singular value should be found by the SVD procedure. The SVD procedure finds the largest singular value first and then successively finds the next largest. The magnitudes of the singular values are used in one of two different procedures to estimate the rank. The concept that is used is that the singular values should go to zero when the rank of the matrix is exceeded. For theoretical data, this happens exactly. For measured data, because of random errors and small inconsistencies in the data, the singular values do not become zero but become very small. Therefore, the rate of change of the singular values rather than the absolute values is used as an indicator. In one approach, each singular value is divided by the first (largest) to form a normalized ratio. This normalized ratio is treated much like the error chart, and the appropriate rank (model order) is chosen when the normalized ratio approaches an asymptote. In another similar approach, each singular value is divided by the previous singular value, forming a normalized ratio that is approximately equal to 1 if the successive singular values are not changing in magnitude. When a rapid decrease in the magnitude of the singular value occurs, the ratio of successive singular values drops (or peaks if the inverse of the ratio is plotted) as an indicator of rank (model order) of the system. Figure 21.25 shows examples of these rank estimate procedures.

Residuals. Continuous systems have an infinite number of degrees-of-freedom, but, in general, only a finite number of modes can be used to describe the dynamic behavior of a system. The theoretical number of degrees-of-freedom can be reduced by using a finite frequency range. Therefore, for example, the frequency response

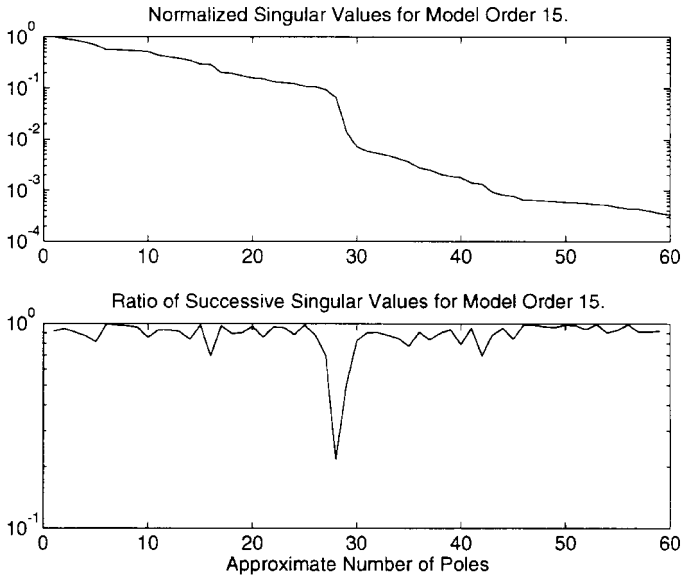

FIGURE 21.25 Model order determination: rank estimation.

can be broken up into three partial sums, each covering the modal contribution corresponding to modes located in the frequency ranges.

In the frequency range of interest, the modal parameters can be estimated to be consistent with Eq. (21.60). In the lower and higher frequency ranges, residual terms can be included to account for modes in these ranges. In this case, Eq. (21.60) can be rewritten for a single frequency response function as

$$H_{pq}(\omega) = R_{F_{pq}} + \sum_{r=1}^{n} \frac{A_{pqr}}{j\omega - \lambda_r} + \frac{A_{pqr}^*}{j\omega - \lambda_r^*} + R_{I_{pq}}(\omega) \tag{21.90}$$

where $R_{F_{pq}}$ = residual flexibility
$R_{I_{pq}}(s)$ = residual inertia

The residual term that compensates for modes below the minimum frequency of interest is called the *inertia restraint,* or *residual inertia.* The residual term that compensates for modes above the maximum frequency of interest is called the *residual flexibility.* These residuals are a function of each frequency response function measurement and are not global properties of the frequency response function matrix. Therefore, residuals cannot be estimated unless the frequency response function is measured. In this common formulation of residuals, both terms are real-valued quantities. In general, this is a simplification; the residual effects of modes below and/or above the frequency range of interest cannot be completely represented by such simple mathematical relationships. As the system poles below and above the range of interest are located in the proximity of the boundaries, these effects are not the real-valued quantities noted in Eq. (21.90). In these cases, residual modes may be included in the model to partially account for these effects. When this is done, the modal parameters that are associated with these residual poles have no physical significance but may be required in order to compensate for strong dynamic influences from out-

side the frequency range of interest. Using the same argument, the lower and upper residuals can take on any mathematical form that is convenient as long as the lack of physical significance is understood. Mathematically, power functions of frequency (zero, first, and second order) are commonly used within such a limitation. In general, the use of residuals is confined to frequency response function models. This is primarily due to the difficulty of formulating a reasonable mathematical model and solution procedure in the time domain for the general case that includes residuals.

MODAL IDENTIFICATION ALGORITHMS (SDOF)

For any real system, the use of single degree-of-freedom algorithms to estimate modal parameters is always an approximation since any realistic structural system has many degrees-of-freedom. Nevertheless, in cases where the modes are not close in frequency and do not affect one another significantly, single degree-of-freedom algorithms are very effective. Specifically, single degree-of-freedom algorithms are quick, rarely involving much mathematical manipulation of the data, and give sufficiently accurate results for most modal parameter requirements. Naturally, most multiple degree-of-freedom algorithms can be constrained to estimate only a single degree-of-freedom at a time if further mathematical accuracy is desired. The most commonly used single degree-of-freedom algorithms involve using the information at a single frequency as an estimate of the modal vector.

Operating Vector Estimation. Technically, when many single degree-of-freedom approaches are used to estimate modal parameters, sufficient simplifying assumptions are made that the results are not actually modal parameters. In these cases, the results are often referred to as *operating vectors* rather than modal vectors. This term refers to the fact that if the structural system is excited at this frequency, the resulting motion is a linear combination of the modal vectors rather than a single modal vector. If one mode is dominant, then the operating vector is approximately equal to the modal vector. The approximate relationships that are used in these cases are represented in the following two equations:

$$H_{pq}(\omega_r) \approx \frac{A_{pqr}}{j\omega_r - \lambda_r} + \frac{A_{pqr}^*}{j\omega_r - \lambda_r^*} \tag{21.91}$$

$$H_{pq}(\omega_r) \approx \frac{A_{pqr}}{-\sigma_r} \tag{21.92}$$

For these less complicated methods, the damped natural frequencies ω_r are estimated by observing the maxima in the frequency response functions. The damping factors σ_r are estimated using half-power methods.[1] The residues A_{pqr} are then estimated from Eq. (21.91) or (21.92) using the frequency response function data at the damped natural frequency.

Complex Plot (Circle Fit). The circle-fit method utilizes the concept that the data curve in the vicinity of a modal frequency looks circular. In fact, the diameter of the circle is used to estimate the residue once the damping factor is estimated. More importantly, this method utilizes the concept that the distance along the curve between data points at equidistant frequencies is a maximum in the neighborhood of the modal frequency. Therefore, the circle-fit method is the first method to detect closely spaced modes.

This method can give erroneous answers when the modal coefficient is near zero. This occurs essentially because, when the mode does not exist in a particular frequency response function (either the input or the response degree-of-freedom is at a node of the mode), the remaining data in the frequency range of the mode are strongly affected by the next higher or lower mode. Therefore, the diameter of the circle that is estimated is a function of the modal coefficient for the next higher or lower mode. This can be detected visually but is somewhat difficult to detect automatically. The approximate relationship that is used in this case is represented in the following equation:

$$H_{pq}(\omega_r) \approx R_{pq} + \frac{A_{pqr}}{j\omega_r - \lambda_r} + \frac{A_{pqr}^*}{j\omega_r - \lambda_r^*} \tag{21.93}$$

Two-Point Finite Difference Formulation. The difference method formulations are methods that are based upon comparing adjacent frequency information in the vicinity of a resonance frequency. When a ratio of this information, together with information from the derivative of the frequency response function at the same frequencies, is formed, a reasonable estimation of the modal frequency and residue for each mode can be determined under the assumption that modes are not too close together. This method can give erroneous answers when the modal coefficient is near zero. This problem can be detected by comparing the predicted modal frequency to the frequency range of the data used in the finite difference algorithm. As long as the predicted modal frequency lies within the frequency band, the estimate of the residue (modal coefficient) should be valid.

The approximate relationships that are used in this case are represented in the following equations. The frequencies noted in these relationships are as follows: ω_1 is a frequency near the damped natural frequency ω_r, and ω_p is the peak frequency close to the damped natural frequency ω_r.

Modal frequency (λ_r):

$$\lambda_r \approx \frac{j\omega_p H_{pq}(\omega_p) - j\omega_1 H_{pq}(\omega_1)}{H_{pq}(\omega_p) - H_{pq}(\omega_1)} \tag{21.94}$$

Residue (A_{pqr}):

$$A_{pqr} \approx \frac{j(\omega_1 - \omega_p) H_{pq}(\omega_1) H_{pq}(\omega_p)}{H_{pq}(\omega_p) - H_{pq}(\omega_1)} \tag{21.95}$$

Since both of the equations that are used to estimate modal frequency λ_r and residue A_{pqr} are linear equations, a least squares solution can be formed by using other frequency response function data in the vicinity of the resonance. For this case, additional equations can be developed using $H_{pq}(\omega_2)$ or $H_{pq}(\omega_3)$ in the above equations instead of $H_{pq}(\omega_1)$.

MODAL IDENTIFICATION ALGORITHMS (MDOF)

All multiple degree-of-freedom equations can be represented in a unified matrix polynomial approach. The methods that are summarized in the following sections are listed in Tables 21.3 and 21.4.

High-Order Time-Domain Algorithms. The algorithms that fall into the category of high-order time-domain algorithms include the algorithms most commonly used to determine modal parameters. The least squares complex exponential (LSCE) algorithm is the first algorithm to utilize more than one frequency response function, in the form of impulse-response functions, in the solution for a global estimate of the modal frequency. The polyreference time-domain (PTD) algorithm is an extension to the LSCE algorithm that allows multiple references to be included in a meaningful way so that the ability to resolve close modal frequencies is enhanced. Since both the LSCE and PTD algorithms have good numerical characteristics, these algorithms are still the most commonly used today. The only limitations for these algorithms are the cases involving high damping. As these are high-order algorithms, more time-domain information is required than for low-order algorithms.

First-Order Time-Domain Algorithms. The first-order time-domain algorithms include several well-known algorithms such as the Ibrahim time-domain (ITD) algorithm and the eigensystem realization algorithm (ERA). These algorithms are essentially a state-space formulation with respect to the second-order time-domain algorithms. The original development of these algorithms is quite different from that presented here, but the resulting solution of linear equations is the same regardless of development. There is a great body of published work on both the ITD and ERA algorithms, much of which discusses the various approaches for condensing the overdetermined set of equations that results from the data (least squares, double least squares, singular value decomposition). The low-order time-domain algorithms require very few time points in order to generate a solution because of the increased use of spatial information.

Second-Order Time-Domain Algorithms. The second-order time-domain algorithm has not been reported in the literature previously but is simply modeled after the second-order matrix differential equation with matrix dimension N_o. Since an impulse-response function can be thought to be a linear summation of a number of complementary solutions to such a matrix differential equation, the general second-order matrix form is a natural model that can be used to determine the modal parameters. This method is developed by noting that it is the time-domain equivalent to a frequency-domain algorithm known as the polyreference frequency-domain (PFD) algorithm. The low-order time-domain algorithms require very few time points in order to generate a solution because of the increased use of spatial information.

High-Order Frequency-Domain Algorithms. The high-order frequency-domain algorithms, in the form of scalar coefficients, are the oldest multiple degree-of-freedom algorithms utilized to estimate modal parameters from discrete data. These are algorithms like the rational fraction polynomial (RFP), power polynomial (PP), and orthogonal polynomial (OP) algorithms. These algorithms work well for narrow frequency bands and limited numbers of modes but have poor numerical characteristics otherwise. While the use of multiple references reduces the numerical conditioning problem, the problem is still significant and not easily handled. In order to circumvent the poor numerical characteristics, many approaches have been used (frequency normalization, orthogonal polynomials), but the use of low-order frequency-domain models has proven more effective.

Orthogonal Polynomial Concepts. The fundamental problem with using a rational fraction polynomial (power polynomial) method can be highlighted by looking at the characteristics of the data matrices. These matrices involve power

polynomials that are functions of increasing powers of $s = j\omega$. These matrices are of the Vandermonde form and are known to be ill-conditioned for cases involving wide frequency ranges and high-ordered models.

VANDERMONDE MATRIX FORM:

$$\begin{bmatrix} (j\omega_1)^0 & (j\omega_1)^1 & (j\omega_1)^2 & \dots & (j\omega_1)^{2m-1} \\ (j\omega_2)^0 & (j\omega_2)^1 & (j\omega_2)^2 & \dots & (j\omega_2)^{2m-1} \\ (j\omega_3)^0 & (j\omega_3)^1 & (j\omega_3)^2 & \dots & (j\omega_3)^{2m-1} \\ \dots & \dots & \dots & \dots & \dots \\ (j\omega_i)^0 & (j\omega_i)^1 & (j\omega_i)^2 & \dots & (j\omega_i)^{2m-1} \end{bmatrix} \tag{21.96}$$

Ill-conditioning, in this case, means that the accuracy of the solution for the matrix coefficients α_m is limited by the numerical precision of the available arithmetic of the computer. Since the matrix coefficients α_m are used to determine the complex-valued modal frequencies, this presents a serious limitation for the high-order frequency-domain algorithms. The ill-conditioning problem can be best understood by evaluating the condition number of the Vandermonde matrix. The *condition number* measures the sensitivity of the solution of linear equations to errors, or small amounts of noise, in the data. The condition number gives an indication of the accuracy of the results from matrix inversion and/or linear equation solution. The condition number for a matrix is computed by taking the ratio of the largest singular value to the smallest singular value. A good condition number is a small number close to unity; a bad condition number is a large number. For the theoretical case of a singular matrix, the condition number is infinite.

The ill-conditioned characteristic of matrices that are of the Vandermonde form can be reduced, but not eliminated, by the following:

- Minimizing the frequency range of the data
- Minimizing the order of the model
- Normalizing the frequency range of the data (0,2) or (–2,2)
- Use of orthogonal polynomials

Several orthogonal polynomials have been applied to the frequency-domain modal parameter estimation problem, such as

- Forsythe polynomials
- Chebyshev polynomials
- Legendre polynomials
- Laguerre polynomials

First-Order Frequency-Domain Algorithms. Several algorithms have been developed that fall into the category of first-order frequency-domain algorithms, including the simultaneous frequency-domain (SFD) algorithm and the multiple reference simultaneous frequency-domain algorithm. These algorithms are essentially frequency-domain equivalents to the ITD and ERA algorithms and effectively involve a state-space formulation when compared to the second-order frequency-domain algorithms. The state-space formulation utilizes the derivatives of the

frequency response functions as well as the frequency response function in the solution. These algorithms have superior numerical characteristics compared to the high-order frequency-domain algorithms. Unlike the low-order time-domain algorithms, though, sufficient data from across the complete frequency range of interest must be included in order to obtain a satisfactory solution.

Second-Order Frequency-Domain Algorithms. The second-order frequency-domain algorithms include the polyreference frequency-domain (PFD) algorithms. These algorithms have superior numerical characteristics compared to the high-order frequency-domain algorithms. Unlike the low-order time-domain algorithms, though, sufficient data from across the complete frequency range of interest must be included in order to obtain a satisfactory solution.

Residue Estimation. Once the modal frequencies and modal participation vectors have been estimated, the associated modal vectors and modal scaling (residues) can be found with standard least squares methods in either the time or the frequency domain. The most common approach is to estimate residues in the frequency domain utilizing residuals, if appropriate:

$$\{H_{pq}(\omega)\}_{N_s \times 1} = \left[\frac{1}{j\omega - \lambda_r}\right]_{N_s \times (2n+2)} \{A_{pqr}\}_{(2n+2) \times 1} \tag{21.97}$$

where N_s = number of spectral lines $\geq 2n + 2$

$$\left[\frac{1}{j\omega - \lambda_r}\right] = \begin{bmatrix} \frac{1}{j\omega_1 - \lambda_1} & \frac{1}{j\omega_1 - \lambda_2} & \frac{1}{j\omega_1 - \lambda_3} & \cdots & \frac{1}{j\omega_1 - \lambda_{2n}} & \frac{-1}{\omega_1^2} & 1 \\ \frac{1}{j\omega_2 - \lambda_1} & \frac{1}{j\omega_2 - \lambda_2} & \frac{1}{j\omega_2 - \lambda_3} & \cdots & \frac{1}{j\omega_2 - \lambda_{2n}} & \frac{-1}{\omega_2^2} & 1 \\ \frac{1}{j\omega_3 - \lambda_1} & \frac{1}{j\omega_3 - \lambda_2} & \frac{1}{j\omega_3 - \lambda_3} & \cdots & \frac{1}{j\omega_3 - \lambda_{2n}} & \frac{-1}{\omega_3^2} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{j\omega_{Ns} - \lambda_1} & \frac{1}{j\omega_{Ns} - \lambda_2} & \frac{1}{j\omega_{Ns} - \lambda_3} & \cdots & \frac{1}{j\omega_{Ns} - \lambda_{2n}} & \frac{-1}{\omega_{Ns}^2} & 1 \end{bmatrix}$$

$$\{A_{pqr}\} = \begin{Bmatrix} A_{pq1} \\ A_{pq2} \\ A_{pq} \\ \cdots \\ A_{pq2n} \\ R_{I_{pq}} \\ R_{F_{pq}} \end{Bmatrix}$$

$$\{H_{pq}(\omega)\} = \begin{Bmatrix} H_{pq}(\omega_1) \\ H_{pq}(\omega_2) \\ H_{pq}(\omega_3) \\ \cdots \\ H_{pq}(\omega_{N_s}) \end{Bmatrix}$$

The above equation is a linear equation in terms of the unknown residues once the modal frequencies are known. Since more frequency information N_s is available from the measured frequency response function than the number of unknowns $2n + 2$, this system of equations is normally solved using the same least squares methods discussed previously. If multiple-input frequency response function data are available, the above equation is modified to find a single set of $2n$ residues representing all of the frequency response functions for the multiple inputs and a single output.

MODAL DATA PRESENTATION/VALIDATION

Once the modal parameters are determined, there are several procedures that allow the modal model to be validated. Some of the procedures that are used are

- Measurement synthesis
- Visual verification (animation)
- Finite element analysis
- Modal vector orthogonality
- Modal vector consistency (modal assurance criterion)
- Modal modification prediction
- Modal complexity
- Modal phase colinearity and mean phase deviation

All of these methods depend upon the evaluation of an assumption concerning the modal model. Unfortunately, the success of the validation method defines only the validity of the assumption; the failure of the modal validation does not generally define what the cause of the problem is.

MEASUREMENT SYNTHESIS

The most common validation procedure is to compare the data synthesized from the modal model with the measured data. This is particularly effective if the measured data are not part of the data used to estimate the modal parameters. This serves as an independent check of the modal parameter estimation process. The visual match can be given a numerical value if a correlation coefficient, similar to coherence, is estimated. The basic assumption is that the measured frequency response function and the synthesized frequency response function should be linearly related (unity) at all frequencies.

Synthesis correlation coefficient (SCC):

$$\mathrm{SCC}_{pq} = \Gamma_{pq}^2 = \frac{\left| \sum_{\omega=\omega_1}^{\omega_2} H_{pq}(\omega)\hat{H}_{pq}^*(\omega) \right|^2}{\sum_{\omega=\omega_1}^{\omega_2} H_{pq}(\omega)H_{pq}^*(\omega) \sum_{\omega=\omega_1}^{\omega_2} \hat{H}_{pq}(\omega)\hat{H}_{pq}^*(\omega)} \tag{21.98}$$

where $H_{pq}(\omega)$ = measured frequency response function
$\hat{H}_{pq}(\omega)$ = synthesized frequency response function

VISUAL VERIFICATION

Another common method of modal model validation is to evaluate the modal vectors visually. While this can be accomplished from plotted modal vectors superimposed upon the undeformed geometry, the modal vectors are normally animated (superimposed upon the undeformed geometry) in order to quickly assess the modal vector. In particular, modal vectors are evaluated for physically realizable characteristics such as discontinuous motion or out-of-phase problems. Often, rigid body modes of vibration are evaluated to determine scaling (calibration) errors or invalid measurement degree-of-freedom assignment or orientation. Naturally, if the system under test is believed to be proportionally damped, the modal vectors should be normal modes, and this characteristic can be quickly observed by viewing an animation of the modal vector.

FINITE ELEMENT ANALYSIS

The results of a finite element analysis of the system under test can provide another method of validating the modal model. While the problem of matching the number of analytical degrees-of-freedom N_a to the number of experimental degrees-of-freedom N_e causes some difficulty, the modal frequencies and modal vectors can be compared visually or through orthogonality or consistency checks. Unfortunately, when the comparison is not sufficiently acceptable, the question of error in the experimental model versus error in the analytical model cannot be easily resolved. Generally, assuming minimal errors and sufficient analysis and test experience, reasonable agreement can be found in the first ten deformable modal vectors, but agreement for higher modal vectors is more difficult. Finite element analysis is discussed in detail in Chap. 28, Part II.

MODAL VECTOR ORTHOGONALITY

Another method that is used to validate an experimental modal model is the weighted orthogonality check. In this case, the experimental modal vectors are used together with a mass matrix normally derived from a finite element model to evaluate orthogonality. The experimental modal vectors are scaled so that the diagonal terms of the modal mass matrix are unity. With this form of scaling, the off-diagonal values in the modal mass matrix are expected to be less than 0.1 (10 percent of the diagonal terms).

Theoretically, for the case of proportional damping, each modal vector of a system is orthogonal to all other modal vectors of that system when weighted by the mass, stiffness, or damping matrix. In practice, these matrices are made available by way of a finite element analysis, and normally the mass matrix is considered to be the most accurate. For this reason, any further discussion of orthogonality is made with respect to mass matrix weighting. As a result, the orthogonality relations can be stated as follows:

Orthogonality of modal vectors:

$$\{\psi_r\}[M]\{\psi_s\} = 0 \qquad r \neq s \tag{21.99}$$

$$\{\psi_r\}[M]\{\psi_s\} = M_r \qquad r = s \tag{21.100}$$

Experimentally, the result of zero for the cross orthogonality [Eq. (21.99)] can rarely be achieved, but values up to one-tenth of the magnitude of the generalized mass of each mode are considered to be acceptable. It is a common procedure to form the modal vectors into a normalized set of mode shape vectors with respect to the mass matrix weighting. The accepted criterion in the aerospace industry, where this confidence check is made most often, is for all of the generalized mass terms to be unity and all cross-orthogonality terms to be less than 0.1. Often, even under this criterion, an attempt is made to adjust the modal vectors so that the cross-orthogonality conditions are satisfied.[36–38]

In Eqs. (21.99) and (21.100) the mass matrix must be an $N_o \times N_o$ matrix corresponding to the measurement locations on the structure. This means that the finite element mass matrix must be modified from whatever size and distribution of grid locations are required in the finite element analysis to the $N_o \times N_o$ square matrix corresponding to the measurement locations. This normally involves some sort of reduction algorithm as well as interpolation of grid locations to match the measurement situation.[39, 40]

When Eq. (21.99) is not sufficiently satisfied, one (or more) of three situations may exist. First, the modal vectors can be invalid. This can be due to measurement error or problems with the modal parameter estimation algorithms. This is a very common assumption and many times contributes to the problem. Second, the mass matrix can be invalid. Since the mass matrix is not easily related to the physical properties of the system, this probably contributes significantly to the problem. Third, the reduction of the mass matrix can be invalid. This can certainly be a realistic problem and cause severe errors. One example of this situation occurs when a relatively large amount of mass is reduced to a measurement location that is highly flexible, such as the center of an unsupported panel. In such a situation the measurement location is weighted very heavily in the orthogonality calculation of Eq. (21.99) but may represent only incidental motion of the overall modal vector.

In all probability, all three situations contribute to the failure of cross-orthogonality criteria on occasion. When the orthogonality conditions are not satisfied, this result does not indicate where the problem originates. From an experimental point of view, it is important to try to develop methods that provide confidence that the modal vector is or is not part of the problem.

MODAL VECTOR CONSISTENCY

Since the residue matrix contains redundant information with respect to a modal vector, the consistency of the estimate of the modal vector under varying conditions

such as excitation location or modal parameter estimation algorithms can be a valuable confidence factor to be utilized in the process of evaluation of the experimental modal vectors.

The common approach to estimation of modal vectors from the frequency response function matrix is to measure a complete row or column of the frequency response function matrix. This gives reasonable definition to those modal vectors that have a nonzero modal coefficient at the excitation location and can be completely uncoupled with the forced normal mode excitation method. When the modal coefficient at the excitation location of a modal vector is zero (very small with respect to the dynamic range of the modal vector) or when the modal vectors cannot be uncoupled, the estimation of the modal vector contains potential bias and variance errors. In such cases, additional rows and/or columns of the frequency response function matrix are measured to detect such potential problems.

In these cases, information in the residue matrix corresponding to each pole of the system is evaluated to determine separate estimates of the same modal vector. This evaluation consists of the calculation of a complex modal scale factor (relating two modal vectors) and a scalar modal assurance criterion (measuring the consistency between two modal vectors). The function of the *modal scale factor* (*MSF*) is to provide a means of normalizing all estimates of the same modal vector. When two modal vectors are scaled similarly, elements of each vector can be averaged (with or without weighting), differenced, or sorted to provide a best estimate of the modal vector or to provide an indication of the type of error vector superimposed on the modal vector. In terms of multiple-reference modal parameter estimation algorithms, the modal scale factor is a normalized estimate of the modal participation factor between two references for a specific mode of vibration. The function of the *modal assurance criterion* (*MAC*) is to provide a measure of consistency between estimates of a modal vector. This provides an additional confidence factor in the evaluation of a modal vector from different excitation locations. The modal assurance criterion also provides a method of determining the degree of causality between estimates of different modal vectors from the same system.[41] The *modal scale factor* is defined, according to this approach, as follows:

$$\mathrm{MSF}_{cdr} = \frac{\{\psi_{cr}\}^H \{\psi_{dr}\}}{\{\psi_{dr}\}^H \{\psi_{dr}\}} \tag{21.101}$$

Equation (21.70) implies that the modal vector d is the reference to which the modal vector c is compared. In the general case, modal vector c can be considered to be made up of two parts. The first part is the part correlated with modal vector d. The second part is the part that is not correlated with modal vector d and includes contamination from other modal vectors and any random contribution. This error vector is considered to be noise. The *modal assurance criterion* is defined as a scalar constant relating the portion of the automoment of the modal vector that is linearly related to the reference modal vector as follows:

$$\mathrm{MAC}_{cdr} = \frac{\left|\{\psi_{cr}\}^H \{\psi_{dr}\}\right|^2}{\{\psi_{cr}\}^H \{\psi_{cr}\}\ \{\psi_{dr}\}^H \{\psi_{dr}\}} = \frac{\left(\{\psi_{cr}\}^H \{\psi_{dr}\}\right)\left(\{\psi_{dr}\}^H \{\psi_{cr}\}\right)}{\{\psi_{cr}\}^H \{\psi_{cr}\}\{\psi_{dr}\}^H \{\psi_{dr}\}} \tag{21.102}$$

The modal assurance criterion is a scalar constant relating the causal relationship between two modal vectors. The constant takes on values from 0, representing no consistent correspondence, to 1, representing a consistent correspondence. In this manner, if the modal vectors under consideration truly exhibit a consistent relation-

ship, the modal assurance criterion should approach unity and the value of the modal scale factor can be considered to be reasonable.

The modal assurance criterion can indicate only consistency, not validity. If the same errors, random or bias, exist in all modal vector estimates, this is not delineated by the modal assurance criterion. Invalid assumptions are normally the cause of this sort of potential error. Even though the modal assurance criterion is unity, the assumptions involving the system or the modal parameter estimation techniques are not necessarily correct. The assumptions may cause consistent errors in all modal vectors under all test conditions verified by the modal assurance criterion.

Coordinate Modal Assurance Criterion (COMAC). An extension of the modal assurance criterion is the *coordinate modal assurance criterion* (*COMAC*).[42] The COMAC attempts to identify which measurement degrees-of-freedom contribute negatively to a low value of MAC. The COMAC is calculated over a set of mode pairs, analytical versus analytical, experimental versus experimental, or experimental versus analytical. The two modal vectors in each mode pair represent the same modal vector, but the set of mode pairs represents all modes of interest in a given frequency range. For two sets of modes that are to be compared, there is a value of COMAC computed for each (measurement) degree-of-freedom.

The coordinate modal assurance criterion (COMAC) is defined as follows:

$$\text{COMAC}_p = \frac{\left| \sum_{r=1}^{N} \psi_{pr}\phi_{pr} \right|^2}{\sum_{r=1}^{N} \psi_{pr}\psi_{pr}^* \sum_{r=1}^{N} \phi_{pr}\phi_{pr}^*} \tag{21.103}$$

where ψ_{pr} = modal coefficient from (measured) degree-of-freedom p and modal vector r from one set of modal vectors
ϕ_{pr} = modal coefficient from (measured) degree-of-freedom p and modal vector r from a second set of modal vectors

The above formulation assumes that there is a match for every mode in the two sets. Only those modes that match between the two sets are included in the computation.

MODAL MODIFICATION PREDICTION

The use of a modal model to predict changes in modal parameters caused by a perturbation (modification) of the system is becoming more of a reality as more measured data are acquired simultaneously. In this validation procedure, a modal model is estimated based upon a complete modal test. This modal model is used as the basis to predict a perturbation to the system that is tested, such as the addition of a mass at a particular point on the structure. Then, the mass is added to the structure and the perturbed system is retested. The predicted and measured data or modal model can be compared and contrasted as a measure of the validity of the underlying modal model.

MODAL COMPLEXITY

Modal complexity is a variation on the use of sensitivity analysis in the validation of a modal model. When a mass is added to a structure, the modal frequencies either should be unaffected or should shift to a slightly lower frequency. Modal overcom-

plexity is a summation of this effect over all measured degrees-of-freedom for each mode. Modal complexity is particularly useful for the case of complex modes in an attempt to quantify whether the mode is genuinely a complex mode, a linear combination of several modes, or a computational artifact. The mode complexity is normally indicated by the *mode overcomplexity value* (*MOV*), which is the percentage of the total number of response points that actually cause the damped natural frequency to decrease when a mass is added. A separate MOV is estimated for each mode of vibration, and the ideal result should be 1.0 (100 percent) for each mode.

MODAL PHASE COLINEARITY AND MEAN PHASE DEVIATION

For proportionally damped systems, the modal coefficients for a specific mode of vibration should differ by 0° or 180°. The *modal phase colinearity* (*MPC*) is an index expressing the consistency of the linear relationship between the real and imaginary parts of each modal coefficient. This concept is essentially the same as the ordinary coherence function with respect to the linear relationship of the frequency response function for different averages or the modal assurance criterion (MAC) with respect to the modal scale factor between modal vectors. The MPC should be 1.0 (100 percent) for a mode that is essentially a normal mode. A low value of MPC indicates a mode that is complex (after normalization) and is an indication of a nonproportionally damped system or errors in the measured data and/or modal parameter estimation.

Another indicator that defines whether a modal vector is essentially a normal mode is the *mean phase deviation* (*MPD*). This index is the statistical variance of the phase angles for each mode shape coefficient for a specific modal vector from the mean value of the phase angle. The MPD is an indication of the phase scatter of a modal vector and should be near 0° for a real, normal mode.

REFERENCES

1. Tse, F. S., I. E. Morse, Jr., and R. T. Hinkle: "Mechanical Vibrations: Theory and Applications," 2d ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1978.
2. Craig, R. R., Jr.: "Structural Dynamics: An Introduction to Computer Methods," John Wiley & Sons, Inc., New York, 1981.
3. Ewins, D.: "Modal Testing: Theory and Practice," John Wiley & Sons, Inc., New York, 1984.
4. Bendat, J. S., and A. G. Piersol: "Random Data: Analysis and Measurement Procedures," 3d ed., John Wiley & Sons, Inc., New York, 2000.
5. Bendat, J. S., and A. G. Piersol: "Engineering Applications of Correlation and Spectral Analysis," 2d ed., John Wiley & Sons, Inc., New York, 1993.
6. Himmelblau, H., A. G. Piersol, J. H. Wise, and M. R. Grundvig: "Handbook for Dynamic Data Acquisition and Analysis," I.E.S. Recommended Practice RP-DTE 012.1, Institute of Environmental Sciences, Mount Prospect, Ill., 1994.
7. Dally, J. W., W. F. Riley, and K. G. McConnell: "Instrumentation for Engineering Measurements," John Wiley & Sons, Inc., New York, 1984.
8. Strang, G.: "Linear Algebra and Its Applications," 3d ed., Harcourt Brace Jovanovich Publishers, San Diego, 1988.
9. Lawson, C. L., and R. J. Hanson: "Solving Least Squares Problems," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974.
10. Jolliffe, I. T.: "Principal Component Analysis," Springer-Verlag, New York, 1986.
11. Allemang, R. J., D. L. Brown, and R. W. Rost: "Dual Input Estimation of Frequency Response Functions for Experimental Modal Analysis of Automotive Structures," *SAE Paper* No. 820193, 1982.

12. Potter, R. W.: *J. Acoust. Soc. Amer.,* **66**(3):776 (1977).
13. Brown, D. L., G. Carbon, and R. D. Zimmerman: "Survey of Excitation Techniques Applicable to the Testing of Automotive Structures," *SAE Paper* No. 770029, 1977.
14. Halvorsen, W. G., and D. L. Brown: *Sound and Vibration,* November 1977, pp. 8–21.
15. Allemang, R. J., D. L. Brown, and W. Fladung: *Proc. Intern. Modal Analysis Conf.,* 1994, p. 501.
16. Brown, D. L., R. J. Allemang, R. D. Zimmerman, and M. Mergeay: "Parameter Estimation Techniques for Modal Analysis," *SAE Paper* No. 790221, *SAE Transactions,* **88**:828 (1979).
17. Vold, H., J. Kundrat, T. Rocklin, and R. Russell: *SAE Transactions,* **91**(1):815 (1982).
18. Vold, H., and T. Rocklin: *Proc. Intern. Modal Analysis Conf.,* 1982, p. 542.
19. Ibrahim, S. R., and E. C. Mikulcik: *Shock and Vibration Bull.,* **47**(4):183 (1977).
20. Fukuzono, K.: "Investigation of Multiple-Reference Ibrahim Time Domain Modal Parameter Estimation Technique," M.S. Thesis, Dept. of Mechanical and Industrial Engineering, University of Cincinnati, 1986.
21. Juang, Jer-Nan, and R. S. Pappa: *AIAA J. Guidance, Control, and Dynamics,* **8**(4):620 (1985).
22. Longman, R. W., and Jer-Nan Juang: *AIAA J. Guidance, Control, and Dynamics,* **12**(5):647 (1989).
23. Zhang, L., H. Kanda, D. L. Brown, and R. J. Allemang: "A Polyreference Frequency Domain Method for Modal Parameter Identification," *ASME Paper* No. 85-DET-106, 1985.
24. Lembregts, F., J. Leuridan, L. Zhang, and H. Kanda: *Proc. Intern. Modal Analysis Conf.,* 1986, pp. 589–598.
25. Lembregts, F., J. L. Leuridan, and H. Van Brussel: *Mech. Systems and Signal Processing,* **4**(1):65 (1989).
26. Coppolino, R. N.: "A Simultaneous Frequency Domain Technique for Estimation of Modal Parameters from Measured Data," *SAE Paper* No. 811046, 1981.
27. Craig, R. R., A. J. Kurdila, and H. M. Kim: *J. Analytical and Experimental Modal Anal.,* **5**(3): 169 (1990).
28. Richardson, M., and D. L. Formenti: *Proc. Intern. Modal Analysis Conf.,* 1982, p. 167.
29. Vold, H., "Orthogonal Polynomials in the Polyreference Method," *Proc. Intern. Seminar on Modal Analysis,* Katholieke University of Leuven, Belgium, 1986.
30. Van der Auweraer, H., and J. Leuridan: *Mechanical Systems and Signal Processing,* **1**(3):259 (1987).
31. Shih, C. Y., Y. G. Tsuei, R. J. Allemang, and D. L. Brown: *Mechanical Systems and Signal Processing,* **2**(4):349 (1988).
32. Shih, C. Y., Y. G. Tsuei, R. J. Allemang, and D. L. Brown: *Mechanical Systems and Signal Processing,* **2**(4):367 (1988).
33. Dippery, K. D., A. W. Phillips, and R. J. Allemang: *Proc. Intern. Modal Analysis Conf.,* 1994.
34. Dippery, K. D., A. W. Phillips, and R. J. Allemang: *Proc. Intern. Modal Analysis Conf.,* 1994.
35. Williams, R., J. Crowley, and H. Vold: *Proc. Intern. Modal Analysis Conf.,* 1985, p. 66.
36. Gravitz, S. I.: *J. Aero/Space Sci.,* **25**:721 (1958).
37. McGrew, J.: *AIAA J.,* **7**(4):774 (1969).
38. Targoff, W. P.: *AIAA J.,* **14**(2):164 (1976).
39. Guyan, R. J.: *AIAA J.,* **3**(2):380 (1965).
40. Irons, B.: *AIAA J.,* **3**(5):961 (1965).
41. Allemang, R. J., and D. L. Brown: *Proc. Intern. Modal Analysis Conf.,* 1982, p. 110.
42. Lieven, N. A. J., and D. J. Ewins: *Proc. Intern. Modal Analysis Conf.,* 1988, p. 690.