

Vibration Testing, With Modal Analysis and Health Monitoring

Joseph C. Slater¹

¹ ©2002 Joseph C. Slater. All rights reserved. This work may not be translated or copied in part or in whole without the written permission of the author, Joseph C. Slater, Wright State University, Dayton, OH 45435, except as allowed by law.

Contents

Preface	ix
1 Review of Fundamentals: Introduction to Vibration Analysis	1
1.1 Application of Complex Exponentials to Solving Linear Differential Equations	2
1.1.1 The Euler Relations	2
1.1.2 The Phasor Representation of a Periodic Function	3
1.1.3 Solution for the Homogeneous (free response) Solution	4
1.1.4 Particular (Steady State) Solution to a Harmonic Excitation	7
1.2 Laplace Transform	10
1.3 Fourier Series	11
1.4 Fourier Transform	16
1.5 Properties of the Fourier/Laplace Transforms	19
1.5.1 Linearity	19
1.5.2 Time Shift	20
1.5.3 Fourier Transform of the Derivative of a Function	20
1.5.4 Fourier Transform of the Integral of a Function	21
1.5.5 Convolution in the Time Domain	21
1.5.6 Convolution in the Frequency Domain	22
1.5.7 Parseval's Formula	22
1.6 Frequency Response Functions	22
1.6.1 Complex Modulus/Stiffness	27
1.7 Response to an Arbitrary Excitation	28
1.7.1 Impulse Response of a Viscously Damped System	28
1.7.2 Duhamel's Integral	31
1.8 Characteristics of Nonlinear Systems	32
1.8.1 Linear "Look"	32
1.8.2 Duffing Equation	33
1.8.3 Duffing Equation With Damping	33
1.8.4 Limit Cycles	33
1.9 Self-Adjoint Multiple Degree of Freedom (MDOF) Systems	33
1.10 Problems	45
2 Introduction to Probability Theory	49
2.1 Random Processes	50
2.2 Probability Distribution and Density Functions	52

2.3	Moment Generating Functions	54
2.4	Transformation of distribution coordinates	56
2.5	Normal (Gaussian) Distribution	58
2.5.1	Joint Distributions	58
2.5.2	Distribution of the Sum of Normal Variables	61
2.5.3	Sample (Calculated) Mean	62
2.6	Chi-square Distribution (χ^2)	64
2.6.1	Sample (Calculated) Variance	66
2.7	Rayleigh Process	66
2.8	Central Limit Theorem	67
2.9	Stationary Signals	67
2.10	Ergodic Signals	69
2.11	Cross Correlation of Signals	69
2.12	Correlation of Finite Time Signals	72
2.13	Detrending data	74
2.13.1	Detrending by FFT Correction (frequency domain)	74
2.13.2	Detrending by Curve Fit Extraction (time domain)	75
2.14	Nonlinear Calibration	76
2.15	Problems	76
3	Introduction to Frequency Domain Random Signal Analysis	79
3.1	Estimation of Spectrum Densities and Frequency Response Functions	79
3.2	Frequency Response Functions for Multiple Input Systems	88
3.3	The Discrete and Fast Fourier Transforms	93
3.4	Aliasing	96
3.5	Decimation of Data and Practical Application of Anti-Aliasing Filters	97
3.6	Multiplexing	99
3.7	Estimation of Correlations Using the Discrete Fourier Transform	99
3.8	Quantization Errors	99
3.9	Accounting for the Impact of Signal Errors: Shifting, Smoothing, and Averaging	99
3.9.1	Bias and Drift	99
3.9.2	Averaging: Time Domain (Ensemble Averaging)	100
3.10	Averaging: Frequency Domain	100
3.10.1	The Welch Method for Estimating Auto- and Cross-Spectrum Densities	101
3.11	Windowing and Leakage	101
3.12	Time-Frequency Analysis	105
3.13	Problems	105
4	Equipment in Vibration Testing	109
4.1	Data Acquisition System Settings	109
4.2	Actuators	110
4.2.1	Impulse Hammers	110
4.2.2	Electrodynamic Shakers	113
4.2.3	Piezoceramic patches	113
4.2.4	Acoustic	113

4.2.5	Hydraulic	113
4.3	Displacement Sensors	113
4.3.1	Laser Vibrometers	113
4.3.2	Potentiometers	113
4.3.3	Linear Variable Differential Transducers (LVDTs)	113
4.3.4	Capacitive Sensors	113
4.3.5	Optical Sensors	113
4.3.6	Videogrammetry	113
4.4	Velocity Sensors	113
4.4.1	Doppler Lasers	113
4.4.2	Tachometers	113
4.5	Accelerometers	113
4.5.1	Mounting Accelerometers	113
4.6	Force Sensors	114
4.6.1	Strain Gages	114
4.6.2	Load Cells	114
4.6.3	PVDF washers	114
4.7	Some Notes On Cables	114
5	Basic Vibration Testing: Survey Tests	117
5.1	Analytical Models for Use in Experiment Design	118
5.1.1	Beams	118
5.1.2	Plates	118
5.1.3	Rods	118
5.1.4	Rigid Masses	118
5.1.5	Back of the Envelope Estimates	118
5.1.6	Finite Element Models	118
5.2	Location, Location, Location: Sensor Placement and Where to Hit the Structure	118
5.3	Obtaining Parameter Estimates from Frequency Response Functions	118
5.3.1	Noise	118
5.3.2	Log Decrement	118
5.3.3	Natural Frequency Estimation	118
5.3.4	Half-Power Point Damping Estimation	118
5.3.5	Nyquist Circle Fitting	124
5.3.6	Some Mistakes and Misunderstandings That Can Ruin an Otherwise Good Test	124
5.4	Obtaining Parameter Estimates From Time Domain Data Response Functions	125
5.4.1	Natural Frequency Estimation	125
5.4.2	Extracting Data When Multiple Frequencies Exist	125
5.5	Estimating the Number of Modes	125
5.5.1	Multivariate Mode Indicator Function	125
5.5.2	Complex Mode Indicator Function	125
6	Multiple Degree of Freedom (MDOF) Systems	127
6.1	Non-proportionally Damped, Gyroscopic and Circulatory Systems	127
6.2	State-Space Analysis	128

6.2.1	Realizations	131
6.2.2	Second Order Modal Coordinates	133
6.3	Visualization of Modes	134
6.3.1	Controllability and Observability	134
6.3.2	Impulse Response of a State-Space Model	136
6.3.3	Frequency Response Functions of State-Space Systems	137
6.3.4	Frequency Response Functions of State-Space Systems in Modal Coordinates.	138
6.4	Discrete Time State-Space Models	139
6.4.1	Solution of the Discrete Equations	140
6.4.2	Markov Parameters	140
6.4.3	Controllability of Discrete-Time Models	141
6.4.4	Observability of Discrete-Time Models	142
6.4.5	Additional Comments on Controllability and Observability	143
6.5	Nonlinearities in MDOF Systems	143
6.6	Continuous Systems	143
6.7	Nonlinearities in Continuous Systems	143
6.8	Problems	143
7	Time Domain System Identification	147
7.1	Complex Exponential Method	147
7.2	Least Squares Complex Exponential Method	147
7.3	Polyreference Complex Exponential Method	147
7.4	Ibrahim Time Domain Method (ITD)	147
7.5	Ho-Kalman System Identification	147
7.6	Eigensystem Realization Algorithm (ERA)	151
7.6.1	Modal Amplitude Coherence	151
7.6.2	Mode Singular Value	151
7.6.3	Mode Stabilization Diagram	151
7.7	Minimum Model Error Method	151
7.8	Autoregressive Moving Average Model (ARMA)	151
7.9	Autoregressive Moving Average with Exogenous Variables (ARMAX)	151
7.10	Autoregressive Moving Average Vector (ARMAV)	151
7.11	Random Decrement Technique	151
7.12	Maximum Entropy Method	151
7.13	Problems	151
8	Frequency Domain Modal Analysis	153
8.1	Curve Fitting Single Degree of Freedom Models	153
8.2	Residuals	154
8.3	Rational Fractional Polynomial Method, SISO	159
8.4	Global Least Squares Fitting	160
8.5	MDOF Curve Fitting - Single Input Collocated with Sensor	161
8.6	Reduction of DOFs by Singular Value Decomposition - Single Input	162
8.7	Reduction of DOFs by Singular Value Decomposition - Multiple Input	163
8.8	Polyreference Method (MIMO)	164

8.9	Polyreference Method in Second Order Form	166
8.10	Free Response Methods	168
8.10.1	Ibrahim Time Domain Method	168
8.10.2	PolyMAX Method	168
8.11	Problems	168
9	Model Correction	169
9.1	Identification of Input forces from System Response	169
9.2	Comparison of Finite Element Models and Identified Models	169
9.2.1	Finite Element Model Reduction Methods	169
9.2.2	Normalized Cross Orthogonality	170
9.2.3	Modal Assurance Criterion (MAC)	171
9.2.4	Normalized Modal Difference	171
9.2.5	Coordinate Modal Assurance Criterion	171
9.2.6	Complex Mode Indicator Function	171
9.3	Model Correction	171
9.3.1	Parameter Corrections	172
9.3.2	Boundary Condition Errors	172
9.3.3	Correcting Theoretical Mode Shapes	172
9.3.4	Correcting the Theoretical Stiffness Matrix	172
9.3.5	Model Updating Using Modal Sensitivities	172
10	Machine Health Monitoring	175
10.1	Time Domain Analysis Methods	176
10.1.1	Kurtosis	176
10.1.2	Time Domain/Ensemble Averaging	178
10.2	Frequency Domain Analysis Methods	179
10.2.1	Auto Spectrum Density	179
10.2.2	Cepstrum	180
10.3	Joint Time Domain/Frequency Domain Analysis Methods	180
10.3.1	Spectral Maps (WaterFall Plots)	180
10.3.2	Campbell Diagram	180
10.3.3	Wavelets	180
10.4	Failure Identification	180
10.4.1	Unbalance	180
10.4.2	Rubbing	180
10.4.3	Misalignment	180
10.4.4	Bearing Faults	180
A	Decibels	181
B	Linear Algebra	183
B.1	Basics of Matrices and Vectors	183
B.2	Multiplication of Vectors and Matrices	184
B.3	Multiplication of Matrices with Matrices	184
B.4	Multiplication of Vectors	185
B.5	Linear Independence and Orthogonality of Vectors	185

B.5.1 Example	186
B.5.2 Example	186
B.6 Multiplication of Block Matrices	186
B.7 Derivatives of Products of Matrices and Vectors	187
B.8 The Eigenvalue Problem	188
B.9 Singular Value Decomposition - Principle Component Analysis	189
B.10 Pseudo-Inverse	191
B.11 Linear Least Squares: Curve Fitting Polynomials	192
B.12 The Cayley-Hamilton Theorem	193
C Variable Definitions	195
D Tables	199
D.1 Normal (Gauss) Distribution	199
D.2 χ^2 Distribution (Chi-square)	200
Bibliography	203

Preface

1

Review of Fundamentals: Introduction to Vibration Analysis

Vibration is the exchange of potential and kinetic energy, usually in some periodic, if complex, fashion, and usually without significant energy dissipation. The simplest representation of such a system with some energy dissipation and an external excitation is the single degree of freedom, or *SDOF*, system of Figure 1.1 represented by the differential equation

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = F(t) \quad (1.1)$$

where m is the mass, c is the linear damping coefficient, k is the linear spring stiffness, and $F(t)$ is the externally applied load. Equation (1.1) can be derived by Newton's law, Lagrange's equation, or a variety of other methods. Solution of this equation is standard fare in introductory differential equations, linear systems, and introductory vibrations courses. Often the emphasis is on assuming a solution in real form. In the following sections we focus on solution using complex exponentials as they lend themselves better to solutions of more complex equations.

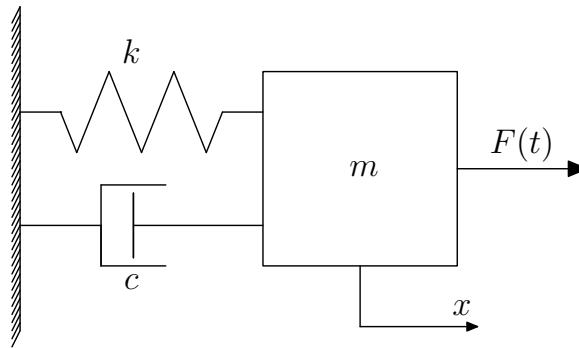


Figure 1.1 Linear forced single degree of freedom system.

1.1 Application of Complex Exponentials to Solving Linear Differential Equations

1.1.1 The Euler Relations

Euler's relation states that

$$e^{\beta j} = \cos(\beta) + j \sin(\beta) \quad (1.2)$$

where $j = \sqrt{-1}$. This is illustrated in Figure 1.2. The vector $e^{\beta j}$ can be plotted in the complex plane and has unit length. From trigonometry one can obtain all necessary relationships between the magnitude, phase (β), the real, and the imaginary parts. Multiplying the complex number by a real constant makes the vector longer without changing its direction (since both the imaginary and real parts would increase proportionally to the constant).

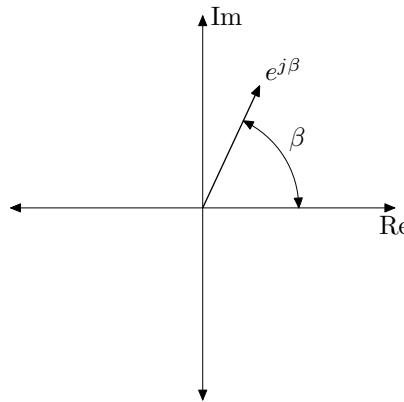


Figure 1.2 Phasor

For example, if we have a complex number, $A^* = A' + jA''$, it can also be represented as $A^* = Ae^{\beta j}$ where $A = \sqrt{A'^2 + A''^2}$ and $\beta = \arctan\left(\frac{A''}{A'}\right)$. Here care must be taken to use a two argument inverse tangent because a positive ratio $\frac{A''}{A'}$ could mean β is either in quadrant 1 or quadrant 3. Likewise, a negative ratio could mean β is in either quadrant 2 or quadrant 4. From basic trigonometry one can also see that $A' = A \cos(\beta)$ and $A'' = A \sin(\beta)$. The convenience of the complex exponential method is derived primarily from the identity $e^{\alpha j}e^{\beta j} = e^{(\alpha+\beta)j}$.

Using Equation (1.2),

$$e^{\gamma j} + e^{-\gamma j} = \cos(\gamma) + j \sin(\gamma) + \cos(-\gamma) + j \sin(-\gamma) \quad (1.3)$$

where γ is some arbitrary angle. Noting that $\cos(-\gamma) = \cos(\gamma)$ and $\sin(-\gamma) = -\sin(\gamma)$, this can be simplified to

$$e^{\gamma j} + e^{-\gamma j} = 2 \cos(\gamma) \quad (1.4)$$

Solving for $\cos(\gamma)$ yields the identity

$$\cos(\gamma) = \frac{e^{\gamma j} + e^{-\gamma j}}{2} \quad (1.5)$$

Alternatively, subtracting the two exponentials yields,

$$e^{\gamma j} - e^{-\gamma j} = \cos(\gamma) + j \sin(\gamma) - \cos(-\gamma) - j \sin(-\gamma) = 2j \sin(\gamma) \quad (1.6)$$

which results in the identity

$$\sin(\gamma) = \frac{e^{\gamma j} - e^{-\gamma j}}{2j} \quad (1.7)$$

1.1.2 The Phasor Representation of a Periodic Function

The advantage of using the Euler relations in vibration analysis is their use in representing periodic functions. Instead of writing a periodic function of time as

$$x(t) = A \cos(\omega t) + B \sin(\omega t) = X \cos(\omega t + \phi), \quad (1.8)$$

where

$$\phi = \arctan\left(\frac{-B}{A}\right) \quad (1.9)$$

we can use equations (1.5) and (1.7) to write it as

$$x(t) = A \frac{e^{\omega t j} + e^{-\omega t j}}{2} + B \frac{e^{\omega t j} - e^{-\omega t j}}{2j} = X^* e^{j\omega t} + \bar{X}^* e^{-j\omega t} = x^*(t) + \bar{x}^*(t) \quad (1.10)$$

Here

$$X^* = \frac{A - B j}{2} \quad (1.11)$$

and \bar{X}^* means complex conjugate of X^* , i.e. $\bar{X}^* = \frac{A + B j}{2}$. The term $x^*(t) = X^* e^{j\omega t}$ is often referred to as a *phasor*. It has a magnitude of $|X| = \frac{1}{2} \sqrt{A^2 + B^2}$ and a phase of $\angle(X^*) = \arctan\left(\frac{-B}{A}\right)$. The derivative of a phasor is

$$\frac{d}{dt} x^*(t) = \frac{d}{dt} X^* e^{j\omega t} = j\omega X^* e^{j\omega t} \quad (1.12)$$

which illustrates the convenience of phasors: the time derivative of a phasor is simply j times the frequency of the phasor (ω for our phasor) times the phasor itself. What this does is turns a calculus operation (a time derivative) into a simple algebraic operation. A derivative increases the amplitude by ω (or decreases the magnitude if $\omega < 1$) and adds a phase of 90° . Performing a time derivative on $x(t)$ of equation (1.8) in real form is a significantly more complex operation. For large complex systems, where many derivatives are involved, the savings from this approach in terms of understanding and simplicity becomes significant. The convenient properties of the exponential also make multiplying and dividing functions written as phasors easy. For example, consider $f^*(t) = F^* e^{j\Omega t}$ multiplied by $x^*(t)$.

$$f^*(t) x^*(t) = F^* e^{j\Omega t} X^* e^{j\omega t} = F^* X^* e^{j(\omega + \Omega)t} \quad (1.13)$$

or consider

$$\frac{x^*(t)}{f^*(t)} = \frac{X^* e^{j\omega t}}{F^* e^{j\Omega t}} = \frac{X^*}{F^*} e^{j(\omega - \Omega)t} \quad (1.14)$$

Both of these calculations are much easier using phasors than their real-functional form counterparts.

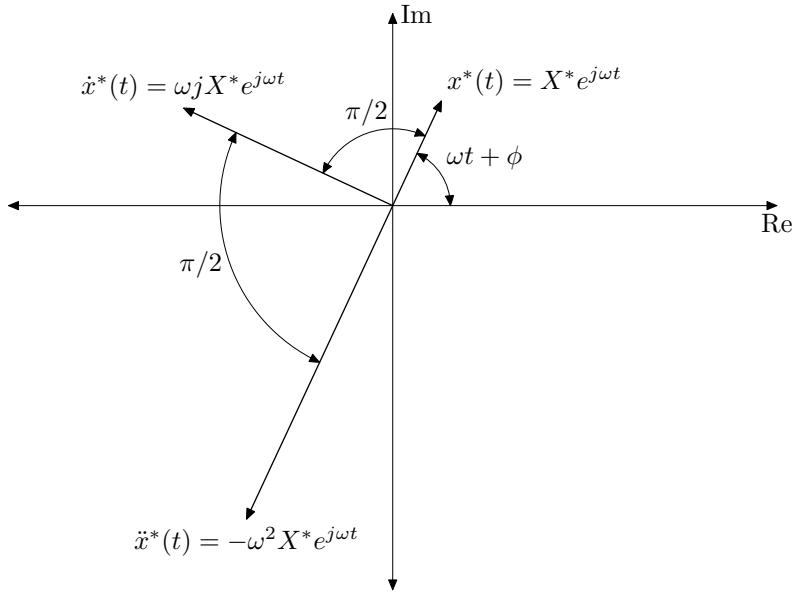


Figure 1.3 Representation of $x(t)$ and its derivatives as phasors.

1.1.3 Solution for the Homogeneous (free response) Solution

Consider the homogeneous equation of motion

$$m\ddot{x} + c\dot{x} + kx = 0 \quad (1.15)$$

The solution can be written as

$$x(t) = X_h e^{\lambda t} \quad (1.16)$$

where the constants X_h and λ need to be determined. Substituting the solution into the equation of motion yields the result

$$(\lambda^2 m + \lambda c + k) X_h e^{\lambda t} = 0 \quad (1.17)$$

Presuming the nontrivial solution by neglecting $X_h e^{\lambda t} = 0$ yields the characteristic equation

$$\lambda^2 m + \lambda c + k = 0 \quad (1.18)$$

the roots of which are

$$\lambda = \frac{-c}{2m} \pm \frac{\sqrt{c^2 - 4mk}}{2m} = -\zeta \omega_n \pm \omega_n \sqrt{1 - \zeta^2} j = -\zeta \omega_n \pm \omega_d j \quad (1.19)$$

In the case of no damping, i.e. $c = 0$, this yields $\lambda = \pm \sqrt{\frac{k}{m}} j$. Noting that two solutions exist

$$x(t) = X_{h1} e^{j\sqrt{\frac{k}{m}}t} + X_{h2} e^{-j\sqrt{\frac{k}{m}}t} = X_{h1} e^{j\omega_n t} + X_{h2} e^{-j\omega_n t} \quad (1.20)$$

where $\omega_n = \sqrt{\frac{k}{m}}$.

This is illustrated in Figure 1.4. The unit phasors $e^{j\omega_n t}$ of Figure 1.4(a) drives the individual solution $x_{h1}(t)$ counter-clockwise in time. The coefficient X^* can be factored into its magnitude, $|X^*|$, and its own constant phasor, $e^{j\phi}$. The phasor results in a constant phase lead (as shown) or lag relative to the time-varying phasor, while the magnitude increases or decreases the apparent length of the phasor representing the solution, $x_{h1}(t)$. When we add the conjugate part of the solution as shown in Figure 1.4(b) the result is twice the real part of either solution, with a phase lead of ϕ .

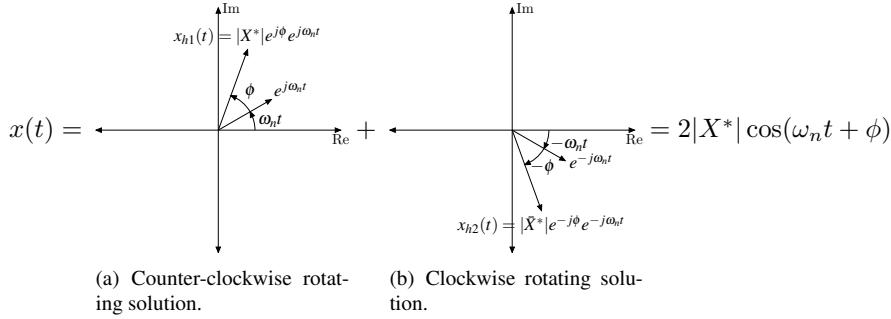


Figure 1.4 Clockwise and counter-clockwise rotating phasor representation of $x(t)$.

Using the notation $X = X' + X''j$ to represent the real and imaginary parts of variables, substituting, and expanding, this can be written, as

$$\begin{aligned} x(t) &= X'_{h1} \cos(\sqrt{\frac{k}{m}}t) + X'_{h1}j \sin(\sqrt{\frac{k}{m}}t) + X''_{h1}j \cos(\sqrt{\frac{k}{m}}t) - X''_{h1} \sin(\sqrt{\frac{k}{m}}t) \\ &\quad + X'_{h2} \cos(\sqrt{\frac{k}{m}}t) - X'_{h2}j \sin(\sqrt{\frac{k}{m}}t) + X''_{h2}j \cos(\sqrt{\frac{k}{m}}t) + X''_{h2} \sin(\sqrt{\frac{k}{m}}t) \end{aligned} \quad (1.21)$$

after using Equation (1.2), $\cos(-\gamma) = \cos(\gamma)$, and $\sin(-\gamma) = -\sin(\gamma)$. Noting that $x(t)$ must be a real function, the imaginary part must be equal to zero yielding

$$\begin{aligned} 0 &= X'_{h1}j \sin(\sqrt{\frac{k}{m}}t) + X''_{h1}j \cos(\sqrt{\frac{k}{m}}t) - X'_{h2}j \sin(\sqrt{\frac{k}{m}}t) + X''_{h2}j \cos(\sqrt{\frac{k}{m}}t) \\ &= (X'_{h1} - X'_{h2}) \sin(\sqrt{\frac{k}{m}}t) + (X''_{h1} + X''_{h2}) \cos(\sqrt{\frac{k}{m}}t) \end{aligned} \quad (1.22)$$

For this expression to be satisfied for all time the coefficients of $\sin(\sqrt{\frac{k}{m}}t)$ and $\cos(\sqrt{\frac{k}{m}}t)$ must be zero, i.e.

$$X'_{h1} = X'_{h2} \quad (1.23)$$

and

$$X''_{h1} = -X''_{h2} \quad (1.24)$$

6 REVIEW OF FUNDAMENTALS: INTRODUCTION TO VIBRATION ANALYSIS

which means that X_{h1} and X_{h2} are complex conjugates ($X_{h1} = \bar{X}_{h2}$). Noting this, Equation (1.21) can be written as

$$\begin{aligned} x(t) &= 2X'_{h1} \cos\left(\sqrt{\frac{k}{m}}t\right) - 2X''_{h1} \sin\left(\sqrt{\frac{k}{m}}t\right) \\ &= A \cos\left(\sqrt{\frac{k}{m}}t\right) + B \sin\left(\sqrt{\frac{k}{m}}t\right) \\ &= C \cos\left(\sqrt{\frac{k}{m}}t + \phi\right) \end{aligned} \quad (1.25)$$

where the third form of the solution can be obtained using trigonometric identities. It is in this form that one notes that the period of oscillation is $\sqrt{\frac{k}{m}}$ which is usually represented as ω_n , the natural frequency. It is often referred to as the undamped natural frequency since in the damped case the frequency of oscillation is different.

In the case of damping, the solution is then given by

$$x(t) = X_{h1}e^{\lambda_1 t} + X_{h2}e^{\lambda_2 t} \quad (1.26)$$

where λ_i , $i = 1, 2$, are given by equation (1.19). Depending upon the value of c , one of the following three cases results: underdamped, overdamped, and critically damped. The level of damping is quantified by the non-dimensional value

$$\zeta = \frac{c}{2\sqrt{km}} \quad (1.27)$$

known as the *damping ratio*.

Overdamped, $\zeta > 1$

If $c > 2\sqrt{km}$ then both roots, λ , are real. As long as $c > 0$, then they are also both negative. This is important as substituting into the solution, Equation (1.26), yields a purely exponential response. If either root is positive, then the corresponding term grows without bound in what is called an unstable response. In the overdamped case, it is not possible for the solution to cross $x(t) = 0$ more than once. No oscillation will be observed, and the coefficients X_{h1} and X_{h2} can be determined by evaluating $x(t)$ and its first time derivative at $t = 0$, and comparing to the initial displacement and velocity.

Critically, $\zeta = 1$

When

$$c = 2\sqrt{km} \quad (1.28)$$

then both roots, λ , are real and equal (i.e. $\lambda_1 = \lambda_2$). This value of c is called *critical damping*, and denoted as c_{cr} . In this case, since both solutions are the same but two are required,¹ the solution becomes

$$x(t) = X_{h1}e^{\lambda t} + X_{h2}te^{\lambda t} \quad (1.29)$$

¹Discussion of why is left to a differential equations text. Pick your favorite, and look up Wronskian.

Again, it is not possible for oscillation to occur and the coefficients X_{h1} and X_{h2} can be determined by evaluating $x(t)$ and its first time derivative at $t = 0$, and comparing to the initial displacement and velocity.

Underdamped, $\zeta < 1$

If $c < 2\sqrt{km}$ then both roots, λ , are complex conjugates. Noting that $\zeta = \frac{c}{2\sqrt{km}}$, and $\omega_n = \sqrt{\frac{k}{m}}$, the roots can be written

$$\lambda = -\zeta\omega_n \pm \omega_n\sqrt{1 - \zeta^2}j \quad (1.30)$$

Substituting for λ_i in Equation (1.26) gives

$$\begin{aligned} x(t) &= X_{h1}e^{\lambda_1 t} + X_{h2}e^{\lambda_2 t} \\ &= X_{h1}e^{(-\zeta\omega_n + \omega_n\sqrt{1 - \zeta^2}j)t} + X_{h2}e^{(-\zeta\omega_n - \omega_n\sqrt{1 - \zeta^2}j)t} \\ &= \left(X_{h1}e^{\omega_n\sqrt{1 - \zeta^2}jt} + X_{h2}e^{-\omega_n\sqrt{1 - \zeta^2}jt} \right) e^{-\zeta\omega_n t} \end{aligned} \quad (1.31)$$

Comparing the coefficient of $e^{-\zeta\omega_n t}$ to Equation (1.25), the total solution can be written

$$x(t) = Ae^{-\zeta\omega_n t} \sin(\omega_n\sqrt{1 - \zeta^2}t - \phi) \quad (1.32)$$

Here the coefficient A and the phase angle ϕ are determined by evaluating $x(t)$ and its first time derivative at $t = 0$, and comparing to the initial displacement and velocity. The quantity $\omega_n\sqrt{1 - \zeta^2}$ is generally referred to as the damped natural frequency and is written as ω_d .

Undamped, $\zeta = 0$

When damping is exceptionally light, expected to be very light, or where peak response at resonance is not a critical factor, the undamped solution

$$x(t) = A \sin(\omega_n t - \phi) \quad (1.33)$$

is often used.

1.1.4 Particular (Steady State) Solution to a Harmonic Excitation

Consider the equation

$$m\ddot{x} + c\dot{x} + kx = Y \cos(\omega_{dr}t) \quad (1.34)$$

The solution can be written as

$$x(t) = X_p \cos(\omega_{dr}t + \phi) \quad (1.35a)$$

or

$$x(t) = A \cos(\omega_{dr}t) + B \sin(\omega_{dr}t) \quad (1.35b)$$

where either the pair X_p and ϕ of solution (1.35a) or the pair A and B of solution (1.35b) must be determined by substituting the assumed form of the solution $x(t)$ into equation (1.34).

8 REVIEW OF FUNDAMENTALS: INTRODUCTION TO VIBRATION ANALYSIS

The solution can be particularly cumbersome, but can be obtained by substituting solution (1.35b) into equation (1.34), solving for the unknown constants A and B , then using the equivalence of the solutions (1.35) along with trigonometric identities to obtain X_p and ϕ . Consider instead the equation

$$m\ddot{x} + c\dot{x} + kx = Y \sin(\omega_{dr}t) \quad (1.36)$$

for which the solution is similarly obtained. An alternative method of solution is to solve both equation (1.34) and equation (1.36) simultaneously. In order to keep track of the two solutions, the solution of equation (1.34) will be referred to as x_1 , while the solution to equation (1.36) will be referred to as x_2 . Multiplying equation (1.36) by $j = \sqrt{-1}$, and adding it to equation (1.34) yields

$$m(\ddot{x}_1 + \ddot{x}_2j) + c(\dot{x}_1 + \dot{x}_2j) + k(x_1 + x_2j) = Y(\cos(\omega_{dr}t) + j \sin(\omega_{dr}t)) \quad (1.37)$$

or more simply as

$$m\ddot{x} + c\dot{x} + kx = Y(\cos(\omega_{dr}t) + j \sin(\omega_{dr}t)) \quad (1.38)$$

where

$$x = x_1 + jx_2 \quad (1.39)$$

Using the Euler relation, the right side of equation (1.38) can be written as

$$Y(\cos(\omega_{dr}t) + j \sin(\omega_{dr}t)) = Ye^{j\omega_{dr}t} \quad (1.40)$$

The particular solution to equation (1.38) is

$$x(t) = X_p^* e^{j\omega_{dr}t} \quad (1.41)$$

where the unknown constant coefficient X_p^* is complex (as the solution will demonstrate).

Using $\frac{d}{dt}e^{j\omega_{dr}t} = j\omega_{dr}e^{j\omega_{dr}t}$ and $\frac{d^2}{dt^2}e^{j\omega_{dr}t} = -\omega_{dr}^2e^{j\omega_{dr}t}$, then substituting (1.41) into equation (1.38) yields

$$mX_p^*(-\omega_{dr}^2)e^{j\omega_{dr}t} + cX_p^*j\omega_{dr}e^{j\omega_{dr}t} + kX_p^*e^{j\omega_{dr}t} = Ye^{j\omega_{dr}t} \quad (1.42)$$

factoring out $e^{j\omega_{dr}t}$ and solving for X_p^* yields

$$X_p^* = \frac{Y}{-\omega_{dr}^2m + j\omega_{dr}c + k} = \frac{(k - m\omega_{dr}^2) - j\omega_{dr}c}{(k - m\omega_{dr}^2)^2 + (c\omega_{dr})^2} Y = X_p e^{j\theta} \quad (1.43)$$

where X_p^* has been expressed in complex form with a magnitude X_p and a phase $\theta = \arctan\left(\frac{-c\omega_{dr}}{k - m\omega_{dr}^2}\right)$ using Euler's equation. Note that θ found here is identical to ϕ found from solving equation (1.34). Substituting for X_p in equation (1.41) gives

$$\begin{aligned} x(t) &= X_p e^{j\theta} e^{j\omega_{dr}t} \\ &= X_p e^{j(\omega_{dr}t + \theta)} \\ &= X_p (\cos(\omega_{dr}t + \theta) + j \sin(\omega_{dr}t + \theta)) \end{aligned} \quad (1.44)$$

Recalling equation (1.39) yields

$$x_1 = X_p \cos(\omega_{dr}t + \theta) \quad (1.45a)$$

$$x_2 = X_p \sin(\omega_{dr}t + \theta) \quad (1.45b)$$

Thus the solution to equation (1.34), represented by equation (1.45a) can more easily be obtained by:

1. Assuming the excitation is in the form of a complex exponential.
2. Assume the response is in the form of equation (1.41).
3. Substitute into the equation of motion. For this example, the equation of motion is given by equation (1.34), (1.36), or (1.38). Your equation will likely be different.
4. Obtain the amplitude of the response, X_p , and the phase, θ by factoring out $e^{j\omega_{dr}t}$ from each term and solving for X_p^* .
5. The solution will have the same trigonometric function as the excitation, will have an amplitude X_p , and a relative phase of θ .

Alternatively, a solution to Equation (1.34) can be obtained by using Equation (1.5). Substituting $\cos(\omega_{dr}t) = (e^{\omega_{dr}tj} + e^{-\omega_{dr}tj})/2$ into Equation (1.34) yields

$$m\ddot{x} + c\dot{x} + kx = Y \left(\frac{e^{\omega_{dr}tj} + e^{-\omega_{dr}tj}}{2} \right) = \frac{Y}{2}e^{\omega_{dr}tj} + \frac{Y}{2}e^{-\omega_{dr}tj} \quad (1.46)$$

for which the solution is

$$x_p(t) = \frac{1}{2}X_p^*e^{\omega_{dr}tj} + \frac{1}{2}\bar{X}_p^*e^{-\omega_{dr}tj} \quad (1.47)$$

where \bar{X}_p^* means complex conjugate of X_p^* , and X_p^* is given by Equation (1.43). Introducing the angle $\phi = \arctan\left(\frac{X_p''}{X_p'}\right)$,

$$\begin{aligned} x_p(t) &= \frac{1}{2}X_p e^{\phi j} e^{\omega_{dr}tj} + \frac{1}{2}X_p e^{-\phi j} e^{-\omega_{dr}tj} \\ &= \frac{1}{2}X_p e^{(\omega_{dr}t+\phi)j} + \frac{1}{2}X_p e^{-(\omega_{dr}t+\phi)j} \\ &= X_p \cos(\omega_{dr}t + \phi) \end{aligned} \quad (1.48)$$

Example 1.1.1 Consider a single degree of freedom system attached to a sinusoidally moving base via a spring and a dashpot. The governing equation of motion for the system is

$$m\ddot{x} + c\dot{x} + kx = ky(t) + cy(t)$$

where $y(t) = Y \cos(\omega_{dr}t)$ is the motion of the base. Find $x(t)$ as a function of m , c , k , and Y .

Solution:

If we instead use $y(t) = Ye^{\omega_{dr}tj} = Y(\cos(\omega_{dr}t) + j\sin(\omega_{dr}t))$, we can write the equation of motion as

$$m\ddot{x} + c\dot{x} + kx = Ye^{j\omega_{dr}t}(k + cj\omega_{dr})$$

Assuming a form of the solution $x(t) = Xe^{j\omega_{dr}t}$, and substituting into the equation of motion, gives

$$X = Y \frac{k + cj\omega_{dr}}{k - m\omega_{dr}^2 + cj\omega_{dr}} = Y \frac{k^2 + c^2\omega_{dr}^2 - km\omega_{dr}^2 - mc\omega_{dr}^3 j}{(k - m\omega_{dr}^2)^2 + c^2\omega_{dr}^2}$$

The magnitude and phase of the response are $|X|$ and $\angle X$ respectively, and the solution for $x(t)$ can now be written as

$$x(t) = Y \left(\frac{k^2 + (c\omega_{dr})^2}{(k - m\omega_{dr}^2)^2 + (c\omega_{dr})^2} \right)^{\frac{1}{2}} \cos(\omega_{dr}t + \phi)$$

where

$$\phi = \arctan\left(\frac{c\omega_{dr}}{k}\right) - \arctan\left(\frac{c\omega_{dr}}{k - m\omega_{dr}^2}\right) = \arctan\left(\frac{-mc\omega_{dr}^3}{k^2 + c^2\omega_{dr}^2 - km\omega_{dr}^2}\right)$$

1.2 Laplace Transform

A convenient way of solving linear differential equations is the *Laplace transform*. It does this by transforming differential equations into algebraic ones in the *s-domain*, which in turn can be solved using algebraic tools. The solution is then transformed back into the time domain. The one-sided Laplace transform of a function $x(t)$ is defined as

$$X(s) = \mathcal{L}\{x(t)\} = \int_0^\infty x(t)e^{-st}dt \quad (1.49)$$

where s is a complex variable. It should be noted that there is also a two-sided Laplace transform defined as

$$X(s) = \mathcal{L}\{x(t)\} = \int_{-\infty}^\infty x(t)e^{-st}dt \quad (1.50)$$

The form of equation (1.49), called the *one-sided Laplace transform* is commonly used in solving differential equations because we cannot possibly know the force, displacement, or its derivatives from $t = -\infty$. However, in section 2.3 we will see that the two-sided Laplace transform is useful in calculating means, variances, and other statistical properties.

Using the Laplace transform method, it is presumed that the current states, or initial conditions, are known, and thus integrating before $t = 0$ is unnecessary. To demonstrate how a differential equation is turned into an algebraic one, it is instructive to consider the Laplace transform of $\dot{x}(t)$. Applying the definition of the Laplace transform, and using integration by parts,

$$\begin{aligned} \mathcal{L}\left\{\frac{dx(t)}{dt}\right\} &= \int_0^\infty \frac{dx(t)}{dt} e^{-st} dt \\ &= x(t)e^{-st} \Big|_{t=0}^{t=\infty} - \int_0^\infty x(t)(-s)e^{-st} dt \\ &= -x(0) + s\mathcal{L}\{x(t)\} \end{aligned} \quad (1.51)$$

Thus, applying the Laplace transform to the single degree of freedom forced system gives

$$\begin{aligned} \mathcal{L}\{m\ddot{x} + c\dot{x} + kx\} &= \mathcal{L}\{f(t)\} \\ m(s^2X(s) - sx(0) - \dot{x}(0)) + c(sX(s) - x(0)) + kX(s) &= F(s) \end{aligned} \quad (1.52)$$

The solution for $x(t)$ is then obtained by taking the inverse of $X(s)$, or

$$X(s) = \frac{F(s) + msx(0) + m\dot{x}(0) + cx(0)}{ms^2 + cs + k} \quad (1.53)$$

If we ignore the transient response by setting $x(0) = \dot{x}(0) = 0$ we can solve for the *transfer function*, $T(s)$,

$$T(s) = \frac{X(s)}{F(s)} = \frac{1}{ms^2 + cs + k} \quad (1.54)$$

Table 1.1 Functions and their Laplace transforms

$y(t)$	$Y(s) = \mathcal{L}\{y(t)\}$
1	$\frac{1}{s}$
t	$\frac{1}{s^2}$
t^2	$\frac{1}{s^3}$
$\frac{t^n}{n!}$	$\frac{1}{s^{n+1}}$
$e^{-\zeta\omega t}$	$\frac{1}{s+\zeta\omega}$
$\sum_{i=1}^{\infty} \delta(t - iT)$	$\frac{1}{1 - e^{-Ts}}$
$\sin(\omega t)$	$\frac{\omega}{s^2 + \omega^2}$
$\cos(\omega t)$	$\frac{s}{s^2 + \omega^2}$
$e^{-\zeta\omega t} \sin(\omega t)$	$\frac{\omega}{(s + \zeta\omega)^2 + \omega^2}$
$e^{-\zeta\omega t} \cos(\omega t)$	$\frac{s + \zeta\omega}{(s + \zeta\omega)^2 + \omega^2}$
$\frac{-\omega_n^2}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2} t - \theta)$, $\theta = \cos^{-1}(\zeta)$, $(\zeta < 1)$	$\frac{s\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2}$

1.3 Fourier Series

Many excitation forces are not harmonic, but instead are a combination of harmonic excitations. In fact, over a finite period of time any excitation can be represented by a combination of harmonic excitations. This fact will be used when we perform signal processing. Applying Fourier Series analysis to a periodic signal demonstrates that a random signal can indeed be represented by a series of harmonic functions. Consider a function $f(t)$ with a periodicity T . Assume that an approximation for that function can be represented by

$$f(t) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \right) \quad (1.55)$$

12 REVIEW OF FUNDAMENTALS: INTRODUCTION TO VIBRATION ANALYSIS

where n is an integer. Integrating both sides over one period gives

$$a_0 = \frac{1}{T} \int_0^T f(t) dt \quad (1.56)$$

Multiplying both sides by $\cos\left(\frac{2\pi mt}{T}\right)$ where m is also an integer gives

$$\begin{aligned} f(t) \cos\left(\frac{2\pi mt}{T}\right) &= a_0 \cos\left(\frac{2\pi mt}{T}\right) \\ &+ \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{2\pi nt}{T}\right) \cos\left(\frac{2\pi mt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \cos\left(\frac{2\pi mt}{T}\right) \right) \end{aligned} \quad (1.57)$$

If both sides are integrated over any period T , say $0 < t < T$, only the term $n = m$ will remain on the right hand side. This is due to the orthogonality properties of sines and cosines. Any sine wave multiplied by any cosine wave and integrated over a time that is an integer multiple of the period of both functions will yield an answer of zero. In addition, any sine wave multiplied by any other sine wave and integrated over a time that is an integer multiple of the period of both functions will yield an answer of zero. This statement also holds true for two cosine waves *except when they both have the same period*. This property is easily referenced in a calculus text, or can be inferred by reason by a) sketching two such trig functions, b) sketching them multiplied by one another, c) marking and noting corresponding positive and negative regions, d) noting that there is as much negative area as there is positive. Also note that the first term, a_0 , also drops out of the expression because the integral of a sine or cosine function over any number of complete periods yields zero. Many engineers tend to have forgotten how to integrate the non-zero case, so here is a simple method for performing it. Consider the following:

$$M = \int_0^T \cos^2\left(\frac{2\pi mt}{T}\right) dt \quad (1.58)$$

Since the integral is over a number of complete cycles, we note that we would obtain the same result if the cosines had been sines, i.e.

$$M = \int_0^T \cos^2\left(\frac{2\pi mt}{T}\right) dt = \int_0^T \sin^2\left(\frac{2\pi mt}{T}\right) dt \quad (1.59)$$

Consider the sum of the two integrals, and noting that they are equal

$$\begin{aligned} &\int_0^T \cos^2\left(\frac{2\pi mt}{T}\right) dt + \int_0^T \sin^2\left(\frac{2\pi mt}{T}\right) dt \\ &= \int_0^T \cos^2\left(\frac{2\pi mt}{T}\right) dt + \int_0^T \cos^2\left(\frac{2\pi mt}{T}\right) dt \\ &= M + M \end{aligned} \quad (1.60)$$

Also, noting that

$$\begin{aligned}
 & \int_0^T \cos^2 \left(\frac{2\pi mt}{T} \right) dt + \int_0^T \sin^2 \left(\frac{2\pi mt}{T} \right) dt \\
 &= \int_0^T \cos^2 \left(\frac{2\pi mt}{T} \right) + \sin^2 \left(\frac{2\pi mt}{T} \right) dt \\
 &= \int_0^T 1 dt = T
 \end{aligned} \tag{1.61}$$

therefore

$$M = \int_0^T \cos^2 \left(\frac{2\pi mt}{T} \right) dt = \frac{T}{2} \tag{1.62}$$

Thus evaluating equation (1.57) integrated over $0 < t < T$ for $m \neq 0$ yields

$$a_n = \frac{2}{T} \int_0^T f(t) \cos \left(\frac{2\pi nt}{T} \right) dt, n = 0, \dots, \infty \tag{1.63}$$

Likewise

$$b_n = \frac{2}{T} \int_0^T f(t) \sin \left(\frac{2\pi nt}{T} \right) dt, n = 0, \dots, \infty \tag{1.64}$$

This is the traditional form of the Fourier series as generally taught in introductory vibrations texts. Consider another form of the Fourier series where instead of using trigonometric functions we take advantage of Euler's relations and write

$$f(t) = \sum_{n=-\infty}^{\infty} F_n e^{j2\pi nt/T} \tag{1.65}$$

Note that the limits have been changed in the summation. The reason for this will become apparent later. Following the development of the sine based Fourier series, we will multiply both sides of equation (1.65) by $e^{-j2\pi mt/T}$ and integrate from $0 < t < T$ yielding

$$\int_0^T f(t) e^{-j2\pi mt/T} dt = \int_0^T \sum_{n=-\infty}^{\infty} F_n e^{j2\pi nt/T} e^{-j2\pi mt/T} dt = \int_0^T \sum_{n=-\infty}^{\infty} F_n e^{j2\pi(n-m)t/T} dt \tag{1.66}$$

Applying the Euler equation (1.2)

$$\int_0^T f(t) e^{-j2\pi mt/T} dt = \int_0^T \sum_{n=-\infty}^{\infty} F_n (\cos(2\pi(n-m)t/T) + j \sin(2\pi(n-m)t/T)) dt \tag{1.67}$$

After integration, all terms on the right hand side are zero except that of $n = m$ resulting in

$$\int_0^T f(t) e^{-j2\pi nt/T} dt = \int_0^T F_n 1 dt = F_n T \tag{1.68}$$

therefore

$$F_n = \frac{1}{T} \int_0^T f(t) e^{-j2\pi nt/T} dt \tag{1.69}$$

14 REVIEW OF FUNDAMENTALS: INTRODUCTION TO VIBRATION ANALYSIS

Applying Euler's equation to this expression yields

$$F'_n + jF''_n = \frac{1}{T} \left(\int_0^T f(t) \cos(2\pi nt/T) dt - j \int_0^T f(t) \sin(2\pi nt/T) dt \right) \quad (1.70)$$

where F' is the real part of F , and F'' is the imaginary part. It is clear then that

$$F_n = F'_n + jF''_n = \frac{a_n}{2} - j \frac{b_n}{2} \quad (1.71)$$

Also, noting equation (1.70),

$$F_{-n} = \frac{a_n}{2} + j \frac{b_n}{2} = \bar{F}_n \quad (1.72)$$

where \bar{F}_n represents the complex conjugate of F_n .

Example 1.3.1 Find the Fourier series of the function $x(t)$ with a period of 4 seconds where

$$x(t) = \begin{cases} 1, & 0 < t < 2 \\ 0, & 2 < t < 4 \\ \vdots & \end{cases}$$

Solution:

Substituting $x(t)$ into equation (1.69) yields

$$\begin{aligned} X_n &= \frac{1}{4} \int_0^2 1 e^{-j2\pi nt/4} dt + \frac{1}{4} \int_2^4 0 e^{-j2\pi nt/4} dt \\ &= \frac{1}{4} \frac{-4}{2\pi nj} e^{-jn\pi t/2} \Big|_0^2 \\ &= \frac{j}{2n\pi} (e^{-jn\pi} - 1) \\ &= \begin{cases} \frac{1}{2} & n = 0 \\ \frac{-j}{\pi n} & n \text{ odd} \\ 0 & n \text{ even} \end{cases} \end{aligned}$$

Note that in obtaining X_n for $n = 0$, $n = 0$ must be substituted prior to integrating to avoid a divide by zero error. Substituting X_n into equation (1.65), applying Euler's equation, and simplifying gives

$$\begin{aligned} x(t) &= \frac{1}{2} + \sum_{n=-\infty}^{\infty} \frac{-j}{n\pi} e^{j2\pi nt/4}, \quad n = 1, 3, 5, \dots \\ &= \frac{1}{2} + \sum_{n=1}^{\infty} \frac{2}{\pi n} \sin\left(\frac{n\pi t}{2}\right), \quad n = 1, 3, 5, \dots \end{aligned}$$

A short cut could have been taken by noting equation (1.71) and directly writing the Fourier expansion in real form. Figure 1.5 illustrates the Gibbs effect. The Gibbs effect is the observation that Fourier series do not perform well at sharp corners due to inevitable truncation

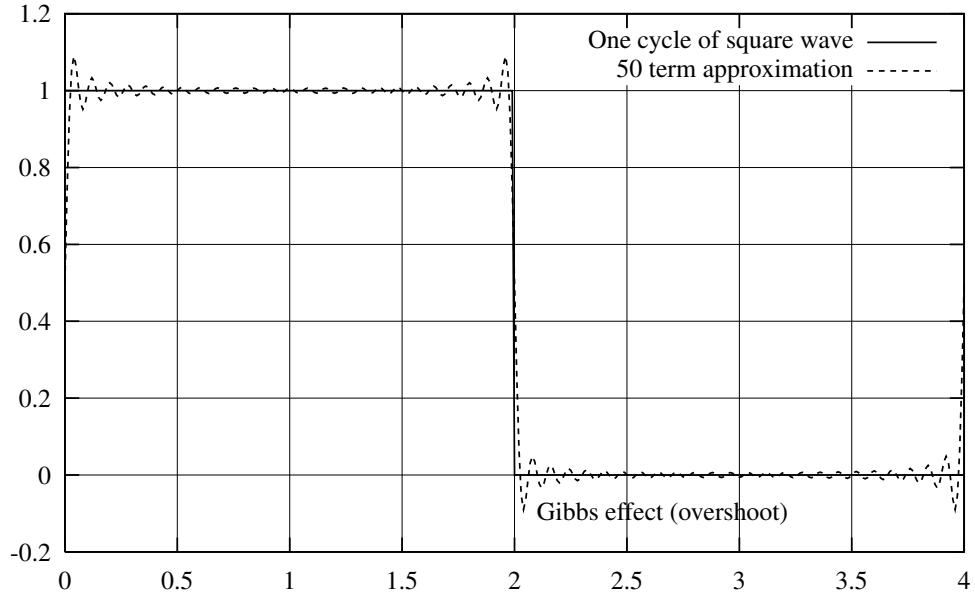


Figure 1.5 50-term Fourier series representation of a square wave.

errors. Certainly, additional terms will improve the approximation, but later in the text when we discuss the practical aspects of signal processing real limitations on the maximum frequency available exist.

Example 1.3.2 Find the Fourier series of²

$$f(t) = \sum_{m=-\infty}^{\infty} \delta(t - m)$$

where

$$\delta(t) = 0, \quad t \neq m \quad (1.73)$$

and

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (1.74)$$

The period of the repeated function is $T = 1$. Using equation (1.69) yields

$$F_n = \frac{1}{1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \delta(t) e^{-j2\pi nt} dt = e^0 = 1$$

Then, substituting into equation (1.65)

$$f(t) = \sum_{n=-\infty}^{\infty} e^{j2\pi nt}$$

²This represents an impulse occurring once per second. Periodic impulses are characteristic of certain types of failures in bearings.

For the term $n = 0$, $e^{j2\pi 0t} = 1$ and for terms $\pm n$, $e^{j2\pi nt} + e^{j2\pi(-n)t} = 2 \cos 2\pi nt$

$$\therefore f(t) = 1 + \sum_{n=1}^{\infty} 2 \cos 2\pi nt$$

The advantages of this form of the Fourier series are twofold: first, the integrations tend to be much easier for those who don't like "difficult" integrations, and second, it is directly related to the *discrete Fourier transform* (DFT) that is the mainstay of signal processing. Without the DFT and its more expedient algorithm, the *fast Fourier transform* (FFT), modern vibration testing as we know it wouldn't exist.

1.4 Fourier Transform

Consider substituting equation (1.69) into equation (1.65) yielding

$$f(t) = \sum_{-\infty}^{\infty} \left(\frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-j2\pi mt/T} dt \right) e^{j2\pi nt/T} \quad (1.75)$$

Now $1/T = \Delta f = \Delta\omega/2\pi$ is the change in frequency from one term to the next in the Fourier series, and likewise $n/T = f$ from equation (1.55). Be careful to note that f is used here for frequency, while $f(t)$ refers to a time signal(a common notation for force). If we consider applying this relation to a function that doesn't repeat, we must then consider $T \rightarrow \infty$. As $T \rightarrow \infty$, $\Delta\omega \rightarrow d\omega$ and $n/T \rightarrow \omega/2\pi$. In performing the integral, we also can't forget the earlier non-repeating part of the function prior to $t = 0$ (there is no well-defined time $t = 0$). We then change the limits of the integral to be from $-\infty$ to ∞ , and replace the summation by an integral over $d\omega$ giving

$$f(t) = \int_{-\infty}^{\infty} \left(\frac{d\omega}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \right) e^{j\omega t} \quad (1.76)$$

or

$$f(t) = \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \right) e^{j\omega t} d\omega \quad (1.77)$$

There are multiple ways to parse this expression. In mechanical engineering, the term $1/2\pi$ is traditionally factored to the front of the expression leaving

$$f(t) = \mathcal{F}^{-1} \{F(j\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega) e^{j\omega t} d\omega \quad (1.78)$$

where

$$F(j\omega) = \mathcal{F} \{f(t)\} = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (1.79)$$

Here equation (1.79)³ is commonly called the *forwards Fourier transform* (commonly abbreviated to *Fourier Transform*), and equation (1.78) is commonly referred to as the *backward*

³For users of Mathematica, the default behavior does not match these definitions (it instead matches definitions common to physics). The options {a,b} should be set to {1,-1}. See the Mathematica manual for further information

or *inverse Fourier transform*. Together they form the *Fourier transform pair*. The term $1/2\pi$ could have been placed in either expression, as it is in many other fields, and the pair of expressions would have been equivalent. While there is no fixed convention in vibration analysis, this form of the expressions is the predominant form. An argument can be made for the rationale of this form in that the forward Fourier transforms represented by equation (1.79) and (1.78) can be written as

$$f(t) = \int_{-\infty}^{\infty} F(f)e^{j2\pi ft} df \quad (1.80)$$

and

$$F(f) = \int_{-\infty}^{\infty} f(t)e^{-j2\pi ft} dt \quad (1.81)$$

simply by recognizing that $f = \omega/2\pi$. Thus in using equations (1.81) and (1.80) there is no extraneous 2π to cause confusion.

The Fourier transform of a function exists for most, but not all, functions. A sufficient, but not necessary, condition is that the function satisfy the *Dirichlet conditions*. To satisfy the Dirichlet conditions a function must:

- be absolutely integrable: i.e. $\int_{-\infty}^{\infty} |f(t)| dt < \infty$, and
- have a finite number of discontinuities and a finite number of minima and maxima in any arbitrary time interval.

The end result is that almost all practical engineering functions satisfy these conditions. Two important exceptions are the sine and cosine because they are not absolutely integrable. However, the Dirichlet conditions are sufficient, not necessary, and we will see how these too are Fourier transformable.

Example 1.4.1 Find the Fourier Transform of the Dirac delta function $\delta(t)$.

Solution:

Recall that by definition

$$\delta(t) = 0, \quad t \neq 0 \quad (1.82)$$

and

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (1.83)$$

The Fourier Transform of $\delta(t)$ is obtained using (1.79).

$$\Delta(\omega) = \int_{-\infty}^{\infty} \delta(t)e^{-j\omega t} dt \quad (1.84)$$

Noting that $\delta(t) = 0$ for $t \neq 0$, the limits of the integral can equivalently be changed to arbitrary non-zero values, say $t = \pm b$.

$$\Delta(\omega) = \int_{-b}^b \delta(t)e^{-j\omega t} dt \quad (1.85)$$

Integrating by parts yields

$$\Delta(\omega) = u(t)e^{-j\omega t} \Big|_{-b}^b - \int_{-b}^b u(t)(-j\omega)e^{-j\omega t} dt \quad (1.86)$$

where $u(t)$ is the unit step function (more formally referred to as the Heaviside function). Evaluating the first term and breaking the integral into two parts,

$$\Delta(\omega) = e^{-j\omega b} + \int_{-b}^0 u(t)(j\omega)e^{-j\omega t} dt + \int_0^b u(t)(j\omega)e^{-j\omega t} dt \quad (1.87)$$

The first integral is zero, since $u(t) = 0$ for $t < 0$. Over the interval $0 < t < b$, $u(t) = 1$. Therefore $\Delta(\omega)$ becomes

$$\Delta(\omega) = e^{-j\omega b} - e^{-j\omega t} \Big|_0^b = e^{-j\omega b} - (e^{j\omega b} - e^0) = 1 \quad (1.88)$$

Since the value of $\Delta(\omega)$ cannot depend on b , $\Delta(\omega) = 1$. A better argument is that since the integrand is equal to 0 for $t \ll b$, evaluation of the upper limit cannot be of consequence, and should never even have been performed. Integration by parts didn't clear up the solution process, but instead muddled things up. There is seemingly no value of b that doesn't show up in the solution. There are a number of other seemingly unsatisfactory explanations for the second term of (1.88). This problem will be revisited when we consider the problem of leakage in discrete Fourier transforms. In the continuous case, there is a simple solution. Considering the definition of the Dirac delta function given by equations (1.82) and (1.83), since $e^{-j\omega t}$ is relatively slow varying as compared to $\delta(t)$, we can consider it to be constant over the entire period of time that $\delta(t) \neq 0$. This allows us to factor it outside of the integral and evaluate it at $t = 0$. Thus equation (1.84) becomes

$$\Delta(\omega) = e^{-j\omega 0} \int_{-\infty}^{\infty} \delta(t) dt = 1 \int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (1.89)$$

Thus we get the same result with much less effort. This approach can be applied in general when a Dirac delta function is inside the integrand so that

$$\int_{\tau-b}^{\tau+b} \delta(t-\tau) f(t) dt = f(\tau) \quad (1.90)$$

as long as $f(t)$ is bounded and is often referred to as the sifting (Chen 1994) property of the impulse because it effectively sifts out the value of $f(t)$ at $t = \tau$.

Example 1.4.2 Find the Fourier transform of $f(t) = \sin(\Omega t)$.

Solution:

Following the steps of the previous problem, the Fourier Transform of $\sin(\Omega t)$ is obtained using (1.79).

$$F(\omega) = \int_{-\infty}^{\infty} \sin(\Omega t) e^{-j\omega t} dt \quad (1.91)$$

In this case, it's easier to apply the inverse Fourier transform than the forward Fourier transform, except that we don't yet know the answer. Let's first then consider a guess, $\tilde{F}(\omega) = \delta(\omega - \Omega)$. If this is correct, we will get $\tilde{f}(t) = \sin(\Omega t)$

$$\tilde{f}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\omega - \Omega) e^{j\omega t} d\omega = \frac{1}{2\pi} e^{j\Omega t} = \frac{1}{2\pi} (\cos(\Omega t) + j \sin(\Omega t)) \quad (1.92)$$

This doesn't give us the correct answer. However, since $f(t) = \frac{2\pi}{2j} (\tilde{f}(t) - \tilde{f}(-t))$

$$\begin{aligned}
f(t) &= \frac{2\pi}{2j} \left(\tilde{f}(t) - \tilde{f}(-t) \right) \\
&= \frac{2\pi}{2j} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\omega - \Omega) e^{j\omega t} d\omega - \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\omega - \Omega) e^{-j\omega t} d\omega \right)
\end{aligned} \tag{1.93}$$

Noting that

$$\int_{-\infty}^{\infty} \delta(\omega - \Omega) e^{-j\omega t} d\omega = \int_{-\infty}^{\infty} \delta(\omega + \Omega) e^{j\omega t} d\omega \tag{1.94}$$

then the second integral can be replaced and $f(t)$ can be simplified to

$$\begin{aligned}
f(t) &= \frac{1}{2\pi} \left(\frac{\pi}{j} \int_{-\infty}^{\infty} \delta(\omega - \Omega) e^{j\omega t} d\omega - \frac{\pi}{j} \int_{-\infty}^{\infty} \delta(\omega + \Omega) e^{j\omega t} d\omega \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \underbrace{\frac{\pi}{j} (\delta(\omega - \Omega) - \delta(\omega + \Omega)) e^{j\omega t} d\omega}_{F(j\omega) = \mathcal{F}\{f(t)\}}
\end{aligned} \tag{1.95}$$

Thus

$$\mathcal{F}\{f(t)\} = \frac{\pi}{j} (\delta(\omega - \Omega) - \delta(\omega + \Omega)) \tag{1.96}$$

is the Fourier transform of $\sin(\Omega t)$.

1.5 Properties of the Fourier/Laplace Transforms

(Weaver et al. 1990) The power of the Fourier transform is not only that it adds insight into vibration problems, but also that it provides significant mathematical simplifications. The following properties, illustrated for the Fourier Transform, also apply to the two-sided Laplace Transform, with s being substituted for $j\omega$.

1.5.1 Linearity

The Fourier transform is a linear operation. This is illustrated by considering the Fourier transform of the sum of two functions

$$\begin{aligned}
\mathcal{F}\{ax_1(t) + bx_2(t)\} &= \int_{-\infty}^{\infty} (ax_1(t) + bx_2(t)) e^{-j\omega t} dt \\
&= a \int_{-\infty}^{\infty} x_1(t) d\omega + b \int_{-\infty}^{\infty} x_2(t) d\omega \\
&= aX_1(j\omega) + bX_2(j\omega)
\end{aligned} \tag{1.97}$$

1.5.2 Time Shift

The Fourier transform of a time shifted function is the same function with a frequency adjusted phase shift

$$\begin{aligned}
 \mathcal{F}\{x(t + \tau)\} &= \int_{-\infty}^{\infty} (x(t + \tau)) e^{-j\omega t} dt \\
 &= \int_{-\infty}^{\infty} (x(t)) e^{-j\omega(t-\tau)} d(t - \tau) \\
 &= \int_{-\infty}^{\infty} (x(t)e^{j\omega\tau}) e^{-j\omega t} dt \\
 &= \int_{-\infty}^{\infty} (x(t)) e^{-j\omega t} dt e^{j\omega\tau} \\
 &= X(j\omega)e^{j\omega\tau} = X(j\omega)e^{j\phi}
 \end{aligned} \tag{1.98}$$

1.5.3 Fourier Transform of the Derivative of a Function

The Fourier transform of the time derivative of a function is the Fourier transform of the function multiplied by $j\omega$.

$$\begin{aligned}
 \mathcal{F}\left\{\frac{dx(t)}{dt}\right\} &= \mathcal{F}\left\{\frac{d}{dt}\frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t} d\omega\right\} \\
 &= \mathcal{F}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)j\omega e^{j\omega t} d\omega\right\} \\
 &= j\omega\mathcal{F}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t} d\omega\right\} \\
 &= j\omega X(j\omega)
 \end{aligned} \tag{1.99}$$

and extending to higher orders

$$\mathcal{F}\left\{\frac{d^n x(t)}{dt^n}\right\} = (j\omega)^n X(j\omega) \tag{1.100}$$

The Laplace Transform of the derivative of a function was given by equation (1.51). It is common for practitioners to substitute $j\omega$ for s to convert the Laplace transform of a function to the Fourier transform. However, this isn't correct as the former is a one-sided transformation, where the second is two-sided. There are times, however, in system identification where the one-sided Fourier transform is required. Specifically, the proper Fourier transform of the response of a system (not necessarily at rest) to an impulse response (as given by equation (1.121)) should recognize that the function has a value of zero for $t < 0$. That is

$$\mathcal{F}(x(t)u(t)) = \int_0^{\infty} x(t)e^{-j\omega t} dt \tag{1.101}$$

where

$$u(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \end{cases} \tag{1.102}$$

is the *Heaviside* or *step function*. This is necessary to ensure values of zero for times preceding the impulse. In these circumstances, the Laplace transform and Fourier transform yield analogous results so that the *one-sided Fourier transform* of the derivative of a function is given by

$$\mathcal{F} \left\{ \frac{d(x(t)u(t))}{dt} \right\} = j\omega \mathcal{F} \{ x(t) \} - x(0) \quad (1.103)$$

analogous to equation (1.51).

1.5.4 Fourier Transform of the Integral of a Function

It follows directly from the previous section that

$$\mathcal{F} \left(\int_{-\infty}^{\infty} x(t)dt \right) = \frac{1}{j\omega} X(j\omega) \quad (1.104)$$

1.5.5 Convolution in the Time Domain

The Fourier transform of a convolution integral in the time domain is the product of the Fourier transforms of the two functions:

$$\begin{aligned} \mathcal{F}^{-1} (X(f)F(f)) &= \int_{-\infty}^{\infty} X(f)F(f)e^{j2\pi ft} df \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau)e^{-j2\pi f\tau} d\tau F(f)e^{j2\pi ft} df \\ &= \int_{-\infty}^{\infty} x(\tau) \int_{-\infty}^{\infty} F(f)e^{j2\pi f(t-\tau)} df d\tau \\ &= \int_{-\infty}^{\infty} x(\tau)f(t-\tau)d\tau \end{aligned} \quad (1.105)$$

Therefore,

$$X(f)F(f) = \mathcal{F} \left(\int_{-\infty}^{\infty} x(\tau)f(t-\tau)d\tau \right) = \mathcal{F} \left(\int_{-\infty}^{\infty} x(\tau-t)f(-\tau)d\tau \right) \quad (1.106)$$

Similarly,

$$\begin{aligned} X(f)\bar{F}(f) &= \mathcal{F} \left(\int_{-\infty}^{\infty} x(\tau)f(\tau-t)d\tau \right) = \mathcal{F} \left(\int_{-\infty}^{\infty} x(\tau+t)f(\tau)d\tau \right) \\ &= \mathcal{F} \left(\int_{-\infty}^{\infty} x(\tau+t)f(\tau)d\tau \right) = \mathcal{F} \left(\int_{-\infty}^{\infty} x(\tau)f(\tau-t)d\tau \right) \end{aligned} \quad (1.107)$$

1.5.6 Convolution in the Frequency Domain

The inverse Fourier transform of a convolution integral in the frequency domain is the product of the inverse Fourier transforms of the two functions:

$$\begin{aligned}
 \mathcal{F}(x(t)f(t)) &= \int_{-\infty}^{\infty} x(t)f(t)e^{-j2\pi ft}dt \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X(v)e^{j2\pi vt}dv f(t)e^{-j2\pi ft}dt \\
 &= \int_{-\infty}^{\infty} X(v) \int_{-\infty}^{\infty} f(t)e^{-j2\pi(f-v)t}dt dv \\
 &= \int_{-\infty}^{\infty} X(v)F(f-v)dv
 \end{aligned} \tag{1.108}$$

1.5.7 Parseval's Formula

The inverse Fourier transform of equation (1.107) is

$$\int_{-\infty}^{\infty} (X(f)\bar{F}(f)e^{j2\pi ft}) df = \int_{-\infty}^{\infty} x(\tau+t)f(\tau)d\tau \tag{1.109}$$

Substituting x for f (and likewise X for F) and setting $t = 0$ yields Parseval's formula:

$$\int_{-\infty}^{\infty} (X(f)\bar{X}(f)) df = \int_{-\infty}^{\infty} x^2(\tau)d\tau \tag{1.110}$$

For practical cases, this is often bounded in time. In order to alleviate this, Parseval's formula is often modified to be

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{\infty} (X(f)\bar{X}(f)) df = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(\tau)d\tau = x_{RMS}^2 \tag{1.111}$$

needs fix

1.6 Frequency Response Functions

Consider now applying the Fourier Transform to the model of a spring-mass-damper system given by equation (1.34). Using equation (1.100), equation (1.34) becomes

$$(-\omega^2 m + j\omega c + k)X(j\omega) = F(j\omega) \tag{1.112}$$

solving for $X(j\omega)/F(j\omega)$ we obtain the *frequency response function*, or *FRF*

$$\frac{X(j\omega)}{F(j\omega)} = H(j\omega) = \frac{1}{-\omega^2 m + j\omega c + k} \tag{1.113}$$

The principle uses of the FRF are to obtain a quick understanding of the response of a system to harmonic excitations, or, when obtained experimentally, to be able to identify system

Table 1.2 Fourier transforms of selected time functions.

$f(t)$	$\mathcal{F}\{f(t)\}$
$e^{-\zeta\omega_n t}u(t)$	$\frac{1}{\zeta\omega_n + j\omega}$
$e^{\zeta\omega_n t}u(-t)$	$\frac{1}{\zeta\omega_n - j\omega}$
$\delta(t)$	1
1	$2\pi\delta(\omega)$
$u(t)$	$\pi\delta(\omega) + \frac{1}{j\omega}$
$\text{sgn}(t)$	$\frac{2}{j\omega}$
$\sin(\omega_n t)$	$j\pi(\delta(\omega + \omega_n) - \delta(\omega - \omega_n))$
$\cos(\omega_n t)$	$\pi(\delta(\omega - \omega_n) + \delta(\omega + \omega_n))$
$\sin(\omega_n t)u(t)$	$\frac{-\pi}{2j}(\delta(\omega + \omega_n) - \delta(\omega - \omega_n)) + \frac{\omega_n}{\omega_n^2 - \omega^2}$
$\cos(\omega_n t)u(t)$	$\frac{\pi}{2}(\delta(\omega - \omega_n) + \delta(\omega + \omega_n)) + \frac{j\omega}{\omega_n^2 - \omega^2}$
$e^{-\zeta\omega_n t}\sin(\omega_n t)u(t)$	$\frac{\omega_n}{(\zeta\omega_n + j\omega)^2 + \omega_n^2}$
$e^{-\zeta\omega_n t}\cos(\omega_n t)u(t)$	$\frac{\zeta\omega_n + j\omega}{(\zeta\omega_n + j\omega)^2 + \omega_n^2}$
$\sum_{n=-\infty}^{\infty} \delta(t - n\Delta t)$	$\frac{2\pi}{\Delta t} \sum_{n=-\infty}^{\infty} \delta(\omega - \frac{2\pi n}{\Delta t})$

parameters. A quick observation of the FRF shows that if c is small, its greatest magnitude will occur at $k - \omega^2 m = 0$, which gives the natural frequency equation $\omega_n = \sqrt{k/m}$. Alternatively, recognizing that the FRF has dimensions of $1/k$, multiplying both sides by k gives the non-dimensional FRF

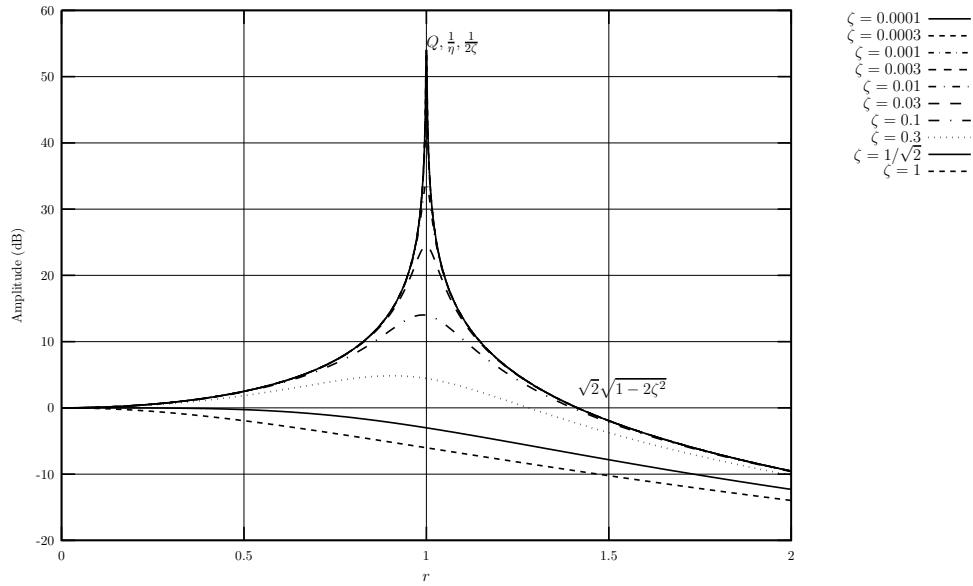
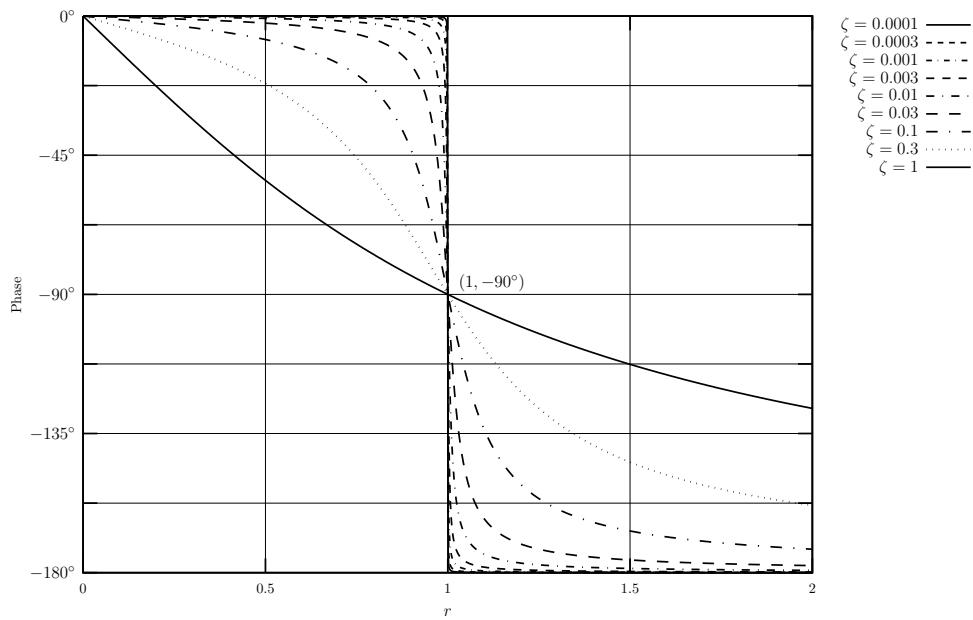
$$\frac{X(j\omega)k}{F(j\omega)} = H(j\omega) = \frac{1}{1 - \omega^2 \frac{m}{k} + j \frac{c}{k}} = H_n(r) = \frac{1}{(1 - r^2) + 2\zeta r j} \quad (1.114)$$

where $r = \omega/\omega_n$ is the normalized (non-dimensionalized) excitation frequency, and $H_n(r)$ is the non-dimensional frequency response function. Figures 1.6 and 1.7 show values of the non-dimensionalized single degree of freedom (SDOF) FRF as a function of r for a wide variety of damping ratios, ζ .

From equation (1.114), one should note the following:

- The magnitude of the FRF at $r = 0$ is 0 dB⁴.
- The phase of the FRF at $r = 0$ is 0.
- The magnitude of the FRF at $r = 1$ is $\approx 1/2\zeta$.
- The phase of the FRF at $r = 1$ is $-\pi/2$.
- The peak amplitude of the FRF occurs at $r = \sqrt{1 - 2\zeta^2}$.
- For light enough damping, the peak response can be assumed to occur at $r = 1$, with a value of $1/2\zeta$.
- The response is attenuated for all values of $r > \sqrt{2}\sqrt{1 - 2\zeta^2}$.

⁴See Appendix A.

Figure 1.6 Non-dimensional FRF amplitude as a function of r Figure 1.7 Non-dimensional FRF phase as a function of r

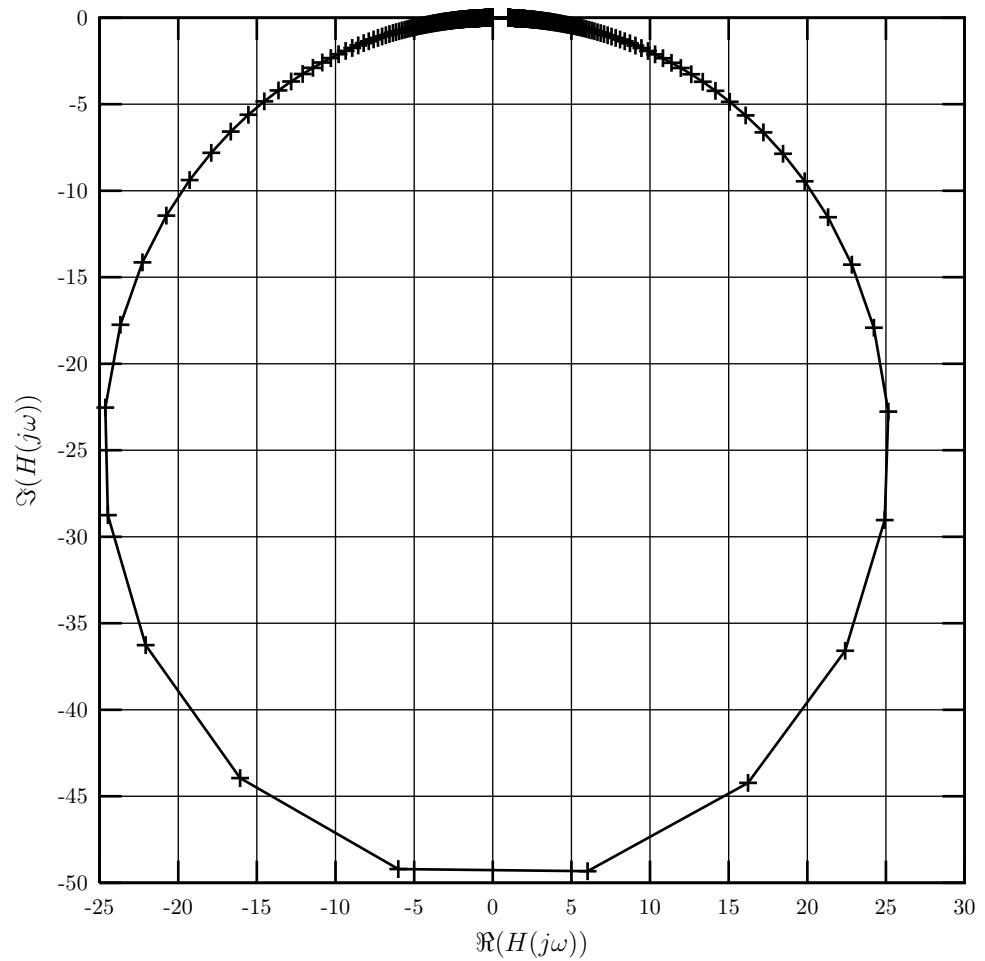


Figure 1.8 Non-dimensional compliance FRF Nyquist plot, $\zeta = 0.1$

Example 1.6.1 Find $x_p(t)$ for the system defined by

$$4\ddot{x} + 0.01\dot{x} + 300x = 7 \sin(12t) \quad (1.115)$$

Solution:

From equation (1.113),

$$\begin{aligned} \frac{X(j\omega)}{F(j\omega)} &= \frac{1}{-\omega^2 m + j\omega c + k} \\ &= \frac{1}{-12^2 \times 4 + j \times 12 \times 0.01 + 300} \\ &= -3.62 \times 10^{-3} - 1.57 \times 10^{-6} j \\ &= 3.62 \times 10^{-3} e^{-3.14j} \end{aligned} \quad (1.116)$$

Therefore, the amplification (attenuation in this case), and the phase of $x(t)$ relative to $f(t)$ are 3.62×10^{-3} and -3.14 radians respectively, and

$$x(t) = 7 \times (3.62 \times 10^{-3}) \sin(12t - 3.14) = 2.54 \times 10^{-2} \sin(12t - 3.14) \quad (1.117)$$

An alternate form of the frequency response function, useful for modal analysis, is obtained using partial fraction expansion (Reid 1983) as

$$\begin{aligned} \frac{X(j\omega)}{F(j\omega)} &= H(j\omega) = \frac{a}{-\omega^2 + 2\zeta\omega_n j + \omega_n^2} \\ &= \frac{b}{j\omega - \lambda} + \frac{\bar{b}}{j\omega - \bar{\lambda}} \end{aligned} \quad (1.118)$$

where $a = \frac{1}{m}$, λ are given by equation (1.19), and $b = \frac{1}{m(\lambda - \bar{\lambda})}$. Both a and b are referred to as the *residues* of the transfer function (Ewins 1992; Maia and E Silva 1998), the form to be used depending on the form of the frequency response function being used. While this distinction and terminology isn't significant in the single degree of freedom case, it is central to modal analysis as discussed later in chapter 6. Considering the most commonly of interest underdamped case, substituting $\lambda = -\zeta\omega_n - j\omega_n\sqrt{1 - \zeta^2}$ into the expression for b yields

$$\begin{aligned} b &= \frac{a}{(\lambda - \bar{\lambda})} \\ &= \frac{1}{m((- \zeta\omega_n - j\omega_n\sqrt{1 - \zeta^2}) - (-\zeta\omega_n + j\omega_n\sqrt{1 - \zeta^2}))} \\ &= \frac{1}{m2j\omega_n\sqrt{1 - \zeta^2}} = \frac{1}{2jm\omega_d} \end{aligned} \quad (1.119)$$

Taking the inverse Fourier transform of equation (1.118) for the case of $F(j\omega) = 1$ gives

$$x(t) = \mathcal{F}^{-1}\{X(j\omega)\} = be^{\lambda t} + \bar{b}e^{\bar{\lambda}t} \quad (1.120)$$

Substitution of $\lambda = -\zeta\omega_n - j\omega_n\sqrt{1 - \zeta^2}$ and $b = \frac{1}{2jm\omega_d}$, and applying the Euler relations results in the under-damped impulse response function given in equation (1.143) as

$$x(t) = h(t) = \frac{1}{m\omega_d} e^{-\zeta\omega_n t} \sin \omega_d t \quad (1.121)$$

for $\tau = 0$.

Often, the frequency response function obtained is between the force input and the velocity output and is called the *mobility* and is commonly obtained when using sensors such as Doppler-based sensors such as certain laser motion sensors. Applying equation (1.100) to equation (1.113) gives

$$\frac{j\omega X(j\omega)}{F(j\omega)} = j\omega H(j\omega) = \frac{j\omega}{-\omega^2 m + j\omega c + k} \quad (1.122)$$

Further, when accelerometers are used, the frequency response function is between the acceleration and the force and is called the *inertance*. Applying equation (1.100) to equation (1.122) gives

$$\frac{-\omega^2 X(j\omega)}{F(j\omega)} = -\omega^2 H(j\omega) = \frac{-\omega^2}{-\omega^2 m + j\omega c + k} \quad (1.123)$$

By its very nature, since the limit as $\omega \rightarrow \infty = \frac{1}{m}$, inertance data tends to be much better at capturing high frequency responses.

1.6.1 Complex Modulus/Stiffness

An alternate model of damping that is somewhat successful in modeling damping in materials is called the *complex modulus*. With this model of damping, the energy dissipated per radian (averaged over a cycle) divided by the maximum potential energy of the spring is defined as the loss factor

$$\eta = \frac{\Delta U}{2\pi \left(\frac{1}{2} k X^2 \right)} \quad (1.124)$$

For a system with viscous damping, the energy dissipated per cycle is

$$\begin{aligned} \Delta U &= \oint_{F_d} dx \\ &= \int_0^{2\pi/\omega} c \dot{x} \frac{dx}{dt} dt \\ &= \int_0^{2\pi/\omega} c \dot{x}^2 dt \end{aligned} \quad (1.125)$$

Assuming a harmonic response

$$x(t) = X \sin \omega t \quad (1.126)$$

where we have adjusted the time $t = 0$ so that the phase $\phi = 0$. Substituting $\dot{x}(t) = X\omega \cos \omega t$ into equation (1.125) the energy dissipated per cycle is

$$\begin{aligned} \Delta U &= \int_0^{2\pi/\omega} c (X\omega \cos \omega t)^2 dt \\ &= c\omega^2 X^2 \int_0^{2\pi/\omega} (\cos \omega t)^2 dt \\ &= c\omega^2 X^2 \frac{\pi}{\omega} \\ &= c\omega X^2 \pi \end{aligned} \quad (1.127)$$

Thus for a viscously damped system, the loss factor is given by

$$\begin{aligned}\eta &= \frac{c\omega X^2 \pi}{2\pi \frac{1}{2} k X^2} \\ &= \frac{c\omega}{k}\end{aligned}\quad (1.128)$$

giving an effective damping coefficient, c , of

$$c_{eq} = \frac{\eta k}{\omega} \quad (1.129)$$

Substituting this value of c into equation (1.112) gives

$$\begin{aligned}(-\omega^2 m + j\omega \frac{\eta k}{\omega} + k)X(j\omega) &= F(j\omega) \\ (-\omega^2 m + j\eta k + k)X(j\omega) &= F(j\omega) \\ (-\omega^2 m + k^*)X(j\omega) &= F(j\omega)\end{aligned}\quad (1.130)$$

where $k^* = k(1 + \eta j) = k + jk'$ is called the *complex stiffness*. Thus the frequency response function is

$$\frac{X(j\omega)}{F(j\omega)} = h(j\omega) = \frac{1}{-\omega^2 m + k(1 + j\eta)} \quad (1.131)$$

The usefulness of this model is when damping is measured in a material as a *hysteresis* loop as shown in Figure 1.9. Loss factors for materials are generally available in reference books, while loss factors for viscoelastic materials (materials that have high amounts of damping, with typically frequency dependent loss factor and storage modulus) are generally produced by manufacturers. Loss factors for metals tend to be quite small, on the order of a fraction of a percent, and relatively constant over wide frequency ranges. This model is not generally suitable for transient analysis, but Nashif, Jones, and Henderson show that with proper consideration, it is indeed proper (Nashif et al. 1985). The primary requirement is that $\eta = 0$ for $\omega = 0$, otherwise the solution becomes non-causal. For further details, the reader is directed to the text by Inman (Inman 2000).

1.7 Response to an Arbitrary Excitation

1.7.1 Impulse Response of a Viscously Damped System

Consider equation (1.1) with initial conditions $x(0) = x_0$ and $\dot{x}(0) = v_0$ excited by the force

$$F(t) = \hat{F}\delta(t) \quad (1.132)$$

where $\delta(t)$ is the *Dirac delta function* or unit impulse, as defined by equations (1.82) and (1.83), and \hat{F} is a constant. The Dirac delta function is defined as

$$\delta(t) = 0, \quad t \neq 0 \quad (1.133)$$

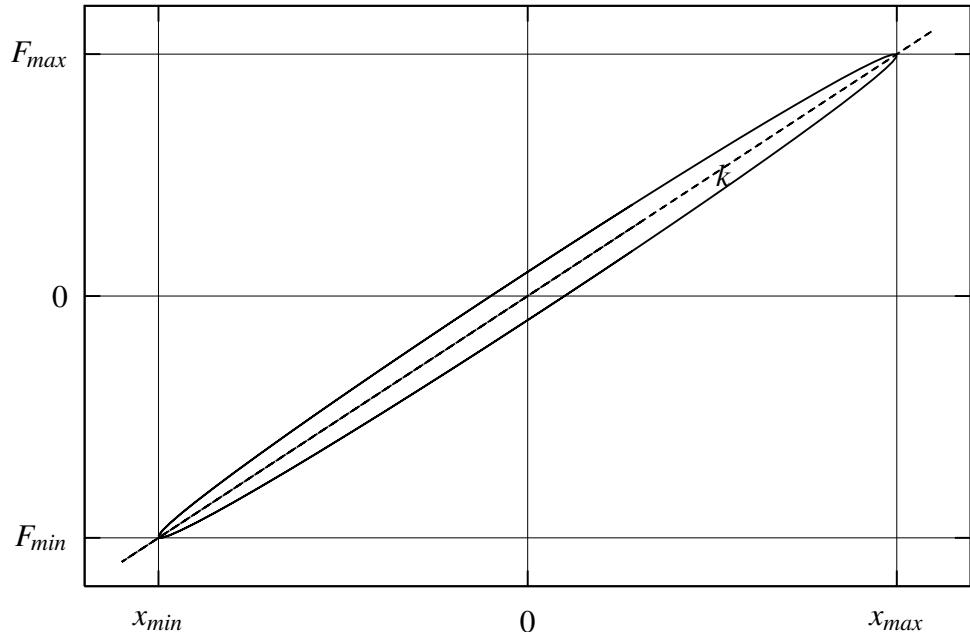


Figure 1.9 Example hysteresis loop.

but such that

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (1.134)$$

In a practical situation, let's assume that the force is defined such that

$$F(t) = \begin{cases} 0, & t < 0 \\ \frac{\hat{F}}{\Delta\tau}, & 0 < t < \Delta\tau \\ 0, & \Delta\tau < t \end{cases} \quad (1.135)$$

Although the ideal delta function never occurs in practice, it is a useful tool for approximating sharp impacts as well as for deriving responses to arbitrary excitations. Considering Newton's law

$$F(t) = \frac{d(mv)}{dt} \quad (1.136)$$

and integrating with respect to time, and assuming constant mass,

$$\begin{aligned} \int_0^{\Delta\tau} F(\tau) d\tau &= \int_0^{\Delta\tau} \frac{d(mv)}{d\tau} d\tau \\ \hat{F} &= m(v(\Delta\tau) - v_0) \end{aligned} \quad (1.137)$$

the result obtained is the *principle of impulse and momentum*. Here we have neglected forces from the spring and dashpot for the time being. Assuming a constant acceleration over the

period $\Delta\tau$, the velocity during the response is

$$v(t) = v_0 + \frac{\hat{F}}{m\Delta\tau}t = v_0 + \frac{v(\Delta\tau) - v_0}{\Delta\tau}t \quad (1.138)$$

$$x(t) = x_0 + v_0t + \frac{v(\Delta\tau) - v_0}{\Delta\tau}\frac{t^2}{2} \quad (1.139)$$

and the approximate error due to neglecting the force in the spring and in the dashpot during the impulsive load is then⁵

$$\begin{aligned} I_{kc} &= \int_0^{\Delta\tau} kx + c\dot{x}dt \\ &= \int_0^{\Delta\tau} k \left(x_0 + v_0t + \frac{v(\Delta\tau) - v_0}{\Delta\tau}\frac{t^2}{2} \right) + c \left(v_0 + \frac{v(\Delta\tau) - v_0}{\Delta\tau}t \right) dt \\ &= k \left(x_0\Delta\tau + v_0\frac{\Delta\tau^2}{2} + (v(\Delta\tau) - v_0)\frac{\Delta\tau^2}{6} \right) + c \left(v_0\Delta\tau + (v(\Delta\tau) - v_0)\frac{\Delta\tau}{2} \right) \\ &\approx (kx_0 + cv_0)\Delta\tau, \quad \text{for small } \Delta\tau \end{aligned} \quad (1.140)$$

In fact, as $\Delta\tau \rightarrow 0$, $I_{kc} \rightarrow 0$ because kx_0 and cv_0 are bounded. This is in contrast to the magnitude of the applied load which was defined to increase to maintain a constant impulse. For a finite value of $\Delta\tau$, the resulting velocity after the impulse is

$$v_f \approx \frac{\hat{F}}{m} + v_0 \quad (1.141)$$

as long as the assumption of a constant acceleration is a reasonable approximation, and $(kx_0 + cv_0) \ll \hat{F}$. For $F(t) = \hat{F}\delta(t)$, this is an exact solution. The impulse can thus be treated as an initial condition of $v_{0+} = \hat{F}/m + v_0$ and the free response can be obtained using equation (1.26), (1.29), (1.32), or (1.33) as appropriate, given $x(0) = x_0$. For the case of zero initial conditions, and an impulse at time $t = \tau$, these four equations can be re-written

$$x(t) = \hat{F}h(t - \tau) \quad (1.142)$$

where

$$h(t) = \begin{cases} \frac{1}{m\omega_n} \sin \omega_n(t - \tau), & 0 = \zeta \\ \frac{1}{m\omega_d} e^{-\zeta\omega_n(t-\tau)} \sin \omega_d(t - \tau), & 0 < \zeta < 1 \\ \frac{1}{m}(t - \tau) e^{\lambda(t-\tau)}, & 1 = \zeta \\ \frac{1}{m} \frac{1}{\lambda_1 - \lambda_2} (e^{\lambda_1(t-\tau)} - e^{\lambda_2(t-\tau)}), & 1 < \zeta \end{cases} \quad (1.143)$$

where the functions $h(t)$ are called the *impulse response functions*, or *IRF*.

⁵Those who have taken a digital control course may note that this is an approximate error due to a zero-order hold. A first-order hold approximation can be shown to give similar, if slightly improved, results. Improved numerical integration methods such as Runge-Kutta use and address these assumptions.

1.7.2 Duhamel's Integral

Although the impulse response is useful in some practical applications, perhaps its greatest importance is its use in *Duhamel's integral* (Duhamel 1834), also known as the *convolution integral*. The convolution integral is often used to determine the response of a system to an arbitrary excitation. Considering an arbitrary force as a sequence of n pulses of duration $\Delta\tau$ as shown in Figure 1.10, the response of the system to the i th impulse is

$$x_i(t) = F(\tau_i)\Delta\tau h(t - \tau_i) \quad (1.144)$$

where $F(\tau_i)$ is the magnitude of the force of the i th pulse force and $h(t - \tau_i)$ is the impulse response as given by equation (1.143). Since the system is linear, the principle of superposi-

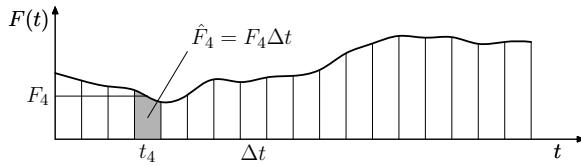


Figure 1.10 Arbitrary excitation represented approximately as impulses.

tion holds true, and the response of the system to the arbitrary force can be approximated by the sum of the responses to impulse excitations, or

$$x(t) = \sum_{i=1}^n F(\tau_i)\Delta\tau h(t - \tau_i) \quad (1.145)$$

In the limit as $\Delta t \rightarrow 0$, the summation is replaced by an integral and $\Delta\tau \rightarrow d\tau$. The response to an arbitrary excitation is then

$$x(t) = \int_{t_0}^t F(\tau)h(t - \tau)d\tau \quad (1.146)$$

where it is necessary that either a) the initial conditions are zero, or b) the initial conditions are known and the response given by equation (1.146) is added, using the principle of superposition, to the free response of the system resulting from the initial conditions. Of importance is the fact that as $\Delta\tau \rightarrow 0$, the error in neglecting the forces of the spring and dashpot during each impulse goes to zero. i.e., from equation (1.140)

$$\lim_{\Delta\tau \rightarrow 0} ((kx_0 + cv_0)\Delta\tau) = 0 \quad (1.147)$$

Thus equation (1.146) is exact. Also, consider substituting $\tau = t - \alpha$, and thus $d\tau = -d\alpha$ for a constant t . Because τ goes from t_0 to t , the limits on α are $\alpha_1 = t - \tau_1 = t - t_0$ to $\alpha = t - \tau_2 = t - t = 0$. Then equation (1.146) then becomes

$$\begin{aligned} x(t) &= \int_{t-t_0}^0 F(t - \alpha)h(\alpha)(-d\alpha) \\ x(t) &= \int_0^{t-t_0} F(t - \alpha)h(\alpha)d\alpha \end{aligned} \quad (1.148)$$

Since α is just a dummy variable of integration,

$$x(t) = \int_{t_0}^t F(\tau)h(t - \tau)d\tau = \int_0^{t-t_0} F(t - \tau)h(\tau)d\tau \quad (1.149)$$

and both forms are equivalent. It's also useful to note that since

$$F(t) = 0, \quad t < 0 \quad (1.150)$$

the lower limits in equations (1.146) and (1.148) can both be changed to $-\infty$. The upper limit can also be changed to ∞ because forces at t greater than any given evaluation time for $x(t)$ do not change the function $x(t)$ for times less than t . Thus equations (1.146) and (1.148) can be written in the alternative forms

$$x(t) = \int_{-\infty}^{\infty} F(\tau)h(t - \tau)d\tau \quad (1.151)$$

and

$$x(t) = \int_{-\infty}^{\infty} F(t - \tau)h(\tau)d\tau \quad (1.152)$$

In this form, it should be apparent that the equivalence of both forms could have been shown using the Fourier convolution principle as shown in Section 1.5.5.

1.8 Characteristics of Nonlinear Systems

The material within this text primarily concerns linear systems. Solution methods, assumptions in signal processing, and system identification techniques all assume that the system transforming an input to an output is linear. However, real-world systems typically exhibit some form of nonlinearity. In fact, it is a fairer statement that the vast majority of systems exhibit nonlinearity, and that we are only able to apply linear analysis techniques by the virtue that engineers tend to design within the realm of linear behavior in order to simplify the design process. Application of linear techniques is subject to the understanding that they lose validity the greater the impact of the nonlinearity on the system response becomes. While understanding experimental error and uncertainty are important, real-world test results and analysis can easily be rendered useless by the careless application of linear techniques to systems showing prominent nonlinear features.(Adams and Allemand 2001)

1.8.1 Linear “Look”

The first evidence that prompts concern from the tester is a frequency response function that doesn't look quite right. While the shape of peaks (resonances) and valleys (anti-resonances) is learned with time, deviation from a frequency response function curve fit can also be used as a means to determine whether significant nonlinearity exists.

Figures 1.6, 8.1, and 8.3 are computer generated from linear models and can provide some sense of the look of resonances and anti-resonances. The resonances are symmetric near the peak and the anti-resonances are symmetric near the bottom. Note, however, that the anti-resonances are often masked by noise in practice and unavailable for observation.

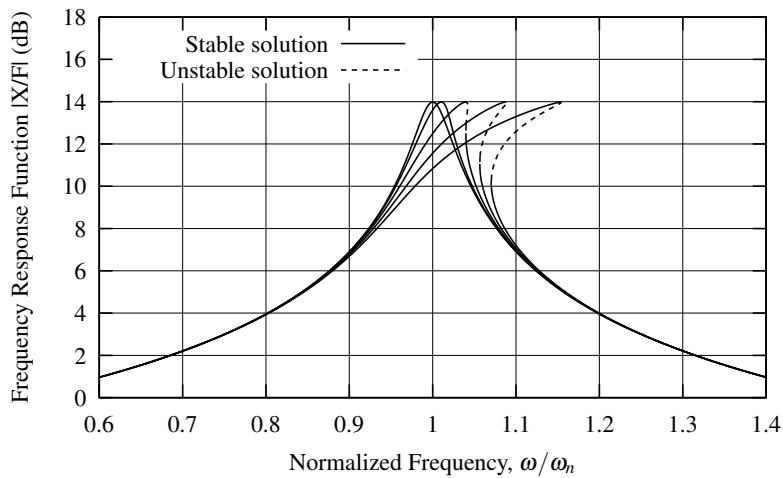


Figure 1.11 Illustrative variation of a FRF peak as α changes.

Figure 1.6 shows that the phase change through a resonance is typically anti-symmetric for damping ratios below 0.1. Note that rapidly fluctuating phase on the order of 360° is usually little more than a small amount of noise combined with a phase wrapping algorithm that is too simplistic as 360° represents no phase change at all. Further, the near coexistence of a resonance and an anti-resonance can cause an unusual looking bump in both the amplitude and phase plots as illustrated in Figures 1.13 and 1.14.

1.8.2 Duffing Equation

$$m\ddot{x} + c\dot{x} + kx + \alpha x^3 = f(t) \quad (1.153)$$

Hardening Spring

Softening Spring

1.8.3 Duffing Equation With Damping

1.8.4 Limit Cycles

1.9 Self-Adjoint Multiple Degree of Freedom (MDOF) Systems

More significant structures require more detailed modeling, the result of which is, in the end, a multiple degree of freedom system. These models can be generated by a variety of means, including lumped mass models, as generally taught in elementary dynamics and systems classes, or more sophisticated energy or weighted residual based models (Kelly 1993;

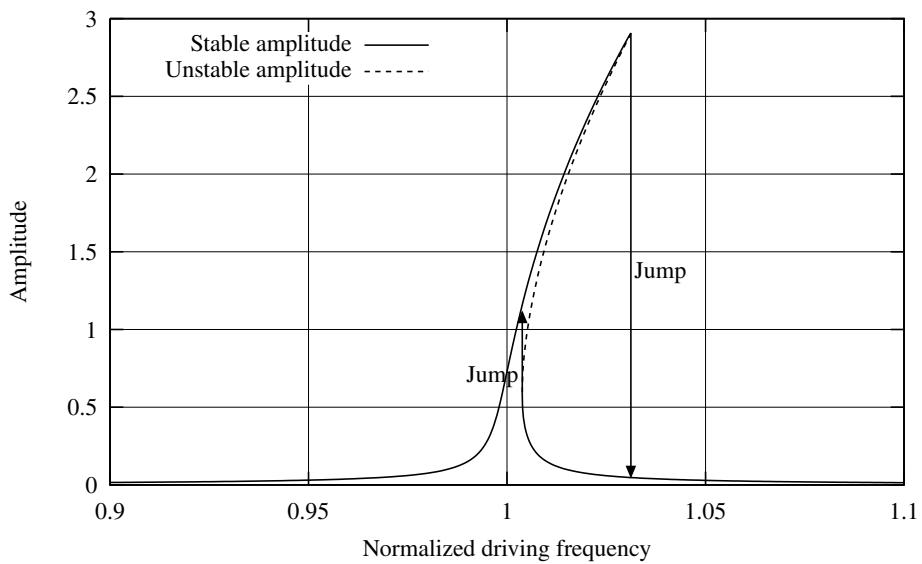


Figure 1.12 Bifurcation behavior of a strongly nonlinear Duffing oscillator near primary resonance.

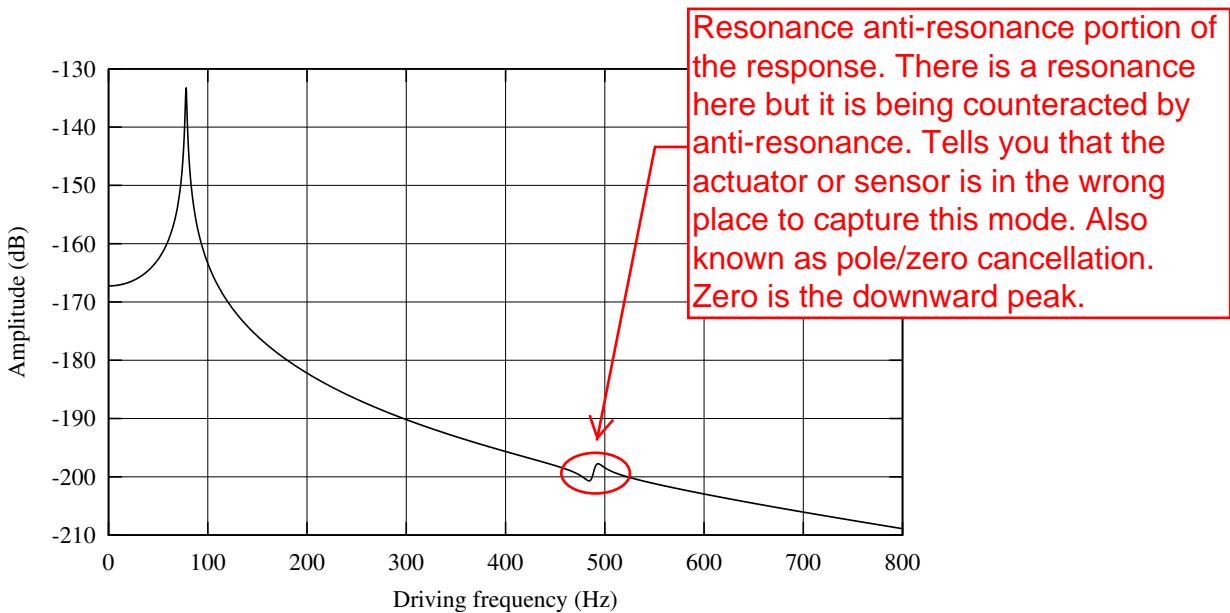
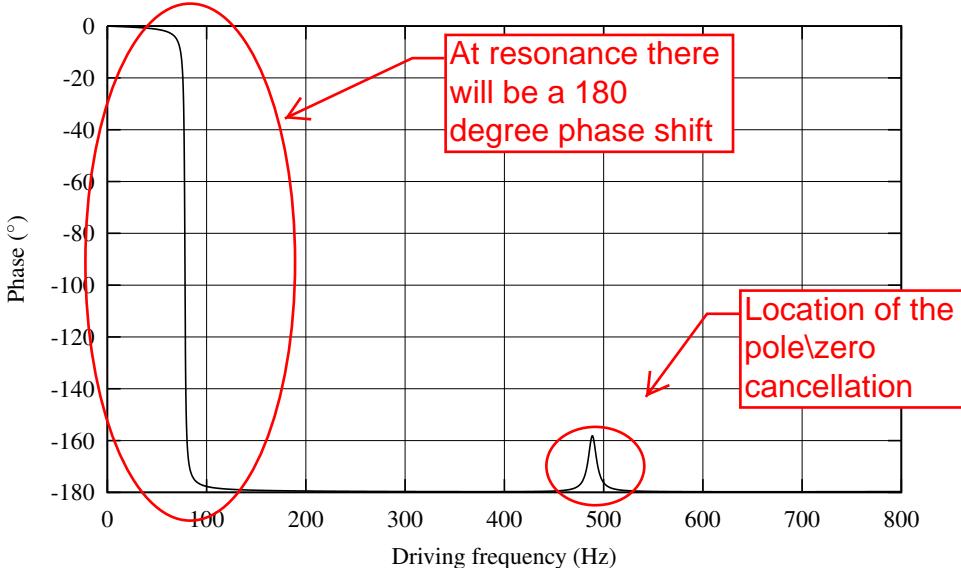


Figure 1.13 Magnitude frequency response function of a beam with near resonance-anti-resonance cancellation.



Part is not spinning, which makes matrices all symmetric. If there is spin then matrices are anti-symmetric.

Figure 1.14 Phase frequency response function of a beam with near resonance-anti-resonance cancellation.

Meirovitch 1997; Weaver et al. 1990). The most common method for generating MDOF models are the *finite element methods*, or FEM (Bathe and Wilson 1976, 1996; Cook et al. 1989; Rao 1999; Zienkiewicz and Taylor 1989). The majority of structures can be modeled as *self-adjoint* systems with n degrees of freedom. These systems are governed by equations of motion of the form

$$M\ddot{\mathbf{x}}(t) + C\dot{\mathbf{x}}(t) + K\mathbf{x}(t) = \mathbf{f}(t) \quad (1.154)$$

where M is a positive-definite $n \times n$ matrix, C and K are positive-semi-definite $n \times n$ matrices, and all are real and symmetric. The displacement vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \end{bmatrix} \quad (1.155)$$

represents displacements of the n degrees of freedom (generalized coordinates), which may be in any direction, x , y , z , or any combination thereof, or a rotation about any unit direction vector in three dimensional space. The force vector

$$\mathbf{f}(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \\ \vdots \\ f_n(t) \end{bmatrix} \quad (1.156)$$

represents forces acting on each of the n generalized coordinates. Such equations of motion are typical of non-rotating structures without inclusion of aerodynamic loading.

Solution of equation (1.154) is most often performed in modal coordinates due to physical insight obtained and numerical advantages. In order to transform into modal coordinates, we first consider the un-forced, or homogeneous, and undamped, $C = 0$, equation of motion given by

$$M\ddot{\mathbf{x}} + K\mathbf{x} = \mathbf{0} \quad (1.157)$$

where for the sake of simplicity dependence on time is no longer explicitly shown. A solution for $\mathbf{x}(t)$ is then assumed to be

$$\mathbf{x} = \psi e^{\lambda t} \quad (1.158)$$

Substituting into equation (1.157) gives

$$(M\lambda^2 + K)\psi = \mathbf{0} \quad (1.159)$$

Pre-multiplying by the M^{-1} and setting $\lambda = j\omega$ results in the eigenvalue problem

$$(M^{-1}K - I\omega^2)\psi = \mathbf{0} \quad (1.160)$$

which is more familiar to students who have taken a basic vibration course covering multiple degree of freedom systems. This is a poor form of the eigenvalue to use in practice due to numerical issues and won't be used here. (Meirovitch 1997)

Equation (1.159) should be recognized as an equivalent eigenvalue problem with the eigenvalue λ^2 and the eigenvector ψ . Since the vector is n elements long, and we also have the unknown λ^2 to solve for, we have only n equations but $n + 1$ unknowns. In addition, the equations are nonlinear in the combined unknowns λ^2 and ψ_i . A first attempt at solving these equations for ψ would be to premultiply equation (1.159) by $(M\lambda^2 + K)^{-1}$ giving

$$\begin{aligned} (M\lambda^2 + K)^{-1}(M\lambda^2 + K)\psi &= (M\lambda^2 + K)^{-1}\mathbf{0} \\ I\psi &= (M\lambda^2 + K)^{-1}\mathbf{0} \\ \psi &= (M\lambda^2 + K)^{-1}\mathbf{0} \end{aligned} \quad (1.161)$$

At first observation, the only solution appears to be $\psi = \mathbf{0}$. However, this is the so-called *trivial solution*. In fact observing equation (1.157), $\mathbf{x} = \mathbf{0}$ is a viable solution, if useless for understanding dynamic response. An alternative solution is the case where $A = (M\lambda^2 + K)^{-1}$ does not exist.

The inverse of a matrix A is defined as

$$A^{-1} = \frac{\text{adj}(A)}{\det(A)} \quad (1.162)$$

where $\text{adj}(A)$ is the *adjoint*⁶ (Golub and Van Loan 1985; Meirovitch 1997) or *classical adjoint* (Horn and Johnson 1999) matrix of A and $\det(A)$ is the determinant of A . Of importance is the denominator. If the $\det(A) = 0$, then the inverse of A is undefined, and the

⁶The i, j element of the adjoint of a matrix is the determinant of the remaining matrix when the j th row and the i th column are removed, multiplied by -1^{i+j} . In practice, other methods are used to calculate the inverse when necessary.

solution in the form of equation (1.161) makes no sense. Thus, since we are expecting solutions other than the trivial solution, finding the value/s of λ^2 such that $\det(M\lambda^2 + K) = 0$ is the only feasible solution.

Taking the determinant of $M\lambda^2 + K$ and setting it equal to zero results in an n th order polynomial in λ^2 , the solution of which results in n real and negative values of λ^2 . Solving for λ results in n imaginary conjugate pairs of values $\lambda_i = \pm j\omega_i$, $i = 1, \dots, n$. For each value of λ_i^2 , a unique (linearly independent from the rest) ψ_i can be found. Even when there are repeated values, i.e. $\lambda_i^2 = \lambda_{i+1}^2$, linearly independent vectors ψ_i and ψ_{i+1} can be found for each occurrence of a solution, λ^2 . In the current case, the vectors ψ_i are real and linearly independent. The vectors ψ_i are the mode shape of the structure and the corresponding values ω_i are the natural frequencies in radians/sec⁷.

The methods for obtaining λ_i and ψ_i from a given mass matrix, M , and stiffness matrix, K , pair vary greatly in practice from the elementary methods often learned in introductory courses. Most often methods such as subspace iteration, the QR method, inverse iteration, or other even more sophisticated algorithms are used (Bathe and Wilson 1976; Meirovitch 1997) and are beyond the scope of this book.

It is convenient to *mass normalize* the eigenvectors ψ_i such that

$$\psi_i^T M \psi_i = 1 \quad (1.163)$$

After doing so, the eigenvectors are *mass orthonormal* and *stiffness orthogonal*. Consider the two unique eigenvalue solutions ($l \neq m$), written in a slightly different form

$$K\psi_l = -M\lambda_l^2\psi_l \quad (1.164a)$$

$$K\psi_m = -M\lambda_m^2\psi_m \quad (1.164b)$$

Pre-multiplying each by the alternate eigenvector transposed gives

Identity used in derivation:
 $(A^*B^*C)^T =$
 $C^T B^T A^T$

$$\psi_m^T K \psi_l = -\psi_m^T M \lambda_l^2 \psi_l = -\lambda_l^2 \psi_m^T M \psi_l \quad (1.165a)$$

$$\psi_l^T K \psi_m = -\psi_l^T M \lambda_m^2 \psi_m = -\lambda_m^2 \psi_l^T M \psi_m \quad (1.165b)$$

Subtracting (1.165a) from the transpose of (1.165b), and again noting that M and K are symmetric, gives

$$\begin{aligned} \psi_m^T K \psi_l - \psi_l^T K \psi_m &= -\lambda_m^2 \psi_m^T M \psi_l + \lambda_l^2 \psi_l^T M \psi_m \\ 0 &= (\lambda_l^2 - \lambda_m^2) \psi_m^T M \psi_l \end{aligned} \quad (1.166)$$

If $(\lambda_l^2 - \lambda_m^2) \neq 0$, then $\psi_m^T M \psi_l = 0$, and since $\psi_m^T M \psi_m = 1$,

$$\psi_m^T M \psi_l = \delta_{lm} = \begin{cases} 1, & l = m \\ 0, & l \neq m \end{cases} \quad (1.167)$$

⁷It is very rare for a unit of time other than seconds to be used for a structure

Mode shapes (eigenvectors) are orthogonal through either the mass matrix M and stiffness matrix K . They are usually not orthogonal by themselves.

where δ_{lm} is the *Kronecker delta function* and is as defined above. Considering now equation (1.165a)

$$\begin{aligned}\psi_m^T K \psi_l &= -\lambda_l^2 \psi_m^T M \psi_l \\ &= -\lambda_l^2 \delta_{lm} \\ &= \begin{cases} -\lambda_l^2, & l = m \\ 0, & l \neq m \end{cases}\end{aligned}\quad (1.168)$$

In the case where there are repeated solutions, i.e. $\lambda_m^2 = \lambda_{m+1}^2 = \dots$, substituting the coordinate transformation

$$\mathbf{x} = M^{-1/2} \mathbf{q} \quad (1.169)$$

into equation (1.157) and pre-multiplying by $(M^{-1/2})^T$ where $M^{-1/2}$ is defined such that

$$M = \left(M^{1/2} \right)^T M^{1/2}$$

gives

$$\begin{aligned}\left(M^{-1/2} \right)^T M M^{-1/2} \ddot{\mathbf{q}} + \left(M^{-1/2} \right)^T K M^{-1/2} \mathbf{q} &= \left(M^{-1/2} \right)^T \mathbf{f} \\ I \ddot{\mathbf{q}} + \tilde{K} \mathbf{q} &= \left(M^{-1/2} \right)^T \mathbf{f}\end{aligned}\quad (1.171)$$

Look up Choleski decomposition!!!
Maybe used when M or K is not symmetric

where \tilde{K} is the mass normalized stiffness matrix. Solving the homogeneous equation by assuming a solution of the form

$$\mathbf{q}(t) = \mathbf{v} e^{\lambda t} \quad (1.172)$$

yields the single matrix eigenvalue problem

$$(I\lambda^2 + \tilde{K})\mathbf{v} = \mathbf{0} \quad (1.173)$$

Since \tilde{K} is symmetric, its eigenvectors, \mathbf{v}_i , are orthonormal (Golub and Van Loan 1985). Substituting (1.158) and (1.172) into (1.169) gives

$$\psi = M^{-1/2} \mathbf{v} \quad (1.174)$$

Since the vectors \mathbf{v}_i are orthonormal, and $M^{-1/2}$ must be non-singular, the vectors ψ_i must be linearly independent even in the case of repeated eigenvalues.

Given the multiplicity of solutions for ω^2 and ψ , the solution for \mathbf{x} must then be considered the linear summation

$$\mathbf{x} = \sum_{i=1}^n (a_i e^{+j\omega_i t} + \bar{a}_i e^{-j\omega_i t}) \psi_i \quad (1.175)$$

Applying the Euler relation, equation (1.2), this can be written in the more physically intuitive form

$$\mathbf{x} = \sum_{m=1}^n R_i \sin(\omega_i t + \phi_i) \psi_i \quad (1.176)$$

where R_i is a *modal amplitude*. If we further define

$$r_i(t) = R_i \sin(\omega_i t + \phi_i) \quad (1.177)$$

to be our *modal coordinates*, then equation (1.176) can be written in matrix form as

$$\mathbf{x}(t) = \Psi \mathbf{r}(t) \quad (1.178)$$

where

$$\Psi = [\psi_1 \ \psi_2 \ \psi_3 \ \dots \ \psi_n] = \begin{bmatrix} \psi_{1,1} & \psi_{1,2} & \dots & \psi_{1,m} & \dots & \psi_{1,n} \\ \psi_{2,1} & \psi_{2,2} & \dots & \psi_{2,m} & \dots & \psi_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ \psi_{l,1} & \psi_{l,2} & \dots & \psi_{l,m} & \dots & \psi_{l,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{n,1} & \psi_{n,2} & \dots & \psi_{n,m} & \dots & \psi_{n,n} \end{bmatrix} \quad (1.179)$$

Substituting equation (1.178) into equation (1.154),

$$M\Psi\ddot{\mathbf{r}}(t) + K\Psi\mathbf{r}(t) = \mathbf{f}(t) \quad (1.180)$$

Defined in modal coordinates

Pre-multiplying (1.180) by Ψ^T , and using equations (1.167) and (1.168) the equations are now transformed into individual uncoupled *modal equations*

$$I\ddot{\mathbf{r}}(t) + \Omega^2\mathbf{r}(t) = \tilde{\mathbf{f}}(t) \quad (1.181)$$

where I is the identity matrix,

$$\Omega^2 = \begin{bmatrix} \omega_1^2 & 0 & 0 & \dots & 0 \\ 0 & \omega_2^2 & 0 & \dots & 0 \\ 0 & 0 & \omega_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \omega_n^2 \end{bmatrix} \quad (1.182)$$

and

$$\tilde{\mathbf{f}}(t) = \Psi^T \mathbf{f}(t) \quad (1.183)$$

is the modal force vector. The resulting decoupled equations can each then be solved using the methods of sections 1.1-1.7. Applying equation (1.178), the solution can be transformed back into physical coordinates. Also, using equation (1.178), the initial conditions can be transformed into modal coordinates for use in solving equations (1.181) giving

$$\mathbf{r}(0) = \Psi^{-1} \mathbf{x}(0) \quad (1.184a)$$

$$\dot{\mathbf{r}}(0) = \Psi^{-1} \dot{\mathbf{x}}(0) \quad (1.184b)$$

This procedure can also be followed for damped systems under limited cases. In the case of viscous damping, where the governing equation of motion is

$$M\ddot{\mathbf{x}}(t) + C\dot{\mathbf{x}}(t) + K\mathbf{x}(t) = \mathbf{f}(t), \quad (1.185)$$

then if

$$CM^{-1}K = KM^{-1}C \quad (1.186)$$

the damping matrix, C , is diagonalized by $\Psi^T C \Psi = \text{diag}(2\zeta_i \omega_i)$ (Caughey and O'Kelly 1965). Consider substituting

$$x(t) = M^{-1/2}q(t) \quad (1.187)$$

into equation (1.185), but with no forcing vector, where

$$(M^{-1/2})^T M M^{-1/2} = I \quad (1.188)$$

Pre-multiplying by $(M^{-1/2})^T$ yields

$$I\ddot{q}(t) + \tilde{C}\dot{q}(t) + \tilde{K}q(t) = \mathbf{0} \quad (1.189)$$

Now let

$$q(t) = \Upsilon r(t) \quad (1.190)$$

where Υ^8 is the orthonormal matrix of the eigenvectors of \tilde{K} , \mathbf{v}_i , such that

$$\Upsilon^T \Upsilon = I \quad (1.191)$$

$$\Upsilon^T \tilde{K} \Upsilon = \Omega^2$$

substituting (1.190) into equation (1.189) and pre-multiplying by Υ^T yields

$$I\ddot{r}(t) + \Upsilon^T \tilde{C} \Upsilon \dot{r}(t) + \Omega^2 r(t) = \mathbf{0}$$

C is 2*diagonal matrix comprised of zeta and omega

In order for the equations of motion to be decoupled it is necessary that $\Upsilon^T \tilde{C} \Upsilon = 2\text{diag}(\zeta_i \omega_i)$ be diagonal. Two matrices are simultaneously diagonalized by the same eigenvectors, Υ , if and only if they commute (Bellman 1960), i.e.

$$\tilde{C} \tilde{K} = \tilde{K} \tilde{C}$$

Decoupled in this case is matrices C and K being diagonal which allows for the highlighted equation

$$\Upsilon 2\text{diag}(\zeta_i \omega_i) \Upsilon^T \Upsilon \Omega^2 \Upsilon^T = \Upsilon \Omega^2 \Upsilon^T \Upsilon 2\text{diag}(\zeta_i \omega_i) \Upsilon^T \quad (1.194)$$

$$\Upsilon 2\text{diag}(\zeta_i \omega_i) \Omega^2 \Upsilon^T = \Upsilon \Omega^2 2\text{diag}(\zeta_i \omega_i) \Upsilon^T$$

Pre-multiplying by Υ^T and post-multiplying by Υ results in an identity statement showing that if the matrices are diagonalized by the same eigenvectors, Υ , then they must commute.⁹ Since $\tilde{C} = (M^{-1/2})^T C (M^{-1/2})$ and $\tilde{K} = (M^{-1/2})^T K (M^{-1/2})$,

$$\tilde{C} \tilde{K} = \tilde{K} \tilde{C}$$

$$\begin{aligned} (M^{-1/2})^T C (M^{-1/2}) (M^{-1/2})^T K (M^{-1/2}) &= (M^{-1/2})^T C (M^{-1/2}) (M^{-1/2})^T K (M^{-1/2}) \\ (M^{-1/2})^T C M^{-1} K (M^{-1/2}) &= (M^{-1/2})^T C M^{-1} K (M^{-1/2}) \end{aligned} \quad (1.195)$$

Pre-multiplying by $(M^{1/2})^T$ and post-multiplying by $M^{1/2}$ yields the modified condition for complete decoupling of the equations of motion as given by equation (1.186).

⁸Called *Upsilon*

⁹For proof of the converse, please refer to Bellman (1960).

The modal equations then are

$$I\ddot{\mathbf{r}}(t) + 2\text{diag}(\zeta_i\omega_i)\dot{\mathbf{r}}(t) + \Omega^2\mathbf{r}(t) = \tilde{\mathbf{f}}(t) \quad (1.198)$$

decoupled state equation

A special case of this is *Rayleigh damping*, often referred to as *proportional damping*, where

$$C = \alpha M + \beta K \quad (1.197)$$

In this case the damping ratio is given by

$$\zeta_i = \frac{1}{2} \left(\frac{\alpha}{\omega_i} + \beta \omega_i \right) \quad (1.198)$$

From a modeling standpoint, this results in the ability to explicitly define only two damping ratios (as there are only two parameters to define them: α and β). A more general form is the *Extended Rayleigh damping* (Clough and Penzien 2010) which enables the construction of a damping matrix that can be decoupled the undamped modes with prescribed damping ratios for each mode. A first attempt at generalizing proportional damping so that individual modal damping properties is to apply equation (B.42) yielding a damping matrix

$$C = \sum_i^n \psi_i 2\zeta_i \omega_i \psi_i^T$$

there is a diad here which when summed up for diads yields damping matrix C

However, according to Clough and Penzien (2010) this fails in practice. Instead, the known damping ratio ζ_c of the highest natural frequency of interest, ω_c is used to find the stiffness proportional coefficient so that

$$\zeta_c = \frac{1}{2} \beta \omega_c \quad (1.200)$$

resulting in

$$\beta = \frac{2\zeta_c}{\omega_c} \quad (1.201)$$

Applying equation (1.198), the damping ratio of mode i resulting from this is

$$\zeta_i = \frac{\beta \omega_i}{2} = \zeta_c \frac{\omega_i}{\omega_c} \quad (1.202)$$

if we use the simple damping form

$$C = \beta K \quad (1.203)$$

Presuming that these aren't the values we want, we then use equation (B.42) with (1.202) to correct all of the resulting damping ratios using proscribed damping ratios of ζ_i resulting in

$$C = \beta K + \sum_i^{c-1} \psi_i 2 \left(\zeta_i - \zeta_c \frac{\omega_i}{\omega_c} \right) \omega_i \psi_i^T$$

each freq has specific damping this allows for user to select only freq of interest and their corresponding damping ratios and construct the damping matrix C.

A further advantage of this method is that damping ratios of higher modes increase early with frequency which can benefit stability numerical simulation methods where high frequency modes that are not of interest may nevertheless cause numerical instability.¹⁰

¹⁰Referencing Clough and Penzien, they do not use mass normalized mode shapes. Thus the resulting expression is more complex than given here, yet equivalent.

Similarly, for a complex stiffness model of the form

$$M\ddot{\mathbf{x}}(t) + (K + K'j)\mathbf{x}(t) = \mathbf{f}(t), \quad (1.205)$$

if

$$KM^{-1}K' = K'M^{-1}K \quad (1.206)$$

then the imaginary part of the stiffness matrix, K' , is diagonalized by $\Psi^T K' \Psi = \text{diag}(\eta_i \omega_i^2)$ yielding modal equations of motion

$$\begin{aligned} \mathbf{r} &= \text{in time domain} & I\ddot{\mathbf{r}}(t) + (\Omega^2 + \text{diag}(\eta_i)\Omega^2 j)\mathbf{r}(t) &= \tilde{\mathbf{f}}(t) \\ \mathbf{R} &= \text{freq domain} & I\ddot{\mathbf{r}}(t) + \Omega^2(1 + \text{diag}(\eta_i)j)\mathbf{r}(t) &= \tilde{\mathbf{f}}(t) \end{aligned} \quad (1.207)$$

The *modal frequency response functions* are then obtained by taking the Fourier transform of the appropriate modal equation of motion, either (1.181), (1.196), or (1.207), and solving for

$$\tilde{h}_i(j\omega) = \frac{R_i(j\omega)}{\tilde{F}_i(j\omega)} = \begin{cases} \frac{1}{\omega_i^2 - \omega^2}, & \text{undamped} \\ \frac{1}{\omega_i^2 + 2\zeta_i \omega_i \omega j - \omega^2}, & \text{viscous damping} \\ \frac{1}{\omega_i^2(1 + \eta_i j) - \omega^2}, & \text{complex stiffness damping} \end{cases} \quad (1.208)$$

for single freq bc of subscript i

Consider now the full system modal equations in frequency response form

all modes combined into matrix

$$\mathbf{R}(j\omega) = \tilde{H}(j\omega)\tilde{\mathbf{F}}(j\omega)$$

disp. vect. in F domain

where $\tilde{H}(j\omega) = \text{diag}(\tilde{h}_i)$. Premultiplying by Ψ and substituting $\mathbf{X}(j\omega) = \Psi\mathbf{R}(j\omega)$ and $\tilde{\mathbf{F}}(j\omega) = \Psi^T \mathbf{F}(j\omega)$ into equation (1.209) gives

$$\Psi\mathbf{R}(j\omega) = \Psi\tilde{H}(j\omega)\Psi^T \mathbf{F}(j\omega) \quad \text{mass matrix normalized mode shapes, in physical space (matrix)}$$

$$\mathbf{X}(j\omega) = (\Psi\tilde{H}(j\omega)\Psi^T) \mathbf{F}(j\omega) \quad (1.210)$$

$$= H(j\omega)\mathbf{F}(j\omega)$$

Transfer funct. matrix relating input F to output X

F-tilde is the forcing vector in F domain but in modal coordinates. If it was bar and not tilde then its complex conj.

Force vector in F domain but physical coordinates

After doing all the substitutions get this result.

Column of H corresponds to where sample is being forced.
Rows of H corresponds to where displacement is being measured.
H is also symmetric

where $H(j\omega)$ is the matrix of transfer functions between forces and displacements in physical coordinates. Expressing Ψ using equation (1.179), and considering equation (1.210)

$$\begin{aligned}
 H(j\omega) &= [\psi_1 \ \psi_2 \ \psi_3 \ \cdots \ \psi_n] \tilde{H}(j\omega) \begin{bmatrix} \psi_1^T \\ \psi_2^T \\ \psi_3^T \\ \vdots \\ \psi_n^T \end{bmatrix} \\
 &= [\psi_1 \tilde{h}_1(j\omega) \ \psi_2 \tilde{h}_2(j\omega) \ \psi_3 \tilde{h}_3(j\omega) \ \cdots \ \psi_n \tilde{h}_n(j\omega)] \begin{bmatrix} \psi_1^T \\ \psi_2^T \\ \psi_3^T \\ \vdots \\ \psi_n^T \end{bmatrix} \quad (1.211) \\
 &= \sum_{i=1}^n \psi_i \psi_i^T \tilde{h}_i(j\omega) \\
 &= \sum_{i=1}^n {}_i A \tilde{h}_i(j\omega)
 \end{aligned}$$

The variable, ${}_i A = \psi_i \psi_i^T$, represents the outer product of ψ_i with itself and is called the *modal constant* or *residue*. That is,

$${}_i A = \psi_i \psi_i^T = \begin{bmatrix} \psi_{1,i} \psi_{1,i} & \psi_{1,i} \psi_{2,i} & \cdots & \psi_{1,i} \psi_{l,i} & \cdots & \psi_{1,i} \psi_{n,i} \\ \psi_{2,i} \psi_{1,i} & \psi_{2,i} \psi_{2,i} & \cdots & \psi_{2,i} \psi_{l,i} & \cdots & \psi_{2,i} \psi_{n,i} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \psi_{m,i} \psi_{1,i} & \psi_{m,i} \psi_{2,i} & \cdots & \psi_{m,i} \psi_{l,i} & \cdots & \psi_{m,i} \psi_{n,i} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{n,i} \psi_{1,i} & \psi_{n,i} \psi_{2,i} & \cdots & \psi_{n,i} \psi_{l,i} & \cdots & \psi_{n,i} \psi_{n,i} \end{bmatrix} \quad (1.212)$$

Thus an individual frequency response function between an input m and output, l , is

$$H_{m,l} = \sum_{i=1}^n \psi_{m,i} \psi_{l,i} \tilde{h}_i(j\omega) = \sum_{i=1}^n {}_i A_{m,l} \tilde{h}_i(j\omega) \quad (1.213)$$

which is typically written as

$$H_{m,l} = \sum_{i=1}^n \frac{{}_i A_{m,l}}{\omega_i^2 + 2\zeta_i \omega_i \omega_j - \omega^2} \quad (1.214)$$

Using equation (1.118), this may be written as the sum of first order complex FRF

$$H_{m,l} = \sum_{i=1}^n \frac{i B_{m,l}}{j\omega - \lambda} + \frac{i \bar{B}_{m,l}}{j\omega - \bar{\lambda}} \quad (1.215)$$

need subscripts for
lambdas

where, as before, $iB_{m,l} = \frac{iA_{m,l}}{\lambda - \lambda}$ and $\lambda = -\zeta\omega_n - j\omega_n\sqrt{1 - \zeta^2}$. An alternative representation of the frequency response function can be obtained directly from equations (1.205) or (1.185). Taking the Fourier transform of each equation,

$$(-\omega^2 M + j\omega C + K) \mathbf{X}(j\omega) = \mathbf{F}(j\omega), \quad \text{viscous} \quad (1.216a)$$

$$(-\omega^2 M + (K + K'j)) \mathbf{X}(j\omega) = \mathbf{F}(j\omega), \quad \text{hysteretic} \quad (1.216b)$$

Since the analysis is the same for either form of damping, we show only the viscous case for the sake of clarity. The frequency response function matrix is then

$$\mathbf{X}(j\omega) = (-\omega^2 M + j\omega C + K)^{-1} \mathbf{F}(j\omega) = H(j\omega) \mathbf{F}(j\omega), \quad \text{viscous} \quad (1.217)$$

Since

$$H(j\omega) = (-\omega^2 M + j\omega C + K)^{-1} \quad (1.218)$$

using the definition of the inverse of a matrix of

$$H_{l,m}(j\omega) = \frac{\text{adj}(-\omega^2 M + j\omega C + K)_{l,m}}{\det(-\omega^2 M + j\omega C + K)} \quad (1.219)$$

used to determine where 0's are.

Where

$$\text{adj}(-\omega^2 M + j\omega C + K)_{l,m} = 0 \quad (1.220)$$

the frequency response function will exhibit a *zero* or *anti-resonance* (Mottershead 1998). Recall that the m, l element of the adjoint of a matrix is the determinant of the remaining matrix when the l th row and the m th column are removed, multiplied by -1^{l+m} , called the l, m cofactor. When $l = m$, this is equivalent to constraining the equations of motion such that $x_l = 0$.

Example 1.9.1 A system is defined by

$$M\ddot{\mathbf{x}} + C\dot{\mathbf{x}} + K\mathbf{x} = \mathbf{0} \quad (1.221)$$

where

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 4 \end{bmatrix} \quad (1.222)$$

Find the poles and zeros of the frequency response function, as well as the FRF itself, between the first and third degrees of freedom.

Solution:

The poles (natural frequencies) of the system are given by the solution of

$$\begin{aligned} \det(-M\omega^2 + K) &= 0 \\ -\omega^6 + 9\omega^4 - 23\omega^2 + 13 &= 0 \end{aligned} \quad (1.223)$$

the solution of which is

$$\omega_n \approx \pm 0.8864, \pm 1.8813, \pm 2.1622 \quad (1.224)$$

If you want to know that system response will be if we add a fixture to a point on the structure we can just excite system at that point and sense at that point and it will let us know what the structure natural frequencies will be if we fixed it at that location

collocation is where excitation is applied at a point and the response is also measured at that location. In other words we are looking at $H(i,i)$, both indicies are same

The zeros of this specific frequency response function are given by

$$\begin{aligned} \text{adj}_{1,3}(K - \omega^2 M) &= 0 \\ \det \left(\begin{bmatrix} -1 & -1 \\ 3 & -1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \omega^2 \right) &= 0 \\ 4 - \omega^2 &= 0 \end{aligned} \quad (1.225)$$

Solving for ω gives zeros at

$$\omega = \pm 2 \quad (1.226)$$

The mass normalized mode shapes are

$$\psi_1 = \begin{bmatrix} 0.7558 \\ 0.5207 \\ 0.3971 \end{bmatrix}, \quad \psi_2 = \begin{bmatrix} 0.6318 \\ -0.7392 \\ -0.2332 \end{bmatrix}, \quad \text{and} \quad \psi_3 = \begin{bmatrix} 0.1721 \\ 0.4271 \\ -0.8877 \end{bmatrix} \quad (1.227)$$

Using equation (1.213), the frequency response function is

$$\begin{aligned} H_{1,3}(j\omega) &= \sum_{i=1}^3 \psi_{1,i} \psi_{3,i} \tilde{h}_i(j\omega) \\ &= \frac{0.3001}{0.7857 - \omega^2} + \frac{-0.1473}{3.5392 - \omega^2} + \frac{-0.1528}{4.6751 - \omega^2} \end{aligned} \quad (1.228)$$

Alternatively, applying equation (1.219), the FRF can also be represented as

$$H_{1,3}(j\omega) = \frac{-\omega^2 + 4}{-\omega^6 + 9\omega^4 - 23\omega^2 + 13} \quad (1.229)$$

1.10 Problems

1. Use the Euler equations to represent $x(t) = Ae^{-\zeta\omega_n t} \sin \omega_d t$ in complex form.
2. Given $x(0) = x_0$ and $\dot{x}(0) = v_0$, obtain the free response of a single degree of freedom under damped system in real form.
3. Given $x(0) = x_0$ and $\dot{x}(0) = v_0$, obtain the free response of a single degree of freedom under damped system in complex exponential form.
4. A single degree of freedom system with $m = 10 \text{ kg}$, $c = 20 \text{ kg/sec}$, and $k = 1000 \text{ N/m}$ is excited by a force of 10 N with a frequency of 10 Hz . What is the amplitude of the response and the phase lag?
5. Find the frequency response function for the system defined by the differential equation

$$10\ddot{x} + .01\dot{x} + 1000x = f(t)$$

6. A system is governed by the following equation of motion:

$$m\ddot{x} + c\dot{x} + kx = cy$$

where $y(t) = Y \sin(\omega_b t)$. Given $m = 10 \text{ kg}$, $c = 1.0 \text{ kg/s}$, and $k = 10,000 \text{ kg/s}^2$.

46 REVIEW OF FUNDAMENTALS: INTRODUCTION TO VIBRATION ANALYSIS

- (a) Sketch the FRF $\frac{X(j\omega)}{Y(j\omega)}$ for $0 < \frac{\omega_b}{\omega_n} < 5$ (both magnitude and phase). Use decibels ($20 \log_{10}$) for the magnitude. *Label limits and key values*

- (b) Fill in the following table (*beware of units!*):

Y (m)	f_b (Hz)	$ X $ (m)	$\angle X$ (deg)
10	0		
2	5		
1	10		

7. Determine the Fourier Series representation of the function for which a single cycle is defined by

$$f(t) = \begin{cases} 0, & 0 < t < 2 \\ 1, & 2 < t < 3 \end{cases}$$

8. Determine the Fourier Series representation of the function for which a single cycle is defined by

$$f(t) = \delta(t), \quad -1 < t < 1$$

9. Determine the Fourier Series representation of the function for which a single cycle is defined by

$$f(t) = \delta\left(t - \frac{\epsilon}{2}\right) + \delta\left(t + \frac{\epsilon}{2}\right), \quad -1 < t < 1$$

10. Find the Fourier series for a ramp $f(t) = t$ for $0 < t < 1$, repeating every 1 second.

11. Find the Fourier series for $f(t) = t$ for $-1/2 < t < 1/2$, repeating every 1 second.

12. Prove that

$$\frac{1}{T} \int_{-\infty}^{\infty} f(t)^2 dt = \sum_{n=-\infty}^{\infty} F_n \bar{F}_n$$

13. Find the Fourier transform of $\cos(\Omega t)$.

14. Determine whether the compliance, mobility, or inertance frequency response functions form a circle in the complex plane.

15. Given

$$M = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \text{ and } K = \begin{bmatrix} 35 & -35 \\ -35 & 35 \end{bmatrix}$$

find the natural frequencies and mode shapes of the system. Mass normalize the mode shapes. Describe the natural motions of the mode/frequency pairs.

16. Given

$$M = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 4 \end{bmatrix}, \text{ and } K = \begin{bmatrix} 35 & -35 & 0 \\ -35 & 45 & -10 \\ 0 & -10 & 20 \end{bmatrix}$$

find the natural frequencies and mode shapes of the system. Mass normalize the mode shapes. Describe the natural motions of the mode/frequency pairs.

17. For the system of problem 16 determine H_{11} , H_{12} , and H_{13} using equation (1.213). Plot the FRF for a frequency range including all 3 natural frequencies.
18. For the system of problem 16 determine H_{11} , H_{12} , and H_{13} using equation (1.219). Plot the FRF for a frequency range including all 3 natural frequencies.
19. For the system of problem 16 with

$$C = \begin{bmatrix} 0.02 & 0.01 & 0 \\ 0.01 & 0.03 & 0.01 \\ 0 & 0.01 & 0.04 \end{bmatrix}$$

determine H_{11} , H_{12} , and H_{13} using equation (1.213). Plot the FRF for a frequency range including all 3 natural frequencies.

20. Determine the zeros of the system of problem 16 for H_{11} , H_{12} , and H_{13} .
21. Calculate the residue matrices for problem 16.
22. A MDOF system at rest with

$$\Psi = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

no damping, and natural frequencies of 3 rad/sec and 10 rad/sec is excited by a force of

$$\mathbf{F}(t) = \begin{bmatrix} \frac{1}{\sqrt{2}}\delta(t) \\ \frac{1}{\sqrt{2}}\delta(t) \end{bmatrix}$$

Find $\mathbf{x}(t)$.

23. Write a computer code to plot the magnitude and phase of a MDOF proportionally damped system using the method of equation (1.213). Apply this code to the system defined in example 1.9.1.
24. Write a computer code to plot the magnitude and phase of a MDOF proportionally damped system using the method of equation (1.219). Apply this code to the system defined in example 1.9.1.
25. Given the matrix

$${}_1A = \boldsymbol{\psi}_1 \boldsymbol{\psi}_1^T = \begin{bmatrix} 0.08450 & 0.00357 & 0.27811 \\ 0.00357 & 0.00015 & 0.01175 \\ 0.27811 & 0.01175 & 0.91535 \end{bmatrix} \quad (1.230)$$

find the first mode shape.

2

Introduction to Probability Theory

The foundation of vibration testing lies in probability theory, in which signals containing useful system information must be processed appropriately to allow the extraction of key model parameters. In the analytical case, where we know the forcing function and solve for the response, or vice versa, statistical methods are not necessary, since everything of importance can be read directly or easily obtained via observation of time histories. Real behavior includes sensor error, noise, system nonlinearities, etc. that preclude such basic approaches. As a result, probability theory is applied to wash out noise, and enhance “true” information in the collected data.

The simplest statistical values used in random vibration analysis are the ‘calculated’, ‘estimated’ or ‘sample’, mean value, given by

$$\bar{x} = \hat{\mu}_x = \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt, \quad \text{or} \quad \frac{1}{N} \sum_{i=1}^N x_i \quad (2.1)$$

the variance, $\hat{\sigma}_x^2$,¹ given by

$$v_x = \hat{\sigma}_x^2 = \frac{1}{T} \int_{-T/2}^{T/2} (x(t) - \hat{\mu}_x)^2 dt, \quad \text{or} \quad \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu}_x)^2 \quad (2.2)$$

the skewness,

$$s_x = \frac{1}{T\sigma_x^3} \int_{-T/2}^{T/2} (x(t) - \hat{\mu}_x)^3 dt, \quad \text{or} \quad \frac{1}{N\sigma_x^3} \sum_{i=1}^N (x_i - \hat{\mu}_x)^3 \quad (2.3)$$

and the kurtosis value, $\hat{\kappa}_x$, given by

$$\hat{\kappa}_x = \frac{1}{T\sigma_x^4} \int_{-T/2}^{T/2} (x(t) - \hat{\mu}_x)^4 dt, \quad \text{or} \quad \frac{1}{N\sigma_x^4} \sum_{i=1}^N (x_i - \hat{\mu}_x)^4 \quad (2.4)$$

¹ σ_x is commonly referred to as the standard deviation.

These represent the center, the spread, the skewness, and the extremeness of the data, respectively. These are referred to as ‘sample’, ‘estimated’, or ‘calculated’ values because they pertain to the set of data being analyzed, not the *process* that generated the data. The $\hat{\cdot}$ symbol is used to denote a calculated quantity. The mean value, $\hat{\mu}_x$, represents an *estimate* of *what* or *where* the signal is over a period of time. It is an estimate because if you perform the experiment again, you won’t likely get the same value. It is an estimate the center of the signal, analogous to the centroid of an object. The right side of equation (2.1) is the *first moment* of $x(t)$ divided by the period T or number of samples N .²

The root mean square value, $\hat{\sigma}_x$, is an estimate of how spread out the signal is, and is related to the amplitude of the signal. It is the *second moment* of $x(t)$ divided by the period T or number of samples N . For a sine wave of amplitude A , $\hat{\sigma}_x = 1/\sqrt{2}A = 0.7071A$, while for a square wave, $\hat{\sigma}_x = A^3$. The kurtosis value is a measure of how “spiky” the signal is and is a better measure of how much of the signal is extreme in nature.⁴ For a sine wave of amplitude A , $\hat{\kappa}_x = 1.5$, while for a square wave of amplitude A , $\hat{\kappa}_x = 1$. The relatively higher value of $\hat{\kappa}_x = 1.5$ for the sine wave as compared to the square wave is an indication that the sine wave contains more extreme values as compared to the square wave (for which no subset of the function can be considered extreme in relation to the rest of the function). Thus the kurtosis value is often used as an automated measure of extreme occurrences in a signal, being more robust than simply observing the maximum values because of the effect of averaging over time. Its primary use in vibration testing is in health monitoring, as will be discussed in chapter 10.

Because these examples are known functions (not random) with known periods, the $\hat{\cdot}$ are not necessary; the values can be determined precisely. After all, measured signals are usually known or we wouldn’t be measuring them. Estimates of the mean, variance, and kurtosis values are more consistent when T or N are increased in the appropriate formulae.

2.1 Random Processes

Dictionary definitions of randomness are similar to “lacking a definite plan, purpose, or pattern” (Anonymous 1981). This is, however, inadequate from a mathematical perspective in that it does not address that the variability occurs in both time as well as *at* time. For example, consider a random sequence in time generated by picking a number between 1 and 10 (assuming that a person can behave randomly, a dubious assumption at best). Eventually a list of numbers would be generated such that there would be no pattern, discernible or not, between one number and the next (or separated by an index greater than 1). This is a random sequence in, or *over* time and a concept that is quite familiar. Random *at* time seems similar, but is in fact quite different. In order to understand random at time, it is useful to invoke the concept of a time machine so we can start our experiment under precisely the same conditions for each experiment. Consider now generating a set of numbers by picking a number between 1 and 10, resetting time to $t = 0$, and repeating. The set of numbers generated thusly have no

²Recall that to find a center of area, you divide the first moment of area by the total area.

³The $\hat{\cdot}$ has been left off because this is an actual value that can be derived using the probability density function and equation (2.17).

⁴This is a fourth moment quantity. The third moment quantity that we skipped is a measure of the skewness of the data and is relevant for systems with non-symmetric nonlinearities such as quadratic stiffnesses and discontinuous stiffnesses.

temporal relation to one another, and as a class appear to be unrelated to one another. That is, each experiment does not influence the next or any subsequent (but not later!) experiments. We've effectively generated a 5th dimension, 3 in space, 1 in time, and the 5th being parallel (to time) *realizations* of the process⁵. Each time we reset time and perform the experiment, we are generating an additional realization. If each of these realizations are unrelated, then the process is mathematically random. If each of these realizations are the same, then we have a process that is only *temporally* random. Creating a process that yields mathematically random results is impossible without our time machine. However, the concept of randomness across realizations is an important one when we consider signal noise. We will often use this concept, along with others, to reset the clock on our experiments for use as a poor-man's time machine.

This set of realizations is collectively referred to as an *ensemble*. If we have an ensemble of signals, we can process them to determine properties such as the ensemble mean, ensemble standard deviation, and more critically, the ensemble probability distribution function, $p_x(x)$. The ensemble mean can be calculated at any time t such that

$$\hat{\mu}_x(t) = \frac{1}{N} \sum_{i=1}^N x_i(t) \quad (2.5)$$

This is also an estimate of the *expected value*, denoted as $E[\cdot] = \mu_x(t)$ and is called the *sample mean* or *calculated mean*. The variance (the square of the standard deviation) can likewise be estimated as

$$\hat{\sigma}_x^2(t) = \hat{\sigma}_x^2(t) = \frac{1}{N-1} \sum_{i=1}^N (x_i(t) - \hat{\mu}_x(t))^2 \quad (2.6)$$

and is called the *sample variance* or *calculated variance*. Other properties can likewise be determined using their discrete definitions, at any given time. This would be sufficient except for the problem that we cannot go back in time and generate additional realizations for use in these expressions. As a result, we instead will, by some means, presume or obtain the *probability density function* (PDF), $p_x(x)$.

Having the PDF allows us to determine the expected value of a signal without having all of its realizations (the real number of which is ∞ anyway). Using the PDF as a weighting function, the *expected value* of $x(t)$ at time t is

$$\mu_x(t) = E[x(t)] = \int_{-\infty}^{\infty} p_x(x) x(t) dx \quad (2.7)$$

This is not a measured value, but the true mean of the process generating x , although it is a value that we never know for a real system. Similarly, the expected value of any function $f(x)$ can be obtained using

$$E[f(x(t))] = \int_{-\infty}^{\infty} p_x(x) f(x(t)) dx \quad (2.8)$$

Needs to go elsewhere.

⁵This concept is prevalent in modern science fiction in the form of “parallel universes”.

2.2 Probability Distribution and Density Functions

Consider some function, random or not, x . For our discussion, x may be a function of time, or a sequence of values. Both will be considered in future developments. The likelihood that x at any instance (time or index) has a value less than some arbitrary value a is denoted as

$$P_x(a) = \text{Prob}[x \leq a] \quad (2.9)$$

The function P_x is called the *cumulative distribution function* (CDF) or *probability distribution function*.

Example 2.2.1 Consider, for example, rolling a die where the face value of the die is x . Then $P_x(1) = \frac{1}{6}$, $P_x(2) = \frac{2}{6}, \dots$, $P_x(6) = \frac{6}{6} = 1$. This is an example of a discrete CDF since it cannot take on values other than $\frac{x}{6}$ where x must be an integer. The CDF is a monotonically increasing function with a minimum value of 0, and a maximum value of 1. Thus

$$\begin{aligned} P_x(-\infty) &= 0 \\ P_x(\infty) &= 1, \quad \text{and} \\ P_x(x_2) &> P_x(x_1) \quad \text{where} \quad x_2 > x_1. \end{aligned} \quad (2.10)$$

Example 2.2.2 Consider now the case where $0 < x < 1$, and the likelihood of x having a value less than a is a itself, e.g. there is an 80% chance that $x(i) \leq 0.8$. Thus

$$P_x(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & 1 < x \end{cases} \quad (2.11)$$

Consider now the probability density function (PDF) defined as

$$p_x(x) = \frac{dP_x(x)}{dx} \quad (2.12)$$

For our current function $P_x(x)$ we get

$$p_x(x) = \begin{cases} 0, & x < 0 \\ 1, & 0 \leq x \leq 1 \\ 0, & 1 < x \end{cases} \quad (2.13)$$

The PDF of this example is in form of a *uniform* probability density function as illustrated in table 2.1 with x limited such that $0 \leq x \leq 1$. As a result of the properties of the CDF,

$$p_x(x) \geq 0 \quad (2.14a)$$

$$\int_{-\infty}^{\infty} p_x(x) dx = 1, \text{ and} \quad (2.14b)$$

$$P_x(x) = \int_{-\infty}^x p_x(\xi) d\xi. \quad (2.14c)$$

Table 2.1 Common Probability Distribution Functions

Type	PDF	CDF	Definition
Gaussian			$p_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$
Truncated Gaussian			$p_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \text{ Const, } a \leq x \leq b, \text{ where Const is defined}$ so $\int_a^b p_x(x) dx = 1$
Rayleigh			$p_x(x) = \frac{x}{\sigma_y^2} e^{-\frac{x^2}{2\sigma_y^2}}$
Chi-Square (χ_n^2)			$p_{\chi_n^2}(y) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-y/2} y^{(n/2)-1},$ $y = \chi_n^2 = \sum_{i=1}^n z_i^2$
t			$p_t(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} \left[1 + \frac{t^2}{n}\right]^{-(n+1)/2},$ where $t = \frac{z}{\sqrt{y/n}}$, z is zero-mean Gaussian, y is χ_n^2
Uniform			$p_x(x) = \frac{1}{b-a}, \quad a \leq x \leq b$
Sine/Cosine			$p_x(x) = \frac{1}{\pi\sqrt{A^2-x^2}}, \quad x < A$

Of course, we will need data to generate our PDF. Assumptions regarding the *form* of the PDF allow significant understanding to be obtained via the use of only a subset of parameters, instead of using raw data. Use of such an approach can result in much better results than could realistically be obtained using raw data (which is generally costly to generate) provided that the type of distribution is known, and that estimates of the key parameters describing the distribution can be obtained accurately. Some distributions commonly encountered in vibration are shown in Table 2.1.

Having the PDF allows us to determine the *expected value* of a signal, defined as

$$\mu_x = E[x] = \int_{-\infty}^{\infty} p_x(\xi) \xi d\xi, \quad (2.15)$$

where ξ is used here as a dummy variable. The PDF acts as a weighting function, representing the likelihood of x having a given value. Calculation of the mean can be recognized as analogous to calculation of a geometrical centroid. Thus the integral of equation 2.12 is the first moment of the distribution of $p_x(x)$ in place of area. Because the integral of $p_x(x)$ is 1, the first moment is equivalent to the mean, or center value. Similarly, the expected value of any function $f(x)$ can be obtained using

$$E[f(x)] = \int_{-\infty}^{\infty} p_x(\xi) f(\xi) d\xi \quad (2.16)$$

Using this definition, the variance of a process is defined as

$$v_x = \sigma_x^2 = E[(x - \mu_x)^2] = \int_{-\infty}^{\infty} p_x(\xi) (\xi - \mu_x)^2 d\xi \quad (2.17)$$

where σ_x can be recognized as analogous to the radius of gyration for an area represented by $p_x(x)$. The variance is calculated from the second moment of $p_x(x)$ relative to its mean.

Consider finding these properties pertaining to the sine function. The PDF for a sine function is given in Table 2.1. Of course, a sine wave is not random. We know exactly what value to expect as long as we know the frequency and phase (e.g. $x = A \sin(\omega t + \phi)$). Applying equation (2.16)

$$\mu_x = E[x] = \int_{-\infty}^{\infty} p_x(\xi) \xi d\xi = \int_{-\infty}^{\infty} \frac{1}{\pi \sqrt{A^2 - \xi^2}} \xi d\xi = 0 \quad (2.18)$$

and

$$\begin{aligned} \sigma_x^2 &= E[(x - \mu_x)^2] = \int_{-\infty}^{\infty} \frac{1}{\pi \sqrt{A^2 - \xi^2}} \xi^2 d\xi \\ &= \int_{-A}^A \frac{1}{\pi \sqrt{A^2 - \xi^2}} \xi^2 d\xi = \frac{A^2}{2} \end{aligned} \quad (2.19)$$

These are not nearly as easy to evaluate as equations (2.1) and (2.2), but then again, expected values are not intended for deterministic cases where such equations are available. They are, however, more accurate in the sense that they better represent the properties of the process generating the data that would be used in equations (2.1) and (2.2).

2.3 Moment Generating Functions

An alternate method of obtaining the mean and variance of a PDF is to apply the Laplace transform. Instead of using equations (2.15) and (2.17), we will apply equation (1.50) to the PDF. For consistency, we will not integrate over time, so the Laplace transform of a PDF $p_x(x)$ is

$$\mathcal{M}_x(s) = \int_{-\infty}^{\infty} e^{sx} p_x(x) dx = E[e^{sx}] \quad (2.20)$$

Laplace transform
def.

If we recognize

$$e^{sx} = \sum_{k=0}^{\infty} \frac{1}{k!} (sx)^k = 1 + sx + \frac{1}{2!} (sx)^2 + \dots \quad (2.21)$$

then we can recognize that

$$\begin{aligned} \mathcal{M}_x(s) &= \int_{-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \frac{1}{k!} (sx)^k \right) p_x(x) dx \\ &= \sum_{k=0}^{\infty} \left(\frac{1}{k!} \int_{-\infty}^{\infty} (sx)^k p_x(x) dx \right) \\ &= 1 + sE[x] + \frac{1}{2!} s^2 E[x^2] + \dots \end{aligned} \quad (2.22)$$

taylor series
expansion

Thus, to find $E[x]$, once we have $\mathcal{M}_x(s)$, we can apply

$$\mu_x = E[x] = \frac{d\mathcal{M}_x(s)}{ds} \Big|_{s=0} \quad (2.23)$$

The value $\mu_x = E[x]$ is called the *first moment* of x . The second moment is calculated using

$$E[x^2] = \frac{d^2 \mathcal{M}_x(s)}{ds^2} \Big|_{s=0} \quad (2.24)$$

It is important to recognize, though, that $v_x = \sigma_x^2 = E[x^2]$ only if $\mu_x = 0$. Otherwise

$$\begin{aligned} v_x = \sigma_x^2 &= E[(x - \mu_x)^2] = E[x^2 - 2\mu_x x + \mu_x^2] = E[x^2] - E[2x\mu_x] + E[\mu_x^2] \\ &= E[x^2] - 2\mu_x E[x] + E[\mu_x^2] = E[x^2] - 2\mu_x \mu_x + \mu_x^2 = E[x^2] - \mu_x^2 \end{aligned} \quad (2.25)$$

Having $\mathcal{M}_x(s)$, we can thus obtain v_x as

$$v_x = \frac{d^2 \mathcal{M}_x(s)}{ds^2} \Big|_{s=0} - \left(\frac{d\mathcal{M}_x(s)}{ds} \Big|_{s=0} \right)^2 \quad (2.26)$$

Because of this simplification in finding expected values, $\mathcal{M}_x(s)|_{s=0}$ is referred to as the *moment generating function* (Bendat and Piersol 2000; Wirsching et al. 1995).

Further, note that the moment generating function gives us a simplification in understanding the sum of variables (Parzen 1972). Consider a variable y defined as $y = x_1 + x_2$ where x_1 and x_2 are independent. Then using equation (2.20)

$$\mathcal{M}_y(y) = E[e^{ys}] = E[e^{x_1 s} e^{x_2 s}] = E[e^{x_1 s}] + E[e^{x_2 s}] \quad (2.27)$$

This fact is illustrated in section 2.5.2 for Gaussian distributions.

Example 2.3.1 *Find the moment generating function for a normal (Gaussian) distribution. A normal distribution has a probability density function of*

$$p_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (2.28)$$

The moment generating function is

$$\mathcal{M}_z(s) = E[e^{sz}] = \int_{-\infty}^{\infty} e^{sz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sz - \frac{z^2}{2}} dz \quad (2.29)$$

Noting that

$$sz - \frac{z^2}{2} = -\frac{1}{2}(z - s)^2 + \frac{s^2}{2} \quad (2.30)$$

equation (2.29) becomes

$$\mathcal{M}_z(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{s^2}{2}} e^{-\frac{1}{2}(z-s)^2} dz = e^{\frac{s^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-s)^2} dz = e^{\frac{s^2}{2}} \quad (2.31)$$

because

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-s)^2} dz = 1 \quad (2.32)$$

is just the area under the PDF of a normal distribution with a mean value of s .

2.4 Transformation of distribution coordinates

Consider a PDF written in terms of variables x_1, x_2, \dots for which it is desired to write the PDF in terms of variables y_1, y_2, \dots . Because the coordinates y_i are stretched/compressed relative to the coordinates x_i , we cannot simply substitute $x_i = f(y_1, y_2, \dots)$ without rescaling the CDF. Recall that the CDF, with the single variable definition given by equation (2.14c), is akin to a volumetric integral. The entire integral under the PDF must be equal to 1. Thus if the coordinates aren't equally scaled in length, then the magnitude of the PDF must be adjusted to compensate.

Consider the PDF of a uniform distribution in terms of two variables, x and y , illustrated in Figure 2.1. Clearly from the figure $y = \frac{1}{2}x$. If the area under the PDF for each infinitesimal

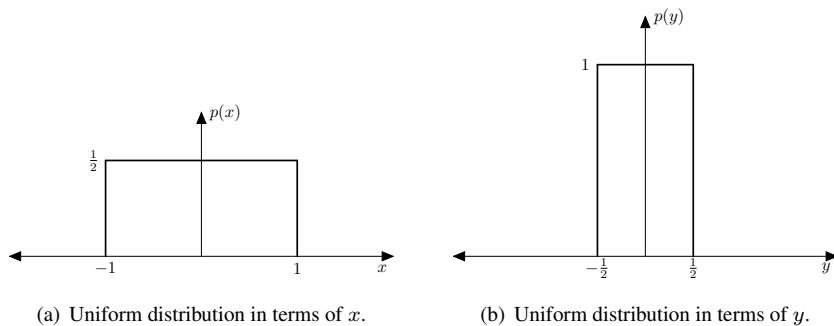


Figure 2.1 A uniform PDF as a function of variable x and y .

dy corresponding to a segment dx is to be equal, then

$$p(x)dx = p(y)dy \quad (2.33)$$

Thus,

$$p(y) = p(x) \frac{dx}{dy} = \frac{1}{2}2 = 1 \quad (2.34)$$

Here the term $\frac{dx}{dy}$ is called the *Jacobian*, J , and has a wide range of applications including such topics as finite elements, CFD, and optimization.

Two dimensions are sufficient to illustrate the more general n -variable case. Consider then the situation of transforming a PDF in two variables x_1 and x_2 to two other variables, y_1 and y_2 . Since the volumes under the PDF surfaces must be piecewise equal

column volume

→

p(x_1, x_2) $dx_1 dx_2$ = p(y_1, y_2) $dy_1 dy_2$

column volume

(2.35)

The relationship between the infinitesimal areas $dx_1 dx_2$ and $dy_1 dy_2$ is given by (Jeffrey 2002)

$$dx_1 dx_2 = |J| dy_1 dy_2 \quad (2.36)$$

where

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} \quad (2.37)$$

Thus, substituting equation (2.36) and (2.37) into equation (2.35) yields

$$p_{y_1 y_2 \dots}(y_1, y_2, \dots) = p_{x_1 x_2 \dots}(x_1, x_2, \dots) |J| \quad (2.38)$$

Consider the univariate case where it is desired to obtain the PDF of function $y(x)$ where x is a random variable. The probability of y falling within the region Δy is the same as the probability of x falling within the associated range Δx . That is,

$$P_y(y(x_2)) - P_y(y(x_1)) = P_x(x_2) - P_x(x_1) \quad (2.39)$$

Using equation (2.55c)

$$\int_{y(x_1)}^{y(x_2)} p_y(y(x)) dy = \int_{x_1}^{x_2} p_x(x) dx \quad (2.40)$$

If we consider the limit as $x_2 - x_1 \rightarrow 0$, this becomes

$$p_y(y(x)) dy = p_x(x) dx \quad (2.41)$$

and thus

$$p_y(y) = \frac{p_x(x)}{|dy/dx|} \quad (2.42)$$

Here the absolute value is imposed because PDFs cannot take on negative values. Special care must be taken to account for when $\frac{dy}{dx}$ is negative, as shown in the following example. To complete the process, we must then substitute $x = x(y)$ into the right hand side of equation (2.42).

Example 2.4.1 Consider finding $p_y(y)$ where $y = \sin(\omega t + \phi)$ and $p_t(t) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-\frac{(t)^2}{2\sigma_t^2}}$. Hint:Bendat and Piersol

Example 2.4.2 Consider finding $p_y(y)$ where $y = x^2$ and $p_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_x^2}}$.

$$\frac{dy}{dx} = 2x \quad (2.43)$$

Then using equation (2.42)

$$p_y(y) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_x^2}} \frac{1}{2x} \quad (2.44)$$

Note that this can only be true for $x > 0$ because a PDF must be non-negative for all values of the variable. Substituting $x = \sqrt{y}$ yields

$$p_y(y) = \frac{1}{2\sigma_x \sqrt{2\pi} \sqrt{y}} e^{-\frac{y}{2\sigma_x^2}} \quad (2.45)$$

where $0 \leq y$ and for $0 \leq x$. If we allow x to be negative as well, then since $p_y(x^2) = p_y((-x)^2)$, $p(y)$ doubles to become

$$p_y(y) = \frac{1}{\sigma_x \sqrt{2\pi} \sqrt{y}} e^{-\frac{y}{2\sigma_x^2}} \quad (2.46)$$

due to the contribution from $x < 0$.

2.5 Normal (Gaussian) Distribution

A *Gaussian random process* is one for which the ensemble probability density function is always Gaussian, i.e.

$$p_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \quad (2.47)$$

A zero mean can always be achieved through a change of variables so that $y = x - \mu_x$ is zero mean Gaussian if x is Gaussian with a mean of μ_x . Further, a normalized Gaussian distribution can be obtained by defining

$$z = \frac{x - \mu_x}{\sigma_x} \quad (2.48)$$

Then the normalized Gaussian distribution is

$$p_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (2.49)$$

Appendix D.1 provides a coarse table of PDF and CDF values of the normal distribution for quick reference. More detailed tables can be generated by modern mathematical tools such as Octave (Eaton 2002). The value of z corresponding to a particular value of its cumulative distribution function is typically denoted as z_α . Thus

$$\text{Prob}[z \leq z_\alpha] = P(z_\alpha) = 1 - \alpha \quad (2.50)$$

Example 2.5.1 Find the value of z which 95% of all values of z are below.

This value is the $z_{.05}$ value. Since there is no closed form expression for $P(z)$, we must use

$$P(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi \quad (2.51)$$

Using the command `norminv` in MATLAB, `normal_inv` in Octave, or the table of Appendix D.1 yields $z_{.05} = 1.65$.

2.5.1 Joint Distributions

Very often the measured outcomes of an experiment involves two variables. We can thus consider the distribution of each not only independently, but with respect to one another. The likelihood that $x \leq a$ and $y \leq b$ is denoted as

$$P_{xy}(a, b) = \text{Prob}[x \leq a \cap y \leq b] \quad (2.52)$$

where \cap simply means that both conditions must be met and is meant to be read as “and”. The function P_{xy} is called the *joint cumulative distribution function*, and is defined such that

$$\lim_{a \rightarrow \infty, b \rightarrow \infty} P_{xy}(a, b) = 1 \quad (2.53)$$

The joint density function defined as

$$p_{xy}(x, y) = \frac{\partial^2 P_{xy}(x, y)}{\partial x \partial y} \quad (2.54)$$

and similar to the univariate case of section 2.2

$$p_{xy}(x, y) \geq 0 \quad (2.55a)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{xy}(x, y) dx dy = 1, \text{ and} \quad (2.55b)$$

$$P_{xy}(x, y) = \int_{-\infty}^y \int_{-\infty}^x p_x(\xi_1, \xi_2) d\xi_1 d\xi_2 \quad (2.55c)$$

In addition, we can define the *marginal* PDF as a function of $p_{xy}(x, y)$ for either variable. The marginal PDF is the univariate PDF in either x or y independent of the other variable and can be obtained from

$$p_x(x) = \int_{-\infty}^{\infty} p_{xy}(x, y) dy \quad (2.56)$$

Example 2.5.2 Consider the PDF represented by

$$p_{r,\theta}(r, \theta) = \frac{r}{2\pi\sigma_r^2} e^{\frac{-r^2}{2\sigma_r^2}} \quad (2.57)$$

Obtain the marginal PDF $p_r(r)$.

Recognizing that this is a polar coordinate system, we will integrate θ from 0 to 2π :

$$p_r(r) = \int_0^{2\pi} p_{r,\theta}(r, \theta) d\theta = \int_0^{2\pi} \frac{r}{2\pi\sigma_r^2} e^{\frac{-r^2}{2\sigma_r^2}} d\theta = \frac{r}{\sigma_r^2} e^{\frac{-r^2}{2\sigma_r^2}} \quad (2.58)$$

The resulting PDF is called the Rayleigh distribution and is common in random vibration analysis.

Consider presuming that two variables x and y are related by

$$y = mx \quad (2.59)$$

where we will presume for the sake of simplicity that x and y have been shifted so that their expected values are both equal to 0. The squared error of this assumption in the y direction is given by

$$E[(y - mx)^2] = E[y^2] - 2mE[xy] + m^2E[x^2] = \sigma_y^2 - 2m\varsigma_{xy} + m^2\sigma_x^2 \quad (2.60)$$

where

$$\varsigma_{xy} = E[xy] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p_{xy}(x, y) dx dy \quad (2.61)$$

is called the *covariance*. Finding the minimum of $E[(y - mx)^2]$ with respect to m gives

$$\frac{dE[(y - mx)^2]}{dm} = -2\varsigma_{xy} + 2m\sigma_x^2 = 0 \quad (2.62)$$

which gives a best ‘fit’ line slope of

$$m = \frac{\varsigma_{xy}}{\sigma_x^2} \quad (2.63)$$

Substituting back into equation (2.59) gives

$$y = \frac{\varsigma_{xy}}{\sigma_x^2} x \quad (2.64)$$

Dividing both sides by σ_y and reorganizing yields

$$\begin{aligned} \frac{y}{\sigma_y} &= \frac{\varsigma_{xy}}{\sigma_x \sigma_y} \frac{x}{\sigma_x} \\ &= \rho_{xy} \frac{x}{\sigma_x} \end{aligned} \quad (2.65)$$

where

$$\rho_{xy} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sqrt{E[(x - \mu_x)^2]} \sqrt{E[(y - \mu_y)^2]}} \quad (2.66)$$

is the *correlation coefficient* or *normalized covariance*. It is shown on page 70 that $|\rho_{xy}| \leq 1$.

If we define

$$z_1 = \frac{x - \mu_x}{\sigma_x}, \quad \text{and} \quad z_2 = \frac{y - \mu_y}{\sigma_y} \quad (2.67)$$

then we can write instead that

$$\rho_{xy} = \rho_{z_1 z_2} = \frac{E[z_1 z_2]}{\sqrt{E[z_1^2]} \sqrt{E[z_2^2]}} \quad (2.68)$$

If it is possible then to write

$$p_{z_1 z_2}(z_1, z_2) = p_{z_1}(z_1) p_{z_2}(z_2) \quad (2.69)$$

so that $E[z_i] = 0$ and $E[z_i^2] = 1$, then substituting into equation (2.68) gives

$$\begin{aligned} \rho_{xy} &= \frac{E[z_1 z_2]}{\sqrt{E[z_1^2]} \sqrt{E[z_2^2]}} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_1 z_2 p_{z_1}(z_1) p_{z_2}(z_2) dz_1 dz_2}{\sqrt{E[z_1^2]} \sqrt{E[z_2^2]}} \\ &= \frac{\int_{-\infty}^{\infty} z_1 p_{z_1}(z_1) dz_1 \int_{-\infty}^{\infty} z_2 p_{z_2}(z_2) dz_2}{1 \cdot 1} \\ &= E[z_1] E[z_2] = 0 \end{aligned} \quad (2.70)$$

this expression only valid when x and y are truly independent. otherwise cant multiply density functions for x and y to get joint probability function

The most commonly considered 2-D distribution function is the joint normal distribution. This distribution represents the likelihood of two variables, x and y , occurring as a function of their values and is given by (Newland 1993; Wirsching et al. 1995)

$$p_{xy}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} e^{\frac{-1}{2(1-\rho_{xy}^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho_{xy} \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right]} \quad (2.71)$$

Alternatively, if we use normalized variables,

$$p_{z_1 z_2} = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{-(z_1^2 - 2\rho z_1 z_2 + z_2^2)}{2(1-\rho^2)}} \quad (2.72)$$

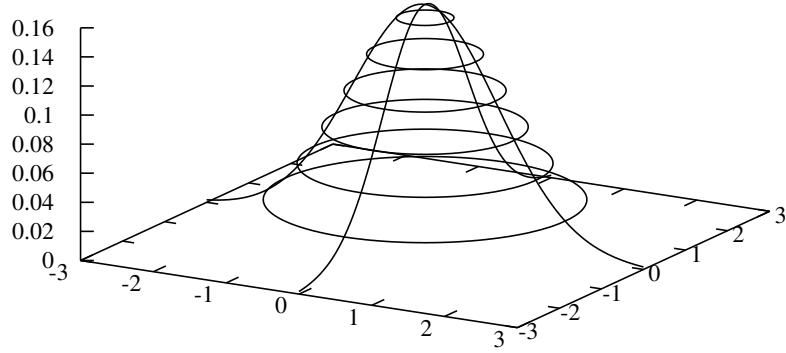


Figure 2.2 Joint Normal Distribution, uncorrelated variables.

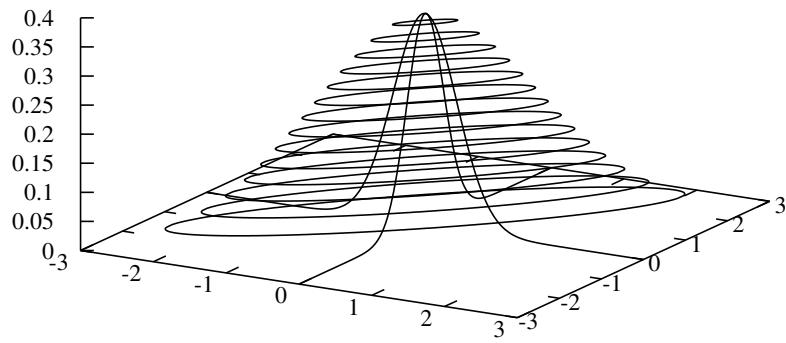


Figure 2.3 Joint Normal Distribution, correlated variables.

2.5.2 Distribution of the Sum of Normal Variables

Consider a situation where two independent variables, x_1 , and x_2 , with normal distributions are added. It is desired to know the PDF, $p_y(y)$, (PDF) of the combined variable, $y = x_1 + x_2$. However, observing equation (2.36), the PDF transformation requires that we transform from two variables to two new variables. To satisfy this, we redefine the original variables to be

$$x_2 = y - x_1, \quad x_1 = x_1 \quad (2.73)$$

and the new variables to be

$$y = x_1 + x_2, \quad x_1 = x_1 \quad (2.74)$$

Calculating J using equation (2.37) gives

$$J = \det \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = 1 \quad (2.75)$$

Noting that

$$p_{x_1 x_2}(x_1, x_2) = p_{x_1}(x_1)p_{x_2}(y - x_1) \quad (2.76)$$

because x_1 and x_2 are independent (uncorrelated),

$$p_{x_1y}(x_1, y) = J p_{x_1}(x_1) p_{x_2}(y - x_1) = p_{x_1}(x_1) p_{x_2}(y - x_1) \quad (2.77)$$

Noting that $p_y(y)$ is a marginal PDF of $p_{x_1,y}(x_1, y)$, integrating both sides over x_1 gives

$$p_y(y) = \int_{-\infty}^{\infty} p_{x_1y}(x_1, y) dx_1 = \boxed{\int_{-\infty}^{\infty} p_{x_1}(x_1) p_{x_2}(y - x_1) dx_1} \quad (2.78)$$

Note that the right side is a convolution. Applying the Laplace Transform to both sides and using the convolution property of the Laplace Transform⁶, this becomes

$$M_y(s) = M_{x_1}(s) M_{x_2}(s) \quad (2.79)$$

Applying equation (2.27),

$$M_y(s) = e^{\mu_{x_1}s + \frac{1}{2}\sigma_{x_1}^2s^2} e^{\mu_{x_2}s + \frac{1}{2}\sigma_{x_2}^2s^2} = e^{(\mu_{x_1} + \mu_{x_2})s + \frac{1}{2}(\sigma_{x_1}^2 + \sigma_{x_2}^2)s^2} \quad (2.80)$$

which is the moment generating function for a normal distribution with a mean of

$$\mu_y = \mu_{x_1} + \mu_{x_2} \quad (2.81)$$

and variance of

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 \quad (2.82)$$

Distributions obtained by other mathematical operations combining two variables are illustrated in chapter 6 by Papoulis and Pillai (Papoulis and Pillai 2002). For variables resulting from the sum of N Gaussian distributions, equations (2.79), (2.81) and (2.82) generalize to

$$M_y(s) = \prod_{i=1}^N M_{x_i}(s), \quad (2.83)$$

$$\mu_y = \sum_{i=1}^N \mu_{x_i} \quad (2.84)$$

and

$$\sigma_y^2 = \sum_{i=1}^N \sigma_{x_i}^2 \quad (2.85)$$

In fact, equations (2.83-2.85) are a special result of the *central limit theorem* (Section 2.8).

2.5.3 Sample (Calculated) Mean

The values for the mean, variance, and kurtosis of a set of data points as defined by equations (2.1-2.4) are estimates of the true values which are hidden in the random nature of the process. If we have enough data points, we hope to be able to converge to the true expected values. We further would like to quantify how much error we should expect, and how the expected

⁶The convolution property of the Laplace Transform is equivalent to equation (1.106) with s substituted for $j2\pi f$

convolution, so we can go to laplace domain to simplify by turning the integral into product

error changes when we include additional data points. To verify the equations of section 2, consider the expected value of the calculated mean,

$$\begin{aligned}
 E[\hat{\mu}_x] &= E\left[\frac{1}{N} \sum_{i=1}^N x_i\right] = \frac{1}{N} E\left[\sum_{i=1}^N x_i\right] \\
 &= \frac{1}{N} \sum_{i=1}^N E[x_i] = \frac{1}{N} \sum_{i=1}^N \mu_x = \mu_x
 \end{aligned}$$

calculated mean, uses the hat

population average also know as process average

Note that we have used here that the expected value of any individual sample, x_i , is equal to μ_x by equation (2.111). As a result, using the calculated estimate of the mean value, we expect to get the mathematically defined expected value. Generally, we won't, but it's all the information that exists. Now, the *expected error* in the estimate of the expected value is

$$\begin{aligned}
 E[(\hat{\mu}_x - \mu_x)^2] &= E\left[\left(\frac{1}{N} \sum_{i=1}^N x_i - \mu_x\right)^2\right] \equiv \frac{1}{N^2} E\left[\left(\sum_{j=1}^N (x_j - \mu_x)\right)^2\right] \\
 &= \frac{1}{N^2} E\left[\sum_{i=1}^N \sum_{j=1}^N (x_i - \mu_x)(x_j - \mu_x)\right] \\
 &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(x_i - \mu_x)(x_j - \mu_x)]
 \end{aligned} \tag{2.87}$$

Since the samples x_i and x_j are independent, and their expected values have been removed, the expected value $E[(x_i - \mu_x)(x_j - \mu_x)] = 0$ for $i \neq j$. This allows us to simplify equation (2.87) to

$$E[(\hat{\mu}_x - \mu_x)^2] = \frac{1}{N^2} \sum_{i=1}^N E[(x_i - \mu_x)^2] \tag{2.88}$$

If we assume that x is ergodic⁷, then for any values i and j and we can write $E[(x_i - \mu_x)^2]$ as

$$\begin{aligned}
 E[(x_i - \mu_x)^2] &= E[(x_j - \mu_x)^2] \\
 &= E[(x - \mu_x)^2]
 \end{aligned} \tag{2.89}$$

and thus in equation (2.88)

$$\sum_{i=1}^N E[(x_i - \mu_x)^2] = N E[(x - \mu_x)^2] \tag{2.90}$$

Thus applying equation (2.17) we get

$$E[(\hat{\mu}_x - \mu_x)^2] = \frac{1}{N} \sigma_x^2 \tag{2.91}$$

⁷Ergodic means that since each sample, x_i , is random, with the same distributions as the others, say x_j , then we can just use the generic x to represent the process generating the values x_i . A more in depth discussion is presented in section 2.10.

which illustrates that the expected error of an estimate of the mean value of a process diminishes as $\frac{1}{N}$. It's vital to understand that this is not the actual error, but only the average error if we performed the 'experiment' an infinite number of times.

This is best illustrated using the example of rolling a dice. The expected value is the unachievable value of 3.5. In a single role, the calculated mean will clearly not match the expected value. However, after 10 rolls, the calculated mean, being a random value itself, has a variance of approximately $3.5/20 = .15$ (obtained using the PDF), the variance of x (the random values of the die), divided by the number of rolls. This means that the odds are in your favor that the calculated mean will be closer to $\mu_x = 3.5$ than any of the realizable values.

2.6 Chi-square Distribution (χ^2)

Consider now a process that is represented by

$$\chi_N^2 = \sum_{i=1}^N z_i^2 \quad (2.92)$$

where z_i are normalized Gaussian distribution variables. This is a common distribution resulting from estimation of the variance in equation (2.2) and auto-spectral densities in equation (3.89), with the exception that the set of data is not usually normalized Gaussian data. It is also unfortunately not normalized by N in practice, making side-by-side comparison for different values of N uncommon.

The mean and variance can be shown using equation (2.16) to be

$$\mu_{\chi_N^2} = N \quad (2.93)$$

and

$$v_{\chi_N^2} = \sigma_{\chi_N^2}^2 = 2N \quad (2.94)$$

while if non-normalized, but still zero mean, Gaussian distributed variables, x_i are used, then

$$\mu_{\chi_N^2} = N\sigma_x^2 \quad (2.95)$$

and

$$v_{\chi_N^2} = \sigma_{\chi_N^2}^2 = 2N\sigma_x^4 \quad (2.96)$$

The PDF for the χ_N^2 distribution is given by

$$p_{\chi_N^2}(x) = \frac{1}{2^{N/2}\Gamma(N/2)} e^{-x/2} x^{(N/2)-1} \quad (2.97)$$

Figure 2.4 shows the shapes of the χ_N^2 for a range of values of N . Both axes have been normalized to emphasize the shapes of the PDFs. Since dividing by N is a normal part of calculating averaging, it's most appropriate to do this so that the constriction of the χ_N^2 distribution can be observed for an increased number of samples. It is further necessary to scale the magnitude of the PDF so that the wide range of amplitudes appear in the figure.

The Figure 2.4 shows that the dominant shape of the PDF is exponential for low values of N and nearly Gaussian for high values of N . Further, as $N \rightarrow \infty$, the distribution of $\sqrt{2\chi_N^2}$

approaches a Gaussian distribution with a mean of $\sqrt{2N - 1}$ and variance of $\sigma^2 = \frac{2}{\sqrt{2\chi_N^2}} = 1$ (Guttman et al. 1982), and the χ_N^2 distribution itself also approaches a Gaussian distribution (Wirsching et al. 1995), the thresholds at which a Gaussian distribution is an acceptable approximation being subjective and left to the reader to use their personal judgment. Figure 2.4 provides some visual information for making a qualitative decision.

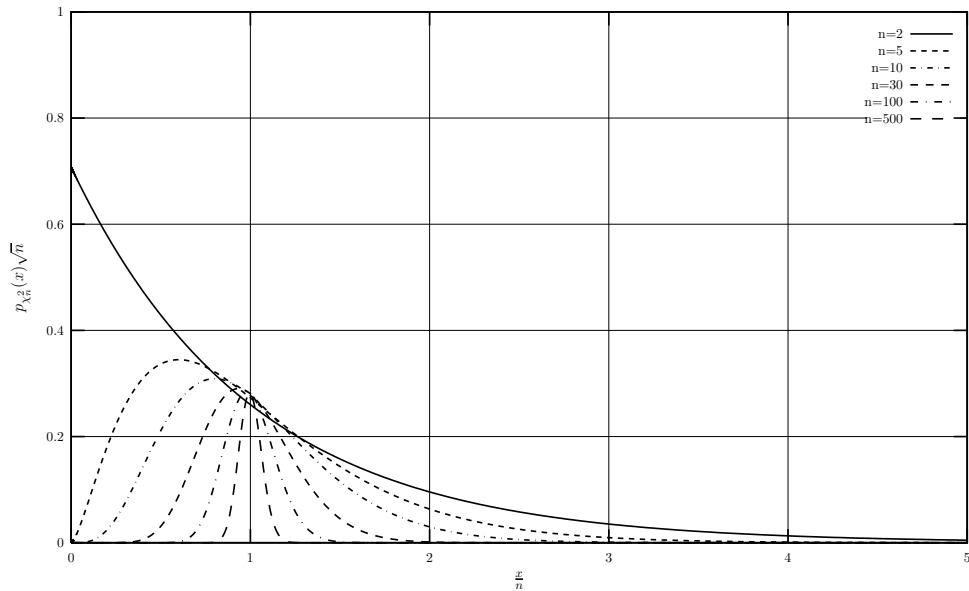


Figure 2.4 Normalized χ_N^2 distributions versus $\frac{x}{N}$ for example values of n .

To observe the constriction that occurs with an increase in N , considering Appendix D.2 on page 200. For $N = 10$, a 95% *confidence interval* results in bounds of $3.25 < x < 20.48$. However, with $N = 100$, this interval is tightened to $74.22 < x < 129.56$.⁸ When normalized by N it is clear that using more samples significantly reduced the relative error. The abbreviated table listed in Appendix D.2 is sufficient for most work. However, more detailed distributions can always be obtained using the command `norminv` in MATLAB, `normal_inv` in Octave, or a table in a probability text.

2.6.1 Sample (Calculated) Variance

Next consider the expected value of the variance estimate given by equation (2.2)

$$\begin{aligned}
 E[\hat{\sigma}_x^2] &= E\left[\frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu}_x)^2\right] = \frac{1}{N-1} E\left[\sum_{i=1}^N (x_i - \hat{\mu}_x)^2\right] \\
 &= \frac{1}{N-1} E\left[\sum_{i=1}^N (x_i - \mu_x - \hat{\mu}_x + \mu_x)^2\right] \\
 &= \frac{1}{N-1} E\left[\sum_{i=1}^N (x_i - \mu_x)^2 - 2 \sum_{i=1}^N (x_i - \mu_x)(\hat{\mu}_x - \mu_x) + \sum_{i=1}^N (\mu_x - \hat{\mu}_x)^2\right] \\
 &= \frac{1}{N-1} \left(\sum_{i=1}^N E[(x_i - \mu_x)^2] - 2E\left[\left(\sum_{i=1}^N x_i - \mu_x\right)(\hat{\mu}_x - \mu_x)\right] + N E[(\mu_x - \hat{\mu}_x)^2] \right)
 \end{aligned} \tag{2.98}$$

See page
308 Wirsh-
ing

Noting equation (2.89) and applying equation (2.1) to substitute for the second summation,

$$\begin{aligned}
 E[\hat{\sigma}_x^2] &= \frac{1}{N-1} (N E[(x - \mu_x)^2] - 2N E[(\hat{\mu}_x - \mu_x)(\hat{\mu}_x - \mu_x)] + N E[(\mu_x - \hat{\mu}_x)^2]) \\
 &= \frac{1}{N-1} (N E[(x - \mu_x)^2] - N E[(\mu_x - \hat{\mu}_x)^2])
 \end{aligned} \tag{2.99}$$

Substituting for $E[(x - \mu_x)^2]$ using equation (2.17) and substituting for $E[(\mu_x - \hat{\mu}_x)^2]$ using equation (2.91)

$$\begin{aligned}
 E[\hat{\sigma}_x^2] &= \frac{1}{N-1} \left(N \sigma_x^2 - N \frac{\sigma_x^2}{N} \right) \\
 &= \sigma_x^2
 \end{aligned} \tag{2.100}$$

when expanded
it is 1 times the
middle
expression so

after summation

This means that as we obtain more sample points, the estimate of the variance will converge to the expected value of the variance, σ_x^2 . If we had used a denominator of N in front of the sample variance calculation, the estimate would have been biased.

Comparing equation (2.98), or alternatively equation (2.2), to equation (2.92), the sample variance is a χ^2 distributed variable derived from non-normalized variables, divided by $N - 1$. Thus as $N \rightarrow \infty$, from equations (2.95) and (2.96), the expected value of the variance estimate could have been shown to be σ_x^2 while the expected error in the estimate is

standard deviation
of variance
estimation

$$E[(\hat{\sigma}_x^2 - \sigma_x^2)^2] = \frac{2}{N} \sigma_x^4 \tag{2.101}$$

2.7 Rayleigh Process

Consider two independent Gaussian processes x and y with $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = \sigma$. The measures x and y can be thought of as positions on a target, with their distributions representing the degree

⁸In practice the confidence interval is typically, and incorrectly, considered to be a bound on the quantity being estimated. However, it is actually a bound on the estimation of that parameter. For example, consider calculating a 95% confidence interval for the calculated mean from 100 samples. In 95% of similar experiments the calculated mean should be between those bounds. Keep in mind, however, that the confidence interval is itself an estimated quantity. So, there is variability in the confidence interval estimates as well.

of accuracy in hitting the center of the target. Presuming x and y to be independent, then by equation (2.69), the joint probability density function is

$$p_{xy}(x, y) = p_x(x)p_y(y) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}} \right) \left(\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-y^2}{2\sigma^2}} \right) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \quad (2.102)$$

If we want to measure accuracy, it is more common in such a situation to consider the distance, r , from the bullseye. To convert from Cartesian to polar coordinates, we recognize that

$$x = r \cos(\theta) \quad (2.103a)$$

$$y = r \sin(\theta) \quad (2.103b)$$

Using equation (2.38)

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{\partial(r \cos(\theta))}{\partial r} & \frac{\partial(r \cos(\theta))}{\partial \theta_2} \\ \frac{\partial(r \sin(\theta))}{\partial r} & \frac{\partial(r \sin(\theta))}{\partial \theta_2} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{vmatrix} = r \quad (2.104)$$

and substituting equations (2.103) into (2.102) gives

$$p_{r\theta}(r, \theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad (2.105)$$

The marginal PDF in r was obtained in example 2.5.2 by integrating over the domain, θ .

$$p_r(r) = \frac{r}{\sigma_r^2} e^{-\frac{r^2}{2\sigma_r^2}} \quad (2.106)$$

2.8 Central Limit Theorem

The *central limit theorem* states that the PDF of the sum of a large number of mutually independent random variables, each with any individual PDF, converges to a Gaussian distribution provided that each random variable contributes a relatively inconsequential amount (Bendat and Piersol 2000; Miller and Freund 1977; Papoulis and Pillai 2002; Parzen 1972). For example, for a variable y defined as (Bendat and Piersol 2000)

$$y = \sum_{i=1}^N \alpha_i x_i \quad (2.107)$$

the mean and variance are

$$\mu_y = \sum_{i=1}^N \alpha_i \mu_{x_i} \quad (2.108)$$

$$\sigma_y^2 = \sum_{i=1}^N \alpha_i^2 \sigma_i^2 \quad (2.109)$$

with a Gaussian distribution.

2.9 Stationary Signals

The concept of a dynamic signal being stationary seems absurd to an intelligent person because it keeps moving. This belies the difference between the common and the mathematical definition of stationary. Calling a signal stationary, using the common definition, is relatively pointless because in that case it can simply be called a value... for example, the weight is 10 N. While you may measure for a long period

of time, it should be constant, and referring to the weight as a signal is silly. Thus the mathematical definition has been formed for which properties of the signal *tend* to be relatively constant over time. A *weakly stationary* signal is defined as one in which the ensemble mean (equation (2.5)) is constant and the ensemble autocorrelation

$$R_{xx}(t, \tau) = \frac{1}{N} \sum_{i=1}^N x_i(t)x_i(t + \tau) \quad (2.110)$$

depends only on the difference in time, τ .

If we were to evaluate and track higher moments (which we haven't defined, but comparing the mean to the autocorrelation should give some idea of how to extrapolate) and deem them to be sufficiently constant over time, then the signal would be deemed *strongly stationary*. In practice, the determination that a signal is weakly stationary is often deemed sufficient to assume strong stationarity. The justification is valid in linear processes where the probability density functions are Gaussian because a linear process operating on a Gaussian input yields a Gaussian output, and a Gaussian distribution can be uniquely represented by its mean and standard deviation, the two measures used to determine weak stationarity.

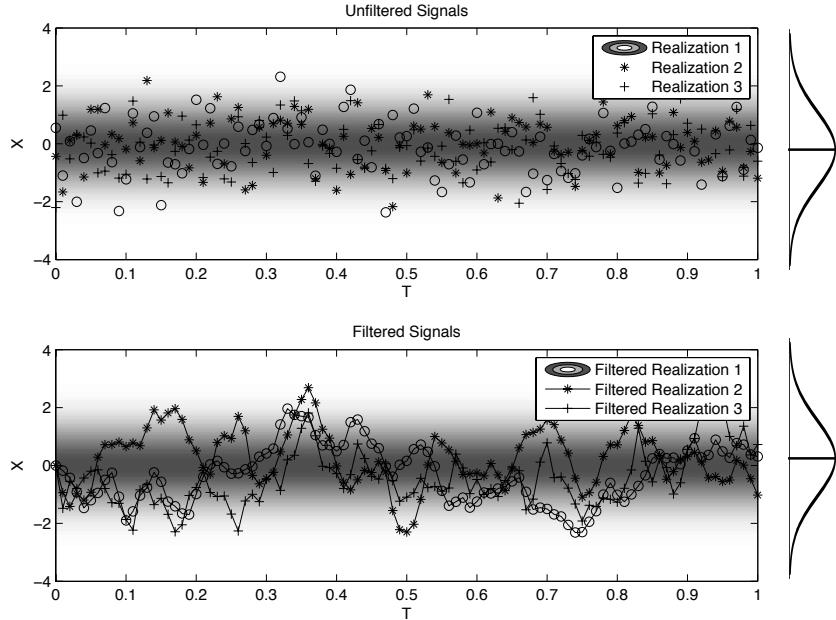


Figure 2.5 High and Low Frequency Random Ergotic Processes. Shading represents value of PDF.

^aGraphic courtesy Oleg Shirayev.

2.10 Ergodic Signals

Consider now the reality that one cannot ever perform the operations to prove that a signal is stationary as prescribed in the previous section.⁹ The first solution to the problem is to assume that the signal is *ergodic*. A signal is ergodic if, for any realization, i , of that signal, the ensemble properties match those of the realization. That is,

$$\mu_x(t, i) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2+t}^{T/2+t} x_i(\tau) d\tau = \mu_x(t) \quad (2.111)$$

where $\mu_x(t)$ was defined by equation (2.5), and

$$\begin{aligned} R_{xx}(t, \tau, i) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2+t}^{T/2+t} x_i(\sigma) x_i(\sigma + \tau) d\sigma \\ &= R_{xx}(t, \tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t) x_i(t + \tau) \end{aligned} \quad (2.112)$$

An example of a random ergodic process is shown in Figure 2.5. The upper illustration shows 3 realizations, sampled at 100Hz, of the same process. The shading represents the likelihood that the signal will be at any given value at any given time. Since the process is stationary, the PDF represented by the shading is constant over time. The lower illustration shows three realizations of a lower bandwidth (reduced maximum frequency content) process sampled at the same frequency. Lines connect the data points only to better illustrate that the process fluctuates much less than the first process, even though it has the same PDF over time.

In Section 2.1 the concept of random in time was introduced. Here the properties are evaluated for a single realization with the index i . The problem with these definitions is that T is undefined. For a continuous random signal with infinite bandwidth (unreasonably high frequencies content), T can be infinitely small since the value of $x(t)$ is changing so quickly that one could never measure $x(t)$, but instead would only be able to approximate its likelihood of being at any given value. Over the period of almost no time the signal is sufficiently rich enough that a complete characterization of the signal could be obtained. This is like the Heisenburg uncertainty principle in some sense, but worse since $x(t)$ is able to move at speeds making light speed look slow. This is, of course a fantasy, and the extreme limiting case, opposite the one using the common definition of stationary.

In practice, the phrase “constant” must be replaced with “fairly constant”, and personal judgment must be used. Also, a bandwidth must be prescribed, above which variation can occur without the signal being deemed non-stationary. For example, consider a sine wave with a constant amplitude and a frequency of 10 Hz. If we choose $T = 1$ sec, then obtain values for the mean and auto-correlation as a function of time, we would find that they vary minimally as compared to the standard deviation and variance (standard deviation squared) respectively of such a signal. Thus we can state that the signal is weakly stationary over a period of 1 second. On the other hand, if we choose a time $T = .01$ sec and perform the same operations, we would determine that the signal is non-stationary. The result is that stationarity of a signal must be considered relative to some time scale. Thus, when the phrase *stationary signal* is used, it should be interpreted as stationary in time. A stationary random signal is generally assumed to be ergodic for the sake of convenience.

2.11 Cross Correlation of Signals

When relating two signals to one another we are often interested in a measure of how well knowledge of the value of one signal tells us the value of the other. Say we have two signals, x and y , and we think

⁹Collecting alternate realizations is not feasible because we cannot go back in time

that they might be related by the equation

$$y = mx \quad (2.113)$$

where the axes have been shifted so that the line intersects the origin. Let's define the measured quantities as \hat{x} and \hat{y} . The vertical error between the line $y = mx$ and the measured values of \hat{y} is

$$\xi_y = \hat{y} - y = \hat{y} - mx \quad (2.114)$$

In general, we won't precisely know x , but for now let's act as if $\hat{x} = x$. The best value of m is then the one that minimizes this error. The function that is most often minimized is the average value of the square of the error, or the variance of the error defined by equation (2.2).¹⁰ The variance of ξ_y is

$$\sigma_{\xi_y}^2 = \frac{1}{N-1} \sum_{i=1}^N (\hat{y}_i - mx_i)^2 = \frac{1}{N-1} \sum_{i=1}^N (\hat{y}_i^2 - 2mx_i\hat{y}_i + m^2x_i^2) = \sigma_{\hat{y}}^2 - 2m\varsigma_{x\hat{y}} + m^2\sigma_x^2 \quad (2.115)$$

where

$$\varsigma_{x\hat{y}} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(\hat{y}_i - \mu_{\hat{y}}) \quad (2.116)$$

is called the *covariance*.¹¹ Finding the minimum of $\sigma_{\xi_y}^2$ with respect to m yields

$$\frac{d\sigma_{\xi_y}^2}{dm} = -2\varsigma_{x\hat{y}} + 2m\sigma_x^2 = 0 \quad (2.117)$$

which yields

$$m = \frac{\varsigma_{x\hat{y}}}{\sigma_x^2} \quad (2.118)$$

The presumed correct values of y are then given by equation (2.113) as

$$y = \frac{\varsigma_{x\hat{y}}}{\sigma_x^2} x \quad (2.119)$$

Dividing both sides of this expression by σ_y and reordering yields

$$\frac{y}{\sigma_y} = \rho_{xy} \frac{x}{\sigma_x} \quad (2.120)$$

where

$$\rho_{xy} = \frac{\varsigma_{x\hat{y}}}{\sigma_x \sigma_y} \quad (2.121)$$

is the *correlation coefficient* or *normalized covariance*¹². The normalization of the covariance by the standard deviations causes the correlation coefficient to be limited to values $-1 \leq \rho_{xy} \leq 1$. To prove this, consider evaluating $\sigma_{\alpha x + \beta y}^2$. By definition (recall equation (2.2))

$$\sigma_{\alpha x + \beta y}^2 \geq 0 \quad (2.122)$$

Expanding the terms of equation (2.122) gives

$$\alpha^2 \sigma_x^2 + 2\alpha\beta\varsigma_{xy} + \beta^2 \sigma_y^2 \geq 0 \quad (2.123)$$

Assuming $\beta \neq 0$, dividing by β^2 , and defining $\frac{\alpha}{\beta} = \epsilon$,

$$f(\epsilon) = \epsilon^2 \sigma_x^2 + 2\epsilon\varsigma_{xy} + \sigma_y^2 \geq 0 \quad (2.124)$$

¹⁰The average value of the error, ξ_y , is zero if the coordinates have been shifted properly.

¹¹Recall that we shifted the coordinates for this derivation so that $\mu_x = 0$ in equation 2.115.

¹²It can be easily shown that the phase angle between two sinusoids represented by x and y is given by $\phi = \arccos(\rho_{xy})$.

This is a quadratic polynomial in ϵ with no real *and* unique roots because it is non-negative¹³. The solution for ϵ is

$$\epsilon = \frac{-2\varsigma_{xy} \pm \sqrt{(2\varsigma_{xy})^2 - 4\sigma_x^2\sigma_y^2}}{2\sigma_x^2} \quad (2.125)$$

In order for the solutions to be either complex or non-unique (repeated and real)

$$(2\varsigma_{xy})^2 - 4\sigma_x^2\sigma_y^2 \leq 0 \quad (2.126)$$

Thus

$$\varsigma_{xy}^2 \leq \sigma_x^2\sigma_y^2 \quad (2.127)$$

which is a form of the Cauchy-Schwartz inequality. Therefore

$$-1 \leq \rho_{xy} = \frac{\varsigma_{xy}}{\sigma_x\sigma_y} \leq 1 \quad (2.128)$$

We would have obtained the same formulae for m , ς_{xy} and ρ_{xy} even if we had presumed $\hat{y} = y$ and $\hat{x} \neq x$. These relations are accurate enough for most applications, one of which is signal processing. However, a better best-fit line is obtained by minimizing the perpendicular distance from the line to each data point. That is, find m such that

$$\text{error}^2 = \frac{1}{N} \sum_{i=1}^N \hat{x}^2 \left(1 - \frac{1}{1+m^2} \right) - \hat{x}\hat{y} \frac{2m}{1+m^2} + \hat{y}^2 \left(1 - \frac{1}{1+m^2} \right) \quad (2.129)$$

is minimized. This minimization is referred to as *double least squares*.

Often, more detailed information is desired relating a signal to itself or to another signal. Consider attempting to ascertain if a signal at time t_1 is related to itself at some later time t_2 . This can be represented by the auto-correlation function given by

$$R_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)dt \quad (2.130)$$

Observing the expression for the autocorrelation at $\tau = 0$,

$$\begin{aligned} R_{xx}(0) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)^2 dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} ((x(t) - \mu_x)^2 + 2x(t)\mu_x - \mu_x^2) dt \\ &= \sigma_x^2 + \mu_x^2 \end{aligned} \quad (2.131)$$

Further, evaluating $R_{xx}(\tau)$ gives

$$\begin{aligned} R_{xx}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} ((x(t) - \mu_x)(x(t+\tau) - \mu_x) + x(t)\mu_x + x(t+\tau)\mu_x - \mu_x^2) dt \\ &= \varsigma_{x(t),x(t+\tau)} + \mu_x^2 \end{aligned} \quad (2.132)$$

¹³The polynomial, for any set of values, represents a parabola that doesn't cross $f(\epsilon) = 0$.

2.12 Correlation of Finite Time Signals

Any real signal will have a finite length. This leads to a dilemma in the evaluation of expressions (2.130) and (2.137). The only value of τ for which they can be evaluated for a finite time signal is $\tau = 0$. There are two potential compromises that make sense.

The first is to assume that the signals have non-zero values only for $-T/2 < t < T/2$. This form of the auto-correlation is often referred to as the *linear auto-correlation* in signal processing texts. The problem with this approach is that it is generally not true. However, in relating two signals, while understanding the effects of the signal truncation, it can still lead to a good understanding of the true auto-correlation.

An alternative, the *circular auto-correlation* function, assumes that the signal repeats itself indefinitely. This corresponds to the assumption necessary for obtaining a Fourier series. While this can be true, or approximately true, in cases such as rotating equipment and simple signals (simple sinusoids, combinations of harmonics...), it is helpful to assure that the signal length is precisely the same length as the single cycle. Often, the actual signal is truncated to the length appropriate for a single cycle rather than trying to set the record length a priori.

Example 2.12.1 Obtain the linear and circular auto-correlations of $\sin(t)$ for $-\pi \leq t \leq \pi$.

Since $R_{xx}(\tau) = R_{xx}(-\tau)$, we only need to evaluate $R_{xx}(\tau)$ for $\tau \geq 0$. The circular auto-correlation is given by

$$\begin{aligned} R_{xx}(\tau) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(t) \sin(t + \tau) dt \\ &= \frac{\cos(\tau)}{2} \end{aligned} \quad (2.133)$$

For the linear auto-correlation, since $\sin(t + \tau) = 0$ for $t + \tau > \pi$, the upper limit of integration must be changed to $\pi - \tau$ giving

$$\begin{aligned} R_{xx}(\tau) &= \frac{1}{2\pi} \int_{-\pi}^{\pi-\tau} \sin(t) \sin(t + \tau) dt \\ &= \left(1 + \frac{\tan(\tau) - \tau}{2\pi}\right) \frac{\cos(\tau)}{2} \end{aligned} \quad (2.134)$$

The tapering to zero of the linear auto-correlation due to truncation of the data is readily apparent. In practice, the circular auto- and cross-correlations are more commonly encountered in part due to their relative ease of computation using the discrete Fourier transform (section 3.3 and the Fourier convolution principle (equation (1.105)). Zero padding (section 3.7) can be used to obtain the linear auto- and cross-correlations via the Fourier convolution principle.

because

$$\rho_{x(t),x(t-\tau)} = \frac{\zeta_{x(t),x(t-\tau)}}{\sigma_{x(t)}\sigma_{x(t-\tau)}} \quad (2.135)$$

where we have proven that $-1 \leq \rho_{xy} \leq 1$ for any x and y , and $\sigma_x = \sigma_{x(t)} = \sigma_{x(t-\tau)}$, $R_{xx}(\tau)$ is bounded such that

$$-\sigma_x^2 + \mu_x^2 \leq R_{xx}(\tau) \leq \sigma_x^2 + \mu_x^2 \quad (2.136)$$

The *cross-correlation* function relates two different signals by

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t + \tau) dt \quad (2.137)$$

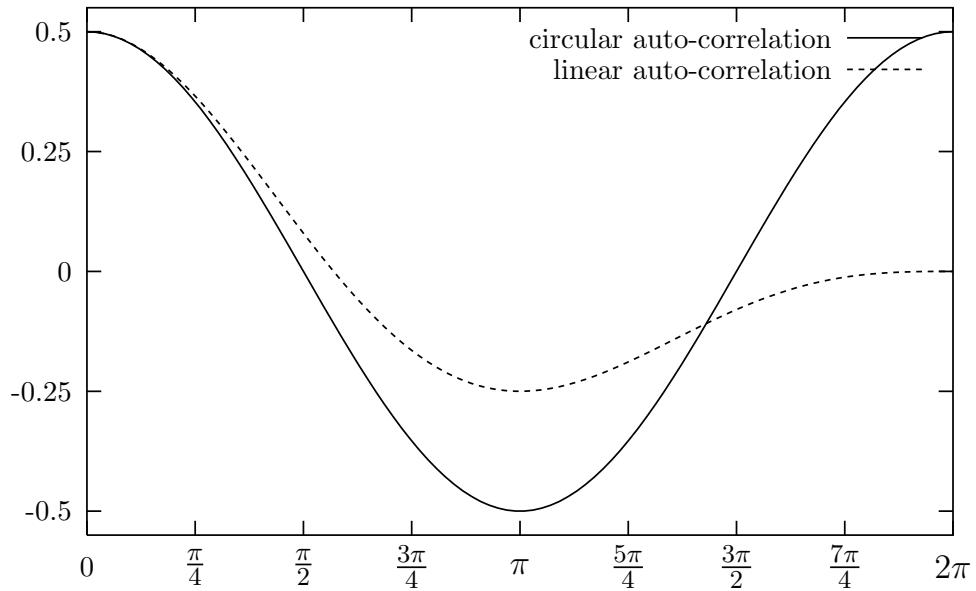


Figure 2.6 Linear and circular auto-correlation functions for $\sin(t)$.

This function can be very beneficial in finding phase lag (time lag) between signals of equal frequency as it is much more robust than simply identifying the time lag between peaks. The primary application is measuring wave speed between two points on a structure such as a bladed disk assembly. By observation,

$$\begin{aligned}
 R_{xy}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t + \tau)dt \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t - \tau)y(t)dt = R_{yx}(-\tau)
 \end{aligned} \tag{2.138}$$

All of the derived quantities of this section are derived *over* time. Later, in section 3.9.2, we will evaluate quantities *at* time, and at that point we will introduce other forms of these quantities for random signals.

Example 2.12.2 Figure 2.7 shows a sample of this polymer being tested for complex modulus as described in section 1.6.1. The goal is to determine the value of k^* for $0\text{Hz} < f < 100\text{Hz}$.¹⁴ In order to do this, the sample is excited at 5 Hz intervals over the frequency range of interest and an appropriate close to the amplitude at which the material is expected to be used. The test fixture shows a load cell approximately measuring force on the specimen as well as an accelerometer measuring dynamic amplitude of response. Other issues such as mass loading, etc., are addressed in ?.

¹⁴While a broad band excitation (random excitation with a wide band of frequencies, see chapter 3) would seem to make sense, that approach often leads to poor reliability of the results.

Figure shows time histories of the acceleration and load.

Figure 2.7 Polymer stress/strain test system.??

Figure 2.8 Force and acceleration time histories.

Figure 2.9 Cross correlation between harmonic load and response (acceleration) of a polymer material. Data courtesy Mostafa El-Ashry, Wright State University.

2.13 Detrending data

(Bendat and Piersol 2000) Almost every experimentalist is put in the situation of having data with a trend in the data needing removal. Such trends come as offsets from sensors that don't output a zero voltage corresponding to a zero measurement, drifts due to time varying behavior of electronics or sensors¹⁵, or other miscellaneous effects.

An example is shown in figure ???. This data was collected to observe the hysteresis in a mechanical system, with a load cell and a strain gage used for measurement. Unfortunately the equipment hadn't reached thermal equilibrium, and the data drifts to the right as time goes on.

Trends in data are slow frequency phenomenon. As a result, they lend themselves to two forms of correction, both of which are add hoc unless a valid model of the trend exists (typically it is insufficient). In order to remove such trends, two reliable methods exist, fft correction, and curve fit extraction.

2.13.1 Detrending by FFT Correction (frequency domain)

Since trend processes are slow moving, they appear in the frequency domain at extremely low frequencies. Referring to section 3.3, the very lowest and highest indices of the fft of the signal correspond to the lowest frequencies. To remove the trend from the data, we take the FFT of the data and observe the magnitudes, especially at the low frequencies. A log scale should be used to obtain reasonable plots. Figures 2.10-2.11 illustrate the process. Figure 2.11) illustrates the undue magnitude of the low and high index components (observe the far left and far right ends of the data) corresponding to the low frequency components of the signals. These values need to be replaced by the ideally correct values. Noting section 3.3, approximately half of these values are the complex conjugates of the other half. For this illustration, a relatively sloppy substitution of zero is used for all values that are unduly high. Better approximations can be had by iterating between substitution and reviewing the resulting FFT. Our adjusted, or corrected, FFT is shown in figure 2.11 compared to the original uncorrected FFT. It is clear that this is not a very good guess but it is sufficient for illustration. The inverse FFT of the data is shown in figure 2.10 and shows a clear improvement in terms of trend reduction. Recognizing that the result should be a relatively nice sine wave, some artifacts of the process can be observed. This is a result of leakage as discussed in section 3.11. A signal prone to leakage will be prone to these distortions. All of the benefits and costs associated with windowing and leakage also apply to this correction technique. This method is most appropriate for an oscillating drift in the data.

¹⁵Manufacturers always advise that sensors and excitors be brought to thermal equilibrium before use.

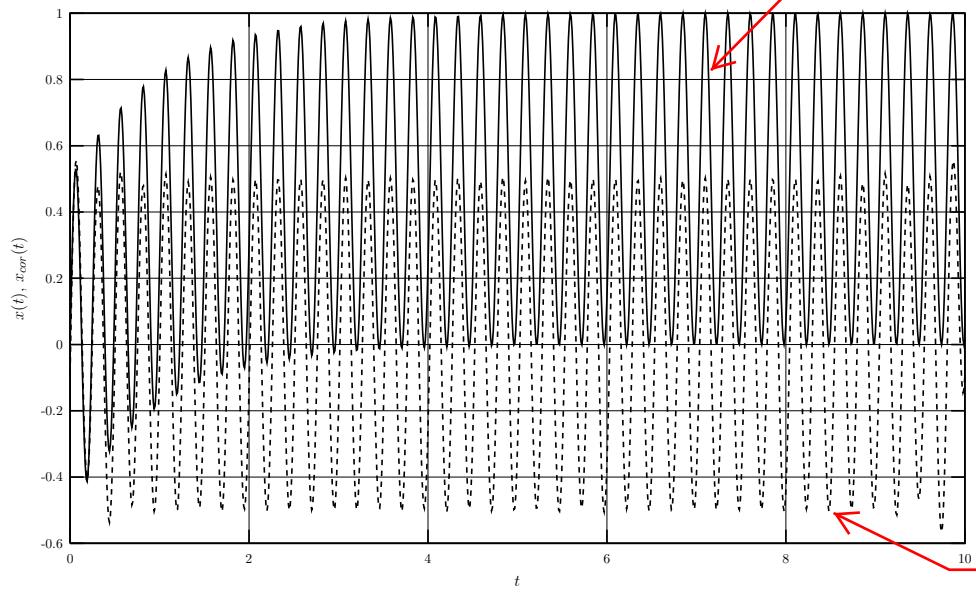


Figure 2.10 Original and detrended signal via FFT method.

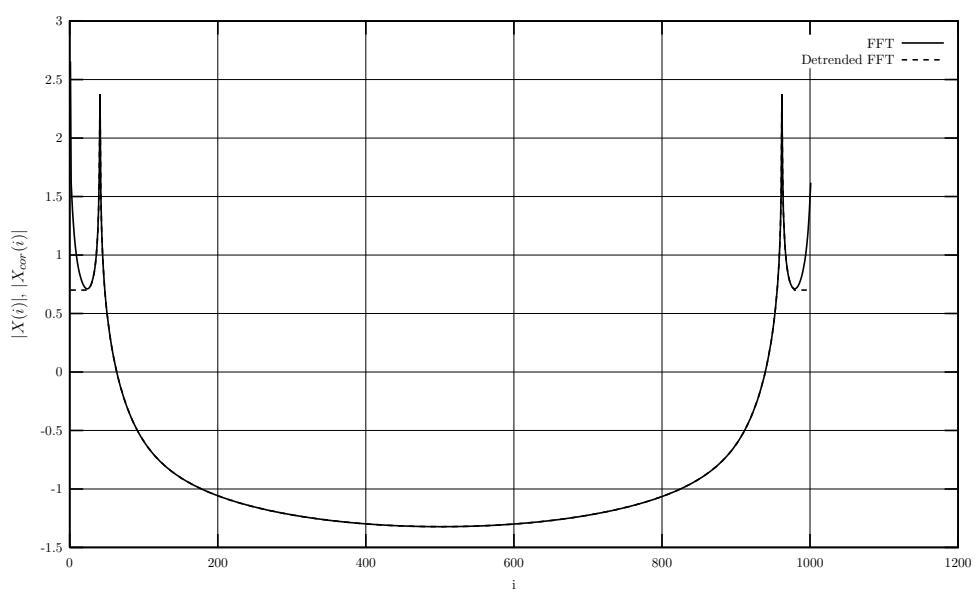


Figure 2.11 Original and detrended magnitude of FFT of signal.

2.13.2 Detrending by Curve Fit Extraction (time domain)

The second way to de-trend data avoids the problems of the FFT correction method, but results in additional concerns that must be carefully addressed. Presuming that the trend is slow, a relatively

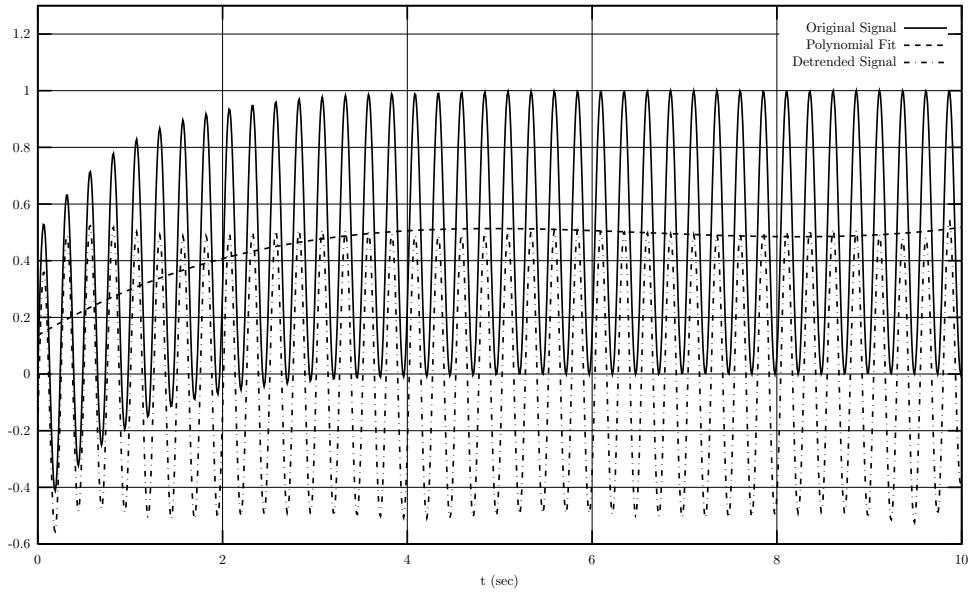


Figure 2.12 Original and detrended signals using polynomial detrending.

smooth polynomial can be subtracted from the signal. Appendix B.11 covers the process of curve fitting polynomials. Using the methods described there, a relatively low order polynomial can be fit to the data as shown in Figure 2.12. A relatively smooth polynomial approximately tracing the drifting mean of the data is plotted. This curve is subtracted from the original data to obtain the de-trended data.

It is important to be aware that de-trending data this way only works for relatively slowly drifting data. Oscillating drift cannot be removed by this method.

2.14 Nonlinear Calibration

2.15 Problems

1. Prove that the kurtosis value of a Gaussian distribution is 3.
2. Prove that the kurtosis value of a *sine* function is 1.5.
3. Derive equation (2.18).
4. Derive equation (2.19).
5. Validate equation (2.46) by using both

$$\sigma_x^2 = \int_{-\infty}^{\infty} p_x(x) x^2 dx$$

and

$$\sigma_x^2 = \int_o^{\infty} p_y(y) y^2 dy$$

to obtain σ_x . Hint: $\int p_x(x)dx = \frac{1}{2}erf\left(\frac{x}{\sqrt{2}\sigma_x}\right) = P(x)$ where erf is the error function, $\int P(x) = \sigma_x^2 p(x) + xP(x)$, and $\frac{d\text{erf}}{dx} = \frac{2e^{-x^2}}{\sqrt{\pi}}$.

6. Derive equation (2.93).

7. Derive equation (2.94).

8. Derive equation (2.95).

9. Derive equation (2.95).

10. Prove that

$$M_x(s) = e^{s\mu_x + \frac{1}{2}\sigma_x^2 s^2} \quad (2.139)$$

for the normally distributed variable x .

11. Prove that using the cross correlation function to determine the phase difference between two harmonic signals is more reliable than observing the relative time shift between the two signals via taking the time difference between two peaks.
12. Derive equations 2.118, 2.120 and 2.121 presuming $\hat{y} = y$ and $\hat{x} \neq x$.
13. Show that the phase angle between two sinusoids represented by x and y is given by $\phi = \text{acos}(\rho_{xy})$.

3

Things to look at when doing vibration testing:
Aliasing
sample rate
duration of data collection
ac-dc coupling
how to swing hammer well
how to run the cables for the sensors
thick wax on sensor vs thin wax on sensor

Introduction to Frequency Domain Random Signal Analysis

The Fast Fourier Transform is the fundamental tool of modern vibration analysis. It allows the time domain data collected during experiments to be quickly and accurately converted to frequency domain data upon which a predominant portion of vibration analysis is based (but not all!). However, discretization of data leads to a variety of additional complications beyond those that would exist if we could work with the purely continuous data. Here we will start by understanding the continuous Fourier analysis, which is more accurate than their discrete counterparts, followed by discrete Fourier analysis.

3.1 Estimation of Spectrum Densities and Frequency Response Functions

Consider the equation of motion of a single degree of freedom system given by

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) \quad (3.1)$$

Taking the Fourier transform of the equation of motion, the equation is then

$$(-m\omega^2 + cj\omega + k) X(j\omega) = \frac{1}{H(j\omega)} F(j\omega) = F(j\omega) \quad (3.2)$$

The frequency response function can then be obtained by solving for $X(j\omega)/F(j\omega)$ yielding

$$H(j\omega) = \frac{X(j\omega)}{F(j\omega)} = \frac{1}{-m\omega^2 + cj\omega + k} \quad (3.3)$$

While this method of determining the frequency response function works when we have an analytical model, often we have no model, but instead have some data in the form of time histories of signals. Let's define our measured signals such that

$$\begin{aligned} \hat{x}(t) &= x(t) + \xi_x(t) \\ \hat{f}(t) &= f(t) + \xi_f(t) \end{aligned} \quad (3.4)$$

error terms

Here $(\hat{\cdot})$ represents a measured quantity, and $\xi_{(\cdot)}(t)$ represents a random function with zero mean. If equation (3.3) is used to obtain and estimate of the FRF using real data, then the measured value of $H(j\omega)$ will be

$$\hat{H}(j\omega) = \frac{\hat{X}(j\omega)}{\hat{F}(j\omega)} = \frac{X(j\omega) + \Xi_x(j\omega)}{F(j\omega) + \Xi_f(j\omega)} \quad (3.5)$$

Expanding as a Taylor series in terms of $\Xi_x(j\omega)$ and $\Xi_f(j\omega)$, and assuming $\Xi_{(\cdot)}(j\omega)$ are small, yields

$$\begin{aligned} H_0(j\omega) &\approx \frac{X(j\omega) + \Xi_x(j\omega)}{F(j\omega) + \Xi_f(j\omega)} \\ &= \frac{X(j\omega)}{F(j\omega)} + \Xi_x(j\omega) \frac{1}{F(j\omega)} - \Xi_f(j\omega) \frac{X(j\omega)}{F(j\omega)^2} + H.O.T. \\ &= H(j\omega) + \Xi_x(j\omega) \frac{1}{F(j\omega)} - \Xi_f(j\omega) \frac{X(j\omega)}{F(j\omega)^2} + H.O.T. \\ &= H(j\omega) \left(1 + \frac{\Xi_x(j\omega)}{X(j\omega)} - \frac{\Xi_f(j\omega)}{F(j\omega)} \right) + H.O.T. \end{aligned} \quad (3.6)$$

noise in output
zeros hard to
determine here
when X is small

noise in input
poles hard to
determine since
force goes to 0
since system
stiffness is very
low at natural
frequencies

where the second and subsequent terms represent the error in the FRF estimate. The first thing one notices is that the error terms are proportional to the Fourier transform of the noise (changing both phase *and* amplitude). The error is also inversely proportional to the Fourier transform of the response, X , (second term) and the Fourier transform of the excitation, F , (third term). Thus the estimate can be improved by decreasing the noise(s) and increasing the true signals, in other words the *signal to noise ratios*. This is a relatively intuitive result, and so far has not yielded significant new insight into how to obtain a better estimate.

The problem with using this approach for estimating the FRF is that it is very sensitive to noise. First consider the case where there is very little response ($X(j\omega) \approx 0$). Under these conditions, the estimate is dominated by the second term, i.e.

$$H_0(j\omega) \approx \Xi_x(j\omega) \frac{1}{F(j\omega)} + H.O.T. \quad (3.7)$$

and its magnitude is linearly dependent on the magnitude of response noise, $\xi_x(t)$.

Next consider the case where there is very little excitation (relatively, $F(j\omega) \approx 0$). This is a common occurrence at resonances, where application of a significant-amplitude force on a structure without doing damage is difficult. In this case, the third, fifth, and other higher order odd-numbered terms dominate the response and can be approximated by

$$H_0(j\omega) \approx H(j\omega) \left(1 - \frac{\Xi_f(j\omega)}{F(j\omega)} \right) + H.O.T. \quad (3.8)$$

Here the second term is dominant at resonance because $F(j\omega) \approx 0$ unless some sort of compensation is used. As a result, in the presence of input and output noise, we will have very poor (terrible!) estimates of the FRF and the “peaks and valleys,” or more formally resonances and anti-resonances, or poles and zeros.

One potential solution to this is averaging over an ensemble of data sets. While for any given trial of an experiment $X(j\omega)$ and $F(j\omega)$ may vary considerably, $H(j\omega)$ often does not.¹ In taking average values of $H(j\omega)$ over an ensemble we will, in the limit as the number of signals in the ensembles goes to ∞ , approach the expected values of $\Xi_x(j\omega)$ and $\Xi_f(j\omega)$, but *only if* $X(j\omega)$ and $F(j\omega)$ remain

¹Some cases where this is not true are vibration tests of deploying systems such as satellites and robots, systems with varying mass like an aircraft or rocket in operation, or systems in which thermal or time effects can change damping properties, like vibration of plastics or bolts. Such systems are not time-invariant, or non-autonomous, and averaging of this sort is not valid.

constant. In practice, this condition is poorly met at best. In addition, for very lightly damped systems, when the noise terms may dominate, although the magnitude of the noise will converge to zero, the phase remains random. Thus, even with significant averaging *and* precise control of the excitation, wild fluctuations of the estimated phase angles will exist. As a result, this method for estimating the frequency response function is never used in practice.

Consider now multiplying equation (3.2) by $\frac{1}{T}\bar{F}(j\omega)$ where $(\bar{\cdot})$ represents the complex conjugate. This yields the relationship

$$\frac{1}{H(j\omega)} \frac{1}{T} X(j\omega) \bar{F}(j\omega) = \frac{1}{T} F(j\omega) \bar{F}(j\omega) = \frac{1}{H(j\omega)} G_{fx}(j\omega) = G_{ff}(j\omega) \quad (3.9)$$

where T is the length of the signals (and can be factored out), $G_{fx}(j\omega) = \frac{1}{T}\bar{F}(j\omega)X(j\omega)$ is the *two-sided cross spectrum density function (CSD)*² and $G_{ff}(j\omega) = \frac{1}{T}\bar{F}(j\omega)F(j\omega)$ is the *two-sided power or two-sided auto-spectrum density (ASD)*. They are also simply the Fourier transforms of the cross-correlation and auto-correlation of equations (2.137) and (2.130), i.e.

$$G_{ff}(j\omega) = \int_{-\infty}^{\infty} R_{ff}(\tau) e^{-j\omega\tau} d\tau \quad \boxed{\text{auto correlation function}} \quad (3.10)$$

$$G_{fx}(j\omega) = \bar{G}_{xf}(j\omega) = \int_{-\infty}^{\infty} R_{fx}(\tau) e^{-j\omega\tau} d\tau \quad \boxed{\text{cross correlation function}}$$

Conversely, applying the inverse Fourier transform

$$R_{ff}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{ff}(j\omega) e^{j\omega\tau} d\omega \quad (3.11a)$$

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{xx}(j\omega) e^{j\omega\tau} d\omega \quad (3.11b)$$

$$R_{xf}(\tau) = \bar{R}_{fx}(-\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{xf}(j\omega) e^{j\omega\tau} d\omega \quad (3.11c)$$

Equations (3.10) and (3.11b) are formally referred to as the *Wiener-Khinchine relations*. These relationships are very important for the theoretical ramifications that can be derived based upon them. However, we will discover that the first method presented for obtaining the spectrum densities is the most practical. The two forms can easily be shown to yield equivalent results. By substituting equation (2.137) for the cross-correlation we obtain

$$\begin{aligned} G_{fx}(f) &= \mathcal{F}(R_{fx}(\tau)) \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{F} \left(\int_{-T/2}^{T/2} f(t) x(t + \tau) dt \right) \end{aligned} \quad (3.12)$$

Applying equation (1.107) then gives

$$G_{fx}(f) = \bar{F}(f) X(f), \quad (3.13a)$$

$$G_{xf}(f) = \bar{X}(f) F(f), \quad (3.13b)$$

$$G_{ff}(f) = \bar{F}(f) F(f) \quad (3.13c)$$

and

$$G_{xx}(f) = \bar{X}(f) X(f) \quad (3.13d)$$

²Need to add section page 146 Bendat, phase and corr max value?

The spectrum densities can be observed to have units of (measurement units)²/Hz. Of course, in a practical situation, T is finite, so we use the finite Fourier transforms

$$X(j\omega) = \int_{-T/2}^{T/2} x(t)e^{-j\omega t} dt \quad (3.14a)$$

$$F(j\omega) = \int_{-T/2}^{T/2} f(t)e^{-j\omega t} dt \quad (3.14b)$$

to obtain *two-sided spectrum density definitions*

$$G_{xf}(j\omega) = \bar{X}(j\omega)F(j\omega) = \bar{G}_{fx}(j\omega), \quad (3.15a)$$

$$G_{xx}(j\omega) = \bar{X}(j\omega)X(j\omega), \quad \text{and} \quad (3.15b)$$

$$G_{ff}(j\omega) = \bar{F}(j\omega)F(j\omega) \quad (3.15c)$$

In practice, we usually use the *one-sided spectrum densities* defined as

$$S_{xf}(j\omega) = \begin{cases} G_{xf}(j\omega), & 0 = \omega \\ 2G_{xf}(j\omega), & 0 < \omega \end{cases} \quad (3.16a)$$

$$S_{xx}(j\omega) = \begin{cases} G_{xx}(j\omega), & 0 = \omega \\ 2G_{xx}(j\omega), & 0 < \omega \end{cases} \quad (3.16b)$$

$$S_{ff}(j\omega) = \begin{cases} G_{ff}(j\omega), & 0 = \omega \\ 2G_{ff}(j\omega), & 0 < \omega \end{cases} \quad (3.16c)$$

by recognizing that $S(j\omega) = \bar{S}(-j\omega)$, and thus no unique information exists for $\omega < 0$.

For the present time, let's consider only the case of infinite time, and two-sided spectrums. Solving for $H(j\omega)$ in equation (3.9) results in

$$\hat{H}_1(j\omega) = \frac{G_{\hat{f}\hat{x}}(j\omega)}{G_{\hat{f}\hat{f}}(j\omega)} \quad (3.17)$$

For notational simplicity, we will refer to derived quantities including error due to noise as (\cdot) , i.e. $G_{\hat{x}\hat{f}}$ will from here on be referred to as $\hat{G}_{xf}(j\omega)$. The question arises regarding the impact of measurement error on $\hat{G}_{xf}(j\omega)$ and $\hat{G}_{ff}(j\omega)$. The estimated cross spectrum density, $\hat{G}_{xf}(j\omega)$, is

$$\begin{aligned} \hat{G}_{xf} &= (\bar{X}(j\omega) + \bar{\Xi}_x(j\omega))(F(j\omega) + \Xi_f(j\omega)) \\ &= (\bar{X}(j\omega)F(j\omega) + \bar{X}(j\omega)\Xi_f(j\omega) + F(j\omega)\bar{\Xi}_x(j\omega) + \bar{\Xi}_x(j\omega)\Xi_f(j\omega)) \\ &= G_{xf}(j\omega) + G_{x\xi_f}(j\omega) + G_{\xi_x f}(j\omega) + G_{\xi_x \xi_f}(j\omega) \end{aligned} \quad (3.18)$$

This lends very little understanding to the influence of noise. If we instead apply the Wiener-Khinchine form of equations (3.10) we get

$$\begin{aligned} \hat{G}_{xf} &= \int_{-\infty}^{\infty} \hat{R}_{xf}(\tau) e^{-j\omega\tau} d\tau \\ &= \lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} \frac{1}{T} \int_{-T/2}^{T/2} (x(t) + \xi_x(t))(f(t - \tau) + \xi_f(t - \tau)) dt e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} (R_{xf}(\tau) + R_{\xi_x f}(\tau) + R_{x\xi_f}(\tau) + R_{\xi_x \xi_f}(\tau)) e^{-j\omega\tau} d\tau \end{aligned} \quad (3.19)$$

Consider now the case where the noise sources are random in the engineering sense³. What is meant by that is that they are unrelated, or more formally, there is no correlation between them, nor either with either true signal. This is a tenuous case, in that it is impossible to find the cross correlation function between the noise in a signal and the true part of the signal since they can never be separated in practice. The best that can be done is to use insight as to what the sources of noise may be, and from that judge whether or not the proposition is justified. If it is, then

$$\begin{aligned} R_{\xi_x f}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \xi_x(t) f(t + \tau) dt = 0, \\ R_{x \xi_f}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) \xi_f(t + \tau) dt = 0, \quad \text{and} \\ R_{\xi_x \xi_f}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \xi_x(t) \xi_f(t + \tau) dt = 0 \end{aligned} \quad (3.20)$$

yielding

$$\hat{G}_{xf}(j\omega) = \int_{-\infty}^{\infty} \hat{R}_{xf}(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} R_{xf}(\tau) e^{-j\omega\tau} d\tau = G_{xf}(j\omega) \quad (3.21)$$

In words, if the input and output noise of the system is uncorrelated to either of the true signals, then the cross spectrum density is unaffected by the presence of noise. The problem with this assumption is that it is not true for finite periods. We will attempt to compensate for this error in section 3.10 when we use *averaging* to reduce the error induced by this assumption. However, error and uncertainties will remain with and are addressed in section 3.10. In practice, most users simply accept the results of averaging. The presumption of uncorrelated noise is not valid, but can be a tolerable assumption.

Consider now the impact of noise on the estimate of the auto spectrum density, $\hat{G}_{ff}(j\omega)$

$$\begin{aligned} \hat{G}_{ff}(j\omega) &= \int_{-\infty}^{\infty} \hat{R}_{ff}(\tau) e^{-j\omega\tau} d\tau \\ &= \lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} \frac{1}{T} \int_{-T/2}^{T/2} (f(t) + \xi_f(t))(f(t - \tau) + \xi_f(t - \tau)) dt e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} (R_{ff}(\tau) + R_{\xi_f f}(\tau) + R_{f \xi_f}(\tau) + R_{\xi_f \xi_f}(\tau)) e^{-j\omega\tau} d\tau \end{aligned} \quad (3.22)$$

Applying again the assumption of uncorrelated noise yields

$$\begin{aligned} \hat{G}_{ff}(j\omega) &= \int_{-\infty}^{\infty} \hat{R}_{ff}(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} (R_{ff}(\tau) + R_{\xi_f \xi_f}(\tau)) e^{-j\omega\tau} d\tau \\ &= G_{ff}(j\omega) + G_{\xi_f \xi_f}(j\omega) \\ &= G_{ff}(j\omega) \left(1 + \frac{G_{\xi_f \xi_f}(j\omega)}{G_{ff}(j\omega)} \right) \end{aligned} \quad (3.23)$$

This illustrates that the error in the ASD estimate of $G_{ff}(j\omega)$ is directly proportional to the inverse of the *signal to noise ratio*, $\frac{G_{ff}(j\omega)}{G_{\xi_f \xi_f}(j\omega)}$.

Similarly

$$\begin{aligned} \hat{G}_{xx}(j\omega) &= G_{xx}(j\omega) + G_{\xi_x \xi_x}(j\omega) \\ &= G_{xx}(j\omega) \left(1 + \frac{G_{\xi_x \xi_x}(j\omega)}{G_{xx}(j\omega)} \right) \end{aligned} \quad (3.24)$$

³Engineering sense: Gaussian, zero mean, regardless of the time frame of the signal.

which is also typically better estimated with a larger response than a smaller response. Our estimate for the frequency response function H_1 is now

$$\hat{H}_1 = \frac{\hat{G}_{fx}}{\hat{G}_{ff}} = \frac{G_{fx}}{G_{ff} + G_{\xi_f \xi_f}} = \boxed{\frac{G_{fx}}{G_{ff}}} \frac{1}{1 + \frac{G_{\xi_f \xi_f}}{G_{ff}}} \quad (3.25)$$

assumed not to have error since lot of averages were assumed for the calculation of the cross spectrum density

this calculation of the response function will give a lower estimate of the true frequency response. this in turn gives the user a lower bound on the response.

where the notation of dependence on $j\omega$ has been dropped for the sake of legibility and should be presumed from here forward. Here the *signal to noise ratio*, $\frac{G_{ff}}{G_{\xi_f \xi_f}}$, shows up (inverted) in the denominator and explicitly indicates that increasing the signal (utilizing a higher input) G_{ff} in the presence of fixed noise $G_{\xi_f \xi_f}$ will result in improved estimates of the FRF. The first thing that one notices is that this estimate is insensitive to noise in the output signal, i.e. $\xi_x(t) \neq 0$ doesn't cause errors. Secondly, it is sensitive to noise in the input signal, especially in regions where $G_{ff} \ll G_{\xi_f \xi_f}$. This is most common at resonance, as was mentioned earlier. The third property of the H_1 estimate is that it is always lower in magnitude than the true value of H , and can be used as a lower bound on the magnitude.

Consider instead now consider multiplying equation (3.2) by $\frac{1}{T} \bar{X}$. This yields

$$\frac{1}{H} X(j\omega) \bar{X}(j\omega) = F \bar{X}(j\omega) = \frac{1}{H} G_{xx} = G_{xf} \quad (3.26)$$

Solving for H in equation (3.26) results in

$$\hat{H}_2 = \frac{\hat{G}_{xx}}{\hat{G}_{xf}} = \frac{G_{xx} + G_{\xi_x \xi_x}}{G_{xf}} = \boxed{\frac{G_{xx}}{G_{xf}}} \left(1 + \frac{G_{\xi_x \xi_x}}{G_{xx}} \right) \quad (3.27)$$

true frequency response function

this estimation of the response function will give upper bound on the estimation
THIS MORE ACCURATE WHEN X IS HIGH

Contrary to the H_1 estimate, the first thing that one notices is that the H_2 estimate is insensitive to noise in the *input* signal, i.e. $G_{\xi_f \xi_f} \neq 0$ doesn't cause error in the FRF estimate. Secondly, it is sensitive to noise in the output signal, especially in regions where $G_{xx} \ll G_{\xi_x \xi_x}$. This is most common at anti-resonance. The third property of the H_2 estimate is that it is always higher in magnitude than the true value of H , and can be used as an upper bound on the magnitude. As a result, the H_2 estimate is the most conservative estimate at resonances (yielding the highest amplitude) and for this reason is typically the default method used in Fourier analyzers.

Thus, for the condition of uncorrelated noise, $|H_1| < |H| < |H_2|$ provides bounds on the estimate of H . If there is no noise in the signals, then $H_1 = H_2$. The *coherence* function quantifying the quality of the FRF estimates can then be defined as

$$\begin{aligned} \gamma_{xf}^2 &= \frac{H_1}{H_2} = \frac{\hat{G}_{fx}}{\hat{G}_{ff}} \frac{\hat{G}_{xf}}{\hat{G}_{xx}} = \frac{|\hat{G}_{xf}|^2}{\hat{G}_{xx} \hat{G}_{ff}} \\ &= \frac{G_{fx} G_{xf}}{(G_{ff} + G_{\xi_f \xi_f})(G_{xx} + G_{\xi_x \xi_x})} \end{aligned} \quad (3.28)$$

where $0 \leq \gamma_{xf}^2 \leq 1$. Note that this equation is only valid when a sufficient number of averages have been taken such that the spectrum densities have approached their expected values (Juang 1994). Using equations (3.9) and (3.26)

$$\tilde{G}_{xx} = \frac{G_{fx} G_{xf}}{G_{ff} + G_{\xi_f \xi_f}} \quad (3.29)$$

is a derived estimate of G_{xx} (as compared to a directly calculated estimate from the signal x). Because $G_{\xi_f \xi_f}$ is positive, it provides a lower bound on G_{xx} . Thus the coherence can also be considered to be the ratio of the derived estimate auto-spectrum density \tilde{G}_{xx} and the directly calculated auto-spectrum

density \hat{G}_{xx} . To assure that a real value for γ_{xf}^2 is obtained in the presence of noise, the coherence is usually calculated as

$$\begin{aligned}\gamma_{xf}^2 &= \frac{|G_{fx}|^2}{(G_{ff} + G_{\xi_f \xi_f})(G_{xx} + G_{\xi_x \xi_x})} \\ &= \frac{|G_{fx}|^2}{G_{ff}G_{xx} + G_{ff}G_{\xi_x \xi_x} + G_{xx}G_{\xi_f \xi_f} + G_{\xi_f \xi_f}G_{\xi_x \xi_x}}\end{aligned}\quad (3.30)$$

Since $\frac{|G_{fx}|^2}{G_{ff}G_{xx}} = 1$ for a linear system, this can be reduced to

$$\gamma_{xf}^2 = \frac{1}{1 + \frac{G_{\xi_f \xi_f}}{G_{ff}} + \frac{G_{\xi_x \xi_x}}{G_{xx}} + \frac{G_{\xi_f \xi_f}G_{\xi_x \xi_x}}{G_{ff}G_{xx}}}\quad (3.31)$$

For the sake of illustration, consider the Taylor series expansion in terms of the noise auto spectrum densities giving

$$\gamma_{xf}^2 = \left(1 - \frac{G_{\xi_x \xi_x}}{G_{xx}} - \frac{G_{\xi_f \xi_f}}{G_{ff}}\right) + H.O.T.\quad (3.32)$$

Observing the second and third terms, when the excitation or response auto spectrum densities are near zero, the coherence is very sensitive to noise. This demonstrates the often observed phenomenon of coherence drops at sharp resonance and anti-resonances, respectively. Using equation (3.31), if no noise exists on the input signal, $f(t)$, then the noise on the output channel can be estimated using

$$G_{\xi_x \xi_x} = (1 - \gamma_{xf}^2) G_{xx}\quad (3.33)$$

and if no noise exists on the output channel, $x(t)$, then the noise on the input signal can be estimated by

$$G_{\xi_f \xi_f} = (1 - \gamma_{xf}^2) G_{ff}\quad (3.34)$$

If noise exists on both channels, then an estimate of the signal noise over the entire frequency range isn't possible. However, applying equation (3.32) in regions where $\frac{G_{\xi_f \xi_f}}{G_{ff}} \approx 0$, i.e. near sharp anti-resonances, the noise in the output signal can be estimated by equation (3.33). Likewise, in regions where $\frac{G_{\xi_x \xi_x}}{G_{xx}} \approx 0$, i.e. near sharp resonances, the noise in the input signal can be estimated by equation (3.34). In practical situations, noise is always non-zero, and such approximations must be made. However, caution must be taken because noise is typically frequency dependent. Consider now estimating only the magnitude of the FRF. Solving equation (3.2) for \hat{X} and using measured signals yields

$$\hat{X} = \hat{H}\hat{F}\quad (3.35)$$

Multiplying equation (3.35) by its complex conjugate gives

$$\hat{X}\hat{X} = \hat{H}\hat{F}\hat{F}^*\hat{H}^*\quad (3.36)$$

If we pre-multiply by $1/T$ then

$$\hat{G}_{xx} = \hat{H}\hat{G}_{ff}\hat{H}^* = \hat{G}_{ff}|\hat{H}|^2\quad (3.37)$$

which yields an estimate of the magnitude of the FRF as

$$|\hat{H}|^2 = \frac{\hat{G}_{xx}}{\hat{G}_{ff}} = \frac{G_{xx} + G_{\xi_x \xi_x}}{G_{ff} + G_{\xi_f \xi_f}}\quad (3.38)$$

SOLUTION FOR AMPLITUDE ONLY



which suffers from the noise sensitivities of both the H_1 and H_2 FRF estimates. Ideally we would like to find the FRF estimate that minimizes the noise. If the true value of $H = \frac{G_{xx}}{G_{xf}}$ were available, the noise in the output signal could be obtained, using equation (3.24), as

$$G_{\xi_x \xi_x} = \hat{G}_{xx} - G_{xx} = \hat{G}_{xx} - G_{xf}H \quad (3.39)$$

Likewise, the noise in the input signal, using $H = \frac{G_{fx}}{G_{xx}}$ is

$$G_{\xi_f \xi_f} = \hat{G}_{ff} - G_{ff} = \hat{G}_{ff} - \frac{G_{fx}}{H} \quad (3.40)$$

Minimizing the sum of these two noise values makes no sense in that they may not even have the same units. In order to obtain a quantity that has consistent units, consider applying equation (3.37) to the input noise, as opposed to the input itself. This is the process leaving to the H_v frequency response estimation derived by Rocklin et al. (1985) and Leuridan et al. (1986) which is more sophisticated than the also improved averaging method proposed by Mitchell (1982). The auto spectrum of $\xi_f(t)$ processed by H is then

$$HG_{\xi_f \xi_f} \bar{H} \quad \text{units of disp. for the multiplication} \quad (3.41)$$

Substituting for $G_{\xi_f \xi_f}$ using equation (3.40) yields

$$HG_{\xi_f \xi_f} \bar{H} = H\hat{G}_{ff} \bar{H} - G_{fx} \bar{H} \quad (3.42)$$

Adding the noise quantities defined by equations (3.39) and (3.42) and calling the result λ yields

$$\begin{aligned} \lambda &= G_{\xi_x \xi_x} + HG_{\xi_f \xi_f} \bar{H} \\ &= \hat{G}_{xx} - G_{xf}H + H\hat{G}_{ff} \bar{H} - G_{fx} \bar{H} \end{aligned} \quad (3.43)$$

The objective of minimizing the error in the FRF estimates is thus one of determining the value $H = H' + jH''$ that minimizes λ . The difficulty lies in the fact that H is complex. Taking $\frac{d\lambda}{dH'} = 0$ and $\frac{d\lambda}{dH''} = 0$ yields two coupled quadratic equations in the two unknowns. The easiest way to solve this is to pose it as an eigenvalue problem. Consider the Hermitian matrix⁴

$$\mathbf{S} = \begin{bmatrix} \hat{G}_{xx} & G_{fx} \\ G_{xf} & \hat{G}_{ff} \end{bmatrix} \quad (3.44)$$

The eigenvalue problem for the matrix \mathbf{S} is equivalent to the minimization of equation (3.64). To see this, consider normalizing the eigenvector corresponding to λ so that its second value is -1 . The eigenvalue solution can then be written as

$$\begin{bmatrix} \bar{H} & -1 \end{bmatrix} \begin{bmatrix} \hat{G}_{ff} & G_{fx} \\ G_{xf} & \hat{G}_{xx} \end{bmatrix} \begin{bmatrix} H \\ -1 \end{bmatrix} = \lambda \quad (3.45)$$

derivation below deals with combining H_1 and H_2 to get optimal estimation of true H

the values in \mathbf{S} are scalar values at some frequency value (say 10Hz) or other values. It is not a matrix of matrices

just H by itself is assumed to be the true frequency response function

this lambda is the resulting error of the noise in the input and output and we want to minimize this value by optimizing as one means

multiplying the expression will yield 3.42
To solve this problem is to calculate eigenvalues of the matrix \mathbf{S} then you will get one min and one max eigenvalue. Take min value and get eigenvector which then will yield value for H .

Carrying out the operations illustrates that equations (3.64) and (3.45) are equivalent. Because \mathbf{S} is 2×2 , there are two solutions to equation (3.45). However, the one we are interested in is clearly the one for which the noise, λ , is lowest. The FRF derived by this means is referred to as the H_v estimate. Its advantage over the H_1 and H_2 estimates are that it will lie between the two estimates, and match the H_1 or H_2 estimate when there is no noise in the output or input signals, respectively.

A fourth alternative to estimation of the FRF is the H_3 estimate. The H_3 estimate takes advantage of the noise-cleaning behavior of the cross spectrum density. If a third signal, $y(t)$, is available, equation (3.2) can be multiplied by \bar{Y} , yielding

$$H_3 = \frac{G_{yx}}{G_{yf}} \quad (3.46)$$

Yes a better estimate is likely obtained using the more sophisticated, and complicated, H_s method by Wicks and Mitchell (1987); Wicks and Vold (1986).

⁴A Hermitian Matrix, A , is one where its conjugate transpose, represented by (\cdot^H) , is equal to itself. The eigenvalues are real, and the eigenvectors satisfy $\mathbf{v}^H \mathbf{A} \mathbf{v} = \lambda$.

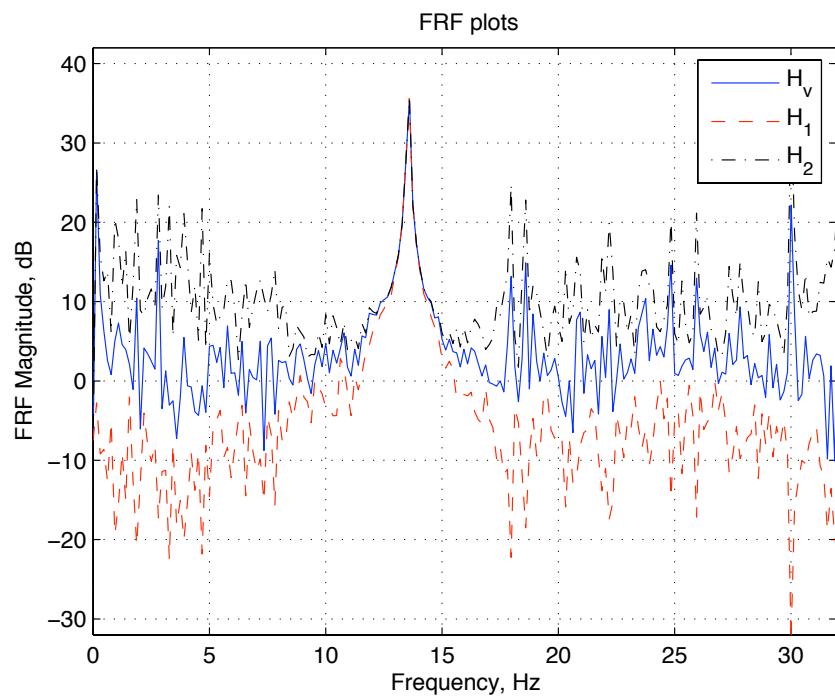


Figure 3.1 Frequency response function estimates using the H_1 , H_2 , and H_v methods.

3.2 Frequency Response Functions for Multiple Input Systems

Following the derivation of Allemand et al. (1982, 1983); Leuridan et al. (1986); Rocklin et al. (1985), consider now the multiple degree of freedom (MDOF) system defined by

$$M\ddot{\mathbf{x}}(t) + C\dot{\mathbf{x}}(t) + K\mathbf{x}(t) = \mathbf{F}(t) \quad (3.47)$$

Taking the Fourier transform of the equation of motion, the equation is then

$$[-M\omega^2 + Cj\omega + K] \mathbf{X} = \mathbf{F} \quad (3.48)$$

The frequency response function can then be obtained by solving for \mathbf{X} gives

$$\begin{aligned} \mathbf{X}(j\omega) &= \begin{bmatrix} X_1 \\ X_2 \\ \vdots \end{bmatrix} = \mathbf{H}\mathbf{F}(j\omega) = [-M\omega^2 + Cj\omega + K]^{-1} \mathbf{F}(j\omega) \\ &= \begin{bmatrix} H_{1,1} & H_{1,2} & \dots \\ H_{2,1} & H_{2,2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \end{bmatrix} \end{aligned} \quad (3.49)$$

where it should be noted that H_{lm} is the frequency response function between input m and output l . Multiplying the i^{th} row of equation (3.49) by \bar{F}_j yields

$$\begin{aligned} \bar{F}_j X_i &= H_{i1} \bar{F}_j F_1 + H_{i2} \bar{F}_j F_2 + \dots \\ &= \sum_{k=1}^n H_{ik} \bar{F}_j F_k \end{aligned} \quad (3.50)$$

After taking the limit as $T \rightarrow \infty$ and applying equations (3.15a) and (3.15c)

$$G_{f_j x_i} = \sum_{k=1}^n H_{ik} G_{f_j f_k} \quad (3.51)$$

For the sake of compactness and clarity the annotation of dependence on $j\omega$ is neglected for the remainder of the section. This can be written in matrix form as

$$\begin{bmatrix} G_{f_1 x_i} \\ G_{f_2 x_i} \\ \vdots \end{bmatrix} = \begin{bmatrix} G_{f_1 f_1} & G_{f_1 f_2} & \dots \\ G_{f_2 f_1} & G_{f_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \quad (3.52)$$

Solution for the H_1 frequency response functions is then given by

$$\begin{aligned}
 \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} &\approx \begin{bmatrix} \hat{H}_{1:i1} \\ \hat{H}_{1:i2} \\ \vdots \end{bmatrix} \\
 &= \begin{bmatrix} \hat{G}_{f_1 f_1} & \hat{G}_{f_1 f_2} & \dots \\ \hat{G}_{f_2 f_1} & \hat{G}_{f_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} \hat{G}_{f_1 x_i} \\ \hat{G}_{f_2 x_i} \\ \vdots \end{bmatrix} \\
 &= \begin{bmatrix} G_{f_1 f_1} + G_{f_1 \xi_{f_1}} + G_{\xi_{f_1} f_1} + G_{\xi_{f_1} \xi_{f_1}} & G_{f_1 f_2} + G_{f_1 \xi_{f_2}} + G_{\xi_{f_1} f_2} + G_{\xi_{f_1} \xi_{f_2}} & \dots \\ G_{f_2 f_1} + G_{f_2 \xi_{f_1}} + G_{\xi_{f_2} f_1} + G_{\xi_{f_2} \xi_{f_1}} & G_{f_2 f_2} + G_{f_2 \xi_{f_2}} + G_{\xi_{f_2} f_2} + G_{\xi_{f_2} \xi_{f_2}} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \\
 &\begin{bmatrix} G_{f_1 x_i} + G_{f_1 \xi_{x_i}} + G_{\xi_{f_1} x_i} + G_{\xi_{f_1} \xi_{x_i}} \\ G_{f_2 x_i} + G_{f_2 \xi_{x_i}} + G_{\xi_{f_2} x_i} + G_{\xi_{f_2} \xi_{x_i}} \\ \vdots \end{bmatrix}
 \end{aligned} \tag{3.53}$$

Here $\hat{H}_{1:i1}$ denotes the H_1 FRF estimate between the 1st input and the i^{th} output. This can break down when the inputs are not mutually independent leading to *singularity* of the auto spectrum matrix. Also of note is that ideally the noise on the input channels f_i are uncorrelated resulting in the off-diagonal noise terms $G_{\xi_{f_i} \xi_{f_j}}$ being zero. In practice this isn't the case. Nevertheless, the diagonal noise terms $G_{\xi_{f_i} \xi_{f_i}}$ are certainly greater than zero for non-zero noise. This causes the H_1 estimate to be low, under-predicting peaks. Under the assumption of uncorrelated noise, the H_1 estimate of the frequency response is then estimated as

$$\begin{aligned}
 \begin{bmatrix} \hat{H}_{1:i1} \\ \hat{H}_{1:i2} \\ \vdots \end{bmatrix} &= \begin{bmatrix} G_{f_1 f_1} + G_{\xi_{f_1} \xi_{f_1}} & G_{f_1 f_2} & \dots \\ G_{f_2 f_1} & G_{f_2 f_2} + G_{\xi_{f_2} \xi_{f_2}} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} G_{f_1 x_i} \\ G_{f_2 x_i} \\ \vdots \end{bmatrix} \\
 &= ([G_{ff}] + [G_{\xi_f \xi_f}])^{-1} [G_{fx_i}]
 \end{aligned} \tag{3.54}$$

If the number of inputs is not equal to the number of outputs, then a pseudo-inverse must be used instead of an inverse. Further, if we assume a case where there is no noise in the input signal, $f_i(t)$, and that there is noise on the output signal x_i has zero correlation with each input signal, then results and characteristics of the estimate will agree per prior arguments regarding the H_1 estimate for single degree of freedom systems.

Given the criticality of identifying the peak response, the H_2 method that follows is more highly regarded near peaks. Consider again the i^{th} row of equation (3.49) but this time premultiplying by \bar{X}_j . This yields

$$\begin{aligned}
 \bar{X}_j X_i &= \bar{X}_j \mathbf{H}_i \mathbf{F} = \mathbf{H}_i \bar{X}_j \mathbf{F} \\
 &= \sum_{k=1}^n H_{ik} \bar{X}_j F_k
 \end{aligned} \tag{3.55}$$

or, taking the limit as $T \rightarrow \infty$ and applying equations (3.15a) and (3.15c)

$$\begin{aligned} G_{x_j x_i} &= \sum_{k=1}^n H_{ik} G_{x_j f_k} \\ &= \mathbf{G}_{x_i f}^T \mathbf{H}_i \\ &= [G_{x_i f_1} \quad G_{x_i f_2} \quad \dots] \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \end{aligned} \quad (3.56)$$

Writing this in matrix form for all output coordinates yields

$$\begin{bmatrix} G_{x_1 x_i} \\ G_{x_2 x_i} \\ \vdots \end{bmatrix} = \begin{bmatrix} G_{x_1 f_1} & G_{x_1 f_2} & \dots \\ G_{x_2 f_1} & G_{x_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \quad (3.57)$$

The H_2 form of the frequency response functions is then

$$\begin{aligned} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} &\approx \begin{bmatrix} \hat{H}_{2;i1} \\ \hat{H}_{2;i2} \\ \vdots \end{bmatrix} \\ &= \begin{bmatrix} \hat{G}_{x_1 f_1} & \hat{G}_{x_1 f_2} & \dots \\ \hat{G}_{x_2 f_1} & \hat{G}_{x_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} \hat{G}_{x_1 x_i} \\ \hat{G}_{x_2 x_i} \\ \vdots \end{bmatrix} \\ &= \begin{bmatrix} G_{x_1 f_1} + G_{x_1 \xi_{f_1}} + G_{\xi_{x_1} f_1} + G_{\xi_{x_1} \xi_{f_1}} & G_{x_1 f_2} + G_{x_1 \xi_{f_2}} + G_{\xi_{x_1} f_2} + G_{\xi_{x_1} \xi_{f_2}} & \dots \\ G_{x_2 f_1} + G_{x_2 \xi_{f_1}} + G_{\xi_{x_2} f_1} + G_{\xi_{x_2} \xi_{f_1}} & G_{x_2 f_2} + G_{x_2 \xi_{f_2}} + G_{\xi_{x_2} f_2} + G_{\xi_{x_2} \xi_{f_2}} & \dots \\ \vdots & \vdots & \ddots \\ G_{x_1 x_i} + G_{x_1 \xi_{x_i}} + G_{\xi_{x_1} x_i} + G_{\xi_{x_1} \xi_{x_i}} & G_{x_2 x_i} + G_{x_2 \xi_{x_i}} + G_{\xi_{x_2} x_i} + G_{\xi_{x_2} \xi_{x_i}} & \dots \end{bmatrix}^{-1} \\ &\quad \vdots \end{aligned} \quad (3.58)$$

Where in the ideal case the noise has zero correlation and thus

$$\begin{aligned} \begin{bmatrix} \hat{H}_{2;i1} \\ \hat{H}_{2;i2} \\ \vdots \end{bmatrix} &= \begin{bmatrix} \hat{G}_{x_1 f_1} & \hat{G}_{x_1 f_2} & \dots \\ \hat{G}_{x_2 f_1} & \hat{G}_{x_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} \hat{G}_{x_1 x_i} \\ \hat{G}_{x_2 x_i} \\ \vdots \end{bmatrix} \\ &= \begin{bmatrix} G_{x_1 f_1} & G_{x_1 f_2} & \dots \\ G_{x_2 f_1} & G_{x_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} G_{x_{<i} x_i} \\ \vdots \\ G_{x_i x_i} + G_{\xi_{x_i} \xi_{x_i}} \\ \vdots \\ G_{x_{>i} x_i} \end{bmatrix} \end{aligned} \quad (3.59)$$

Without a complete set of response measurements, the matrix \mathbf{G}_{xf} requires the use of a pseudo-inverse. To be successful, the number of inputs o must be less than or equal to the number of outputs p . Having fewer output channels than input channels will most likely lead to unacceptable results.

Next, consider the derived output ASD resulting from the input. As in the SISO case, consider the output ASD resulting from the all of the inputs combined. The i^{th} row of (3.48) can be written as

$$\begin{aligned} X_i &= \sum_{k=1}^o F_k H_{ik} \\ &= [F_1 \quad F_2 \quad \dots] \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \end{aligned} \quad (3.60)$$

where o is the number of input forces. Multiplying each side by its own complex conjugate⁵ yields

$$\begin{aligned} \bar{X}_i X_i &= [\bar{H}_{i1} \quad \bar{H}_{i2} \quad \dots] \begin{bmatrix} \bar{F}_1 \\ \bar{F}_2 \\ \vdots \end{bmatrix} [F_1 \quad F_2 \quad \dots] \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \\ &= [\bar{H}_{i1} \quad \bar{H}_{i2} \quad \dots] \begin{bmatrix} \bar{F}_1 F_1 & \bar{F}_1 F_2 & \dots \\ \bar{F}_2 F_1 & \bar{F}_2 F_2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \end{aligned} \quad (3.61)$$

which after applying definitions (3.13) gives

$$G_{x_i x_i} = [\bar{H}_{i1} \quad \bar{H}_{i2} \quad \dots] \begin{bmatrix} G_{f_1 f_1} & G_{f_1 f_2} & \dots \\ G_{f_2 f_1} & G_{f_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \quad (3.62)$$

This is an idealized case, however. In the presence of uncorrelated noise the estimate of the output autospectrum is

$$\begin{aligned} \tilde{G}_{x_i x_i} &= [\bar{H}_{i1} \quad \bar{H}_{i2} \quad \dots] \begin{bmatrix} G_{f_1 f_1} & G_{f_1 f_2} & \dots \\ G_{f_2 f_1} & G_{f_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \\ &+ [\bar{H}_{i1} \quad \bar{H}_{i2} \quad \dots] \begin{bmatrix} G_{\xi_{f_1} \xi_{f_1}} & 0 & \dots \\ 0 & G_{\xi_{f_2} \xi_{f_2}} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} H_{i1} \\ H_{i2} \\ \vdots \end{bmatrix} \\ &= \bar{\mathbf{H}}_i^T \mathbf{G}_{ff} \mathbf{H}_i + \bar{\mathbf{H}}_i^T \mathbf{G}_{\xi_f \xi_f} \mathbf{H}_i \end{aligned} \quad (3.63)$$

which illustrates that if noise is more substantial in the output than input signals, then obtaining the derived output ASD from (3.63) could be advantageous provided the FRFs are known.

⁵Since the equation is a scalar, the right side isn't actually pre-multiplied by its transpose. However, the transpose identity was used to swap the order of the matrices.

The total noise contribution to the output signal can then be written, using equations (3.39), (3.56), (3.63) and (3.52) shown here sequentially) as

$$\begin{aligned}
 \lambda &= G_{\xi_{x_i} \xi_{x_i}} + \bar{\mathbf{H}}_i^T \mathbf{G}_{\xi_f \xi_f} \mathbf{H}_i \\
 &= \hat{G}_{xx} - \mathbf{G}_{x_i f}^T \mathbf{H}_i + \bar{\mathbf{H}}_i^T \mathbf{G}_{\xi_f \xi_f} \mathbf{H}_i \\
 &= \hat{G}_{xx} - \mathbf{G}_{x_i f}^T \mathbf{H}_i + \bar{\mathbf{H}}_i^T (\hat{\mathbf{G}}_{ff} - \mathbf{G}_{ff}) \mathbf{H}_i \\
 &= \hat{G}_{xx} - \mathbf{G}_{x_i f}^T \mathbf{H}_i + \bar{\mathbf{H}}_i^T \hat{\mathbf{G}}_{ff} \mathbf{H}_i - \bar{\mathbf{H}}_i^T \mathbf{G}_{ff} \mathbf{H}_i \\
 &= \hat{G}_{xx} - \mathbf{G}_{x_i f}^T \mathbf{H}_i + \bar{\mathbf{H}}_i^T \hat{\mathbf{G}}_{ff} \mathbf{H}_i - \bar{\mathbf{H}}_i^T \mathbf{G}_{fx}
 \end{aligned} \tag{3.64}$$

This can be written in matrix form as

$$\lambda = [\mathbf{H}_i^H \quad -1] \begin{bmatrix} \hat{\mathbf{G}}_{ff} & \mathbf{G}_{fx_i} \\ \mathbf{G}_{fx_i}^H & \hat{G}_{xx} \end{bmatrix} \begin{bmatrix} \mathbf{H}_i \\ -1 \end{bmatrix} \tag{3.65}$$

where the solution is the familiar eigenvalue problem, the lowest eigenvalue λ representing minimization of the impact of noise. The corresponding eigenvector is then normalized so that the final value is -1 to solve for the H_v estimates of the FRFs.

Multiplying both sides of equation (3.49) by their own corresponding conjugate transposes gives

$$\begin{aligned}
 X_i \bar{X}_i &= \mathbf{H}_1 \mathbf{F} \bar{\mathbf{F}}^T \bar{\mathbf{H}}_1^T \\
 &= [H_{(1)1} \quad H_{(1)2} \quad \dots] \begin{bmatrix} F_1 \\ F_2 \\ \vdots \end{bmatrix} [\bar{F}_1 \quad \bar{F}_2 \quad \dots] \begin{bmatrix} \bar{H}_{(1)1} \\ \bar{H}_{(1)2} \\ \vdots \end{bmatrix} \\
 &= [H_{(1)1} \quad H_{(1)2} \quad \dots] \begin{bmatrix} F_1 \bar{F}_1 & F_1 \bar{F}_2 & \dots \\ F_2 \bar{F}_1 & F_2 \bar{F}_2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \bar{H}_{(1)1} \\ \bar{H}_{(1)2} \\ \vdots \end{bmatrix}
 \end{aligned} \tag{3.66}$$

Applying equations (3.15a) and (3.15c)

$$G_{x_i x_i} = [H_{(1)1} \quad H_{(1)2} \quad \dots] \begin{bmatrix} G_{f_1 f_1} & G_{f_2 f_1} & \dots \\ G_{f_1 f_2} & G_{f_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \bar{H}_{(1)1} \\ \bar{H}_{(1)2} \\ \vdots \end{bmatrix} \tag{3.67}$$

Consider now instead multiplying equation (??) by \bar{X}_i and taking the limit as $T \rightarrow \infty$

$$G_{x_1 x_1} = [H_{11}(j\omega) \quad H_{12}(j\omega) \quad \dots] \begin{bmatrix} G_{x_1 f_1} \\ G_{x_1 f_2} \\ \vdots \end{bmatrix} \tag{3.68}$$

Further, dividing both sides by $G_{x_1 x_1}$ gives

$$\gamma_{x_1 f_i} = \frac{[H_{11}(j\omega) \quad H_{12}(j\omega) \quad \dots] \begin{bmatrix} G_{x_1 f_1} \\ G_{x_1 f_2} \\ \vdots \end{bmatrix}}{G_{x_1 x_1}} \tag{3.69}$$

Substitute into the above to get rid of H. Illustrate gam-masquared is G_{xx} calculated divded by G_{xx} from raw data (Newland) Note why this must be less than 1. Compare to SDOF system

Noting equation (??), we can substitute for the unknown frequency response functions and obtain the *multiple coherence* as

$$\gamma_{x_1 f_i} = \frac{\begin{bmatrix} G_{f_1 x_{(i)}} & G_{f_2 x_{(i)}} & \dots \end{bmatrix} \begin{bmatrix} G_{f_1 f_1} & G_{f_1 f_2} & \dots \\ G_{f_2 f_1} & G_{f_2 f_2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} G_{x_i f_1} \\ G_{x_i f_2} \\ \vdots \end{bmatrix}}{G_{x_1 x_1}} \quad (3.70)$$

Provided that the inputs, f_i , are uncorrelated, the matrix of input auto-spectrum densities will be diagonal. Then equation (3.70) can be written as a summation of *partial coherences*. When they are correlated, the dependence must be accounted for, or the summation of the partial coherences can be greater than 1 (Bendat and Piersol 2000).

3.3 The Discrete and Fast Fourier Transforms

The Discrete Fourier Transform (DFT) is the foundation of computational frequency domain analysis while the Fast Fourier Transform (FFT) algorithm makes calculation of the DFT feasible for large data sequences. Uses of the DFT include estimation of frequency response functions, power- and cross-spectrum densities, auto- and cross-correlations, and impulse response functions from frequency domain data. These calculations allow the vibration analyst to identify natural frequencies, damping ratios, mode shapes and system models from experimental data. However, issues such as sampling rates, leakage, and aliasing must be addressed to obtain quality data.

The practical application of the DFT is the Fast Fourier Transform, or FFT. FFT algorithms result in precisely the same result as a DFT, but in a much more efficient manner, with exceptional computational savings. Since FFTs are available in numerous software packages and available subroutines, FFT algorithms will not be covered here. Instead, the more illustrative, and simpler, DFT will be considered, with the understanding that the mathematical operations here are illustrative in nature, but do not represent how real calculations are performed. For the details of FFT algorithms, the interested reader can consult the text of Newland (Newland 1993).

Fourier's theorem states that any periodic function can be represented by its harmonic components (see section 1.3). The frequency spacing between each harmonic component is $\Delta f = 1/T$ where T is the period of the repeating function. As the period of the repeating function increases, the spacing between adjacent harmonics of the Fourier series diminishes. As the period continues to increase, the functions become non-repeating and the distribution of the harmonics becomes continuous. Such functions can no longer be represented by a discrete Fourier series but must instead be represented by a function of ω as was derived in section 1.4. Here we return to a finite period. Discretizing equation (1.80) yields

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} X(\omega) e^{j2\pi f t} df \\ &= \lim_{N \rightarrow \infty} \Delta f \sum_{n=-N/2+1}^{N/2} X_n e^{j\frac{2\pi n}{T} t} \\ &= \lim_{N \rightarrow \infty} \Delta f \sum_{n=0}^{N-1} X_n e^{j\frac{2\pi n}{T} t} \end{aligned} \quad (3.71)$$

where we will prove the equivalence of changing the limits later in this section. Consider the case where we have a set of discrete data points, x_m where m is an integer $0 \leq m \leq N - 1$. Substituting

$$t = t_m = \frac{mT}{N} \quad (3.72)$$

into equation (3.71) gives the *inverse discrete Fourier transform* (IDFT)

$$x_m = \Delta f \sum_{n=0}^{N-1} X_n e^{j \frac{2\pi n}{N} m}, \quad m = 0, 1, 2, \dots, N-1 \quad (3.73)$$

The Fourier coefficients, X_n , can be obtained from equation (1.81)

$$X_n = \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} n}, \quad n = 0, 1, 2, \dots, N-1 \quad (3.74)$$

by noting that $\frac{\Delta t}{T} = \frac{1}{N}$ if $x(t)$ is represented as the discrete set of data points x_m . This is called the forward *discrete Fourier transform*, or DFT. Noting equation (3.72),

$$x_m = x(t_m) = x\left(\frac{mT}{N}\right) = \Delta f \sum_{n=0}^{N-1} X_n e^{j \frac{2\pi n}{N} m} \quad (3.75)$$

Also, comparing the discrete summation of equation (3.71) to its continuous counterpart (also equation (3.71)), the change in frequency from one term to the next (due to incrementing n) is

$$\Delta f = \frac{1}{T} \quad (3.76)$$

and thus we can write equation (3.74) as

$$\begin{aligned} X_n = X(\omega_n) &= X(n\Delta f) = \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} n} \\ &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j 2\pi m \frac{\Delta t}{T} n} \\ &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j 2\pi(m\Delta t)(n\Delta f)}, \quad n = 0, 1, 2, \dots, N-1 \end{aligned} \quad (3.77)$$

This yields some very important observations. From equation (3.76), in order to have a very fine frequency resolution (small Δf), the temporal length of the signal, T , should be long. Next consider evaluating $X_{N/2-l} + X_{N/2+l}$. Applying equation (3.77)

$$\begin{aligned} X_{N/2-l} + X_{N/2+l} &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} (N/2-l)} + \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} (N/2+l)} \\ &= \Delta t \sum_{m=0}^{N-1} x_m \left(e^{-j \frac{2\pi m}{N} (N/2-l)} + e^{-j \frac{2\pi m}{N} (N/2+l)} \right) \\ &= \Delta t \sum_{m=0}^{N-1} x_m \left(e^{-j(\pi m - \frac{2\pi ml}{N})} + e^{-j(\pi m + \frac{2\pi ml}{N})} \right) \\ &= \Delta t \sum_{m=0}^{N-1} x_m \left(e^{-j\pi m} \left(e^{j \frac{2\pi ml}{N}} + e^{-j \frac{2\pi ml}{N}} \right) \right) \end{aligned} \quad (3.78)$$

Applying Euler's equation (1.2) and equation (1.4), this becomes

$$\begin{aligned} X_{N/2-l} + X_{N/2+l} &= \Delta t \sum_{m=0}^{N-1} x_m \left((\cos(\pi m) + j \sin(\pi m)) 2 \cos\left(\frac{2\pi ml}{N}\right) \right) \\ &= \Delta t \sum_{m=0}^{N-1} x_m \left(\cos(\pi m) 2 \cos\left(\frac{2\pi ml}{N}\right) \right) \end{aligned} \quad (3.79)$$

Since every term in the summation is real, we have proven that the imaginary parts of $X_{N/2-l}$ and $X_{N/2+l}$ are complex conjugates. Similarly, $X_{N/2-l} - X_{N/2+l}$ can be shown to be imaginary, proving that the quantities $X_{N/2-l}$ and $X_{N/2+l}$ are indeed complex conjugates. No restrictions were placed on N or l , except that $N/2$ is an integer (N is even, we'll cover N odd). With this in mind, consider the case where $l = N/2 - p$ where $p < N/2$. Evaluating $X_{N/2+l}$ yields

$$\begin{aligned} X_{N/2+l} &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} (N/2+l)} \\ X_{N/2+(N/2-p)} &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} (N/2+N/2-p)} \\ X_{N-p} &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j 2\pi m} e^{-j \frac{2\pi m}{N} (-p)} \\ X_{N-p} &= \Delta t \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} (-p)} \\ &= X_{-p} \end{aligned} \tag{3.80}$$

The end result is that the coefficients X_n repeat every N indices, and thus we only need to calculate N of them. However, as was illustrated in equation (3.79), only $N/2 + 1$ of them are unique (The $n = 0$ and $n = N/2$ terms are real and unique). Further, because $X_{N/2-l} = \bar{X}_{N/2+l}$, and $X_{-p} = X_{N-p}$, the terms X_n for $N/2 + 1 < n < N - 1$ correspond to $X(-\omega_n)$ where $1 > n > N/2 - 1$. These are the same complex conjugate terms we were required to include in Fourier series analysis in equation (1.65). Without them, the summation yielding x_n will not be real. What this means is that we have terms corresponding to positive and negative frequency values, with the maximum magnitude observed frequency magnitude being $f_{N/2} = N/2T$. Noting that the sampling rate frequency is $f_s = N/T$, we have derived what is called the Nyquist frequency, $f_n = f_s/2$. This is Shannon's sampling theorem: *the maximum frequency about which information can be obtained using a discrete Fourier transform is 1/2 the sampling frequency*. Also note that this allows the change of indices that we made in equation (3.71). Secondly, because the maximum frequency is at $X_{N/2} = X(\frac{N}{2T})$, in order to obtain values for higher frequencies, T should be short, and N should be high. The practical implementation of the DFT and IDFT in code usually results in the ratio $\frac{1}{N}$ in front of the forwards Fourier transform instead of Δf . This is done by multiplying the DFT, equation (3.74), by $\Delta t = \frac{T}{N}$, and dividing the inverse DFT, equation (3.73), by Δt . This allows the DFT and IDFT to be performed with no information other than the length of the data sequence and the sequence itself. Users of software packages such as MATLAB, Octave and other software libraries should be aware of how this factor is included and how it must be compensated. When this is done, the backward and forwards DFTs take the form

$$x_m = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{X}_n e^{j \frac{2\pi n}{N} m}, \quad m = 0, 1, 2, \dots, N - 1 \tag{3.81}$$

and

$$\tilde{X}_n = \sum_{m=0}^{N-1} x_m e^{-j \frac{2\pi m}{N} n}, \quad n = 0, 1, 2, \dots, N - 1 \tag{3.82}$$

respectively where

$$\tilde{X}_n = \frac{1}{\Delta t} X_n \tag{3.83}$$

Example 3.3.1 Consider a data sequence of four points ($N = 4$). The Fourier representation of a discrete function x_m may be written as

$$\begin{aligned} x_m &= \frac{1}{4} \left(\tilde{X}_0 e^{j \frac{2\pi m}{4} 0} + \tilde{X}_1 e^{j \frac{2\pi m}{4} 1} + \tilde{X}_2 e^{j \frac{2\pi m}{4} 2} + \tilde{X}_3 e^{j \frac{2\pi m}{4} 3} \right) \\ &= \frac{1}{4} \left(\tilde{X}_0 + \tilde{X}_1 e^{j \frac{\pi m}{2}} + \tilde{X}_2 e^{j \pi m} + \tilde{X}_3 e^{j \frac{3\pi m}{2}} \right) \end{aligned} \quad (3.84)$$

If we write this equation for each value of m , and put the resulting equations in matrix form, the result is

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & e^{j \frac{2\pi}{4}} & e^{j \frac{4\pi}{4}} & e^{j \frac{6\pi}{4}} \\ 1 & e^{j \frac{4\pi}{4}} & e^{j \frac{8\pi}{4}} & e^{j \frac{12\pi}{4}} \\ 1 & e^{j \frac{6\pi}{4}} & e^{j \frac{12\pi}{4}} & e^{j \frac{18\pi}{4}} \end{bmatrix} \begin{bmatrix} \tilde{X}_0 \\ \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix} \begin{bmatrix} \tilde{X}_0 \\ \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \end{bmatrix} \quad (3.85)$$

Solving for the Fourier coefficients, \tilde{X}_n , gives the DFT

$$\begin{bmatrix} \tilde{X}_0 \\ \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & e^{j \frac{-2\pi}{4}} & e^{j \frac{-4\pi}{4}} & e^{j \frac{-6\pi}{4}} \\ 1 & e^{j \frac{-4\pi}{4}} & e^{j \frac{-8\pi}{4}} & e^{j \frac{-12\pi}{4}} \\ 1 & e^{j \frac{-6\pi}{4}} & e^{j \frac{-12\pi}{4}} & e^{j \frac{-18\pi}{4}} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (3.86)$$

One advantage that DFTs have over their continuous counterparts is that for any repeating sequence of N data points, an exact fit can be found using N Fourier coefficients. However, a significant drawback is that, like the Fourier Series, the DFT is defined only for repeating functions. In practice, it is rare that a set of data points can be taken that represent exactly one complete cycle. This can lead to errors due to a phenomenon called leakage.

3.4 Aliasing

A deleterious effect that can occur when performing discrete sampling of a signal is aliasing. If a high frequency signal exists above the Nyquist frequency of the sample ($2/T$) then it will be mistaken for a lower frequency signal because the discretization process allows only the inclusion of frequencies as defined in equation (3.73). For instance, consider the sampled sinusoid of Figure 3.2. Since $X((N/2 - l)/T) = \tilde{X}((N/2 + l)/T) = \tilde{X}((-N + l)/T)$, a signal with a frequency, $f = (N/2 + l)/T$, above $f_s = N/2T$, appears to be at $f = (-N + l)/T$. This is referred to as aliasing. Once a higher frequency has been sampled at a slower rate, it is impossible to distinguish it from a lower frequency signal. For instance, a 10.4 Hz signal is indistinguishable from a 9.6 Hz signal when sampling at 20 Hz. Once the continuous signal has been discretized, there is no way to determine which frequency is the true frequency. The lowest is assumed to be true, unless rather unique filtering of the signal⁶ is involved. As a result, it is imperative that the experimenter know the frequency range of interest before performing the most significant portion of an experiment. A sampling rate of f_s cannot be used to distinguish any frequencies above $f_s/2$. The sampling rate f_s is commonly referred to as the *Nyquist frequency*.

There are two ways of avoiding the problem of aliasing. The first is to sample at a sufficiently high rate such that no frequency components exist above half the sampling rate. This is impractical, since higher signals will almost always exist (noise will always exist in some form or another). The second is to use an analog filter to prevent higher frequency data from being sampled. Many data acquisition boards and Fourier analyzers have *anti-aliasing filters* built-in. Modern anti-aliasing filters are digitally

⁶Bandpass filtering can be used for this purpose

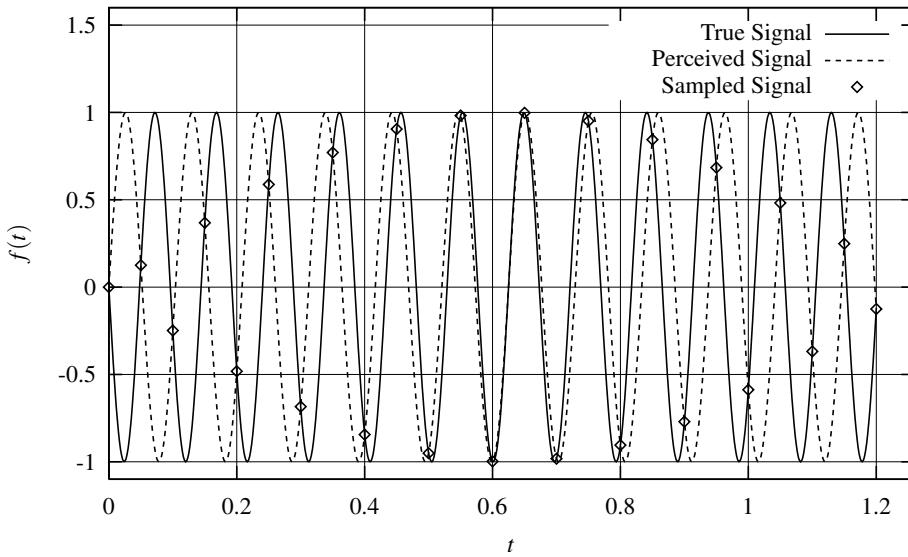


Figure 3.2 Illustration of aliased signal.

controlled analog filters with the cut-off set to approximately 80-90% of 1/2 of the sampling frequency. In displaying usable frequency content only up to 1/2.5 times the sampling frequency, the result is an *effective Nyquist Frequency* of $2.5f$. The reason for the effective Nyquist Frequency is that the analog filters roll-off in the unusable portion of the frequency signal. This data is corrupted in both phase and magnitude. The bright side of this is that modern anti-aliasing filters have a negligible effect on the frequency content of the signal over the usable range (40% of the Nyquist Frequency) and can reduce the aliased content of the resulting signal by more than 90 dB?.

Power supplies and amplifiers for many types of sensors also have low-pass filters built-in to reduce the high frequency content of the sampled signals. However, they are generally not suitable for anti-aliasing since they are usually not directly tied to the sampling rate. Figure 3.3 illustrates how signal frequencies relate to the aliased frequencies.

3.5 Decimation of Data and Practical Application of Anti-Aliasing Filters

Decimation of data, or the removal of samples to reduce the effective sampling rate, is commonly performed when a data acquisition system samples at a higher rate than necessary. The problem with this is that done incorrectly, it can reintroduce aliasing.

Consider a data set sampled at 10 kHz. A 1 second record will contain 10 000 data points. In order to prevent aliasing, the cut-off frequency for an anti-aliasing filter must be at or below the one half the Nyquist frequency of 10 kHz, or 5 kHz. Presuming that an anti-aliasing filter was used, there should be no significant signal content above 5 kHz, and we should be ignoring frequency content between 4 kHz and 5 kHz because of magnitude and phase corruption of the data in this range. If we subsequently decide that the data should have been sampled at 1 kHz, selecting every 10th point results in a new Nyquist Frequency of 2 kHz. Unfortunately, our filter was set to remove signals only above 5 kHz.

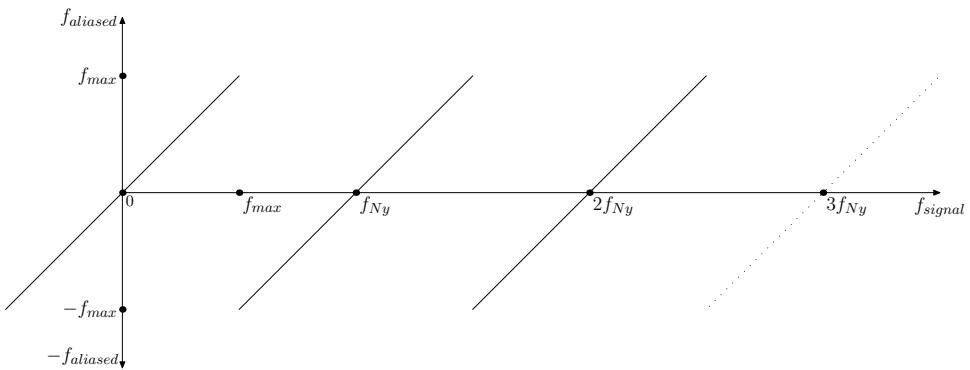


Figure 3.3 Aliased frequencies versus actual signal frequencies. Points equal distances, and opposite sides, of the f_{signal} axis are complex conjugates. The frequency f_{max} is the maximum signal frequency that cannot be confused with a lower signal frequency.

Thus we've allowed captured frequencies between 1 and 5 kHz to corrupt the frequency content of our data between 0 and 1 kHz through aliasing.

Anti-aliasing in modern data acquisition systems is typically implemented in two stages. First, the data is sampled at a very high frequency after passing through an analog anti-aliasing (lowpass) filter with a cut-off frequency of approximately 45% of the sampling rate. This assures that frequency content at 40% of the sampling rate is not degraded significantly in phase or amplitude. It also assures that frequencies above 50% of the sampling rate are sufficiently removed from the data.

In the second stage, a digital lowpass filter is applied with a cutoff frequency appropriate to the sample frequency desired. If the user is interested in data to 1000 Hz, the algorithm will interpret this as 1250 Hz, filtering data above approximately 1125 Hz with a digital filter, and decimating the data to an effective sampling rate of 2500 Hz (the effective Nyquist frequency). Subsequently the data will be processed appropriately using an FFT, and results will be displayed and made available only to a frequency of 1000 Hz, even though the equipment still retains frequency data to 1250 Hz. The reason for this is that the frequency content between 1000 Hz and 1250 Hz is likely to have been damaged by the filter (reduced in amplitude and modified in phase), and is unsuitable for use in representing the system unless the user also wants to observe the filter dynamics of their recording system. While this is often important in identification for control theory, it is less useful in vibration testing. More importantly, the filters involved in the feedback loop should be identified, but those may or may not be the same filters in your data acquisition system. Users must be aware when doing testing for feedback control of what is part of your system, and what is part of your external data acquisition system. For more information on filtering, the reader should consult the following or other text on filtering: Strum and Kirk (1989), Phillips and Nagle (1990), Jackson (1989), Hamming (1983), or Thede (c2005).

3.6 Multiplexing

3.7 Estimation of Correlations Using the Discrete Fourier Transform

A common practice in random signal analysis is zero padding. Zero padding is the appending of zeros to a random signal. Sometimes this is for the purpose of allowing the application of a Fast Fourier Transform algorithm on a signal whose length is not a power of two. The need for this is decreasing as increasing computer speed is making such speed gains less necessary.

Zero padding is also used in the calculation of the linear correlation (see section 2.12) The linear correlation of two data series is the circular correlation of the sufficiently zero-padded data series(Jackson 1989). One of the beneficial side-effects of zero padding is an increase in resolution⁷.

3.8 Quantization Errors

3.9 Accounting for the Impact of Signal Errors: Shifting, Smoothing, and Averaging

It is inevitable that noise will contribute to the recorded signal. The nature of the noise encountered can vary widely. Vankarsen and Allemand (1984) Errors in estimates of the true signal are inevitable. However the degree of understanding of the inherent noise determines how much the effects of the noise can be minimized, as well as improving the understanding of the uncertainty in the estimates of the true signal.

3.9.1 Bias and Drift

Vibration signals are typically zero-mean. In performing vibration analysis, coordinates are typically chosen so that the response has a zero mean value. In actual testing, however, signals often contain a *signal bias*. This is a constant offset error from the true value. In some cases determining this bias is simple. For a velocity or acceleration sensor one can take a very long data sample while the test article is quiet and obtain an average value. If the test article cannot be quieted, the same method can be used, but a longer time sample is warranted in order to improve estimates. It is prudent to estimate the mean values of sub-records in order to obtain confidence in the bias estimate. If possible, it is advisable to invert the transducer and repeat such analysis to determine the effect of gravity on the bias.

It is not atypical for a sensor (or more appropriately sensor systems) to ‘drift’. Sensors can drift in two ways. The first is that the bias can change over time. The second is the calibration factor⁸ can change. When this occurs over a short period of time (measured in minutes), it is usually the results of the user failing to allow the electrical components to warm up sufficiently and reach steady state. When this occurs at a slower rate it is typically environmental in nature. For instance, a sensor on an outdoor object such as a bridge can be expected to be subjected to significant temperature variations. This can cause significant drift in the signal.

Both bias and drift can often be removed from the signal by selecting a ‘AC’ mode. Such a mode is effectively a high pass filter that will try to automate the removal of average values from the signal. This can be especially effective when unpredictable drifting may occur in the signal. Unfortunately, this

⁷needs more understanding of the effects of zero padding

⁸The conversion of the electrical output to physical units.

cannot be used when displacement or strain values are taken without losing the constant values. Often the actual drift of these values over time is the interesting part of the signal. However, in that case, either the actual drifting of the signal must be eliminated by thermally isolating the sensor and electronics, or the dependence of the bias and calibration factors on the temperature must be determined by a separate experiment, and the effects of drifting and bias must be removed after data is collected.

3.9.2 Averaging: Time Domain (Ensemble Averaging)

Averaging in the time domain is rather specific to rotating machinery. Sampling for each sub-record is synchronous with a specific shaft location or other regular position in a rotating machine. Averaging is performed not in time, but across the ensemble

$$\hat{x}(k\Delta T) = \lim_{n_r \rightarrow \infty} \frac{1}{n_r} \sum_{i=1}^{n_r} x_i(k\Delta T) \quad (3.87)$$

and is called *ensemble averaging*. If the noise is assumed to be ergodic and zero-mean, then averaging will help to reduce the effect of noise in the signal, converging to zero as $n_r \rightarrow \infty$. Further, it can reduce the impact of signals whose source is adjacent equipment or shafts.

Consider a sine wave of arbitrary frequency. Presuming that it is asynchronous with the signal of interest, the phase can be assumed to have a uniform distribution. Since such a signal is ergodic with zero mean, continuous averaging will eliminate its effect on the signal of interest.

3.10 Averaging: Frequency Domain

If the assumption is made that the noise is ergodic, then in considering equations (3.23) and (3.24), we can address the terms $G_{\xi_f \xi_f}$ and $G_{\xi_x \xi_x}$ statistically. Ideally, we would like to measure them prior to performing the experiment. Measuring them during and after the experiment is prudent for assuring that the noise is indeed ergodic. If only a single pair of records, $\hat{x}(t)$ and $\hat{f}(t)$, are taken, then there is no method of separating the true parts of the auto spectrum densities from the contributions of noise. However, if the noise is ergodic, and multiple records are obtained, then an average value, as well as other statistical quantities, can be estimated. Thus the spectrum densities are usually estimated as

$$\hat{G}_{xf} = \frac{1}{n_r T} \sum_{i=1}^{n_r} \bar{X}_i \hat{F}_i \quad (3.88)$$

$$\hat{G}_{xx} = \frac{1}{n_r T} \sum_{i=1}^{n_r} \bar{X}_i \hat{X}_i \quad (3.89)$$

and

$$\hat{G}_{ff} = \frac{1}{n_r T} \sum_{i=1}^{n_r} \bar{F}_i \hat{F}_i \quad (3.90)$$

where n_r is the number of records.

As each additional record adds two statistical degrees of freedom, the normalized random error (error relative to the biased estimate) is

$$\epsilon[G_{xx}] = \frac{\sigma_{G_{xx}}}{\hat{G}_{xx}} = \sqrt{\frac{2}{n_r}} \quad (3.91)$$

The $(1 - \alpha)$ confidence interval for G_{xx} is then

$$\left[\frac{2n_r \hat{G}_{xx}}{\chi_{2n_r; \alpha/2}^2} \leq G_{xx} \leq \frac{2n_r \hat{G}_{xx}}{\chi_{2n_r; -\alpha/2}^2} \right] \quad (3.92)$$

with a 95% confidence interval approximated by (Bendat and Piersol 2000)

$$\left[\left(1 - \frac{2}{\sqrt{n_r}} \right) \hat{G}_{xx} \leq G_{xx} \leq \left(1 + \frac{2}{\sqrt{n_r}} \right) \hat{G}_{xx} \right] \quad (3.93)$$

when the normalized random error is small. Since G_{ff} follows the same form as G_{xx} , using equations (3.28) and (3.91),

$$\epsilon[|G_{xf}|] = \sqrt{\frac{2}{\gamma_{xf}^2 n_r}} \quad (3.94)$$

and the variance of the phase⁹ is (Bendat and Piersol 2000)

$$\sigma^2[\theta_{xf}] \approx \frac{1 - \gamma_{xf}^2}{2n_r \gamma_{xf}^2} \quad (3.95)$$

which is also the phase of the $H(j\omega)$.

For the case of no noise in the input signal, the normalized random error of the magnitude of $H(j\omega)$ is (Bendat and Piersol 2000)

$$\epsilon[|\hat{H}_{xf}|] \approx \frac{(1 - \gamma_{xf}^2)^{1/2}}{|\gamma_{xf}| \sqrt{2n_r}} \quad (3.96)$$

which is true for the $H_2(j\omega)$ estimate since noise in the input signal is “washed out.”

Note: for a large number of averages (> 30), since a χ^2 distribution is approximately Gaussian, and presuming small values of ϵ , confidence intervals of a variable f can be estimated using: 68% confidence:

$$[f(1 - \epsilon) \leq f \leq f(1 + \epsilon)] \quad (3.97)$$

and for 95% confidence

$$[f(1 - 2\epsilon) \leq f \leq f(1 + 2\epsilon)] \quad (3.98)$$

3.10.1 The Welch Method for Estimating Auto- and Cross-Spectrum Densities

3.11 Windowing and Leakage

The problem of leakage is one of failure to satisfy the repetition requirement of a Fourier Series (see Section 1.3) in calculation of the DFT. That is, when representing a function with a Fourier Series, the function must repeat indefinitely with a period we have defined as T . If it doesn't, then it can't be represented with a Fourier Series. However, this isn't completely true. If the goal is to represent the function only during a period T , then we can do so at the expense of discontinuities in the extrapolated signal.

Consider two periods of a signal shown in Figure 3.4(a). The circle represent the discrete sampling of the true signal shown with a solid line. Sampling was performed at exactly eight samples per cycle. However, if the sample frequency had been slightly lower, while still taking a total of 16 data points, we would obtain the record of Figure 3.4(b)¹⁰. Figure 3.6(b) illustrates what the extrapolated (assumed continuous) signal of the higher sample rate data looks like. Clearly this is not correct. However, the assumption of the Fourier series, which is built into the DFT, prescribes that the signal must repeat itself. For this simple case we could compensate, with enough care, by recognizing (assuming) that only a single frequency exists in the data, and then accurately obtain amplitude and phase information.

⁹expected value of the error of the phase squared.

¹⁰Most systems don't allow for slight changes in sampling rates and/or number of samples per run. Here it is done for the purpose of illustrating how sampling rate can effect the representation of a signal with a fixed frequency.

However, in a more complex situation where multiple frequencies exist within the signal, this isn't possible.

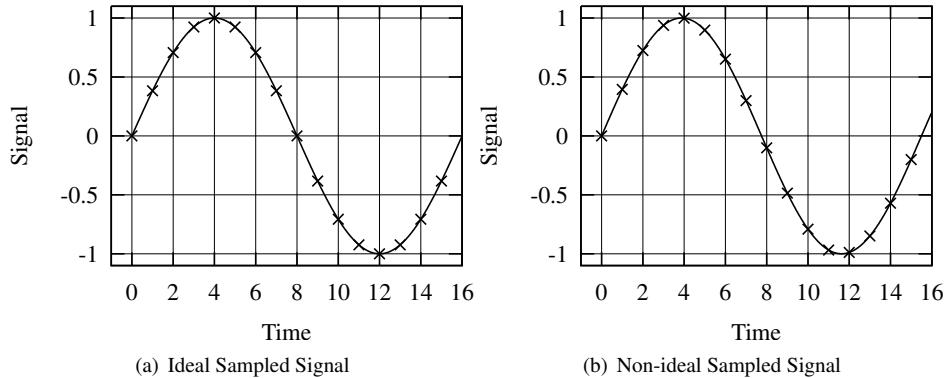


Figure 3.4 Sampled data of signals with different frequencies.

One solution to avoid this discontinuity is to *window* the data. Windowing data means weighting it with a windowing function so that it tapers to zero at the start and end. Figure 3.6(c) show the non-ideal sampled windowed data, as well as its assumed Fourier repetition. The windowed function is then

$$x_w(t) = x(t)w(t) \quad (3.99)$$

where $w(t)$ is the windowing function. For the example of Figure 3.6(c), a *Hanning window*¹¹, as shown in Figure 3.7, is applied. The discontinuity has been minimized (eliminated by using this window). The effects of the inappropriate sampling rate as well as the subsequent application of a Hanning window is illustrated in plots of spectrum densities in Figure 3.6. The solid lines don't actually represent values, but are plotted only to connect associated data points. The ideal case, represented by the circle data points, is actually zero everywhere except at the true signal frequency. The effect of the failure to obtain an appropriate sampling rate is illustrated by the data plotted with asterisks. First of all, because the actual frequency is not a multiple of the total duration of the signal, T , there is not data point at the true signal frequency. This is another illustration of how the combination of non-ideal sampling rate and signal length causes degradation of the information. It is also noticeable that the majority of the non-peak magnitudes have been increased, while the closest non-peak data point magnitude has been decreased. This is leakage. The energy of the signal that should be located at 1 Hz is shifted to other frequencies. While only one frequency exists in the real data, in order to accurately represent the signal over a fractional number of cycles, other frequency components must be included. Application of a Hanning window improves and degrades the auto spectrum density estimate. It has reduced leakage of data to far away frequencies at the cost of increasing leakage to nearby frequencies. This can have a dramatic effect on damping estimates that are performed based upon the sharpness of the peak.

¹¹There is often much confusion regarding the so-called Hanning window. The Hanning window was actually defined first by von Hann. The introduction of the Hamming window, developed by R. W. Hamming (Hamming 1983), seems to have lead to the modern misnaming of the von Hann window as the Hanning window (Otnes and Enochson 1972).

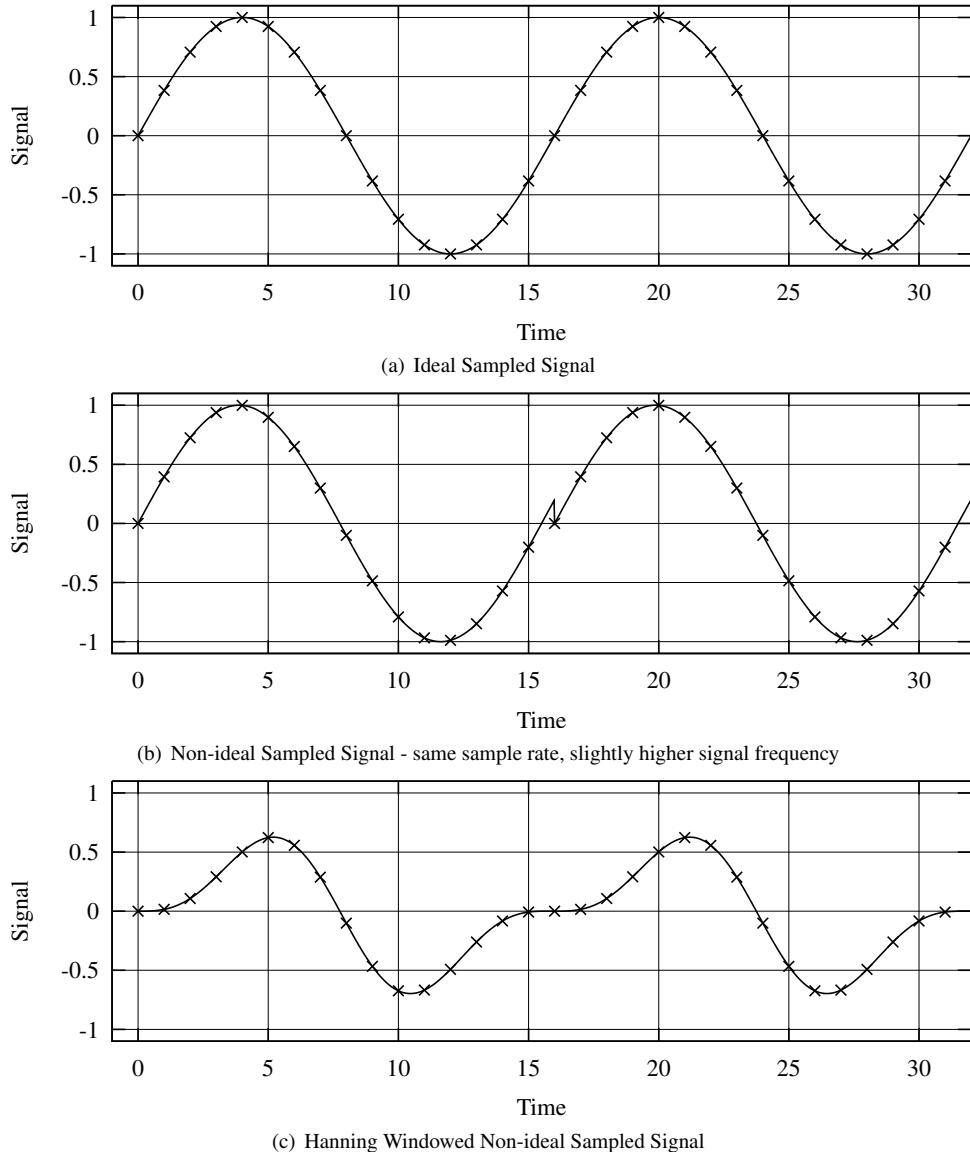


Figure 3.5 Repetition of two sampled cycles. Leakage is caused by the jog in the data between samples 16 and 17 of the non-ideal sampled signal. Windowing is used to taper the ends of the non-ideal sampled signal.

The leakage effect is better understood observing the application of data windows in the frequency domain. Taking the Fourier transform of equation (3.99) and using equation (1.108) gives

$$\mathcal{F}(x_w(t)) = \mathcal{F}(x(t)w(t)) = X_w(f) = \int_{-\infty}^{\infty} X(v)W(f-v)dv \quad (3.100)$$

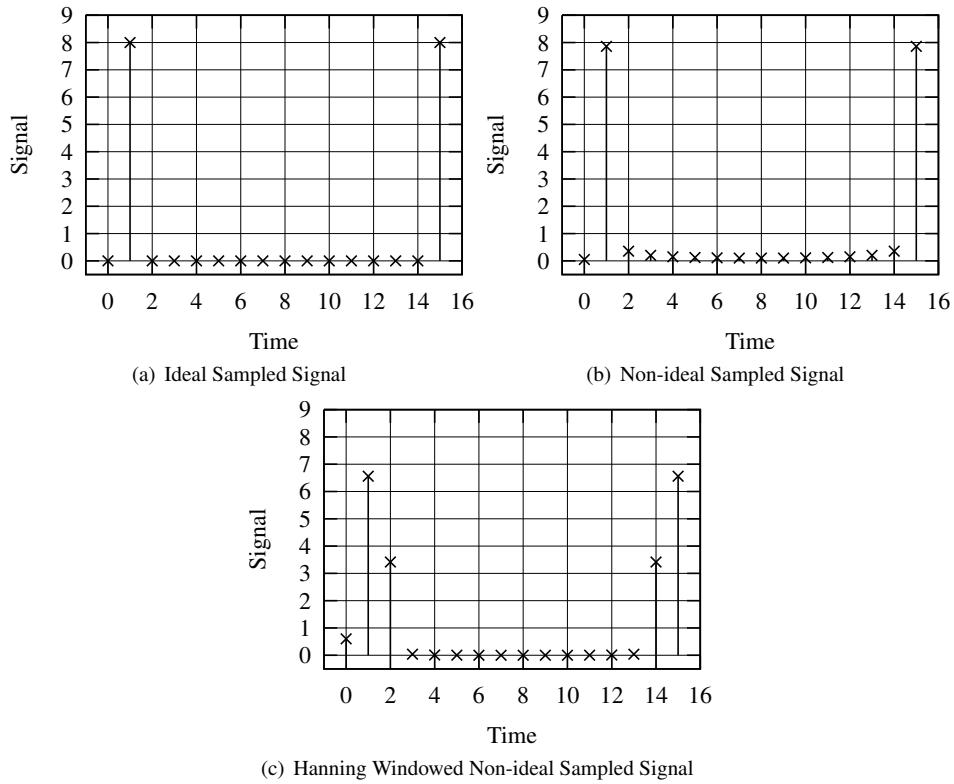


Figure 3.6 Discrete Fourier Transforms of Ideal Sampled Signal, Non-ideal sampled signal, and Windowed Non-ideal Sampled Signal with Hanning window applied. Note that the Hanning window reduces distant leakage at the expense of a blunted peak.

The effect is that the windowed DFT of $x(t)$, $X_w(f)$, at the frequency f , is the complex weighted average of the true, or correct, $X(f)$ over the entire range of f . Consider the value of $X_w(f)$ for our non-ideally sampled signal at $f = .5$ Hz. The frequency step size is $\Delta f = 1/T = .5$ Hz. Observing Figure 3.9, because the ASD of the hanning window is large at a distance of Δf , the value of $X_w(0.5)$ will be strongly influenced by the value of $X(1)$. It will also be strongly influenced by the value of $X(0)$, but since that value is close to zero, its contribution to the error is negligible. Referring to Figure 3.8 now, the value of $X_w(2.8)$ is marginally influenced by the high peak value of $X(1)$ because the frequency difference of $3.6\Delta f$. The auto spectrum density of the windowing function is actually about 45 dB lower than its peak value here.

This high rate of drop off in the magnitude of the *side-lobes* comes at the expense of a wider main lobe. The ideal window wouldn't allow the influence of any frequencies away from the middle frequency, resulting in $X_w(f) = X(f)$. However, this is not possible to achieve. In fact, not choosing to window the data is actually choosing the *boxcar* window. This window doesn't leak much in the immediate vicinity of the signal frequency, but has the negative effect of leaking frequency content much more strongly much further away from the true frequency. The Hamming window is often popular due to its decent far frequency attenuation, while still retaining a narrow main lobe. The small second

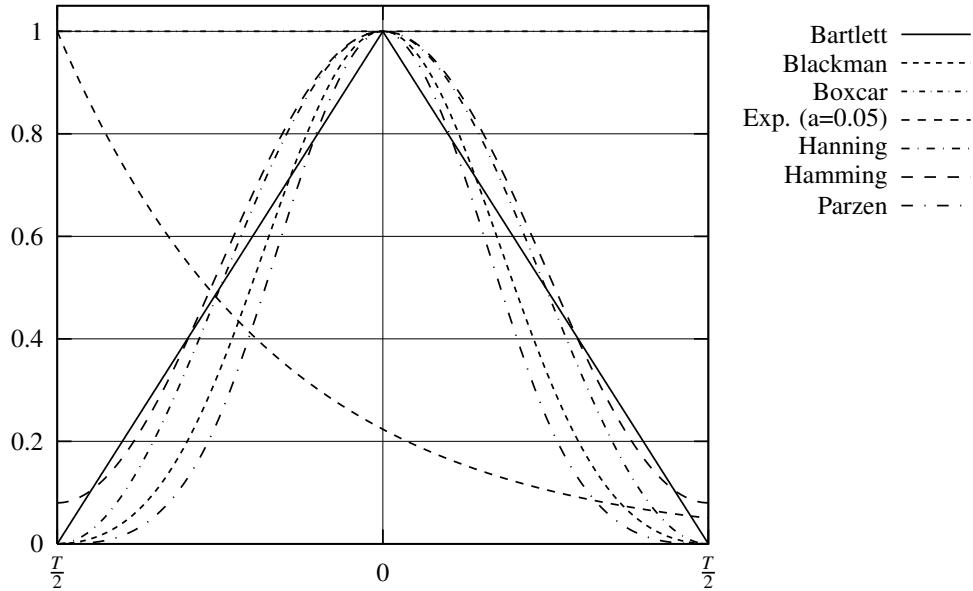


Figure 3.7 A selection of commonly used data windows (Time domain).

side lobe helps attenuation when frequencies are relatively closely spaced.

Use of data windows when estimating spectrum densities requires amplitude compensation because much of the signal energy has been dropped out, as can be observed by the amplitude reductions in Figure 3.7. The fourth column of Table 3.1 contains factors that the estimated spectrum densities should be multiplied by when using the corresponding data window.

Observing Figure 3.6, the leakage has been greatly reduced by the application of the Hanning window. Similar results can be obtained by applying other windows.

3.12 Time-Frequency Analysis

Hilber Huang (SDM51, Palazotto),

3.13 Problems

1. Prove that the eigenvalue, λ obtained in obtaining the H_v FRF estimate is $\leq G_{xx}(1 - \gamma^2)$.
2. Prove $x_{RMS}^2 = \sigma_x^2 = \int_{-\infty}^{\infty} G_{xx}(f)df$ where X_{RMS} is the RMS (root mean square) value of $x(t)$.¹²

¹²This is a horrible way to perform this calculation in practice. Real Fourier Analyzers have anti-aliasing filters that remove all high frequency analog content from the signal. Thus high frequency responses that may contribute significantly are lost. Secondly, because of Nyquist effects, the top 20% of the ASD is typically not reported due to real distortions of the anti-aliasing filters. The only instance where a reasonable answer might be expected is when using a displacement or velocity sensor (not accelerometer) with the anti-aliasing filters turned off. Regardless, calculation of the variance using equation (2.2) is much more accurate.

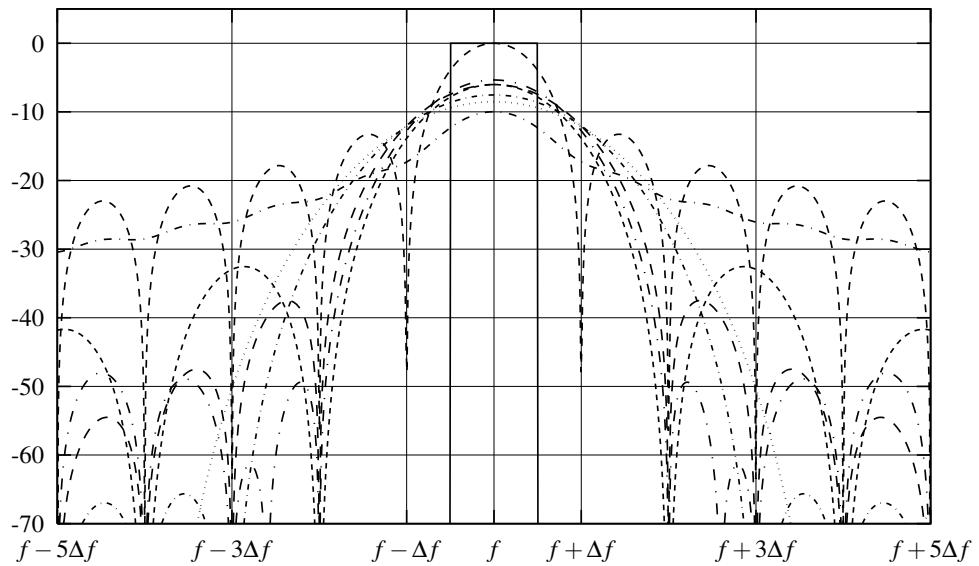


Figure 3.8 Auto-spectrum densities of commonly used data windows (Wide frequency range).

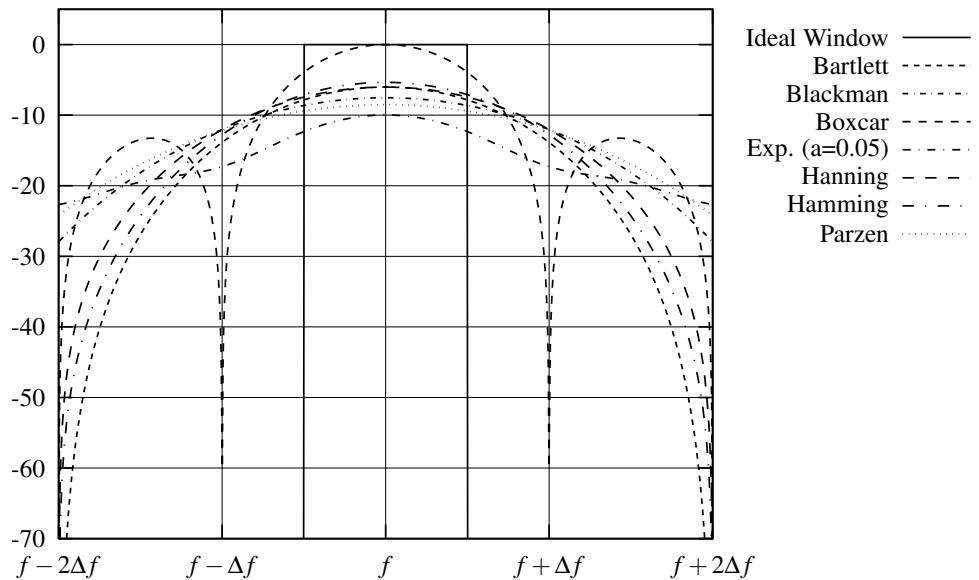


Figure 3.9 Auto-spectrum densities of commonly used data windows (Small frequency range).

^aThe effect of applying a window to a signal prior to calculating spectrum densities can be compensated for by multiplying the resulting spectrum density by the SD correction factor.

^bThe value a for an exponential is the percent of the envelope remaining at the end of the time window. It is not advisable to expect a valid ASD estimate even after correction when using an exponential window.

^c I_0 is the modified Bessel function. For specific information on this window, please see Hamming (Hamming 1983).

Table 3.1 Window functions and their auto-spectrum densities.

Name	Window, $-T/2 < t < T/2$	Window ASD	SD cor. ^a
Bartlett	$1 - \frac{2 t }{T}$	$\frac{(\cos(\pi fT) - 1)^2}{f^4 \pi^4 T^3}$	3
Blackman	$\frac{21}{50} + \frac{1}{2} \cos(\frac{2\pi t}{T}) + \frac{2}{25} \cos(\frac{4\pi t}{T})$	$\frac{9(28 - 3f^2 T^2)^2 \sin(f\pi T)^2}{2500 f^2 \pi^2 T (4 - 5f^2 T^2 + f^4 T^4)^2}$	$\frac{5000}{1523}$
Boxcar	1	$\frac{\sin(\pi fT)^2}{f^2 \pi^2 T}$	1
Exponential ^b	$a^{\frac{t}{T} + \frac{1}{2}}, \quad 0 \leq a < 1$	$\frac{T(1 + a^2 - 2a \cos(2f\pi T))}{4f^2 \pi^2 T^2 + \log(a)^2}$	$\frac{2T \log(a)}{-T + a^2 T}$
Hamming	$\frac{27}{50} + \frac{23}{50} \cos(\frac{2\pi t}{T})$	$\frac{(27 - 4f^2 T^2)^2 \sin(f\pi T)^2}{2500 f^2 \pi^2 T (-1 + f^2 T^2)^2}$	$\frac{5000}{1987}$
Hanning	$\frac{1}{2} (1 + \cos(\frac{2\pi t}{T}))$	$\frac{\sin(f\pi T)^2}{4f^2 \pi^2 T (-1 + f^2 T^2)^2}$	$\frac{8}{3}$
Kaiser-Bessel ^c	$\frac{I_0\left(b\sqrt{1 - \frac{4t^2}{T^2}}\right)}{I_0(b\pi)}$	$\frac{\left \int_{\frac{-T}{2}}^{\frac{T}{2}} e^{2j f \pi t} I_0(b\pi \sqrt{1 - \frac{4t^2}{T^2}}) dt\right ^2}{T I_0(b\pi)^2}$	$\frac{I_0(b\pi)^2}{\int_{\frac{-1}{2}}^{\frac{1}{2}} I_0(b\pi \sqrt{1 - 4t^2})^2 dt}$
Parzen	$1 - \frac{24t^2(T - 2 t)}{T^3}, \quad t < \frac{T}{4}$ $2\left(1 - \frac{2 t }{T}\right)^3, \quad \frac{T}{4} < t < \frac{T}{2}$	$\frac{9216 \sin\left(\frac{f\pi T}{4}\right)^8}{f^8 \pi^8 T^7}$	$\frac{560}{151}$

3. Prove $G_{xx}(j\omega) \geq 0$ for $-\infty < x < \infty$.
4. Show that the DFT and Inverse DFT are conservative in that performing both operations on a signal returns the original signal.
5. Generate a harmonic signal with frequency ω . Add a time varying exponentially decaying bias with a time constant of $\frac{0.1}{\omega}$. Using no knowledge of the involved parameters (only the signal), estimate the time varying bias function, and correct the errant signal.
6. A pure sinusoid is discretely sampled by your DAQ system. However, the period of the sample and sampling rate is such that the correct Fourier transform of the signal lies exactly between two adjacent discrete frequencies as shown in Figure 3.6. Determine what your actual ASD will look like if you use a) a boxcar window, b) a Hanning window.
7. For the previous problem, list methods that can be used to improve the leakage, presuming a) you cannot guarantee that the true frequency will lie on a discrete frequency, and b) you cannot change the window selection.
8. At a particular frequency, $S_{xx} = 1$, $S_{ff} = 2$, and $S_{fx} = 0.95459 - 0.94549j$. Calculate H_1 , H_2 , H_v , and γ_{xf} .

4

Equipment in Vibration Testing

4.1 Data Acquisition System Settings

Modern vibration testing is generally performed using digital data acquisition systems. These systems filter a voltage coming from a sensor to prevent aliasing, then digitize the resulting signal. Basic settings on the data acquisition system are:

1. *Bandwidth (or sampling speed).* The bandwidth will be approximately $\frac{1}{2.5}$ of the sampling speed, usually given in Hz. A higher sampling speed allows the user to observe the sensor signal in more detail, however, generally for a shorter period of time, since signal length is restricted by memory or storage space.
2. *Sample length.* For vibration testing systems, for the convenience of performing the Fast Fourier Transform, the number of samples is generally restricted to powers of 2 (i.e. 256, 512, 1024...). An upper limit is often in place due to practical concerns such as memory available, storage space available, in conjunction with knowledge of how many other channels of data may be expected. As computer speed and storage space become more available, these limitations are being relaxed.
3. *Triggering.* Triggering allows the data acquisition system to start taking data under pre-defined conditions. In vibration testing, this is generally the onset of a large voltage indicating the start of an excitation signal. Depending on the specific experiment, triggering can be performed using either the response or excitation channels. For free response testing, only the response signal is available, so it is generally used.
 - (a) *Voltage level.* A voltage level must be set under which data collection is to commence. This value may be positive or negative.
 - (b) *Positive or negative slope.* Some systems can differentiate between an increasing or decreasing value, allowing more sophisticated settings.
 - (c) *Filtering.* Some systems can pre-filter the data so that triggering will only occur for certain frequencies of the response.
 - (d) *Free Run.* Free run, or oscilloscope, mode allows continuous data collection and display. Only part of the data may be available for collection, but this mode has the advantage of allowing the user to more readily check sensors for signal levels and appropriate responses (for a rough check of performance).

4.2 Actuators

4.2.1 Impulse Hammers

A common form of excitation used for vibration testing is the impulse hammer. This is due to its relatively easy setup and use. Unlike other excitation sources the hammer does not need to be attached to the test item, and as a result can be easily be moved from one point to the next. This advantage makes the impulse hammer very popular for modal analysis, since most often one sensor may be placed in a single location, or very few fixed locations, on the test article, while the hammer may *roam* about the structure, obtaining input-output relationships between the sensor and hammer locations. In addition, the impulse hammer can be useful in assessing the quality of the location chosen for the excitation load, even in cases where the final excitation source may be from another actuator. For example, one would not want to excite a specimen at a node of an important mode. Considering the principle of reciprocity, one can also use the impulse hammer to assess the quality of sensor locations for certain types of sensors... e.g. those that sense linear displacement, velocity, or acceleration. A less commonly considered advantage of using a roaming hammer for testing is that generally fewer sensors need to be mounted on the test specimen. This advantage is minimized, of course, when non-contact sensors are used. The minimization of the cabling corresponding to the motion sensors also reduces parasitic damping, stiffness, and mass observed in the test. Often parasitic effects are reduced by having a fixed actuator, but a roaming sensor (or corresponding sets of each). The problem with roaming sensors occurs when they contribute parasitically. In moving the sensor, the parasitic effect moves with it, creating system inconsistencies from one set of input-output data to the next. These inconsistencies can be considerable, and can, especially in lightly damped systems, confound system identification techniques. In any case, it is important to understand these effects, attempt to mitigate them, and attempt to compensate for them when they are significant.

address meaning of parasitic

Hammer Fundamentals

The manner of use of an impulse hammer should not be confused with that of a typical hammer from your hardware store. They are very similar, however few individuals outside of carpenters can competently swing a nail hammer. Both hammers operate to take advantage of the principle of impulse and momentum. It is imperative for both devices that the user not attempt to apply a force to the object, but instead focus on letting the momentum of the hammer head perform the task. Consider, in either case, the hammer head to be a rock, which you are going to use for pounding, by bouncing it off the nail/test article. The easiest way to control the rock is by attaching a hammer to it. However, let the head bounce off the nail/test article. In the case of a nail hammer, this will reduce the strain and pain felt in the forearm and elbow. In the case of an impulse hammer, this will result in cleaner, more correct hits. Some advantage can be gained by not gripping the hammer too hard, and in some cases, even allowing it to rock in your hand, using only a couple of fingers to hold it. Don't get carried away, though, or you are back to the rock.

If the hit is applied properly, a single bump with short duration should be observed in the signal from the load cell built into the head of the hammer (see figure 4.1). If two or more bumps are observed, this is called a double-hit (or multiple hit), and data collected from such a strike should be discarded (see figure ??). This is because such a hit tends to over excite a narrow range of frequencies, while neglecting to excite others. It is also very hard to repeat, so any theoretical advantage would not be observed in practice. Most vibration test systems designed for use with hammers can automatically detect and discard this data without intervention from the experimenter. Double-hits are usually the result of either not allowing the hammer to rebound back from the structure after the impact or allowing the structure to rebound into the hammer after the initial contact. This can be especially problematic in the more compliant sections of a structure where the rebound force is insufficient to redirect the

hammer before the structure rebounds. The best solution in these cases is to reduce the hammer head mass so that the force necessary to make the hammer head rebound is reduced. Hammers come in a variety of sizes to facilitate this process. Another common feature found in the response is the negative

Good hammer force here plus system response

Figure 4.1 Good impulse hammer force measurement plus system response

force. Assuming that the contact surfaces do not have attractive forces, this is physically impossible. However, this observation regularly occurs due to a) mass effects of the hammer head and b) anti-aliasing. The load applied by the hammer is approximately measured by the hammer head. A large part of that approximation is the assumption that the load measured is that between the hammer and the surface of the test specimen. However, careful observation of any hammer head shows that the load cell does not strike the object. The hammer head tip does, and its mass causes a negative force to be measured if/when the head is pulled back from the structure. It is, in fact, impossible to remove all of this mass, since some casing mass is inevitable covering the sensing part of the load cell. As a result, it is important when using a hammer to attempt to resist the instinct to pull it away from the structure, and to instead let it rebound with as little resistance as possible. Acceleration of the hammer tip can be measured even when no contact is being made. Fortunately this is usually much lower in magnitude than the force applied to the test specimen, and it is also generally very low in frequency, hopefully putting it away from structural frequencies of interest.

double strike hammer force here. plus bad system response

Figure 4.2 Double strike hammer force measurement

Anti-aliasing filters will cause a brief negative force measurement to be observed immediately following the hammer hit (see figure 4.1). This is not typically a problem for frequency domain work, but can be an issue when precise time domain data is desired. The observed negative force is a result of the anti-aliasing filter not allowing high frequency signals through. The sharp corner at the trailing edge thus ends up showing the Gibbs effect. The leading corner does also but the large impulse tends to obscure this observation from the eye.

A selection of tips is generally provided to assist in controlling the duration of the hit. A softer head is used to provide a longer duration force to the test specimen, which results in a lower bandwidth of the excitation signal. Harder heads do the opposite. The bandwidth can be observed by viewing the auto spectral density (ASD), often referred to as the power spectral density (PSD). The ASD should be approximately constant over the frequency range of interest. Results at frequencies above the point at which the ASD drops below 50% of its maximum are highly questionable. Use of a sufficiently soft head can minimize the anti-aliasing filter's effect on the signal by precluding high frequency content in the excitation.

Setting Up For Hammer Use

When preparing to use a hammer, careful experimentation requires calibration before the experiment, to know the performance characteristics of the hammer, and calibration after the experiment, in order to verify that the hammer performance has not varied during the duration of the test. However, most

experimentation is generally performed with less rigor, and calibration is performed only occasionally. However, a few simple tests can often root out failure of the device, while also helping to set up your data acquisition system. These are performed using either an oscilloscope, or your data acquisition system in a *free run* mode, with data being taken continuously.

The first test is simply a force test to make sure that the sensor is taking some data, the cables are transmitting the data, and the amplifier is performing its task. Push on the head of the hammer in a comfortable fashion so that the load cell should be measuring a load. Nothing specific, simply a load that is within the range of the load cell. For a mid size hammer, one can generally hold the handle in your fingers, turn the hammer head tip to face you, and use the thumb of the same hand to apply a load. Adjust your data acquisition system, time and voltage ranges, to observe the load. See that the signal corresponds approximately to your randomly applied load.

Next, use the hammer as described in the preceding paragraph to strike your test article in a variety of locations, a variety of times (focus on one location at first). At this point, the auto spectral density should be observed to assure that it is relatively flat over the frequency range of interest and the appropriate hammer, and hammer head tip, should be chosen. Changing the hammer head after setting the triggering options can lead to difficulty in triggering the system during your test. It is often easiest to start the process by impacting a hard object first, then moving to the test specimen once approximate settings have been made. The maximum expected voltage can be observed during these strikes and your data acquisition system adjusted accordingly. This process is often automated in systems, sometimes successfully. Consult your manual for setting voltage levels automatically. The maximum level that your system can sense should be set to approximately 50%-100% above what you generally observed. You may need to adjust this as you perform the experiment. Once this level is set, turn on triggering, and set the percentage level necessary to trigger the system. The trigger tells the system that a hit has just occurred, and that it should take and record data on all appropriate channels. In fact, by the time the system recognizes that a hit has occurred, part of the impulse has already taken place. To compensate for this, the system constantly is taking data, and it will include data prior to the trigger in the data that it stores. This is known as the *pre-trigger* data. Generally a setting of 10% or so of the data being pre-trigger data yields the best results. Triggering can also be performed using the signal from your motion sensor. This can prove to be very useful when sampling rates are low relative to the duration of the hit, or when hammer hit force levels vary greatly during a test.

Hammer Usage

Stop the test article from vibrating. Any measured motion is presumed to be caused by the excitation. Residual motion from previous hits will cause the response to appear non-causal (a response is non-causal if it occurs before the excitation). Then hit it cleanly with the hammer as described in section 4.2.1. Reduce motion of the hammer during the test. The mass of the hammer head will cause a force proportional to its acceleration to register in the hammer's load cell. As this force is not actually acting on the structure, its measurement will cause slightly erroneous results. Figure 4.1 shows that even after the impulse, some force is still being measured through the load cell. Some vibration test systems have the ability to zero (ignore) any force measurement coming from the hammer after a specified point in the data collection process. It is prudent to use this option to prevent spurious motions of the hammer from registering in the data. Use an exponential or boxcar type window. Hanning, Hamming, and other similar windows are not appropriate for hammer hits. Ideally the amplitude of the response signal, usually a decaying exponential envelope, should be close to the ambient noise level at the end of the data window. An exponential window can be used to expedite this decay, but at the expense of accuracy in damping measurements (Mcconnell 1995).

Examples

The following experiments were performed on a cantilever beam using typical equipment.

4.2.2 Electrodynanic Shakers**4.2.3 Piezoceramic patches****4.2.4 Acoustic****4.2.5 Hydraulic****4.3 Displacement Sensors****4.3.1 Laser Vibrometers****4.3.2 Potentiometers****4.3.3 Linear Variable Differential Transducers (LVDTs)****4.3.4 Capacitive Sensors****4.3.5 Optical Sensors****4.3.6 Videogrammetry****4.4 Velocity Sensors****4.4.1 Doppler Lasers****4.4.2 Tachometers****4.5 Accelerometers****4.5.1 Mounting Accelerometers**

Accelerometers can be fastened in a variety of ways depending on testing as well as practical concerns. Ideally, the sensor would be rigidly attached to the structure. In practice, some compliance always exists in the connection. In order to obtain valid test data, the interaction of the mounting and the accelerometer must be understood. Manufacturers generally provide customers with calibration data pertaining to each attachment method appropriate for the given sensor as well as directions for mounting accelerometers. The accelerometer will generally perform equal-to or worse-than the calibration due to the greater care generally taken by the manufacturer in obtaining the calibration.

Mounting is most often performed using one of three configurations: adhesive (temporary or permanent), magnetic, or stud. For adhesive mounting, a thin layer of glue or Petro Wax ($< 0.004"$) is used to connect the accelerometer to the structure. Wax is often used due to its convenience, as it is easy to attach and remove the accelerometer without tools. The drawback is a lack of reliability, especially at higher acceleration levels ($> 20 \text{ g's}$), due to the weakness of the bonds. As with any bond, both surfaces must be relatively clean so that the wax will stick to both the accelerometer and the structure. Enough wax must be applied to completely fill the gap between the accelerometer and the test article. It is better

to apply too much than too little. The accelerometer is then applied with firm pressure (no tools) and an alternating twisting motion press out excess wax. It is important that the cable already be attached, since the twisting of connecting a cable can easily break the bond of the wax. Cyanoacrylate (super glue) adhesive can also often be used for temporary bonds, but generally are much more sensitive to the mating of the surfaces being glued. Accelerometers mounted with either wax or cyanoacrylate glue can easily be removed with a twist (some accelerometers require the use of a wrench, often manufacturer supplied, when using glue). A more permanent bonding can be obtained with epoxies. Stronger bonds are often necessary at higher frequencies and amplitudes, but cyanoacrylates can be susceptible to debonding if significant strain occurs in the structure at the point of application.

Magnetic bonding is the fastest mounting configuration, and as a result, often the least reliable. It is, of course, limited to only magnetic or ferrous materials. A near-perfect mate is required between the accelerometer and structure surface. Due to the weakness of the bond and the lack of any filler medium such as glue or wax, magnetic bonds often result in poor connections due rocking or bouncing on the highest points of the mating surface. A filler such as a grease may be used to improve coupling. Magnetic bonding is most often used in oily/greasy environments where temporary adhesive bonds aren't feasible.

A stud mounting is the most reliable, but also the most invasive and difficult mounting to apply. In stud mounting, a hole is drilled and tapped in the structure, a stud is screwed into the hole, and the accelerometer is screwed onto the stud. With the proper torque, the attachment quality approaches that of a permanent bond but can be easily removed. The drawback is that if the structure is too thin in the desired mounting location there may not be enough threads available to firmly attach the stud. In addition, an additional hole has been introduced to the test article with all of the stress concentration, fatigue, and appearance ramifications. The studs can be left in place in prepared well-defined positions, preventing measurement errors due to variation of sensor placement common to magnetic and adhesive bonds.

Hybrid attachments are also available where a mounting pad (stud on one side, and flat on the other) is used between the structure and the sensor. By screwing the stud side into the structure, a very flat surface is made available for temporary adhesive bonding, or in some cases, magnetic bonding. Permanently bonding a magnet to a surface allows future use of magnetic bonding in locations where they couldn't otherwise be used.

4.6 Force Sensors

4.6.1 Strain Gages

Foil Strain Gages

Piezoceramic Strain Sensors

4.6.2 Load Cells

4.6.3 PVDF washers

4.7 Some Notes On Cables

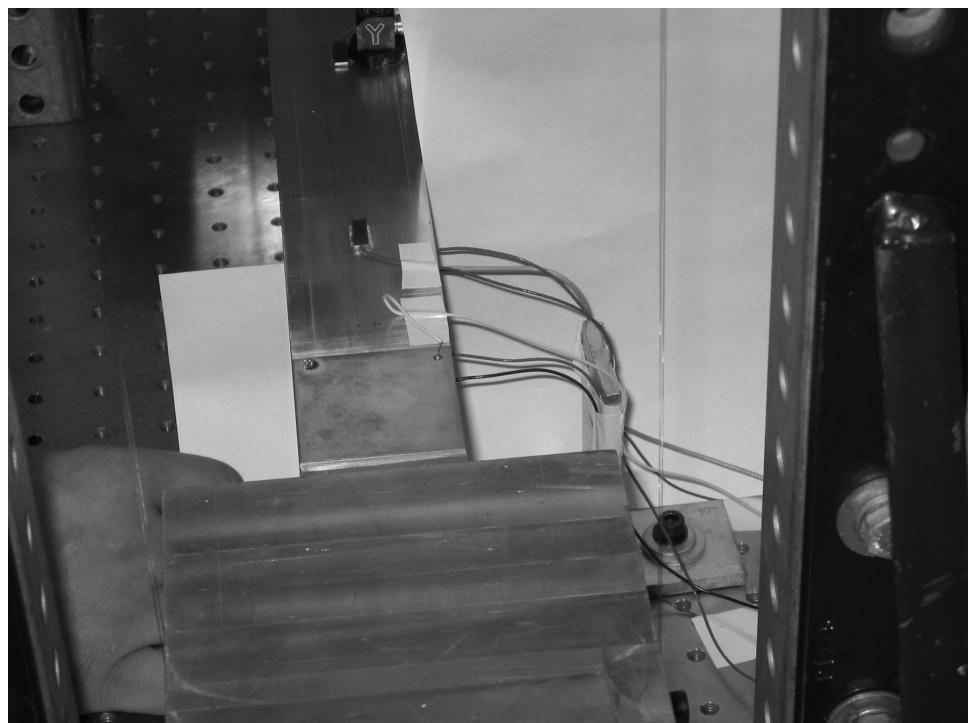


Figure 4.3 Strain relieved cables.

5

Basic Vibration Testing: Survey Tests

The quality of information required from the test drives the fidelity with which the test is performed and often the attention to detail required of the performer. For example, calibration of sensors, or even use of calibration factors, is often unnecessary with the need is simply the acquisition of a single natural frequency of a relatively well understood system. However, even when high fidelity results are desired, it is prudent to perform rough “survey tests” in order to obtain some initial insight into what information more sophisticated methods might glean, what the requirements for your experimental setup will be, and what settings to use when taking data. The purpose of these tests is to obtain rough estimates of such parameters as how many modes exist, what are the approximate natural frequencies (and how many), what parts of the structure have interesting dynamic features, and what is the level damping. This information assists in making decisions regarding the selection of actuation devices, locations of force application, types and locations of sensors to be used, and the algorithm to be used to perform system identification. Sometimes what is referred to here as a survey test was actually intended by the experimenter to be the final comprehensive test, but after considering the issues here, further testing is required. Testing results often inform the tester that further testing is necessary, and this must be considered when scheduling tests, especially on systems with which the experimenter has a relatively poor initial understanding. What constitutes a good understanding *a priori* is defined by understanding the issues addressed in the following sections.

In this chapter, classical methods for parameter excitation are covered including descriptions of their shortcomings. Such methods are ill suited for systems with closely spaced modes, and are often inadequate to address whether or not repeated frequencies exist in the structure. Robustness of the excitation and sensing methods for obtaining the desired information is also addressed. The results of such tests either reveal the model information desired, or direct the investigator to enhanced testing methods perhaps requiring more equipment, greater fidelity, or a re-examination of the assumption of the tests. Often survey tests are performed with impulse hammers due to their convenience. However, as discussed in section 4.2.1, use of impulse hammers requires a great deal of diligence.

5.1 Analytical Models for Use in Experiment Design

The first basic understanding of a system to be tested often comes from a model of some sort, typically a finite element model. However, in many cases, either that model is not performing sufficiently well, or no such model exists. Regardless of the circumstances, an ability to make some reasonable estimates of mode shapes and natural frequencies are invaluable for guiding the initial test, and in some cases may actually preclude the test from being performed.

Some of the resistance to application of basic analytical models is, I believe, derived from the misconception that they are meaningless in modern situations where finite elements have become ubiquitous in practice. However, observation of behavior of finite element models

5.1.1 Beams

5.1.2 Plates

5.1.3 Rods

5.1.4 Rigid Masses

5.1.5 Back of the Envelope Estimates

5.1.6 Finite Element Models

5.2 Location, Location, Location: Sensor Placement and Where to Hit the Structure

5.3 Obtaining Parameter Estimates from Frequency Response Functions

5.3.1 Noise

5.3.2 Log Decrement

5.3.3 Natural Frequency Estimation

5.3.4 Half-Power Point Damping Estimation

The half-power method estimates the loss factor or damping ratio at a natural frequency by observing the sharpness of the peak. The bluntness, defined at $\frac{f_2 - f_1}{2f_n} = \frac{\omega_2 - \omega_1}{2\omega_n} = \frac{r_2 - r_1}{2}$ (see Figure 5.1), is approximately proportional to the damping in the mode. Derivation of the estimating equation is based upon the following assumptions:

1. The damping mechanism is either viscous or linear hysteretic (see equation (1.131)).
2. The damping is light (corrections can be made using the higher order equations).
3. The system is single degree of freedom.
4. The actual peak value can be observed in the FRF.
5. Appropriate selection of Data Window

As with any set of assumptions, for a real system, their validity is always relative to the degree of desired accuracy. Even when they are sufficiently satisfied, a lack of sufficient care in reading values from a plot of the FRF can result in greatly erroneous results.

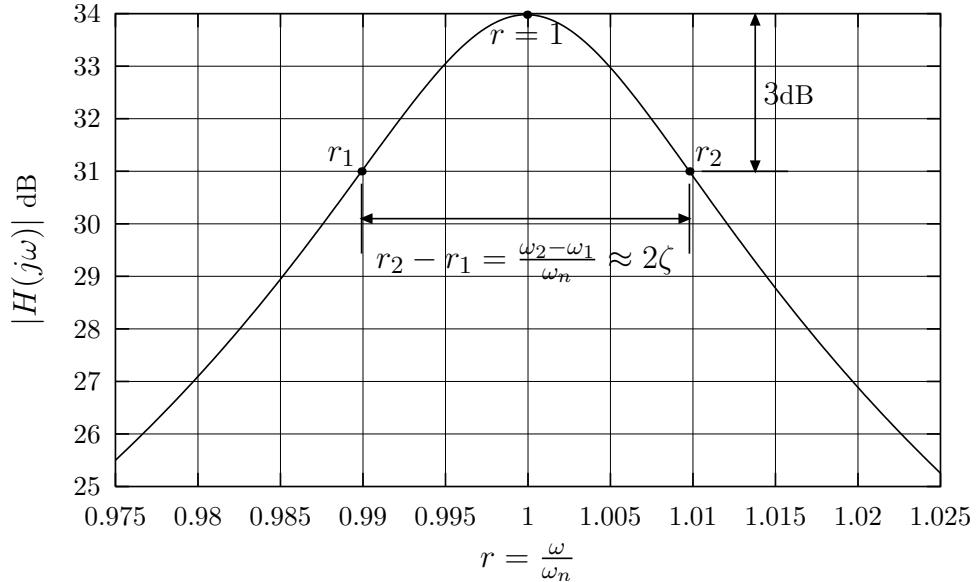


Figure 5.1 Half-power points on frequency normalized FRF.

Derivation of Half-Power Equations, Model Assumptions

For a viscously damped single degree of freedom system, the FRF is given by

$$H(j\omega) = \frac{1}{k - m\omega^2 + cj\omega} \quad (5.1)$$

The magnitude is given by

$$|H(j\omega)| = \frac{1}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}} \quad (5.2)$$

The value of ω at the peak amplitude is given by

$$\omega_p = \omega_n \sqrt{1 - 2\zeta^2} = \sqrt{\frac{k}{m}} \sqrt{1 - 2\zeta^2} \quad (5.3)$$

and by substitution, the peak amplitude is

$$|H(j\omega_p)| = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \quad (5.4)$$

Likewise, the half-power points, $|H(j\omega)| = \frac{1}{\sqrt{2}} |H(j\omega_p)|$ or $|H(j\omega)| \text{ dB} = |H(j\omega_p)| \text{ dB} - 3\text{dB}$, are at

$$\omega_1 = \omega_n \sqrt{1 - 2\zeta^2 - 2\sqrt{\zeta^2 - \zeta^4}} \quad (5.5)$$

and

$$\omega_2 = \omega_n \sqrt{1 - 2\zeta^2 + 2\sqrt{\zeta^2 - \zeta^4}} \quad (5.6)$$

Applying the binomial expansion theorem, or alternatively expanding in a Taylor series about $\zeta = 0$, to equations (5.5) and (5.6), the difference is noted to be approximately

$$\omega_2 - \omega_1 \approx \omega_n 2\zeta \quad (5.7)$$

Thus, dividing by $2\omega_n$ would give us our estimate of the damping ratio. However, in reading values from an FRF magnitude plot, we can only read ω_p . Dividing by $\omega_n \approx \omega_p$ for $\zeta \approx 0$ gives

$$\zeta_{est} = \frac{\omega_2 - \omega_1}{2\omega_p} = \frac{f_2 - f_1}{2f_p} \approx \zeta \quad (5.8)$$

The repeated approximations assuming $\zeta \approx 0$ should lead one to be concerned about the amount of error introduced by this assumption. Figures 5.2 and 5.3 illustrates the degree to which the damping is overestimated for a range of estimated values ζ_{est} . The maximum value of ζ for which half-power points can be found is $\zeta_{max} = \sqrt{\frac{1}{2} - \frac{1}{2\sqrt{2}}} \approx .38\dots$. At this value of ζ , the estimated value is actually $\zeta_{est} = \frac{1}{\sqrt{2}}$, which is approximately 85% high, relative to the true value. The relative errors as shown in Figures 5.2 and 5.3 are normalized by the ζ_{est} for greater utility.

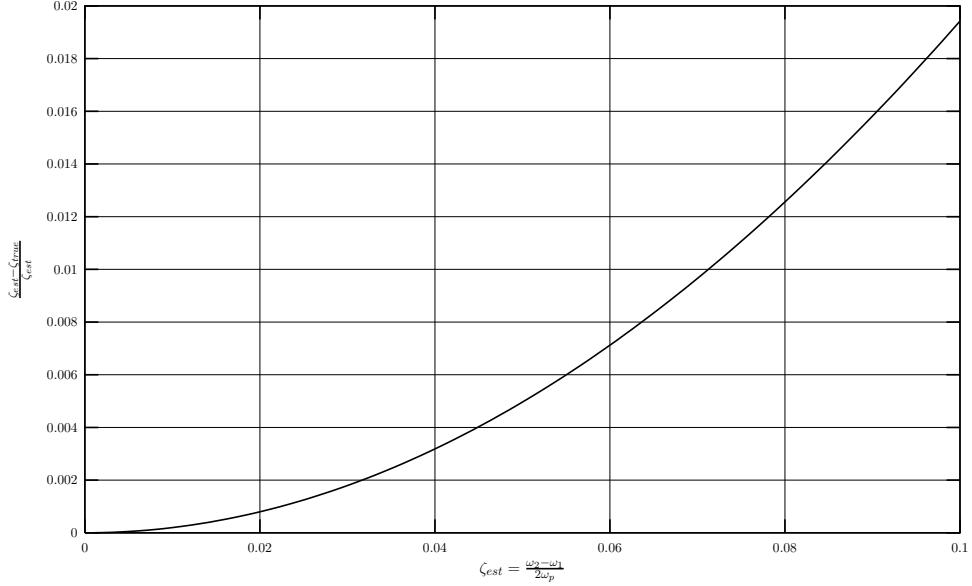


Figure 5.2 Error in estimating ζ using equation (5.8) for low damping.

For a linear-hysteretically damped system, the FRF is given by

$$H(j\omega) = \frac{1}{k - m\omega^2 + \eta j} \quad (5.9)$$

The magnitude is given by

$$|H(j\omega)| = \frac{1}{\sqrt{(k - m\omega^2)^2 + (\eta)^2}} \quad (5.10)$$

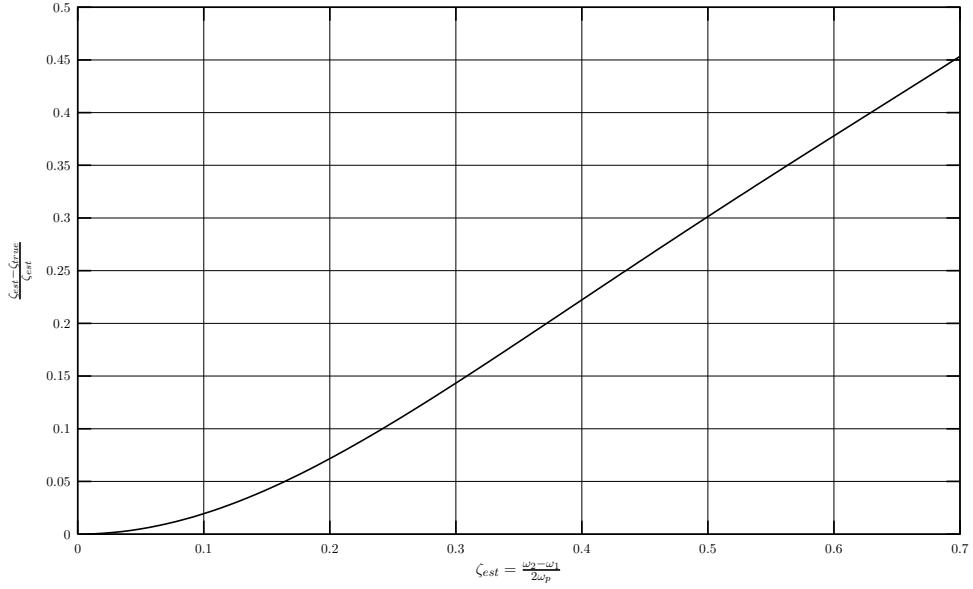


Figure 5.3 Error in estimating ζ using equation (5.8) for high damping.

Unlike the viscously damped case, $\omega_p = \omega_n$. By substitution, the peak amplitude is

$$|H(j\omega_p)| = \frac{1}{\eta} \quad (5.11)$$

Likewise, the half-power points, $|H(j\omega)| = \frac{1}{\sqrt{2}} |H(j\omega_p)|$, are at

$$\omega_1 = \omega_n \sqrt{1 - \eta} \quad (5.12)$$

and

$$\omega_2 = \omega_n \sqrt{1 + \eta} \quad (5.13)$$

These expressions have a much simpler form than equations (5.5) and (5.6), and thus from observation

$$\eta = \frac{\omega_2^2 - \omega_1^2}{2\omega_n^2} = \frac{\omega_2 - \omega_1}{\omega_n} \frac{\omega_2 + \omega_1}{2\omega_n} \quad (5.14)$$

The expression

$$\frac{\omega_2 + \omega_1}{2\omega_n} = \frac{\sqrt{1 + \eta} + \sqrt{1 - \eta}}{2} \approx 1 - \frac{\eta^2}{8} \approx 1 \quad (5.15)$$

for small values of η , so a good approximation for the loss factor, η for low damping is given by

$$\eta_{est} = \frac{\omega_2 - \omega_1}{\omega_n} = \frac{f_2 - f_1}{f_n} \approx \eta \quad (5.16)$$

which is effectively the same as equation (5.8) noting that $\eta \approx 2\zeta$ for light damping. As for the case of viscous damping, this expression over-predicts the amount of damping, especially for high levels. Figures 5.4 and 5.5 show the relative error (here also normalized by the estimated damping). It's readily apparent that the estimate for η using the half-power method is not nearly as bad as that for estimating

ζ . However, for light damping ($\zeta < .1$ or $\eta < .2$), the mathematically induced error in using these expressions is less than 2%. Thus, for their intended purpose, estimates derived from equations (5.16) and (5.8) are quite reasonable. However, when greater levels of damping exist, the figures of this section can be used to adjust the estimates.

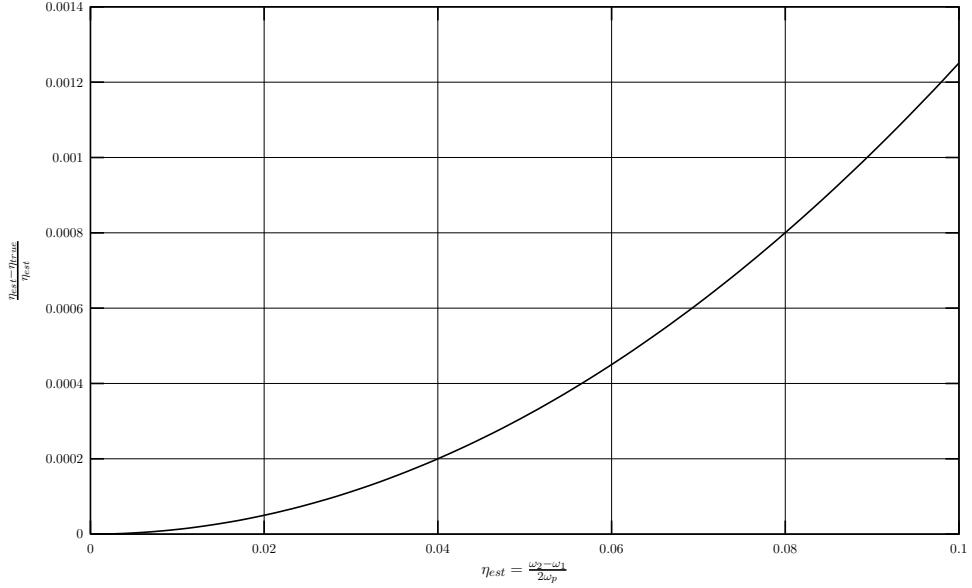


Figure 5.4 Error in estimating η using equation (5.16) for low damping.

Impact of Frequency Resolution

Additional error is introduced if the frequency resolution, Δf , is too large. Equation (3.76) states that

$$\Delta f = \frac{1}{T} \quad (5.17)$$

where T is the sample period. If Δf is large, the estimate of peak magnitude of the actual FRF has a greater propensity to be missed (read too low) due to it not being at one of the discrete frequency values of the FRF. Figure 5.6 gives maximum induced errors due to a coarse frequency resolution (less error can occur with good fortune). Relative error is plotted versus number of points between (inclusively) f_1 and f_2 . For low damping, with at least 8 points, error is relatively small compared to the that from other sources.

Choice of Data Window

A certain degree of leakage of frequency content inevitably occurs between adjacent frequency “bins” as discussed in Section 3.11. Figures 3.8 and 3.9 illustrate that the auto-spectral density at some frequency f also contains some of the amplitudes of adjacent frequencies. For example, consider the Blackman window, and observe Figure 3.9. The auto-spectral density at f is a weighted sum of the

labels of
graphics are
too small

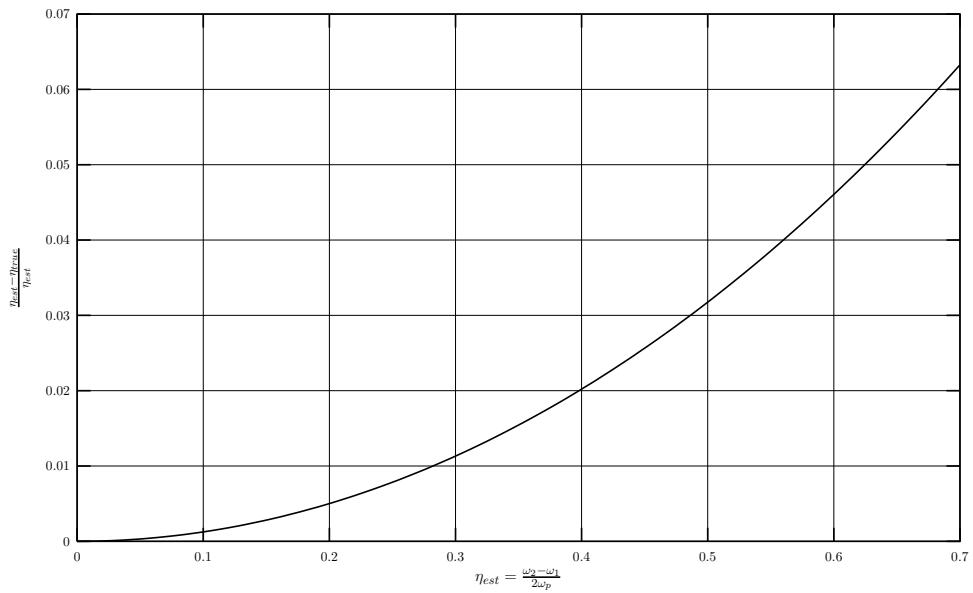


Figure 5.5 Error in estimating η using equation (5.16) for high damping.

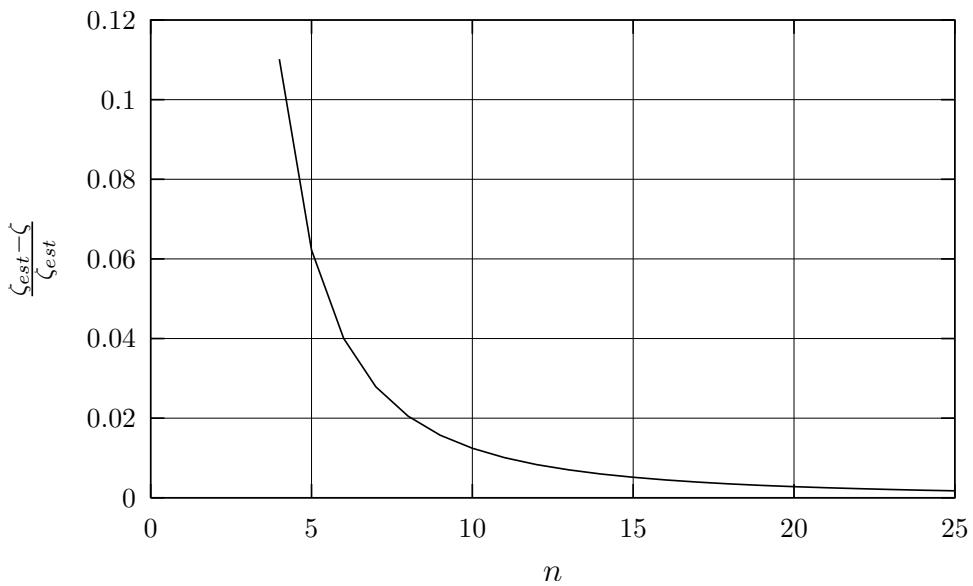


Figure 5.6 Error in estimating η or ζ using discrete FRF estimation, $\zeta < 0.1$ or $\eta < 0.05$.

auto-spectral densities above, at and below f . Frequency content at $f - \Delta f$ is included at a weighting of about 4.5dB below that at f . This is only 1.5dB less than the 3dB points we are looking for. If ASD values on either side and at the peak are in reality, but not measurement, [0 3 0]dB, then applying a Blackman window will result in values of

$$\begin{bmatrix} 10^{\frac{-7.5}{10}} & 10^{\frac{-12}{10}} & 10^{\frac{-28}{10}} \\ 10^{\frac{-12}{10}} & 10^{\frac{-7.5}{10}} & 10^{\frac{-12}{10}} \\ 10^{\frac{-28}{10}} & 10^{\frac{-12}{10}} & 10^{\frac{-7.5}{10}} \end{bmatrix} \begin{bmatrix} 10^{\frac{0}{10}} \\ 10^{\frac{3}{10}} \\ 10^{\frac{0}{10}} \end{bmatrix} \approx \begin{bmatrix} 0.17783 & 0.06310 & 0.00158 \\ 0.06310 & 0.17783 & 0.06310 \\ 0.00158 & 0.06310 & 0.17783 \end{bmatrix} \begin{bmatrix} 1.00000 \\ 1.99526 \\ 1.00000 \end{bmatrix} \quad (5.18)$$

$$= \begin{bmatrix} 0.30531 \\ 0.48100 \\ 0.30531 \end{bmatrix} = \begin{bmatrix} -5.15266 \\ -3.17851 \\ -5.15266 \end{bmatrix} \text{ dB}$$

Equation (5.18) is similar to equation (3.100), but in matrix form, and also for ASDs instead of the DFTs. It is readily apparent that the original 3dB difference in magnitudes has been reduced to a 2dB difference. As shown in Figure 5.6, it is inadvisable to use fewer than 5 points for establishing the location of the half power points, so this example is over-exaggerated. However, choice of window can clearly have the effect of blunting the peak, and causing an overestimate of damping. If no other peak exists near the peak of interest, a Boxcar window actually does the best job in avoiding undue error. Leakage to the peak from distant frequencies is effectively larger, but the same, for all points in/near the peak, and leakage from immediately adjacent frequencies is minimized relative to other windows. Leakage to other frequencies will be significant, but when using Half-Power point estimates, FRF values away from the peak are not of interest.

5.3.5 Nyquist Circle Fitting

5.3.6 Some Mistakes and Misunderstandings That Can Ruin an Otherwise Good Test

Boundary Conditions

Gravity

Over/under loading structure

Impact of Windowing Function Choice

Low Frequency Resolution (mode separation, quadrature estimates)

Impact of Data Length

Impact of Nonlinearities

Also: Exceeding Sensor/Actuator Capabilities (nonlinear/resonance of)

5.4 Obtaining Parameter Estimates From Time Domain Data Response Functions

5.4.1 Natural Frequency Estimation

5.4.2 Extracting Data When Multiple Frequencies Exist

5.5 Estimating the Number of Modes

5.5.1 Multivariate Mode Indicator Function

5.5.2 Complex Mode Indicator Function

6

Multiple Degree of Freedom (MDOF) Systems

More significant structures require more detailed modeling, the result of which is, in the end, a multiple degree of freedom system. These models can be generated by a variety of means, including lumped mass models, as generally taught in elementary dynamics and systems classes, or more sophisticated energy or weighted residual based models (Kelly 1993; Meirovitch 1997; Weaver et al. 1990). The most common method for generating MDOF models are the *finite element methods*, or FEM (Bathe and Wilson 1976, 1996; Cook et al. 1989; Rao 1999; Zienkiewicz and Taylor 1989).

Needs more specific to new chapter.

6.1 Non-proportionally Damped, Gyroscopic and Circulatory Systems

Under a variety of circumstances it is often not possible to solve equation (1.185) as presented in section 1.9. These include non-proportional damping, gyroscopic systems, and aeroelastic systems (Inman 2006). For systems not satisfying equation (1.186), the undamped mode shapes do not decouple the damping matrix. One could attempt to carry on the solution for the eigenvalues and mode shapes by substituting equation (1.158) into the homogeneous form of equation (1.154) yielding

$$M\psi\lambda^2 + C\psi\lambda + K\psi = (M\lambda^2 + C\lambda + K)\psi = 0 \quad (6.1)$$

Following the methodology of section 1.9, the eigenvalues can be obtained from

$$\det(M\lambda^2 + C\lambda + K) = 0 \quad (6.2)$$

Unfortunately, eigensolvers for this form of the eigenvalue expression are not well developed. An alternative form can be obtained by premultiplying equation (8.74) by M^{-1} and rearranging yielding

$$\lambda^2\psi = -M^{-1}K\psi - M^{-1}C\lambda\psi \quad (6.3)$$

Appending this to the identity expression

$$\lambda\psi = I\lambda\psi \quad (6.4)$$

yields

$$\lambda \begin{bmatrix} \psi \\ \lambda\psi \end{bmatrix} = \begin{bmatrix} 0 & I\lambda\psi \\ -M^{-1}K\psi & -M^{-1}C\lambda\psi \end{bmatrix} \quad (6.5)$$

The vector

$$\begin{bmatrix} \psi \\ \lambda\psi \end{bmatrix} \quad (6.6)$$

can be factored out of the right side, so that it appears on both sides, giving

$$\lambda \begin{bmatrix} \psi \\ \lambda\psi \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix} \begin{bmatrix} \psi \\ \lambda\psi \end{bmatrix} \quad (6.7)$$

Defining

$$\phi = \begin{bmatrix} \psi \\ \lambda\psi \end{bmatrix} \quad (6.8)$$

this can now be written as a first-order eigenvalue equation

$$\lambda\phi = A\phi \quad (6.9)$$

where the eigenvalues of A are the eigenvalues of the system and the eigenvectors are as given by (6.8). In the following section the *state space* formulation is derived resulting in the eigen problem and many other benefits that we will take advantage of.

6.2 State-Space Analysis

Often the governing equations themselves are presented and used in what is called *state-space* form. It is then necessary to assemble the equation of motion in state-space form. There are two common forms of the state-space form. The a symmetric form is given by

$$\begin{bmatrix} -K & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & -K \\ -K & -C \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ f(t) \end{bmatrix} \quad (6.10)$$

The advantage of symmetric matrices can be helpful in the eigensolution process. Here the first equation (first row) corresponds to an identity statement, while the original second order equation of motion is represented in the second row. The more commonly found form in system identification is

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ M^{-1}f(t) \end{bmatrix} \quad (6.11)$$

because it is not restricted to second order systems, and the symmetry advantages of equation (6.10) do not exist for gyroscopic or circulatory systems.

Systems for which the governing equations of motion take the form

$$M\ddot{x}(t) + (C + G)\dot{x}(t) + Kx(t) = f(t), \quad (6.12)$$

where G is skew symmetric are called *gyroscopic* systems. The matrix G is a result of T_1 form of kinetic energy, the Coriolis effect often resulting in rotational systems¹. For the undamped case ($C = 0$), if K is positive definite, the system is stable. If in addition the system is damped and C is positive definite, the system is asymptotically stable, meaning that the transient response decays to equilibrium. If C is positive definite, but K is not positive definite (the system has rigid body modes), the system is unstable, meaning that the solution diverges from equilibrium. Further details and cases are examined in Inman (Inman 1989). Another case that can occur in aeroelastic systems is the circulatory case where the system is governed by an equation of motion of the form

$$M\ddot{x}(t) + C\dot{x}(t) + (K + H)x(t) = f(t), \quad (6.13)$$

¹ T_1 kinetic energy is kinetic energy linearly dependent on generalized velocities (Meirovitch 1997).

where H is skew symmetric. The circulatory matrix results from forces nominally tangential to the deforming body. Stability analysis of such systems is much less well developed. In any of these cases, or any of these cases combined, the governing equations of motion can be put into the non-symmetrical state-space form of

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \ddot{\mathbf{x}} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}(K + H) & -M^{-1}(C + G) \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \dot{\mathbf{x}} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ M^{-1}\mathbf{f}(t) \end{bmatrix} \quad (6.14)$$

Models in the form of equation (6.14) are called *state-space models* and are typically presented as

$$\dot{\mathbf{z}} = A\mathbf{z} + B\mathbf{u} \quad (6.15)$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \dot{\mathbf{x}} \end{bmatrix}, A = \begin{bmatrix} 0 & I \\ -M^{-1}(K + H) & -M^{-1}(C + G) \end{bmatrix}, \quad (6.16)$$

$$B = \begin{bmatrix} \mathbf{0} \\ M^{-1}\tilde{\mathbf{B}} \end{bmatrix} \quad (6.17)$$

Here A is called the *state matrix* and B is called the *input matrix* or *input influence matrix*. The variable $\tilde{\mathbf{B}} = \mathbf{I}$ in cases where we are simply performing analysis not including measurement of inputs. If we make this choice, than inevitably some of the inputs $u_i = 0$. Alternatively, for a reduced number of inputs, we can remove columns of $\tilde{\mathbf{B}} = \mathbf{I}$ where no input force is applied.

When doing testing, the inputs, \mathbf{u} , are generally control voltages, and must be translated into forces in the governing equations. Considering the free response case, we again assume a solution

$$\mathbf{x} = \psi e^{\lambda t} \quad (6.18)$$

then

$$\mathbf{z} = \phi e^{\lambda t} = \begin{bmatrix} \psi \\ \lambda\psi \end{bmatrix} e^{\lambda t} \quad (6.19)$$

and

$$\dot{\mathbf{z}} = \lambda\mathbf{z} = \lambda\phi e^{\lambda t} \quad (6.20)$$

Substituting into equation (6.15) with $\mathbf{u} = \mathbf{0}$, and factoring out $e^{\lambda t}$

$$\lambda\phi = A\phi \quad (6.21)$$

which is an eigenvalue problem with eigenvalues λ and eigenvectors ϕ . The eigenvalues of A are either complex conjugate pairs or real pairs as they were for the proportionally damped case. For the system to be stable, it is necessary and sufficient that all eigenvalues have negative real parts. With $2n$ eigenvalues and $2n$ corresponding eigenvectors, the total homogeneous solution is

$$\mathbf{z} = \sum_{i=1}^{2n} a_i \phi_i e^{\lambda_i t} \quad (6.22)$$

where a_i are determined using the initial conditions. The mode shapes, ψ_i , in the general case are complex. The relative phases within each mode shape represent relative phase in modal motion, the best example of which is a traveling wave.

Assuming a solution of

$$\mathbf{z} = \Phi \mathbf{s} \quad (6.23)$$

where

$$\Phi = [\phi_1 \quad \phi_2 \quad \phi_3 \quad \dots \quad \phi_{2n}] \quad (6.24)$$

and \mathbf{s} are the *first-order modal coordinates*. Substituting into equation (6.15), and premultiplying by Φ^{-1} gives

$$\dot{\mathbf{s}} = \Lambda \mathbf{s} + \mathbf{v} \quad (6.25)$$

where

$$\Lambda = \Phi^{-1} A \Phi \quad (6.26)$$

and

$$\mathbf{v} = \Phi^{-1} B \mathbf{u} \quad (6.27)$$

The vector \mathbf{v} are modal inputs. The closed-form solution to the individual modal equations (6.25) is

$$s_i(t) = e^{\lambda_i t} s_i(0) + \int_0^t e^{\lambda_i(t-\tau)} v_i(\tau) d\tau \quad (6.28)$$

which can be proven by substitution. In vector form this is

$$\mathbf{s} = \text{diag}(e^{\lambda_i t}) \mathbf{s}(0) + \int_0^t \text{diag}(e^{\lambda_i(t-\tau)}) \mathbf{v} d\tau \quad (6.29)$$

Substituting $\mathbf{s} = \Phi^{-1} \mathbf{z}$ and pre-multiplying by Φ

$$\begin{aligned} \Phi \Phi^{-1} \mathbf{z}(t) &= \Phi \text{diag}(e^{\lambda_i t}) \Phi^{-1} \mathbf{z}(0) + \int_0^t \Phi \text{diag}(e^{\lambda_i(t-\tau)}) \Phi^{-1} B \mathbf{u} d\tau \\ \mathbf{z}(t) &= e^{At} \mathbf{z}(0) + \int_0^t e^{A(t-\tau)} B \mathbf{u} d\tau \\ &= e^{At} \left(\mathbf{z}(0) + \int_0^t e^{-A\tau} B \mathbf{u} d\tau \right) \end{aligned} \quad (6.30)$$

where

$$e^{At} = \Phi \text{diag}(e^{\lambda_i t}) \Phi^{-1} = \Phi e^{\Lambda t} \Phi^{-1} \quad (6.31)$$

is defined as the *matrix exponential* function, and is simply an abbreviation for decomposing a matrix using eigen-analysis, performing the mathematical operation (an exponential in this case), then reassembling the matrix.

Further, the matrix $\mathcal{S}(\tau) = e^{A\tau}$ is called the *state transition matrix*² and the equation (6.30) can be written in the more succinct form

$$\mathbf{z}(t) = \mathcal{S}(t) \left(\mathbf{z}(0) + \int_0^t \mathcal{S}(-\tau) B \mathbf{u} d\tau \right) \quad (6.32)$$

An alternative method of calculating the matrix exponential is application of the Taylor Series representation of the exponential function

$$e^\alpha = 1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \dots = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} \quad (6.33)$$

Applying this to the matrix At gives

$$e^{At} = \sum_{i=0}^{\infty} \frac{(At)^i}{i!} = \sum_{i=0}^{\infty} \frac{(\Phi \Lambda t \Phi^{-1})^i}{i!} \quad (6.34)$$

²In the more general case of time varying systems, the state transition matrix depends on both start and end times. Thus it is common in the literature to see a definition of the state transition matrix in the form $\mathcal{S}(t_0, t_1) = e^{A(t_1 - t_0)}$, even though absolute time isn't necessary in time-invariant systems.

Since

$$\begin{aligned}
 (\Phi \Lambda t \Phi^{-1})^i &= \prod_{l=1}^i (\Phi \Lambda t \Phi^{-1}) \\
 &= \Phi \Lambda t \Phi^{-1} \times \Phi \Lambda t \Phi^{-1} \times \Phi \Lambda t \Phi^{-1} \dots \\
 &= \Phi(\Lambda t)^i \Phi^{-1}
 \end{aligned} \tag{6.35}$$

then

$$e^{\Lambda t} = \Phi \left(\sum_{i=0}^{\infty} \frac{(\Lambda t)^i}{i!} \right) \Phi^{-1} = \Phi e^{\Lambda t} \Phi^{-1} \tag{6.36}$$

which matches the definition of equation (6.31). The application of equation (6.34) for obtaining the matrix exponential for a fixed value of t has the advantage over (6.31) in that the eigenvalues do not need to be calculated. This is important because a) calculation of the eigenvalues can be error prone, and b) calculation of the eigenvalues tends to be computationally expensive. In practice, calculation using Padé approximation yields more reliable results (Golub and Van Loan 1985).

Example 6.2.1 Obtain the state transition matrix for system with a state matrix

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & 0 \end{bmatrix} \tag{6.37}$$

Solution:

The eigenvalue matrix, Λ is

$$\Lambda = \begin{bmatrix} -j\omega_n & 0 \\ 0 & j\omega_n \end{bmatrix} \tag{6.38}$$

and the eigenvectors are

$$\Phi = \begin{bmatrix} \frac{j}{\omega_n} & \frac{-j}{\omega_n} \\ 1 & 1 \end{bmatrix} \tag{6.39}$$

Calculating the matrix exponential of Λt gives

$$e^{\Lambda t} = \begin{bmatrix} e^{-j\omega_n t} & 0 \\ 0 & e^{j\omega_n t} \end{bmatrix} \tag{6.40}$$

The matrix exponential of $A t$ is then

$$\begin{aligned}
 e^{At} &= \Phi e^{\Lambda t} \Phi^{-1} = \begin{bmatrix} \frac{j}{\omega_n} & \frac{-j}{\omega_n} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-j\omega_n t} & 0 \\ 0 & e^{j\omega_n t} \end{bmatrix} \begin{bmatrix} \frac{-j\omega_n}{j\omega_n^2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\
 &= \begin{bmatrix} \cos(\omega_n t) & \frac{\sin(\omega_n t)}{\omega_n} \\ -\omega_n \sin(\omega_n t) & \cos(\omega_n t) \end{bmatrix}
 \end{aligned} \tag{6.41}$$

6.2.1 Realizations

Consider the state space equations given by (6.15) and (6.83). If we chose to represent \mathbf{z} as

$$\mathbf{z} = P \mathbf{z}' \tag{6.42}$$

we can write the state-space and observation equations as

$$P \dot{\mathbf{z}}' = A P \mathbf{z}' + B \mathbf{u} \tag{6.43}$$

and

$$\begin{aligned}\mathbf{y} &= CP\mathbf{z}' + D\mathbf{u} \\ &= C'\mathbf{z}' + D\mathbf{u}\end{aligned}\tag{6.44}$$

where $C' = CP$. First, consider the case where P is $2n \times 2n$ and non-singular. Then we can pre-multiply equation (6.43) by P^{-1} to yield

$$\begin{aligned}\dot{\mathbf{z}}' &= P^{-1}AP\mathbf{z}' + P^{-1}B\mathbf{u} \\ &= A'\mathbf{z}' + B'\mathbf{u}\end{aligned}\tag{6.45}$$

where $A' = P^{-1}AP$ and $B' = P^{-1}B$. This is called an *equivalent realization* because the input-output relationships for this system are the same as they were for the original system. For example, if we apply equation (6.86) to obtain the frequency response function matrix of the new system, we get

$$\begin{aligned}\mathbf{Y}(j\omega) &= T(j\omega)\mathbf{U}(j\omega) = \left[C' (j\omega I - A')^{-1} B' + D \right] \mathbf{U}(j\omega) \\ &= \left[CP (j\omega I - P^{-1}AP)^{-1} P^{-1}B + D \right] \mathbf{U}(j\omega) \\ &= \left[C (P (j\omega I - P^{-1}AP) P^{-1})^{-1} B + D \right] \mathbf{U}(j\omega) \\ &= [C (j\omega I - A)^{-1} B + D] \mathbf{U}(j\omega)\end{aligned}\tag{6.46}$$

which is equivalent to the result of equation (6.86). Thus the frequency response function relating the inputs to the outputs is invariant with respect to state-space coordinate transformation.

Next consider the case where P is $2n \times q$ where $q < 2n$. The inverse of P can no longer be taken. We instead use *aggregation* (Aoki 1968). Pre-multiplying equation (6.43) by P^T

$$P^T P \dot{\mathbf{z}}' = P^T A P \mathbf{z}' + P^T B \mathbf{u}\tag{6.47}$$

Further, since $P^T P$ is now square and non-singular, we can pre-multiply by $(P^T P)^{-1}$ giving

$$\begin{aligned}(P^T P)^{-1} (P^T P) \dot{\mathbf{z}}' &= (P^T P)^{-1} P^T A P \mathbf{z}' + (P^T P)^{-1} P^T B \mathbf{u} \\ \dot{\mathbf{z}}' &= (P^T P)^{-1} P^T A P \mathbf{z}' + (P^T P)^{-1} P^T B \mathbf{u}\end{aligned}\tag{6.48}$$

In Aoki's paper, he defines the *aggregation matrix*, Q , as

$$Q = (P^T P)^{-1} P\tag{6.49}$$

so that equation (6.48) can be written with some manipulation as

$$\begin{aligned}\dot{\mathbf{z}}' &= QAQ^T (QQ^T)^{-1} \mathbf{z}' + QB\mathbf{u} \\ &= A'\mathbf{z}' + B'\mathbf{u}\end{aligned}\tag{6.50}$$

where $A' = QAQ^T (QQ^T)^{-1}$ and $B' = QB$ now. The output equation is, similarly,

$$\begin{aligned}\mathbf{y} &= CQ^T (QQ^T)^{-1} \mathbf{z}' + D\mathbf{u} \\ &= C'\mathbf{z}' + D\mathbf{u}\end{aligned}\tag{6.51}$$

where $C' = CQ^T (QQ^T)^{-1}$.

The smallest value of q , the number of states, for which the frequency response function of the reduced system defined by equations (6.50) and (6.51) equal to that of equation (6.86) is called a *minimum realization* because it has the minimum number of states necessary to represent the full system input-output characteristics.

6.2.2 Second Order Modal Coordinates

If one selects the transformation matrix $P = \Phi$, the state equations are transformed into first-order modal coordinates as was done in equations (6.25) giving the state equation

$$\dot{\mathbf{s}} = \Lambda \mathbf{s} + \mathbf{v} \quad (6.52)$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \bar{\lambda}_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \bar{\lambda}_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_n & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \bar{\lambda}_n \end{bmatrix} \quad (6.53)$$

when the columns of Φ are ordered according to the corresponding eigenvalues of Λ , \mathbf{v} is defined as $\mathbf{v} = \Phi^{-1} B \mathbf{u}$ by equation (6.27), and the output equation

$$\mathbf{y} = C_m \mathbf{s} + D \mathbf{u} \quad (6.54)$$

where $C_m = C\Phi$. If we make the transformation

$$\mathbf{s} = P_2 \tilde{\mathbf{s}} \quad (6.55)$$

where³

$$P_2 = \begin{bmatrix} -\bar{\lambda}_1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \bar{\lambda}_1 & e^{j(\pi-2\arg(\lambda_1))} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\bar{\lambda}_2 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \bar{\lambda}_2 & e^{j(\pi-2\arg(\lambda_2))} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\bar{\lambda}_n & 1 \\ 0 & 0 & 0 & 0 & \cdots & \bar{\lambda}_n & e^{j(\pi-2\arg(\lambda_n))} \end{bmatrix} \quad (6.56)$$

then the state equations are transformed into *real modal coordinates* and are

$$\dot{\tilde{\mathbf{s}}} = \tilde{\Lambda} \tilde{\mathbf{s}} + B_{rm} \mathbf{u} \quad (6.57)$$

where $\tilde{\Lambda}$ now has the form

$$\tilde{\Lambda} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -\omega_1^2 & -2\zeta_1\omega_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -\omega_2^2 & -2\zeta_2\omega_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -\omega_n^2 & -2\zeta_n\omega_n \end{bmatrix} \quad (6.58)$$

where

$$B_{rm} = P_2^{-1} \Phi^{-1} B \quad (6.59)$$

with the output equation

$$\mathbf{y} = C_{rm} \tilde{\mathbf{s}} + D \mathbf{u} \quad (6.60)$$

where

$$C_{rm} = C\Phi P_2 \quad (6.61)$$

³ $\arg a$ is the angle of the complex argument a . It is important to use the two argument inverse tangent.

Example 6.2.2 MMA file

6.3 Visualization of Modes

Real modes can be plotted and compared to the undeformed structure. Complex modes must be animated by multiplying by $e^{j\omega t}$ and simulated through time.

6.3.1 Controllability and Observability

(Kailath 1980; Skelton 1988)

Controllability

Consider solving equation (6.32) for $\mathbf{u}(t)$ such that from any state $\mathbf{z}(0)$ we can change the system state to $\mathbf{z}(t)$. The closed-form solution to this is

$$\mathbf{u}(\tau) = -B^T \mathcal{S}^T(-\tau) W_c(-t)^{-1} (\mathbf{z}(0) - \mathcal{S}(-t)\mathbf{z}(t)) \quad (6.62)$$

where

$$W_c(t) = \int_0^t \mathcal{S}(\tau) B B^T \mathcal{S}^T(\tau) d\tau \quad (6.63)$$

Proof is obtained by substituting equation (6.62) into equation (6.32). Doing so we get

$$\begin{aligned} \mathbf{z}(t) &= \mathcal{S}(t) \left(\mathbf{z}(0) + \int_0^t \mathcal{S}(-\tau) B \left(-B^T \mathcal{S}(-\tau)^T W_c(-t)^{-1} (\mathbf{z}(0) - \mathcal{S}(-t)\mathbf{z}(t)) \right) d\tau \right) \\ &= \mathcal{S}(t) \left(\mathbf{z}(0) - \int_0^t \mathcal{S}(-\tau) B B^T \mathcal{S}(-\tau)^T d\tau W_c(-t)^{-1} (\mathbf{z}(0) - \mathcal{S}(-t)\mathbf{z}(t)) \right) \\ &= \mathcal{S}(t) (\mathbf{z}(0) - W_c(-t) W_c(-t)^{-1} (\mathbf{z}(0) - \mathcal{S}(-t)\mathbf{z}(t))) \\ &= \mathcal{S}(t) \mathcal{S}(-t) \mathbf{z}(t) \\ &= \mathbf{z}(t) \end{aligned} \quad (6.64)$$

Thus, as long as $W_c(t)$ is non-singular for all t we can find a control input $\mathbf{u}(t)$ to get the system to any desired state. Such a system is called *controllable*. The matrix $\lim_{t \rightarrow \infty} W_c(t)$ is called the *controllability gramian* and is a measure of the degree of system controllability.

From the Cayley-Hamilton theorem, equation (B.78), we can write $\mathcal{S}(t)$ as

$$\mathcal{S}(t) = \sum_{i=0}^{2n-1} g_i(t) A^i \quad (6.65)$$

$$\begin{aligned} W_c(t) &= \int_0^t \left(\sum_{i=0}^{2n-1} g_i(\tau) A^i \right) B B^T \left(\sum_{i=0}^{2n-1} g_i(\tau) (A^i)^T \right) d\tau \\ &= [B \ AB \ A^2 B \ \dots \ A^{2n-1} B] \int_0^t G G^T d\tau \begin{bmatrix} B^T \\ B^T A^T \\ B^T (A^2)^T \\ \vdots \\ B^T (A^{2n-1})^T \end{bmatrix} \quad (6.66) \\ &= \mathcal{C} \int_0^t G G^T d\tau \mathcal{C}^T \end{aligned}$$

where

$$GG^T = \begin{bmatrix} g_0(\tau)I_{2n} \\ g_1(\tau)I_{2n} \\ g_2(\tau)I_{2n} \\ \vdots \\ g_{2n-1}(\tau)I_{2n} \end{bmatrix} \begin{bmatrix} g_0(\tau)I_{2n} & g_1(\tau)I_{2n} & g_2(\tau)I_{2n} & \cdots & g_{2n-1}(\tau)I_{2n} \end{bmatrix} \quad (6.67)$$

Obtaining the inverse of $W_c(t)$ requires the use of the pseudo-inverse of \mathcal{C} such that

$$\begin{aligned} W_c(t)^{-1} &= (\mathcal{C}^T)^{-1} \left(\int_0^t GG^T d\tau \right)^{-1} \mathcal{C}^{-1} \\ &= (\mathcal{C}\mathcal{C}^T)^{-1} \mathcal{C} \left(\int_0^t GG^T d\tau \right)^{-1} \mathcal{C}^T (\mathcal{C}\mathcal{C}^T)^{-1} \end{aligned} \quad (6.68)$$

Noting that the inverse of $\mathcal{C}\mathcal{C}^T$ cannot be singular in order for the inverse of $W_c(t)$ to exist, then the rank of \mathcal{C} must be equal to $2n$ in order for the system to be state-controllable.

Hamdan
and Nayfeh

Modal Controllability

(Hamdan and Nayfeh 1989a,b)

Observability

Consider now the output equation (6.83) with the solution for the states, equation (6.32), substituted for $\mathbf{z}(t)$. This gives an output of

$$\mathbf{y} = C\mathcal{S}(t) \left(\mathbf{z}(0) + \int_0^t \mathcal{S}(-\tau) B \mathbf{u} d\tau \right) + D\mathbf{u} \quad (6.69)$$

Presuming that we know the output \mathbf{y} , and the input, \mathbf{u} , we want to be able to determine the initial state, $\mathbf{z}(0)$. Given this, we can define the unknown initial state's contribution to the output vector to be

$$\tilde{\mathbf{y}} = C\mathcal{S}(t)\mathbf{z}(0) \quad (6.70)$$

where, from equation (6.69), $\tilde{\mathbf{y}}$ is

$$\tilde{\mathbf{y}} = \mathbf{y} - C\mathcal{S}(t) \left(\int_0^t \mathcal{S}(-\tau) B \mathbf{u} d\tau \right) - D\mathbf{u} \quad (6.71)$$

Consider now multiplying equation (6.70) by $S^T(t)C^T$ and integrating from 0 to t giving

$$\begin{aligned} \int_0^t \mathcal{S}^T(\tau) C^T \tilde{\mathbf{y}} d\tau &= \int_0^t \mathcal{S}^T(\tau) C^T C \mathcal{S}(\tau) \mathbf{z}(0) d\tau \\ &= W_o(t) \mathbf{z}(0) \end{aligned} \quad (6.72)$$

where

$$W_o(t) = \int_0^t \mathcal{S}^T(\tau) C^T C \mathcal{S}(\tau) d\tau \quad (6.73)$$

is called the *observability grammian*. The initial state can be determined by

$$\mathbf{z}(0) = W_o^{-1}(t) \int_0^t \mathcal{S}^T(\tau) C^T \tilde{\mathbf{y}} d\tau \quad (6.74)$$

but only if the observability grammian is non-singular. If we consider substituting equation (6.65) into the observability grammian,

$$\begin{aligned}
 W_o(t) &= \int_0^t \left(\sum_{i=0}^{2n-1} g_i(\tau) (A^i)^T \right) C^T C \left(\sum_{i=0}^{2n-1} g_i(\tau) A^i \right) d\tau \\
 &= [C^T \quad A^T C^T \quad (A^2)^T C \quad \dots \quad (A^{2n-1})^T C^T] \int_0^t G G^T d\tau \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{2n-1} \end{bmatrix} \quad (6.75) \\
 &= \mathcal{O}^T \int_0^t G G^T d\tau \mathcal{O}
 \end{aligned}$$

where GG^T is defined by equation (6.67) and

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{2n-1} \end{bmatrix} \quad (6.76)$$

is called the *observability matrix*. Inverting $W_o(t)$ requires the use of the pseudo-inverse of \mathcal{O} such that

$$\begin{aligned}
 W_o(t)^{-1} &= \mathcal{O}^{-1} \left(\int_0^t G G^T d\tau \right)^{-1} (\mathcal{O}^T)^{-1} \\
 &= (\mathcal{O}^T \mathcal{O})^{-1} \mathcal{O}^T \left(\int_0^t G G^T d\tau \right)^{-1} \mathcal{O} (\mathcal{O}^T \mathcal{O})^{-1} \quad (6.77)
 \end{aligned}$$

Noting that the inverse of $\mathcal{O}^T \mathcal{O}$ cannot be singular in order for the inverse of $W_o(t)$ to exist, then the rank of \mathcal{O} must be equal to $2n$ in order for the system to be state-observable.

Modal Observability

6.3.2 Impulse Response of a State-Space Model

The impulse response of the state-space model is a fundamental component of many system identification methodologies. If we consider applying an impulse to a system at rest (zero initial states), equation (6.69) can be applied and simplified as

$$\begin{aligned}
 \mathbf{y}(t) &= \mathbf{h}(t) = C \mathcal{S}(t) \left(\mathbf{z}(0) + \int_0^t \mathcal{S}(-\tau) B \boldsymbol{\delta}(t) d\tau \right) + D \boldsymbol{\delta}(t) \\
 &= C \mathcal{S}(t) B \\
 &= C \mathcal{S}(t) B
 \end{aligned} \quad (6.78)$$

which is valid for $t > 0$. Applying equation (6.31), recalling that $\mathcal{S}(\tau) = e^{\Lambda t}$, this can be rewritten as

$$\mathbf{h}(t) = C \Phi e^{\Lambda t} \Phi^{-1} B \quad (6.79)$$

If the outputs (sensors) represent consistent displacement, velocity, or acceleration, as in a modal test, then we can apply a very physical meaning to sub-terms in equation (6.79) and we write it as

$$\mathbf{h}(t) = \Psi e^{\Lambda t} L \quad (6.80)$$

where $\Psi = C\Phi$ are the (observable) mode shapes and $L = \Phi^{-1}B$ are the *Modal Participation Factors*. In discrete time analysis, $\mathbf{y}(t)$ evaluated at discrete indexed times are defined as the *Markov parameters*

$$\mathcal{M}(k\Delta t) = \Psi e^{\Lambda k\Delta t} L \quad (6.81)$$

6.3.3 Frequency Response Functions of State-Space Systems

The state-space representation of a system also generally includes the *output equation*

$$\mathbf{y} = C_a \ddot{\mathbf{x}} + C_v \dot{\mathbf{x}} + C_d \mathbf{x} \quad (6.82)$$

where C_a , C_v , and C_d are used to translate the actual states into the real measured quantities. Often they simply contain calibration values, amplifier gains, etc. since the real measured quantities are almost always voltages. Solving equation (6.14) for $\ddot{\mathbf{x}}$ and substituting into equation (6.82) gives the more common form of the output equation

$$\mathbf{y} = C\mathbf{z} + D\mathbf{u} \quad (6.83)$$

where

$$C = [C_d - C_a M^{-1}(K + H) \quad C_v - C_a M^{-1}(C + G)], \text{ and } D = [C_a M^{-1} \tilde{B}] \quad (6.84)$$

where C is called the *output matrix*, and D is called the *direct transmission* matrix because direct measurement of an input \mathbf{u} is affected through D . Here it is important to distinguish between the matrix C in the output equation and the matrix C that represents damping in the second order form of the governing equations. While it may be confusing, meaning is readily discerned by context, and unfortunately C is a common standard in both forms of the equations. Further, the other common standard is to use D for the damping matrix instead of C , which doesn't result in any less confusion.

The pair of equations (6.15) and (6.83) are the *continuous-time state-space model*. They represent the real model that we are trying to understand using vibration testing (alternatively to the second order model). For the most part, however, we will instead be focusing on the *discrete-time state-space model* since the data we take will be discretely sampled in time.

The frequency response function of the continuous state-space model is the ratio of the system outputs, \mathbf{y} , to the system inputs, \mathbf{u} . Taking the Fourier transform of equation (6.15) and solving for $Z(j\omega)$ gives

$$Z(j\omega) = (j\omega I - A)^{-1} B U(j\omega) \quad (6.85)$$

Substituting into the Fourier transform of (6.83) gives

$$\mathbf{Y}(j\omega) = T(j\omega) \mathbf{U}(j\omega) = [C(j\omega I - A)^{-1} B + D] \mathbf{U}(j\omega) \quad (6.86)$$

where $T(j\omega)$ is called the *transfer function matrix*. Recognizing that

$$(j\omega I - A)^{-1} = \frac{\text{Adj}(j\omega I - A)}{\det(j\omega I - A)} \quad (6.87)$$

equation (6.86) can be also be written as

$$\mathbf{Y}(j\omega) = T(j\omega) \mathbf{U}(j\omega) = \frac{[C \text{Adj}(j\omega I - A) B + D \det(j\omega I - A)]}{\det(j\omega I - A)} \mathbf{U}(j\omega) \quad (6.88)$$

The poles of the system are then recognized as the solution of

$$\det(j\omega I - A) = 0 \quad (6.89)$$

which are the eigenvalues of A , and the zeros are recognized as the solution of

$$C\text{Adj}(j\omega I - A)B + D\det(j\omega I - A) = 0 \quad (6.90)$$

need words
on poles
and zeros

6.3.4 Frequency Response Functions of State-Space Systems in Modal Coordinates.

Considering equation (6.54) transformed into the frequency domain:

$$\mathbf{Y}(j\omega) = C\Phi\mathbf{S}(j\omega) + D\mathbf{U}(j\omega) \quad (6.91)$$

Writing equation (6.52) in the Fourier domain gives

$$Ij\omega\mathbf{S}(j\omega) = \Lambda\mathbf{S} + \Phi^{-1}B\mathbf{U}(j\omega) \quad (6.92)$$

Solving equation 6.52 for $\mathbf{S}(j\omega)$

$$\mathbf{S}(j\omega) = (j\omega I - \Lambda)^{-1}\Phi^{-1}B\mathbf{U}(j\omega) \quad (6.93)$$

Substituting equation (6.93) into equation (6.91) gives

$$\begin{aligned} \mathbf{Y}(j\omega) &= C\Phi(j\omega I - \Lambda)^{-1}\Phi^{-1}B\mathbf{U}(j\omega) + D\mathbf{U}(j\omega) \\ &= (C\Phi(j\omega I - \Lambda)^{-1}\Phi^{-1}B + D)\mathbf{U}(j\omega) \\ &= (C_m(j\omega I - \Lambda)^{-1}B_m + D)\mathbf{U}(j\omega) \\ &= T(j\omega)\mathbf{U}(j\omega) \end{aligned} \quad (6.94)$$

By observation one should note that this is identical to the result of equation (6.86) but that all matrices have been transformed into modal coordinates. In modal analysis, this equation is typically written as

$$\mathbf{Y}(j\omega) = \Psi(j\omega I - \Lambda)^{-1}L^T + D \quad (6.95)$$

where Ψ are the mode shapes and L are the *Modal Participation Factors*. This expression is analogous to equations (1.213) or (1.214). If the sensor locations and input locations are co-located then $L = \Psi$. Notably this is also the Fourier Transform of equation (6.78). In much of the modal analysis literature the output matrix D is presumed to be zero. This is valid for displacement or velocity sensors, but invalid for acceleration sensors as shown in equation (6.84).

This expression can be written in terms of individual modes as

$$Y_{m,\ell}(j\omega) = \sum_{i=1}^{2n} \frac{\psi_{m,i} L_{\ell,i}^T}{j\omega - \lambda_i} + D_{m,\ell} \quad (6.96)$$

noting that L is a

size of
matrix

6.4 Discrete Time State-Space Models

While the systems that we will perform vibration testing on behave continuously in time, our data is sampled discretely in time. Thus we need to understand discrete-time models, and how to transform between discrete-time models and continuous-time models. Consider equation (6.32) for $t = k\Delta t = kT$ and $t = (k + 1)\Delta t$.

$$\mathbf{z}(k\Delta t) = e^{A k \Delta t} \mathbf{z}(0) + \int_0^{k\Delta t} e^{A(k\Delta t - \tau)} B \mathbf{u}(\tau) d\tau \quad (6.97a)$$

$$\mathbf{z}((k + 1)\Delta t) = e^{A(k+1)\Delta t} \mathbf{z}(0) + \int_0^{(k+1)\Delta t} e^{A((k+1)\Delta t - \tau)} B \mathbf{u}(\tau) d\tau \quad (6.97b)$$

Multiplying equation (6.97a) by $e^{A\Delta t}$ and subtracting it from (6.97b)

$$\mathbf{z}((k + 1)\Delta t) - e^{A\Delta t} \mathbf{z}(k\Delta t) = \int_{k\Delta t}^{(k+1)\Delta t} e^{A((k+1)\Delta t - \tau)} B \mathbf{u}(\tau) d\tau \quad (6.98)$$

Setting $\tau = (k + 1)\Delta t - \sigma$ where σ is used as a dummy variable, and assuming $\mathbf{u}(t) = \mathbf{u}(k\Delta t)$ for $k\Delta t < t < (k + 1)\Delta t$, we can write equation (6.98) as

$$\begin{aligned} \mathbf{z}((k + 1)\Delta t) &= e^{A\Delta t} \mathbf{z}(k\Delta t) + \int_{\Delta t}^0 e^{A\sigma} B(-d\sigma) \mathbf{u}(k\Delta t) \\ \mathbf{z}((k + 1)\Delta t) &= e^{A\Delta t} \mathbf{z}(k\Delta t) + \int_0^{\Delta t} e^{A\sigma} B d\sigma \mathbf{u}(k\Delta t) \\ &= A_d \mathbf{z}(k\Delta t) + B_d \mathbf{u}(k\Delta t) \end{aligned} \quad (6.99)$$

where

$$A_d = e^{A\Delta t} \quad (6.100)$$

and

$$B_d = \int_0^{\Delta t} e^{A\sigma} d\sigma B \quad (6.101)$$

where A and B have been assumed time-invariant. It is important to note that as soon as we factored \mathbf{u} out of the integral we made two assumptions. First, we assumed that it is constant over the period of integration. Further, we assumed that the known value at the beginning of the period is held throughout, *not the actual average value* over the interval. This assumption is commonly referred to as *zero-order hold* and is the most common assumption in the literature regarding discrete time data and analysis.

We have already considered possible ways to calculate $A_d = e^{A\Delta t}$. Considering equation (6.34)

$$A_d = \sum_{i=0}^{\infty} \frac{(A\Delta t)^i}{i!} \quad (6.102)$$

and substituting into equation (6.101) we get

$$\begin{aligned} B_d &= \int_0^{\Delta t} e^{A\sigma} d\sigma B = \left(\int_0^{\Delta t} \sum_{i=0}^{\infty} \frac{(A\sigma)^i}{i!} d\sigma \right) B \\ &= \left(\sum_{i=0}^{\infty} \frac{(A^i \Delta t^{i+1})}{(i+1)!} \right) B = \left(\sum_{i=1}^{\infty} \frac{(A^{i-1} \Delta t^i)}{i!} \right) B, \end{aligned} \quad (6.103)$$

which is a convenient numerical method for obtaining B_d . Comparing equations (6.100) and (6.103), an alternative method for finding B_d is

$$B_d = A^{-1} (A_d - I) B \quad (6.104)$$

However, this expression is valid only if A is non-singular.

6.4.1 Solution of the Discrete Equations

From equation (6.99), noting that

$$\mathbf{z}(k) = A_d \mathbf{z}(k-1) + B_d \mathbf{u}(k-1) \quad (6.105)$$

where the terms Δt have been left out in lieu of using k as an index. Substituting (6.105) into equation (6.99) gives

$$\begin{aligned} \mathbf{z}(k+1) &= A_d (A_d \mathbf{z}(k-1) + B_d \mathbf{u}(k-1)) + B_d \mathbf{u}(k) \\ &= A_d^2 \mathbf{z}(k-1) + A_d B_d \mathbf{u}(k-1) + B_d \mathbf{u}(k) \end{aligned} \quad (6.106)$$

extrapolating back to $k = 0$, the closed form solution is then

$$\mathbf{z}(k) = A_d^k \mathbf{z}(0) + \sum_{i=0}^{k-1} A_d^{k-1-i} B_d \mathbf{u}(i) \quad (6.107)$$

6.4.2 Markov Parameters

Consider the response of equation (6.99) from zero initial conditions to a pulse excitation at input m defined as

$$u_l(0) = \begin{cases} 1, & l = m \\ 0, & l \neq m \end{cases} \quad l = 1 \dots o \quad (6.108)$$

$$u_l(k) = 0, \quad 0 < k \quad l = 1 \dots o \quad (6.109)$$

where o is defined as the number of inputs to the system. The output \mathbf{y} can also be calculated at each time step using equation (6.83). The initial state and the initial output are then

$$\begin{aligned} \mathbf{z}(0) &= \mathbf{0} \\ \mathbf{y}(0) &= D\mathbf{u}(0) \end{aligned} \quad (6.110)$$

Then applying equations (6.99) and (6.83) repeatedly

$$\begin{aligned} \mathbf{z}(1) &= A_d \mathbf{z}(0) + B_d \mathbf{u}(0) = B_d \mathbf{u}(0) \\ \mathbf{y}(1) &= C\mathbf{z}(1) + D\mathbf{u}(1) = CB_d \mathbf{u}(0) \end{aligned} \quad (6.111a)$$

$$\begin{aligned} \mathbf{z}(2) &= A_d \mathbf{z}(1) + B_d \mathbf{u}(1) = A_d B_d \mathbf{u}(0) \\ \mathbf{y}(2) &= C\mathbf{z}(2) + D\mathbf{u}(2) = CA_d B_d \mathbf{u}(0) \end{aligned} \quad (6.111b)$$

$$\begin{aligned} \mathbf{z}(3) &= A_d^2 B_d \mathbf{u}(0) \\ \mathbf{y}(3) &= CA_d^2 B_d \mathbf{u}(0) \end{aligned} \quad (6.111c)$$

⋮

$$\begin{aligned} \mathbf{z}(i) &= A_d^{i-1} B_d \mathbf{u}(0) \\ \mathbf{y}(i) &= CA_d^{i-1} B_d \mathbf{u}(0) \end{aligned} \quad (6.111d)$$

The outputs, $\mathbf{y}(i)$, can be written in the more compact form

$$\begin{bmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \mathbf{y}(2) \\ \vdots \\ \mathbf{y}(i) \end{bmatrix} = \begin{bmatrix} D \\ CB_d \\ CA_d B_d \\ \vdots \\ CA_d^{i-1} B_d \end{bmatrix} \mathbf{u}(0) = \begin{bmatrix} \mathcal{M}(0) \\ \mathcal{M}(1) \\ \mathcal{M}(2) \\ \vdots \\ \mathcal{M}(k) \end{bmatrix} \mathbf{u}(0) \quad (6.112)$$

where the sequence of matrices $\mathcal{M}(0), \mathcal{M}(1), \dots, \mathcal{M}(k)$ are the *Markov parameters* and form the basis of time-domain identification algorithms. The (l, m) element of $\mathcal{M}(k)$, $\mathcal{M}_{l,m}(k)$, represents the values of the l th output, $y_l(k)$ for the system responding to a $t = 0$ pulse excitation in the m th input, $u_m(0) = 1$.

Observing them carefully, it seems that knowing the first three might be sufficient to identify A_d and D , and it also appears that C and B_d cannot be separated independently. This happens to be the case, as will be shown later. We will apply a constraint, in effect, that balances identifying C and B_d in an even-handed fashion. The challenge is how to get the Markov parameters. Simply taking the inverse FFT of the FRF can result in a distorted pulse response function (Juang 1994). A method for obtaining the IRF from the FRF is presented in section 1.6.

Considering a transformation of the state variables as defined by equation (6.42) where P is $2n \times 2n$ and non-singular, we can write $\mathcal{M}(l)$ as

$$\mathcal{M}(l) = C' A_d^{l-1} B_d' = CP(P^{-1} A_d P)^{l-1} P^{-1} B_d = CA_d^{l-1} B_d \quad (6.113)$$

Thus the Markov parameters are also invariant to a conservative change of state variables. Later, when we use the Markov parameters to perform system identification, this fact will result in a non-unique solution for A_d , B_d , and C for which will have to make a selection amongst the infinite number of possible choices that each represent the same input-output relationship of the system.

6.4.3 Controllability of Discrete-Time Models

A state-space model is said to be *controllable* or *state controllable* if any state, $\mathbf{z}(2n)$, can be attained from any other state within a finite time, $k\Delta t$. To prove controllability, an initial state of $\mathbf{z}(0) = \mathbf{0}$ is assumed. If we determine that a system is controllable *from* the zero state, then we can always reverse the process to return to the zero state.⁴ We can write equation (6.107) with zero initial conditions in matrix form as

$$\mathbf{z}(2n) = \underbrace{\begin{bmatrix} B_d & A_d B_d & A_d^2 B_d & \cdots & A_d^{2n-1} B_d \end{bmatrix}}_{2n \times (2n \cdot o)} \underbrace{\begin{bmatrix} \mathbf{u}(2n-1) \\ \mathbf{u}(2n-2) \\ \mathbf{u}(2n-3) \\ \vdots \\ \mathbf{u}(0) \end{bmatrix}}_{(2n \cdot o) \times 1} = \mathcal{C}_d \mathcal{U} \quad (6.114)$$

where \mathcal{C}_d is the *discrete controllability matrix*. There is no usefulness in extending out more than $2n$ time steps because the Cayley-Hamilton theorem (section B.12) shows that polynomial term of A_d greater than $2n - 1$ are linearly dependent on lower power terms. That is, if we include the term A_d^{2n} , we can always reduce the expression for $\mathbf{z}(2n)$ to a form equivalent to that of equation (6.114).

⁴It is left as an exercise to prove that if a system is controllable, the zero state can be achieved from any other state in finite time.

The system is thus controllable if we can find any vector \mathcal{U} such that with the given matrix \mathcal{C}_d , we can obtain any defined $\mathbf{z}(2n)$. If \mathcal{C}_d is square, we can determine \mathcal{U} from

$$\mathcal{U} = \mathcal{C}_d^{-1} \mathbf{z}(2n) \quad (6.115)$$

As long as the inverse of \mathcal{C}_d exists, then we can find a vector \mathcal{U} to obtain any desired $\mathbf{z}(2n)$, and thus the system is controllable. This result is not an optimal control in any sense. In fact, we've set many inputs equal to zero for convenience. If its determinant is non-singular, then the inverse can be found, and the system is controllable. A common check for singularity of a square matrix, rather than determining if it has any eigenvalue equal to 0, is to determine its rank, which is the number of independent columns (or rows) of the matrix. If the rank, $\rho(\mathcal{C}_d)$, of a matrix is equal to its dimension, $2n$ in our case, then it is non-singular.

Unfortunately for this logic, the controllability matrix is generally not square. Consider, however, keeping only the columns of \mathcal{C}_d that are linearly independent. Gauss elimination can be used for this task, but this is only a mental experiment. Note that there cannot be $2n + 1$ independent columns. This is easily demonstrated by selecting any $2n$ of those columns, and assembling them in a matrix we will call $(\tilde{\mathcal{C}}_d)_{1 \rightarrow 2n, 1 \rightarrow 2n}$. Then we can always solve for a vector $\boldsymbol{\alpha}$ such that

$$(\tilde{\mathcal{C}}_d)_{1 \rightarrow 2n, 1 \rightarrow 2n} \boldsymbol{\alpha} = (\tilde{\mathcal{C}}_d)_{1 \rightarrow 2n, 2n+1} \quad (6.116)$$

by inverting $(\tilde{\mathcal{C}}_d)_{1 \rightarrow 2n, 1 \rightarrow 2n}$ where we have used the notation (\rightarrow) to represent a selection over rows or columns.

Consider eliminating the rows of \mathcal{U} corresponding to the discarded columns of \mathcal{C}_d and calling that vector $\tilde{\mathcal{U}}$. Then if we can find a matrix $\tilde{\mathcal{C}}_d$ such that it is non-singular, we can find a $\tilde{\mathcal{U}}$ where

$$\tilde{\mathcal{U}} = \tilde{\mathcal{C}}_d^{-1} \mathbf{z}(2n) \quad (6.117)$$

and the system is thus controllable. In order to determine that a system is controllable, then, all we need to show is that

$$\rho(\mathcal{C}_d) = 2n \quad (6.118)$$

The *discrete controllability gramian* is defined as

$$W_c = \mathcal{C}_d \mathcal{C}_d^T \quad (6.119)$$

derivation
still needed

6.4.4 Observability of Discrete-Time Models

A state-space model is said to be *observable* if the state, $\mathbf{z}(0)$, can be determined from a knowledge of the next $2n$ output vectors, $\mathbf{y}(k)$. To prove observability, we will assume that $\mathbf{u} = \mathbf{0}$. If it is not, we could always calculate $D\mathbf{u}$ and remove its contribution to the output equation from the measured outputs. From equation (6.83),

$$\mathbf{y}(k) = C\mathbf{z}(k) \quad (6.120)$$

Applying equation (6.107) for $\mathbf{u} = \mathbf{0}$

$$\mathbf{y}(k) = CA_d^k \mathbf{z}(0) \quad (6.121)$$

Assembling equation (6.121) for $k = 0, \dots, 2n - 1$

$$\underbrace{\begin{bmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \mathbf{y}(2) \\ \vdots \\ \mathbf{y}(2n-1) \end{bmatrix}}_{(2n \cdot p) \times 1} = \mathcal{Y} = \underbrace{\begin{bmatrix} C \\ CA_d^1 \\ CA_d^2 \\ \vdots \\ CA_d^{2n-1} \end{bmatrix}}_{(2n \cdot p) \times 2n} \mathbf{z}(0) = \mathcal{O}_d \mathbf{z}(0) \quad (6.122)$$

where \mathcal{O}_d is called the *observability matrix*. Following the same logic of the section on controllability, if there are $2n$ independent rows in \mathcal{O}_d , we can form a smaller $2n \times 2n$ matrix $\tilde{\mathcal{O}}_d$ by discarding all other rows. Discarding the corresponding rows in \mathcal{Y} results in the matrix $\tilde{\mathcal{Y}}$. We can then solve for $\mathbf{z}(0)$

$$\mathbf{z}(0) = \tilde{\mathcal{O}}_d^{-1} \tilde{\mathcal{Y}} \quad (6.123)$$

Just as in the controllability case, there is no usefulness in extending out more than $2n$ time steps because the Cayley-Hamilton theorem. Thus, if $\rho(\mathcal{O}_d) = 2n$ the system is observable.

The *discrete observability gramian* is defined as

$$W_o = \mathcal{O}_d^T \mathcal{O}_d \quad (6.124)$$

derivation
still needed

6.4.5 Additional Comments on Controllability and Observability

An additional result is that since equation (6.107) is the closed form solution of a continuous-time system responding to inputs subject to zero-order holds, proving that a discrete-time system is controllable is sufficient to prove that its continuous counterpart is also controllable. It is important to note that the contrary is *not* true. A similar case can be made for observability.

6.5 Nonlinearities in MDOF Systems

Vakakis
Dissertation
and Book

6.6 Continuous Systems

6.7 Nonlinearities in Continuous Systems

Vakakis
Book

6.8 Problems

- Assuming a system has complex modulus damping, knowing the mass normalized mode shapes, the natural frequencies, and the corresponding loss factors, regenerate the second order system model.
- Obtain the state transition matrix, $\mathcal{S}(t)$, for a single degree of freedom *overdamped* system.
- Can the matrix of the previous problem be used to extract the second and third modes? Explain.
- Given the modal matrix

$$\Psi = \begin{bmatrix} 0.91174 & 0.40825 & 0.04546 \\ 0.38697 & -0.81650 & -0.42847 \\ 0.13780 & -0.40825 & 0.90241 \end{bmatrix}$$

and $\omega_1^2 = 1.58$, $\omega_2^2 = 4.00$, and $\omega_3^2 = 11.42$, determine the frequency response function between degree of freedom 1 and degree of freedom 2.

- Obtain the state transition matrix, $\mathcal{S}(t)$, for a single degree of freedom *critically damped* system.
- Prove by substitution that equation (6.28) is the solution to equation (6.25).
- Prove that equations (6.38) and (6.39) are the eigensolution for the matrix A of equation (6.37).
- For the 3 DOF system of Figure 6.1, write the state-space system of equations presuming a displacement sensor is available on each degree of freedom.

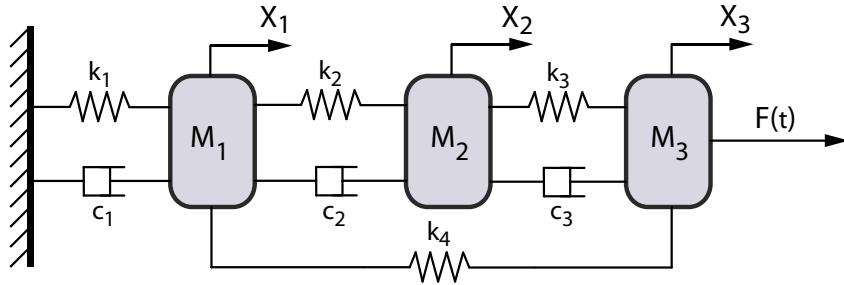


Figure 6.1 Three degree of freedom figure. $m_i = 10$, $c_1 = 1$, $c_2 = c_3 = 0$, $k_1 = k_2 = k_3 = k_4 = 2$, unless otherwise stated.

9. Given the state transition matrix,

$$\mathcal{S}(t) = e^{At} = \begin{bmatrix} \cos(\omega_n t) & \frac{\sin(\omega_n t)}{\omega_n} \\ -\omega_n \sin(\omega_n t) & \cos(\omega_n t) \end{bmatrix}, \quad \text{and } \mathbf{z}(0) = \begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

determine $x(t)$ and $\dot{x}(t)$.

10. Put the following system in state space form:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \ddot{\mathbf{x}} + \begin{bmatrix} 0.1 & -0.1 \\ -0.1 & 0.2 \end{bmatrix} \dot{\mathbf{x}} + \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \sin(\omega t) \quad (6.125)$$

11. Write a computer code to plot the magnitude and phase of a MDOF proportionally damped system using the method of equation (6.86). Apply this code to the system defined in problem 10.
12. Determine the mode shapes and eigenvalues for the system described by equation (6.125), **but with a zero damping matrix**. Use these to generate the state space representation eigenvectors. *You may check your answers with a calculator, but will not receive credit unless you show how the mode shapes of the second order system relate to the eigenvectors of the first order model.*
13. For the 3 DOF system of Figure 6.1, write the state-space system of equations presuming an accelerometer sensor is available on each degree of freedom.
14. For the 3 DOF system of Figure 6.1, obtain the state-space modal realization. What are the mode shapes?
15. For the 3 DOF system of Figure 6.1, obtain the real-form modal realization of the system. Are each of the modes controllable?
16. For the 3 DOF system of Figure 6.1, determine if the system is controllable by using the rank of the controllability matrix.
17. For the 3 DOF system of Figure 6.1, determine if the system is observable by using the rank of the observability matrix.
18. Prove that the system Markov parameters are invariant to coordinate transformation.
19. Prove that if a system is controllable, the zero state can be achieved from any other state in finite time.

20. Derive the conditions under which a continuous-time system is controllable, but its discrete-time counterpart is not.
21. Consider a single degree of freedom system for which $m = 100$, $c = 1$, and $k = 10000$. Generate the discrete state space equations for $\Delta t = \frac{\pi}{20}$, $\Delta t = \frac{\pi}{10}$, and $\Delta t = \frac{2\pi}{10}$ where the single output is a displacement sensor.
22. For systems of problem 21, find the discrete controllability and discrete observability matrices. Also, find the discrete controllability and observability grammians. Describe, physically, why the selection of Δt is important.
23. Use equation (6.102) to determine A_d for the system of problem 21. How many terms in the expansion are necessary to match the matrix exponential function of your favorite code? Plot this as a function of Δt for $0 < \Delta t < \frac{2\pi}{\omega_n}$.
24. Noting equation (6.105), if

$$\begin{aligned} & [z(0) \ z(1) \ z(2) \ z(3) \ z(4)] \\ &= \begin{bmatrix} 1.0000 & 1.0050 & 0.9999 & 0.9849 & 0.9600 \\ 1.0000 & -0.0034 & -1.0066 & -1.9997 & -2.9727 \end{bmatrix} \end{aligned}$$

for free response, and $\Delta t = .01$, estimate the state matrix, A . Hint: If $\mathcal{Z}(0) = [z(0) \ z(1) \ z(2) \ z(3)]$, then $\mathcal{Z}(1) = A_d \mathcal{Z}(0)$.

25. Define the shape of the boundary containing all discrete eigenvalues of A_d for which the continuous system is stable. Hint: Use the fact that the eigenvalues of a stable system must all have negative real parts, the Euler equations, and equation (7.20).
26. Estimate the system model of the system for which the impulse response and corresponding time vector are given in the file vtestingdata1.mat.
27. For the system of Figure 6.1, presuming inputs only as shown, and direct measurement of the displacements, determine
 - (a) the equations of motion in discrete state-space form for $\Delta t = .01$,
 - (b) the transformation necessary to obtain the balanced realization,
 - (c) the controllability and observability grammians in both cases.
28. For the system of equation (6.125), with $C_a = 1$, and $\Delta t = 0.1$, determine the discrete state-space model matrices A_d , B_d , C , and D .
29. Determine the mode shapes of the system defined by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \ddot{x} + \begin{bmatrix} .2 & -.1 \\ -.1 & .2 \end{bmatrix} \dot{x} + \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \sin(\omega t) \quad (6.126)$$

30. Calculate $e^{A\Delta t}$ for the system defined by $\ddot{x} + 10x = 0$ with $\Delta t = .01$.
31. Data is sampled at $f = 1.5915$ Hz. You expect to use Ho-Kalman minimum realization to obtain a system with a natural frequency of approximately 19.9 rad/s. Your identified model has a natural frequency of 9.9 rad/s. Is this possible? If so, demonstrate how this can happen.
32. Consider a the state space system for which

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -3 & 1 & 0 & 0 \\ 1 & -3 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = [1 \ 1 \ 0 \ 0], \quad \text{and} \quad D = [0]$$

Determine whether the system is observable. Use either approach.

33. Determine whether the system of problem 32 is controllable.

in the
directory
/nfs/ecsna1/users/me/jslatec

Time Domain System Identification

7.1 Complex Exponential Method

7.2 Least Squares Complex Exponential Method

7.3 Polyreference Complex Exponential Method

7.4 Ibrahim Time Domain Method (ITD)

7.5 Ho-Kalman System Identification

It was discussed in section 6.4.2 that all of the pertinent system model information for a discrete model is embedded in the Markov parameters. Ho-Kalman (Ho and Kalman 1966) realization theory, upon which the eigensystem realization algorithm (ERA) is based, is a method for extracting this information. It was further noted that the Markov parameters are invariant to coordinate transformations, or *realization independent* parameters. Because of this, any realization extracted based on the Markov parameters is not unique. Thus some decisions will need to be made as to what kind of realization we want.

The first step in the identifying the system matrices from the Markov parameters is to recognize, from equation (6.112) that

$$\mathcal{M}_0 = D \quad (7.1)$$

The remainder of the process is dedicated to identifying the matrices A_d , B_d , and C . We begin by assembling the *Hankel matrix*, defined as

$$\mathcal{H}(k) = \begin{bmatrix} \mathcal{M}_{k+1} & \mathcal{M}_{k+2} & \cdots & \mathcal{M}_{k+b} \\ \mathcal{M}_{k+2} & \mathcal{M}_{k+3} & \cdots & \mathcal{M}_{k+b+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{M}_{k+a} & \mathcal{M}_{k+a+1} & \cdots & \mathcal{M}_{k+a+b-1} \end{bmatrix} \quad (7.2)$$

If we substitute equation (6.112) for the Markov parameters into equation (7.2), for $k = 0$ we get

$$\begin{aligned}\mathcal{H}(0) &= \begin{bmatrix} CB_d & CA_d B_d & \cdots & CA_d^b B_d \\ CA_d B_d & CA_d^2 B_d & \cdots & CA_d^{b+1} B_d \\ \vdots & \vdots & \ddots & \vdots \\ CA_d^a B_d & CA_d^{a+1} B_d & \cdots & CA_d^{a+b} B_d \end{bmatrix} \\ &= \begin{bmatrix} C \\ CA_d \\ CA_d^2 \\ \vdots \\ CA_d^a \end{bmatrix} \begin{bmatrix} B_d & A_d B_d & A_d^2 B_d & \cdots & A_d^b B_d \end{bmatrix} \\ &= \mathcal{O}_d^a \mathcal{C}_d^b\end{aligned}\tag{7.3}$$

where \mathcal{O}_d^a and \mathcal{C}_d^b are near equivalents of the discrete observability and discrete controllability matrices defined in equations (6.122) and (6.114) respectively and are called the *generalized observability matrix* and the *generalized controllability matrix*, respectively. The difference here is that since we don't know the order of the system, we don't know how large a Hankel matrix to construct. This will lead to an iterative process where we try different values for a and b . If we choose $a + b \geq 2n$, then we will have sufficient information in the Hankel matrix to identify the complete $2n \times 2n$ state matrix. Larger values of $a + b$ will generally result in more accurate results, separating the real information from noise. However, choosing $a + b$ too large can result in numerical problems, resulting in a degradation of the identified model. In the ideal case, the rank of $\mathcal{H}(0)$ is the number of states in the system it represents. In the experimental case, however, rank cannot be used because small amounts of noise and computational error lead to the rank being the lesser of $o \times a$ or $p \times b$. From a practical sense, this means that the order of the identified system will be whatever we choose it to be. This doesn't make physical sense. So, instead of using rank as a measure, we use *singular value decomposition*. We represent the singular value decomposing of $\mathcal{H}(0)$ as

$$\mathcal{H}(0) = U \Sigma V^T\tag{7.4}$$

where the matrices U and V satisfy

$$U^T U = I\tag{7.5}$$

and

$$V V^T = I\tag{7.6}$$

And the matrix Σ contains the *singular values* of $\mathcal{H}(0)$ on its diagonal and zeros elsewhere¹. Here a judgment call must be made regarding what is zero and what is not. Real floating point arithmetic does not result in pure zero. Instead, we are looking for values that are so uncharacteristically low as to realistically be considered zero. A threshold of a couple to many orders of magnitude difference can be used as a threshold between the lowest non-zero and the first zero-valued singular value. The work of Juang and Pappa (Juang and Pappa 1986) investigates selection of a threshold based on noise characteristics. The singular values are then sorted in decreasing order so that Σ can be represented as

$$\Sigma = \begin{bmatrix} \Sigma_{2n} & 0 \\ 0 & 0 \end{bmatrix}\tag{7.7}$$

¹Often computer routines will return an equivalent form where Σ is not square. In these cases, both U and V are square. If \mathcal{H} has more rows than columns, then the size of Σ will equal that of \mathcal{H} , and the rows of Σ beyond the number of columns of \mathcal{H} will be all-zero. Further, note that $\mathcal{H}^T = V \Sigma^T U^T$.

where Σ_{2n} is a square diagonal matrix containing the singular values in decreasing order. The rows of U and V are also arranged to the appropriate locations according to their corresponding singular values. The Hankel matrix can then be represented as

$$\mathcal{H}(0) \approx U_{2n} \Sigma_{2n} V_{2n}^T \quad (7.8)$$

where U_{2n} and V_{2n} represent the first $2n$ columns of U and V respectively. This truncation is possible because the columns of U and V corresponding to the zero-valued singular values don't contribute to $\mathcal{H}(0)$ because they are multiplied by the zero singular values in equation (7.4). Thus, a possible decomposition of $\mathcal{H}(0)$ for the \mathcal{O}_d^a and \mathcal{C}_d^b is

$$\mathcal{H}(0) = \mathcal{O}_d^a \mathcal{C}_d^b \approx \left(U_{2n} \Sigma_{2n}^{\frac{1}{2}} \right) \left(\Sigma_{2n}^{\frac{1}{2}} V_{2n}^T \right) \quad (7.9)$$

where the representation is only approximate because we have truncated not only zero-valued parts of the singular value decomposition, but also near-zero parts as well.

Consider next $\mathcal{H}(1)$. Using equation (7.2), and factoring using equations (6.122) and (6.114) gives

$$\begin{aligned} \mathcal{H}(1) &= \begin{bmatrix} CA_d B_d & CA_d^2 B_d & \cdots & CA_d^{b+1} B_d \\ CA_d^2 B_d & CA_d^3 B_d & \cdots & CA_d^{b+2} B_d \\ \vdots & \vdots & \ddots & \vdots \\ CA_d^{a+1} B_d & CA_d^{a+2} B_d & \cdots & CA_d^{a+b+1} B_d \end{bmatrix} \\ &= \begin{bmatrix} C \\ CA_d \\ CA_d^2 \\ \vdots \\ CA_d^a \end{bmatrix} A_d \begin{bmatrix} B_d & A_d B_d & A_d^2 B_d & \cdots & A_d^b B_d \end{bmatrix} \\ &= \mathcal{O}_d^a A_d \mathcal{C}_d^b \end{aligned} \quad (7.10)$$

Substituting for \mathcal{O}_d^a and \mathcal{C}_d^b using the results of equation (7.9), $\mathcal{H}(1)$ can also be written as

$$\mathcal{H}(1) = \left(U_{2n} \Sigma_{2n}^{\frac{1}{2}} \right) A_d \left(\Sigma_{2n}^{\frac{1}{2}} V_{2n}^T \right) \quad (7.11)$$

Post-multiplying by $V_{2n} \Sigma_{2n}^{-\frac{1}{2}}$ and pre-multiplying by $\Sigma_{2n}^{-\frac{1}{2}} U_{2n}^T$ we obtain an estimate for A_d of

$$\hat{A}_d = \Sigma_{2n}^{-\frac{1}{2}} U_{2n}^T \mathcal{H}(1) V_{2n} \Sigma_{2n}^{-\frac{1}{2}} \quad (7.12)$$

and observing equation (7.3), the estimated \hat{B}_d is obtained from the first o columns of

$$\hat{\mathcal{C}}_d^a = \Sigma_{2n}^{\frac{1}{2}} V_{2n}^T \quad (7.13)$$

where o is the number of control inputs and \hat{C} is obtained from the first p rows of

$$\hat{\mathcal{O}}_d^a = U_{2n} \Sigma_{2n}^{\frac{1}{2}} \quad (7.14)$$

where p is the number of outputs.

The identified system matrices represent but one of an infinite number of possible realizations that can generate the Hankel matrices used to identify them. What is special about this form is that it is a *balanced realization*, meaning that it is as controllable as it is observable. This is illustrated by calculating the discrete observability and controllability grammians.

From equation (6.119) the discrete controllability grammian is

$$W_c = \mathcal{C}_d \mathcal{C}_d^T \quad (7.15)$$

Substituting the identified controllability matrix, $\hat{\mathcal{C}}$ from equation (7.13), and using equation (7.5), gives

$$\begin{aligned} W_c &= \Sigma_{2n}^{\frac{1}{2}} V_{2n}^T \left(\Sigma_{2n}^{\frac{1}{2}} V_{2n}^T \right)^T \\ &= \Sigma_{2n}^{\frac{1}{2}} V_{2n}^T V_{2n} \Sigma_{2n}^{\frac{1}{2}} \\ &= \Sigma_{2n} \end{aligned} \quad (7.16)$$

Likewise, from equation (6.124) the discrete observability grammian is

$$W_o = \mathcal{O}^T \mathcal{O} \quad (7.17)$$

Substituting the identified observability matrix, $\hat{\mathcal{O}}$ from equation (7.14), and using equation (7.6), gives

$$\begin{aligned} W_o &= \left(U_{2n} \Sigma_{2n}^{\frac{1}{2}} \right)^T U_{2n} \Sigma_{2n}^{\frac{1}{2}} \\ &= \Sigma_{2n}^{\frac{1}{2}} U_{2n}^T U_{2n} \Sigma_{2n}^{\frac{1}{2}} \\ &= \Sigma_{2n} \end{aligned} \quad (7.18)$$

Thus since the controllability and observability grammians are equivalent, the system realization identified here is balanced.

In order to extract modal parameters from the identified system, we first must transform the discrete model into continuous form. To do this we first take the natural log of equation (6.100) giving

$$A = \frac{\log(A_d)}{\Delta t} \quad (7.19)$$

In order to take the natural log of A_d we reconsider methodology for determining e^{At} covered in section 6.2. Thus

$$\Lambda = \frac{\log(\Lambda_d)}{\Delta t} \quad (7.20)$$

and therefore

$$A = \Phi \frac{\ln(\Lambda_d)}{\Delta t} \Phi^{-1} \quad (7.21)$$

where Φ is the matrix of eigenvectors of A_d (and thus A), and Λ_d are the eigenvalues of A_d .

The continuous-time control matrix can be obtained using equation (6.104) and is

$$B = (A_d - I)^{-1} A B_d \quad (7.22)$$

Note that this method fails when rigid body modes exist because they have discrete eigenvalues of $\lambda_d = 1$. In that case, a series approximation using equation (6.103) must be used. Alternatively, and more efficiently, the inverse of the series of equation (6.103) can be approximated by a series expansion.

The modal damping ratios and natural frequencies can be then obtained using equation (1.30). Further, transforming the state equations into a modal realization (see section 6.2.1) using the transformation matrix

$$P = \Phi \quad (7.23)$$

the state-space equations in modal coordinates become

$$\dot{s} = \Lambda s + B_m u \quad (7.24)$$

where $B_m = \Phi^{-1}B$ as presented earlier in equation (6.25) with the output equation

$$\begin{aligned}\mathbf{y} &= C\Phi\mathbf{s} + D\mathbf{u} \\ &= C_m\mathbf{s} + D\mathbf{u}\end{aligned}\tag{7.25}$$

As the elements of \mathbf{s} now represent modal coordinates, the columns of C_m represent the modes of the system. For instance, if $\mathbf{u} = \mathbf{0}$, and all but one modal coordinate are zero, then the outputs, \mathbf{y} , are the physically observed modal responses. It is important to recognize that, because A_d represents the balanced realization and not the physical one, that the eigenvectors of A_d are *not* the mode shapes of the system.

7.6 Eigensystem Realization Algorithm (ERA)

7.6.1 Modal Amplitude Coherence

7.6.2 Mode Singular Value

7.6.3 Mode Stabilization Diagram

Plots frequency (x axis) and number of states (y axis) to obtain study of sensitivity of number of state selection on values.

7.7 Minimum Model Error Method

7.8 Autoregressive Moving Average Model (ARMA)

7.9 Autoregressive Moving Average with Exogenous Variables (ARMAX)

7.10 Autoregressive Moving Average Vector (ARMAV)

7.11 Random Decrement Technique

7.12 Maximum Entropy Method

7.13 Problems

Note: Data files can be found at <http://www.engineering.wright.edu/~jslater/VibrationTesting>.

1. The impulse response of a system is given in the file `hkdat1.mat`. Using the Ho-Kalman method, **not using EZERA**, to identify the continuous state system matrices. Write out on your exam paper the first 3 rows and the first 4 columns of your Hankel matrix (a 3×4 sub-section of the Hankel matrix).
2. Derive the series expression for calculating equation (7.22).

Frequency Domain Modal Analysis

8.1 Curve Fitting Single Degree of Freedom Models

Consider the SDOF equation of motion

$$m\ddot{x} + c\dot{x} + kx = f \quad (8.1)$$

Transforming into the frequency domain by applying the Fourier transform yields

$$(-m\omega^2 + cj\omega + k)X(j\omega) = F(j\omega) \quad (8.2)$$

Dividing both sides by $F(j\omega)$ yields

$$(-m\omega^2 + cj\omega + k)H(j\omega) = 1 \quad (8.3)$$

If the results of our tests give us the FRF, then we can write this equation for a number of frequencies as

$$\begin{bmatrix} H(j\omega_1) & j\omega_1 H(j\omega_1) & -\omega_1^2 H(j\omega_1) \\ H(j\omega_2) & j\omega_2 H(j\omega_2) & -\omega_2^2 H(j\omega_2) \\ H(j\omega_3) & j\omega_3 H(j\omega_3) & -\omega_3^2 H(j\omega_3) \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} k \\ c \\ m \end{bmatrix} = \mathcal{V}_{0-2,n_s} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \quad (8.4)$$

where

$$\mathcal{V}_{0-2,n_s} = \begin{bmatrix} H(j\omega_1) & j\omega_1 H(j\omega_1) & -\omega_1^2 H(j\omega_1) \\ H(j\omega_2) & j\omega_2 H(j\omega_2) & -\omega_2^2 H(j\omega_2) \\ H(j\omega_3) & j\omega_3 H(j\omega_3) & -\omega_3^2 H(j\omega_3) \\ \vdots & \vdots & \vdots \\ H(j\omega_{n_s}) & j\omega_{n_s} H(j\omega_{n_s}) & -\omega_{n_s}^2 H(j\omega_{n_s}) \end{bmatrix} \quad (8.5)$$

is an augmented *Vandermonde matrix*. Pre-multiplying by \mathcal{V}^T and then by $(\mathcal{V}^T \mathcal{V})^{-1}$ gives

$$\begin{bmatrix} k \\ c \\ m \end{bmatrix} = (\mathcal{V}^T \mathcal{V})^{-1} \mathcal{V}^T \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \quad (8.6)$$

where

$$(\mathcal{V}^T \mathcal{V})^{-1} \mathcal{V}^T = \mathcal{V}^\dagger \quad (8.7)$$

is best obtained using the pseudo-inverse defined by equation (B.63).

This method works fine if there are no other modes and we are using a displacement transducer. If we are using an accelerometer, we re-write the equations

$$(-m\omega^2 + cj\omega + k)(-\omega^2 H(j\omega)) = -\omega^2 \quad (8.8)$$

where the experimentally estimated frequency response function is $H_a(j\omega) = -\omega^2 H(j\omega)$. In matrix form, for a range of frequencies, this becomes

$$\begin{bmatrix} H_a(j\omega_1) & j\omega_1 H_a(j\omega_1) & -\omega_1^2 H_a(j\omega_1) \\ H_a(j\omega_2) & j\omega_2 H_a(j\omega_2) & -\omega_2^2 H_a(j\omega_2) \\ H_a(j\omega_3) & j\omega_3 H_a(j\omega_3) & -\omega_3^2 H_a(j\omega_3) \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} k \\ c \\ m \end{bmatrix} = \mathcal{V}_a \begin{bmatrix} k \\ c \\ m \end{bmatrix} = \begin{bmatrix} -\omega_1^2 \\ -\omega_2^2 \\ -\omega_3^2 \\ \vdots \end{bmatrix} \quad (8.9)$$

Example 8.1.1 Consider a system where the frequency response function data is given by

$$H(j\omega) = \begin{bmatrix} 0.7949 - 0.0636i \\ 0.0000 - 6.6667i \\ -0.5641 - 0.0645i \end{bmatrix}, \text{ for } \omega = \begin{bmatrix} 1.0000 \\ 1.5000 \\ 2.0000 \end{bmatrix} \quad (8.10)$$

Find m , c , and k .

The matrix \mathcal{V} is given by

$$\mathcal{V} = \begin{bmatrix} 0.7949 - 0.0636i & 0.0636 + 0.7949i & -0.7949 + 0.0636i \\ 0.0000 - 6.6667i & 10.0000 & 0.0000 + 15.0000i \\ -0.5641 - 0.0645i & 0.1289 - 1.1281i & 2.2562 + 0.2579i \end{bmatrix} \quad (8.11)$$

and the results of equation (8.6) are

$$\begin{bmatrix} k \\ c \\ m \end{bmatrix} = \begin{bmatrix} 2.250 \\ 0.100 \\ 1.000 \end{bmatrix} \quad (8.12)$$

8.2 Residuals

Typical systems have more than one natural frequency. Each of these frequencies contribute, to some degree, over the entire frequency range. As a result, when we curve fit a single peak while assuming that there is only one mode, the lack of accounting for the other natural frequencies results in curve fitting error. Figure 8.1 shows an example frequency response function. If we consider the FRF written in the modal expansion form of equation (1.213) we can also plot the contribution of each individual mode of the FRF. To compensate for such errors when fitting only a single peak, we introduce *residuals*. Residuals are additional terms added to the FRF function that approximate the contributions of modes outside of the frequency range of interest. The frequency response function of equation (1.113) is instead then written with residuals as

$$H(j\omega) = R_I + \frac{1}{-\omega^2 m + j\omega c + k} + R_K \quad (8.13)$$

where R_I and R_K represent residual contributions from the lower and high frequency modes as *effective inertia* and *effective stiffness*. Any useful function can be used to represent them as long as it is understood that no physical meaning can be assigned to them. A common form used is

$$R_I = \frac{1}{-\omega^2 m_r} \quad (8.14a)$$

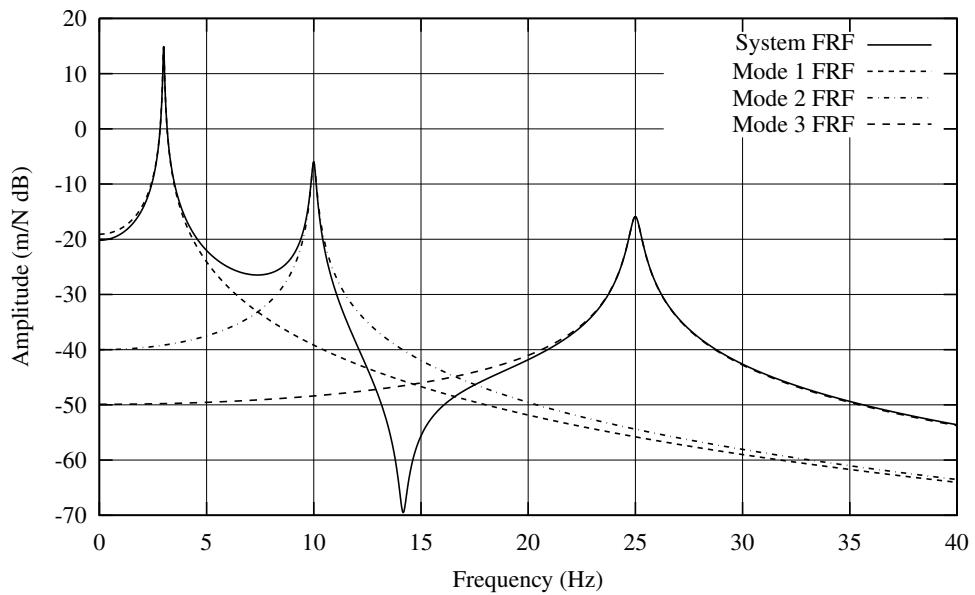


Figure 8.1 Modal Contributions to Frequency Response Function

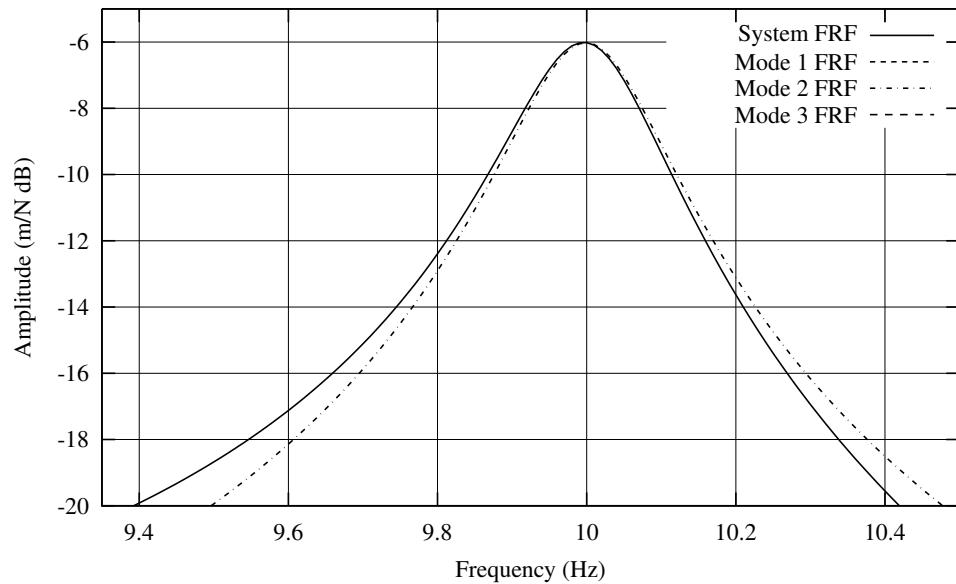


Figure 8.2 Modal Contributions to Frequency Response Function, Zoomed

$$R_K = \frac{1}{k_r} \quad (8.14b)$$

For example, these terms will account for the deviation between the system FRF and the mode 2 FRF when attempting to curve it to a system FRF as shown in Figure 8.2. In the system of Figures 8.1 and 8.2, the modes are well spaced and have marginal impact on the peaks associated with adjacent frequencies. However, the system shown in Figures 8.3 and 8.4 has two closely spaced modes which both smear amplitude content into the frequency range of the adjacent mode. In such a case including residuals is absolutely necessary in order to obtain a fit to the single mode of interest without distortion from the neighboring mode. At some point residuals fail to sufficiently account for the impact of adjacent modes and multi-mode methods, such as those later in this chapter, must be used simply because a sufficiently bandwidth with a single peak is difficult to obtain.

Since they don't have any physical meaning, what name is assigned to each term is irrelevant here. Equation (8.13) can then be written

$$H(j\omega) = \frac{1}{-\omega^2 m_r} + \frac{1}{-\omega^2 m + j\omega c + k} + \frac{1}{k_r} \quad (8.15)$$

With some manipulation, it can be recognized that equation (8.15) can be written as a ratio of two polynomials as

$$H(j\omega) = \frac{\alpha_4(j\omega)^4 + \alpha_3(j\omega)^3 + \alpha_2(j\omega)^2 + \alpha_1 j\omega + \alpha_0}{\beta_4(j\omega)^4 + \beta_3(j\omega)^3 + \beta_2(j\omega)^2} \quad (8.16)$$

Because this is a fraction, we can redefine α_i and β_i to be normalized by β_4 , and we can rewrite this as

$$H(j\omega) = \frac{\alpha_4(j\omega)^4 + \alpha_3(j\omega)^3 + \alpha_2(j\omega)^2 + \alpha_1 j\omega + \alpha_0}{(j\omega)^4 + \beta_3(j\omega)^3 + \beta_2(j\omega)^2} \quad (8.17)$$

Multiplying out terms gives the polynomial equation

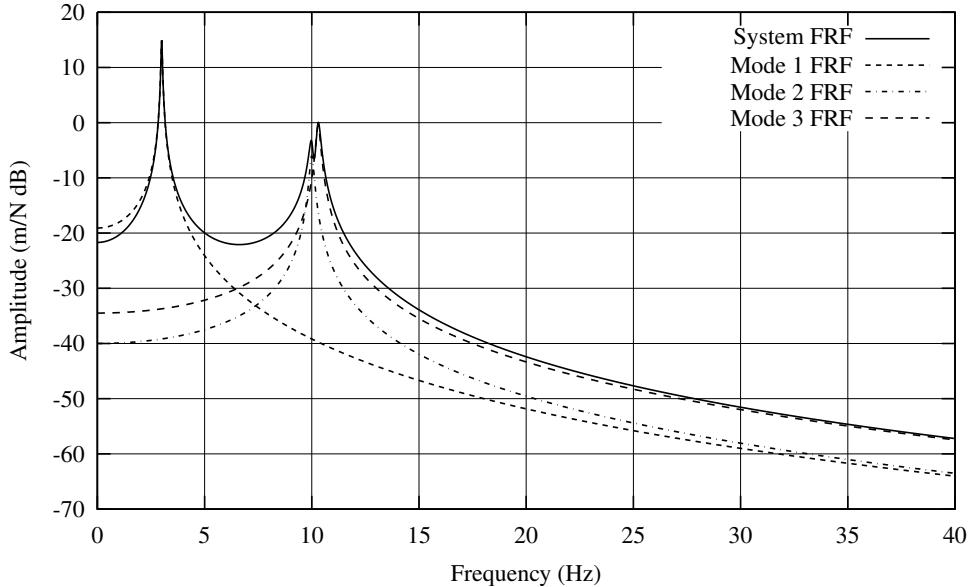


Figure 8.3 Modal Contributions to Frequency Response Function- Closely Spaced Modes

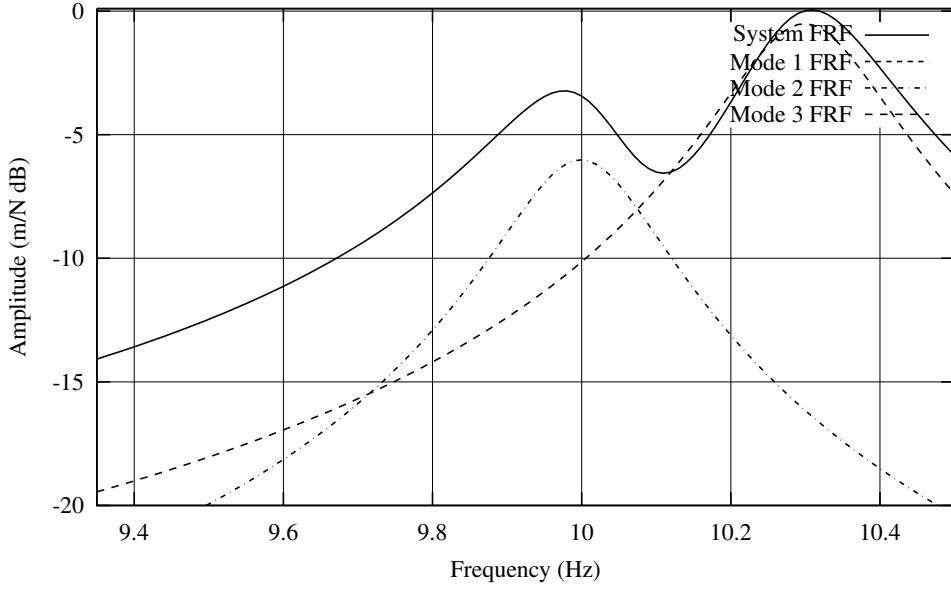


Figure 8.4 Modal Contributions to Frequency Response Function- Closely Spaced Modes, Zoomed

$$\begin{aligned}
 & (j\omega)^4 H(j\omega) + \beta_3(j\omega)^3 H(j\omega) + \beta_2(j\omega)^2 H(j\omega) \\
 & = \alpha_4(j\omega)^4 + \alpha_3(j\omega)^3 + \alpha_2(j\omega)^2 + \alpha_1 j\omega + \alpha_0
 \end{aligned} \tag{8.18}$$

This can be written in matrix form as

$$\begin{bmatrix} (j\omega)^2 H(j\omega) & (j\omega)^3 H(j\omega) & -(j\omega)^0 & \dots & -(j\omega)^4 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = -(j\omega)^4 H(j\omega) \tag{8.19}$$

or

$$\begin{bmatrix} \mathcal{V}_{2-3,1} & -\mathcal{U}_{4,1} \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = -(j\omega)^4 H(j\omega) \tag{8.20}$$

where

$$\mathcal{U}_{4,1} = \begin{bmatrix} (j\omega)^0 & (j\omega)^1 & (j\omega)^2 & (j\omega)^3 & (j\omega)^4 \end{bmatrix} \tag{8.21}$$

is a third order, single value (row), *Vandermonde matrix* in $j\omega$. In order to solve for each of α_i and β_i , at least $n_s > 7$ such equations must be written. Thus we write

$$\begin{bmatrix} \mathcal{V}_{2-3,n_s} & -\mathcal{U}_{4,n_s} \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \mathcal{P} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = - \begin{bmatrix} (j\omega_1)^4 H(j\omega_1) \\ (j\omega_2)^4 H(j\omega_2) \\ (j\omega_3)^4 H(j\omega_3) \\ \vdots \\ (j\omega_{n_s})^4 H(j\omega_{n_s}) \end{bmatrix} \tag{8.22}$$

and the unknown polynomial coefficients are then

$$\begin{bmatrix} \beta \\ \alpha \end{bmatrix} = -\mathcal{P}^\dagger \begin{bmatrix} (j\omega_1)^4 H(j\omega_1) \\ (j\omega_2)^4 H(j\omega_2) \\ (j\omega_3)^4 H(j\omega_3) \\ \vdots \\ (j\omega_{n_s})^4 H(j\omega_{n_s}) \end{bmatrix} \quad (8.23)$$

Since in general \mathcal{P} is not square, we must use the pseudo-inverse of \mathcal{P} defined as

$$\mathcal{P}^\dagger = (\mathcal{P}^T \mathcal{P})^{-1} \mathcal{P}^T \quad (8.24)$$

The two nonzero roots of the polynomial $\lambda^4 + \beta_3\lambda^3 + \beta_2\lambda^2 + 0 + 0 = 0$ yield the poles of the system. This can be followed by a subsequent curve fit to obtain the residue. Noting equations (8.15) and (1.118),

$$H(j\omega) = \frac{1}{-\omega^2 m_r} + \frac{1}{k_r} + \frac{b}{j\omega - \lambda_i} + \frac{\bar{b}}{j\omega - \bar{\lambda}_i} \quad (8.25)$$

Then the residues, b , can be identified using

$$\begin{bmatrix} H(j\omega_1) \\ H(j\omega_2) \\ H(j\omega_3) \\ \vdots \\ H(j\omega_{n_s}) \end{bmatrix} = \begin{bmatrix} \frac{-1}{\omega_1^2} & 1 & \frac{1}{j\omega_1 - \lambda_1} & \frac{1}{j\omega_1 - \bar{\lambda}_1} \\ \frac{-1}{\omega_2^2} & 1 & \frac{1}{j\omega_2 - \lambda_1} & \frac{1}{j\omega_2 - \bar{\lambda}_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{\omega_{n_s}^2} & 1 & \frac{1}{j\omega_{n_s} - \lambda_1} & \frac{1}{j\omega_{n_s} - \bar{\lambda}_1} \end{bmatrix} \begin{bmatrix} \frac{1}{m_r} \\ \frac{1}{k_r} \\ b \\ \bar{b} \end{bmatrix} \quad (8.26)$$

and from that and the poles the remainder of the system properties can be obtain as discussed in section 1.6.

Example 8.2.1 Consider multiple degree of freedom system for which the natural frequency, damping ratio, residues are desired given only a single FRF. Consider a system where the frequency response function data is given by

$$H(j\omega) = \begin{bmatrix} 22.948 - 25.197i \\ 16.335 - 43.478i \\ -11.933 - 47.622i \\ -24.208 - 30.386i \\ -22.909 - 18.321i \end{bmatrix}, \text{ for } \omega = \begin{bmatrix} 0.995 \\ 0.998 \\ 1.001 \\ 1.004 \\ 1.007 \end{bmatrix} \quad (8.27)$$

Beware that this is a textbook example where for the sake of clarity very little data is being used. Typically some 80% of the frequency range of a peak of interest and its corresponding data should be used for the curve fit. This is a qualitative decision based on observation and past experience and becomes challenging to do when frequencies are closely spaced as in the system of Figures 8.3 and 8.4.

Using this data, $[\mathcal{V}_{2-3,n_s} \quad -\mathcal{U}_{4,n_s}]$ is then

$$\begin{bmatrix} -22.72 + 24.94i & -24.82 - 22.60i & -1.00 & -0.99i & 0.99 & 0.99i & -0.98 \\ -16.27 + 43.31i & -43.23 - 16.24i & -1.00 & -1.00i & 1.00 & 0.99i & -0.99 \\ 11.96 + 47.73i & -47.78 + 11.97i & -1.00 & -1.00i & 1.00 & 1.00i & -1.00 \\ 24.41 + 30.64i & -30.77 + 24.52i & -1.00 & -1.00i & 1.01 & 1.01i & -1.02 \\ 23.24 + 18.59i & -18.72 + 23.41i & -1.00 & -1.01i & 1.01 & 1.02i & -1.03 \end{bmatrix} \quad (8.28)$$

where rounding of the data to two decimal places has occurred for the purpose of typesetting. The vector on the right hand side of equation (8.22) is

$$\begin{bmatrix} -22.491 + 24.696i \\ -16.208 + 43.142i \\ 11.987 + 47.838i \\ 24.618 + 30.900i \\ 23.583 + 18.860i \end{bmatrix} \quad (8.29)$$

yielding polynomial coefficients

$$\begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} -0.999 - 0.005i \\ 0.005 - 1.999i \\ -0.485 - 0.787i \\ 0.000 \\ -1.092 - 1.576i \\ 0.000 \\ -0.607 - 0.789i \end{bmatrix} \quad (8.30)$$

which yield a natural frequency of $\omega_n = 0.9999$ and a damping ratio of $\zeta = 0.0048$. Be aware that because four poles exist in the model it is possible to unintentionally identify another peak at another frequency range than the data shows instead of the desired peak. Ironically noise in the data can help to mitigate this possibility. Alternatively, all four roots can be solved for, and the two outside of the frequency range of interest may simply be discarded.

Obtaining the residues of a series of FRF for the same poles by repeatedly applying equation (8.53) would result in an estimate of the corresponding mode shape. This method of this section is not, by itself, recommended for finding results for MDOF systems because there is no enforcement of the consistent value of poles between all of the systems FRFs. Instead, if use of this method is desired, singular value decomposition should be used to reduce the number of FRFs as per section 8.6. Subsequently, identification of a selected pole should nominally take place on the data in which it shows the greatest prominence. Subsequently, equation (8.53) can be applied to each original FRF one at a time to determine the residues and thus the corresponding mode shape.

8.3 Rational Fractional Polynomial Method, SISO

In the case of more than one natural frequency within FRF data, a single FRF can be represented by a *rational polynomial*, as opposed to the *partial fraction* form of equation (1.213). We then can write a single FRF as

$$H_{l,m}(j\omega) = \frac{\alpha_0 + \alpha_1 j\omega + \cdots + \alpha_{n_p-1} (j\omega)^{n_p-1}}{\beta_0 + \beta_1 j\omega + \cdots + \beta_{n_p-1} (j\omega)^{n_p-1} + (j\omega)^{n_p}} \quad (8.31)$$

which is analogous to equation (8.16). Multiplying both sides by the denominator of the right hand side, and rearranging gives

$$[(j\omega)^0 H(j\omega) \quad \dots \quad (j\omega)^{n_p-1} H(j\omega) \quad -(j\omega)^0 \quad \dots \quad -(j\omega)^{n_p-1}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n_p-1} \\ \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n_p-1} \end{bmatrix} = -(j\omega)^{n_p} H(j\omega) \quad (8.32)$$

which can be written as

$$[\mathcal{V}_{n_p-1,1} \quad -\mathcal{U}_{n_p-1,1}] \begin{bmatrix} \beta_{n_p-1} \\ \alpha_{n_p-1} \end{bmatrix} = -(j\omega)^{n_p} H(j\omega) \quad (8.33)$$

Repeating this expression n_s times (once for each frequency sample), it is written

$$[\mathcal{V}_{n_p-1,n_s} \quad -\mathcal{U}_{n_p-1,n_s}] \begin{bmatrix} \beta_{n_p-1} \\ \alpha_{n_p-1} \end{bmatrix} = - \begin{bmatrix} (j\omega_1)^{n_p} H(j\omega_1) \\ (j\omega_2)^{n_p} H(j\omega_2) \\ \vdots \\ (j\omega_{n_s})^{n_p} H(j\omega_{n_s}) \end{bmatrix} \quad (8.34)$$

Solving equation (8.34) for α and β , the natural frequencies and damping ratios can be found as the solutions to $\mathcal{U}_{n_p-1}\beta = 0$. For high n_p the pseudo-inverse of

$$[\mathcal{V}_{n_p-1,n_s} \quad -\mathcal{U}_{n_p-1,n_s}] \quad (8.35)$$

becomes ill-conditioned and cannot be performed in practice. In those cases it becomes necessary to use orthogonal polynomials (Maia and E Silva 1998). Some relief to the ill-conditioning can be obtained by simply normalizing the frequency vector such that the largest and smallest values are -1 and 1 respectively. The obtained natural frequencies must subsequently be rescaled.

8.4 Global Least Squares Fitting

Extending these concepts it is possible to curve fit a large number of FRFs simultaneously obtaining the mode shape, natural frequency, and damping ratio of a selected mode simultaneously. A single term of equation (1.214) can be written as

$${}_i H_{m,\ell} = \frac{{}^i A_{m,\ell}}{\omega_i^2 + 2\zeta_i \omega_i \omega_j - \omega^2} \quad (8.36)$$

where ${}_i H_{m,\ell}$ represents the FRF between the two locations m and ℓ in the frequency range of mode i . Multiplying both sides by the denominator and putting the expression in matrix form as in the previous sections yields

$$[(j\omega)^0 H(j\omega) \quad (j\omega)^1 H(j\omega) \quad -(j\omega)^0] \begin{bmatrix} 2\zeta_i \omega_i \\ \omega_i^2 \\ {}_i A_{m,\ell} \end{bmatrix} = (\omega)^2 {}_i H_{m,\ell}(j\omega) \quad (8.37)$$

This can be applied to any individual FRF. Consider now applying it to a set of FRFs simultaneously. We can write it for an individual ℓ th FRF between an input and an output as

$$[\mathcal{V}_\ell \quad -1] \begin{bmatrix} \boldsymbol{\beta}_i \\ {}_i A_{m,\ell} \end{bmatrix} = (\omega)^2 {}_i H_{m,\ell}(j\omega) \quad (8.38)$$

By recognizing that for any given linear second order system, the characteristic equation represented in the denominator of each FRF, is the same, we can thus assemble these equations into a single matrix for each equation

$$\begin{bmatrix} \mathcal{V}_1 & -1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathcal{V}_2 & \mathbf{0} & -1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{V}_n & \mathbf{0} & \mathbf{0} & \cdots & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_i \\ \psi_{i,\ell} \boldsymbol{\psi}_i \end{bmatrix} = - \begin{bmatrix} (j\omega)^2 \mathbf{H}_1(j\omega) \\ (j\omega)^2 \mathbf{H}_2(j\omega) \\ \vdots \\ (j\omega)^2 \mathbf{H}_n(j\omega) \end{bmatrix} \quad (8.39)$$

where $\psi_{i,\ell}$ is the ℓ term in mode $\boldsymbol{\psi}_i$. Each of the matrices \mathcal{V}, \mathcal{U} , and the vectors \mathbf{H} have length n_s .

8.5 MDOF Curve Fitting - Single Input Collocated with Sensor

Consider now the second order equation of motion in the frequency domain given by

$$M(-\omega^2) \mathbf{X}(j\omega) + Cj\omega \mathbf{X}(j\omega) + K \mathbf{X}(j\omega) = \tilde{B} F(j\omega) \quad (8.40)$$

The system is defined to have a single input at a location where there is also a sensor. Dividing by $F(j\omega)$ gives

$$M(-\omega^2) \mathbf{H}(j\omega) + Cj\omega \mathbf{H}(j\omega) + K \mathbf{H}(j\omega) = \tilde{B} \quad (8.41)$$

Transposing and rearranging this equation, we can write

$$[-\omega^2 \mathbf{H}(j\omega)^T \quad j\omega \mathbf{H}(j\omega)^T \quad \mathbf{H}(j\omega)^T] \begin{bmatrix} M \\ C \\ K \end{bmatrix} = \tilde{B}^T \quad (8.42)$$

Evaluating this over a range of n_s frequencies gives

$$\begin{bmatrix} -\omega_1^2 \mathbf{H}(j\omega_1)^T & j\omega_1 \mathbf{H}(j\omega_1)^T & \mathbf{H}(j\omega_1)^T \\ -\omega_2^2 \mathbf{H}(j\omega_2)^T & j\omega_2 \mathbf{H}(j\omega_2)^T & \mathbf{H}(j\omega_2)^T \\ -\omega_3^2 \mathbf{H}(j\omega_3)^T & j\omega_3 \mathbf{H}(j\omega_3)^T & \mathbf{H}(j\omega_3)^T \\ \vdots & \vdots & \vdots \\ -\omega_{n_s}^2 \mathbf{H}(j\omega_{n_s})^T & j\omega_{n_s} \mathbf{H}(j\omega_{n_s})^T & \mathbf{H}(j\omega_{n_s})^T \end{bmatrix} \begin{bmatrix} M \\ C \\ K \end{bmatrix} = \begin{bmatrix} \tilde{B}^T \\ \tilde{B}^T \\ \tilde{B}^T \\ \vdots \\ \tilde{B}^T \end{bmatrix} \quad (8.43)$$

Which can be solved for the mass, stiffness and damping matrices by

$$\begin{bmatrix} M \\ C \\ K \end{bmatrix} = \begin{bmatrix} -\omega_1^2 \mathbf{H}(j\omega_1)^T & j\omega_1 \mathbf{H}(j\omega_1)^T & \mathbf{H}(j\omega_1)^T \\ -\omega_2^2 \mathbf{H}(j\omega_2)^T & j\omega_2 \mathbf{H}(j\omega_2)^T & \mathbf{H}(j\omega_2)^T \\ -\omega_3^2 \mathbf{H}(j\omega_3)^T & j\omega_3 \mathbf{H}(j\omega_3)^T & \mathbf{H}(j\omega_3)^T \\ \vdots & \vdots & \vdots \\ -\omega_{n_s}^2 \mathbf{H}(j\omega_{n_s})^T & j\omega_{n_s} \mathbf{H}(j\omega_{n_s})^T & \mathbf{H}(j\omega_{n_s})^T \end{bmatrix}^\dagger \begin{bmatrix} \tilde{B}^T \\ \tilde{B}^T \\ \tilde{B}^T \\ \vdots \\ \tilde{B}^T \end{bmatrix} \quad (8.44)$$

where a pseudo-inverse must again be used. This method will work provided that the number of modes appearing in the FRFs is equal to the number of sensors. Alternatively, using the principle of reciprocity,

this method can be used for a single sensor and multiple inputs, such as a hammer ping test. Further, if velocity or acceleration sensors are used, then equation (8.43) can be written as

$$\begin{bmatrix} -\omega_1^2 \mathbf{H}(j\omega_1)^T & j\omega_1 \mathbf{H}(j\omega_1)^T & \mathbf{H}(j\omega_1)^T \\ -\omega_2^2 \mathbf{H}(j\omega_2)^T & j\omega_2 \mathbf{H}(j\omega_2)^T & \mathbf{H}(j\omega_2)^T \\ -\omega_3^2 \mathbf{H}(j\omega_3)^T & j\omega_3 \mathbf{H}(j\omega_3)^T & \mathbf{H}(j\omega_3)^T \\ \vdots & \vdots & \vdots \\ -\omega_{n_s}^2 \mathbf{H}(j\omega_{n_s})^T & j\omega_{n_s} \mathbf{H}(j\omega_{n_s})^T & \mathbf{H}(j\omega_{n_s})^T \end{bmatrix} \begin{bmatrix} M \\ C \\ K \end{bmatrix} = \begin{bmatrix} \tilde{B}^T(j\omega_1)^a \\ \tilde{B}^T(j\omega_2)^a \\ \tilde{B}^T(j\omega_3)^a \\ \vdots \\ \tilde{B}^T(j\omega_{n_s})^a \end{bmatrix} \quad (8.45)$$

where a is 0, 1, or 2 for compliance, mobility, or inertance FRFs respectively.

8.6 Reduction of DOFs by Singular Value Decomposition - Single Input

Consider a set of N frequency response functions from a single input. Presuming reciprocity, the input and output coordinates can be swapped to perform the following analysis. If reciprocity does not apply, a similar derivation can be performed. If the frequency content of those FRFs spans fewer than N modes then an NDOF model with N modes isn't a feasible solution. An alternative is to consider reducing the N FRFs into n_m where n_m is the number of natural frequencies indicated within the frequency range. Since each of the FRFs is a linear combination of their corresponding modal FRFs (See equation (1.211)), it should be possible to reveal only n_m unique FRFs that can be linearly recombined to form each of the N original FRFs, barring the influence of modes outside of the frequency range of interest. This can be done by performing *singular value decomposition* on the set of available FRFs such that

$$\mathcal{V}(j\omega) = \begin{bmatrix} \mathbf{H}_1(j\omega)^T \\ \mathbf{H}_2(j\omega)^T \\ \mathbf{H}_3(j\omega)^T \\ \vdots \\ \mathbf{H}_N(j\omega)^T \end{bmatrix} = U\Sigma V^H \quad (8.46)$$

where $(\cdot)^H$ means *Hermitian* or conjugate-transpose. Here U and V are each complex, while Σ is real-valued. An interpretation of the mathematics is that the rows of V^H correspond to the modal frequency response functions normalized to unit length. The diagonal matrix Σ scales them to adjust for the net amount that they contribute to all FRFs combined, and the matrix U provides the non-dimensional linear recombination of the scaled FRFs to yield the original FRFs. The condensed FRFs are then

$$\mathcal{V}_c(j\omega) = U^H \mathcal{V}(j\omega) = \Sigma V^H \quad (8.47)$$

where either the second or third result, being identical, can be calculated and used. The the new input matrix becomes

$$\tilde{B}_c = U \tilde{B} \quad (8.48)$$

Here it should be noted that we are using the truncated form of the singular value decomposition. The analysis of the previous section can then be performed, with the results being in the coordinates of the transformed equations. The mode shapes can be extracted by transforming them back into physical coordinates via

$$\Psi = U \Psi_c \quad (8.49)$$

or by further refining with a second curve-fitting step. Noting the modal expansion of equation (1.211)

$$H_{l,m} = \sum_{i=1}^n {}_i A_{l,m} \tilde{h}_i(j\omega) \quad (8.50)$$

Write up
appendix on
SVD
Double
check...
should
this not be
 $(V)^T$?

but including residuals this becomes

$$H_{l,m} = \frac{1}{-\omega^2 m_r} + \frac{1}{k_r} + \sum_{i=1}^n {}_i A_{l,m} \tilde{h}_i(j\omega) \quad (8.51)$$

Further, noting equation (1.118),

$$H_{l,m} = \frac{1}{-\omega^2 m_r} + \frac{1}{k_r} + \sum_{i=1}^n \left(\frac{{}_i B_{l,m}}{j\omega - \lambda_i} + \frac{{}_i \bar{B}_{l,m}}{j\omega - \bar{\lambda}_i} \right) \quad (8.52)$$

Then the modal residues can be identified using

$$\begin{bmatrix} H_{l,m}(j\omega_1) \\ H_{l,m}(j\omega_2) \\ H_{l,m}(j\omega_3) \\ \vdots \\ H_{l,m}(j\omega_{n_s}) \end{bmatrix} = \begin{bmatrix} \frac{-1}{\omega_1^2} & 1 & \frac{1}{j\omega_1 - \lambda_1} & \frac{1}{j\omega_1 - \bar{\lambda}_1} & \frac{1}{j\omega_1 - \lambda_2} & \cdots & \frac{1}{j\omega_1 - \bar{\lambda}_{n_m}} \\ \frac{-1}{\omega_2^2} & 1 & \frac{1}{j\omega_2 - \lambda_1} & \frac{1}{j\omega_2 - \bar{\lambda}_1} & \frac{1}{j\omega_2 - \lambda_2} & \cdots & \frac{1}{j\omega_2 - \bar{\lambda}_{n_m}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{\omega_{n_s}^2} & 1 & \frac{1}{j\omega_{n_s} - \lambda_1} & \frac{1}{j\omega_{n_s} - \bar{\lambda}_1} & \frac{1}{j\omega_{n_s} - \lambda_2} & \cdots & \frac{1}{j\omega_{n_s} - \bar{\lambda}_{n_m}} \end{bmatrix} \begin{bmatrix} \frac{1}{m_r} \\ \frac{1}{k_r} \\ {}_1 B_{l,m} \\ {}_1 \bar{B}_{l,m} \\ {}_2 B_{l,m} \\ \vdots \\ {}_{n_m} \bar{B}_{l,m} \end{bmatrix} \quad (8.53)$$

8.7 Reduction of DOFs by Singular Value Decomposition - Multiple Input

The method of the previous section can be used to reduce the number of degrees of freedom for the identification when there are multiple inputs. This is done by assembling the FRFs into a block matrix, each block comprised of $p \times p$ size FRF matrix evaluated in incrementing frequency values. For example, a system with 3 outputs and 2 inputs would look like

$$\begin{aligned} \mathcal{V}_{0,n_s}(j\omega) = \mathcal{V}(j\omega) &= \begin{bmatrix} H(j\omega_1)^T \\ H(j\omega_2)^T \\ \vdots \\ H(j\omega_{n_s})^T \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} H_{1,1}(j\omega_1) & H_{2,1}(j\omega_1) & H_{3,1}(j\omega_1) \\ H_{1,2}(j\omega_1) & H_{2,2}(j\omega_1) & H_{3,2}(j\omega_1) \end{bmatrix} \\ \begin{bmatrix} H_{1,1}(j\omega_2) & H_{2,1}(j\omega_2) & H_{3,1}(j\omega_2) \\ H_{1,2}(j\omega_2) & H_{2,2}(j\omega_2) & H_{3,2}(j\omega_2) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} H_{1,1}(j\omega_{n_s}) & H_{2,1}(j\omega_{n_s}) & H_{3,1}(j\omega_{n_s}) \\ H_{1,2}(j\omega_{n_s}) & H_{2,2}(j\omega_{n_s}) & H_{3,2}(j\omega_{n_s}) \end{bmatrix} \end{bmatrix} \quad (8.54) \end{aligned}$$

where the definition of the form of $H(j\omega)$, shown in equation (8.58), is used for simplicity of derivation in other sections.

8.8 Polypreference Method (MIMO)

The *Polypreference method* Zhang et al. (1986) is a very robust low-order multiple input multiple output method. It is based directly on the MIMO state-space relation

$$h(t) = \Psi e^{\Lambda t} L^T = C_m e^{\Lambda t} B_m \quad (8.55)$$

given by equation (6.80) where the outputs and inputs need not be co-located. Further, $h(t)$ is a matrix of the form

$$h(t) = \begin{bmatrix} h_{1,1}(t) & h_{1,2}(t) & \cdots & h_{1,o}(t) \\ h_{2,1}(t) & h_{2,2}(t) & \cdots & h_{2,o}(t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{p,1}(t) & h_{p,2}(t) & \cdots & h_{p,o}(t) \end{bmatrix} \quad (8.56)$$

where $h_{3,2}(t)$ is the impulse response of the third output (sensor location) to an impulse at the second input location. The matrix Ψ is the *modal matrix* and L is the called the *modal participation matrix*. Taking the Fourier transform of equation (8.55) yields the first order state-space system frequency response function derived in section 6.3.4

$$H(j\omega) = \Psi [j\omega I - \Lambda]^{-1} L^T \quad (8.57)$$

where

$$H(j\omega) = \begin{bmatrix} H_{1,1}(j\omega) & H_{1,2}(j\omega) & \cdots & H_{1,o}(j\omega) \\ H_{2,1}(j\omega) & H_{2,2}(j\omega) & \cdots & H_{2,o}(j\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{p,1}(j\omega) & H_{p,2}(j\omega) & \cdots & H_{p,o}(j\omega) \end{bmatrix} \quad (8.58)$$

Further, taking the one-sided Fourier transform of the first time derivative of equation (8.55) using equation (1.103)

$$\mathcal{F} \left\{ \frac{dh(t)}{dt} \right\} = j\omega H(j\omega) - h(t)|_{t=0} = j\omega H(j\omega) - \Psi L^T \quad (8.59)$$

Performing the same operation by first taking the derivative, then the Fourier transform yields

$$\mathcal{F} \left\{ \frac{dh(t)}{dt} \right\} = \mathcal{F} \left\{ \Psi \Lambda e^{\Lambda t} L^T \right\} = \Psi \Lambda [j\omega I - \Lambda]^{-1} L^T \quad (8.60)$$

because $\mathcal{F} \{ e^{\Lambda t} \} = [j\omega I - \Lambda]^{-1}$. Equating the right hand sides of equations (8.59) and (8.60) yields

$$j\omega H(j\omega) - \Psi L^T = \Psi \Lambda [j\omega I - \Lambda]^{-1} L^T \quad (8.61)$$

Combining equations (8.57) and (8.61)

$$\begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) \end{bmatrix} - \begin{bmatrix} 0 \\ \Psi L^T \end{bmatrix} = \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) - \Psi L^T \end{bmatrix} = \begin{bmatrix} \Psi \\ \Psi \Lambda \end{bmatrix} [j\omega I - \Lambda]^{-1} L^T \quad (8.62)$$

Consider now the eigenvalue problem for the system given by

$$A\Psi = \Psi\Lambda \quad (8.63)$$

It is prudent to be aware that this will only be a portion of the true state matrix as the eigenvectors Ψ are output eigenvectors (of length p), the ramifications of which are discussed later. Equation 8.63 can alternatively be written

$$[-A \quad I] [\Psi \quad \Psi\Lambda] = [0] \quad (8.64)$$

Postmultiplying by $[j\omega I - \Lambda]^{-1} L^T$ and noting equation (8.62) gives

$$[-A \quad I] \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) - \Psi L^T \end{bmatrix} = [0] \quad (8.65)$$

Multiplying out and rearranging gives

$$[A \quad \Psi L^T] \begin{bmatrix} H(j\omega) \\ I \end{bmatrix} = AH(j\omega) + \Psi L^T = j\omega H(j\omega) \quad (8.66)$$

Transposing this for consistency with other analyses in this text gives

$$H^T(j\omega)A^T + L\Psi^T = [H^T(j\omega) \quad I] \begin{bmatrix} A^T \\ L\Psi^T \end{bmatrix} = j\omega H(j\omega)^T \quad (8.67)$$

Evaluated at multiple frequencies and assembled into a single matrix equation

$$\begin{bmatrix} H^T(j\omega_1) & I \\ H^T(j\omega_2) & I \\ \vdots & \vdots \\ H^T(j\omega_{n_s}) & I \end{bmatrix} \begin{bmatrix} A^T \\ L\Psi^T \end{bmatrix} = \begin{bmatrix} j\omega_1 H(j\omega_1)^T \\ j\omega_2 H(j\omega_2)^T \\ \vdots \\ j\omega_{n_s} H(j\omega_{n_s})^T \end{bmatrix} \quad (8.68)$$

or

$$[\mathcal{V}_{0,n_s} \quad \mathcal{U}_{0,n_s}] \begin{bmatrix} A^T \\ L\Psi^T \end{bmatrix} = \mathcal{V}_{1,n_s} \quad (8.69)$$

A and ΨL^T can be extracted from

$$\begin{bmatrix} A^T \\ L\Psi^T \end{bmatrix} = [\mathcal{V}_{0,n_s} \quad \mathcal{U}_{0,n_s}]^\dagger \mathcal{V}_{1,n_s} \quad (8.70)$$

Subsequently, the eigenvalue problem of equation (8.63) can be solved for Λ and Ψ . Once Ψ and ΨL^T are known L^T can be found from

$$L^T = \Psi^\dagger(\Psi L^T) \quad (8.71)$$

This methodology is sensitive to having enough sensors. The size of the matrix of mode vectors, Ψ , is $p \times n_m$ where we recall p is the number of outputs, and n_m the number of system modes within the data.¹ Because of this, and the fact that we are using a first order representation of the system, we will have only partial information identified by the process unless $p \geq 2n_m$.² To illustrate this, we can solve equation (8.57) for $[j\omega I - \Lambda]^{-1} L^T$ using a pseudo-inverse

$$[j\omega I - \Lambda]^{-1} L^T = (\Psi^T \Psi)^{-1} \Psi^T H(j\omega) \quad (8.72)$$

substituting into equation (8.61)

$$\begin{aligned} j\omega H(j\omega) - \Psi L^T &= \Psi \Lambda (\Psi^T \Psi)^{-1} \Psi^T H(j\omega) \\ j\omega H(j\omega) - \Psi L^T &= AH(j\omega) \end{aligned} \quad (8.73)$$

where A is the identified *state matrix*. This can be recognized by comparing to equation (6.50) if one recognizes that a) multiple realizations of the state matrix exist, and b) it may be transformed from an

¹This will be a subjective decision discussed in section 8.6.

²Note if the number of inputs is greater than the number of outputs, for self-adjoint systems we can apply reciprocity and use this method if $\phi \geq 2n_m$.

incomplete modal realization into another, in part. The fact that some mode/eigenvector pairs will be missing simply means that it will represent the portion of the system corresponding only to the included modes. The method of aggregation as described in section 6.2.1 tends to yield somewhat poor results under some circumstances ?. However, when the number of modes within the frequency band of the data satisfies $p \geq 2n_m$ there will be sufficient allowance for A to represent all of the modes within the band of interest yielding better results (Lembregts et al. 1986).

Alternatively, if \mathcal{V}_{0,n_s} are insufficiently independent, the pseudo inverse of equation (8.71) will be numerically ill-conditioned. This will happen when the number of modes substantially contributing to the FRFs is less than $p/2$. Typically noise will fill the role of the additional modes and allow the pseudo inverse to be performed. However, in this case, a number of false modes are identified. Determining the number of modes in the data tends to be qualitative, but can be aided through the observation of the magnitudes of singular values in section 8.7.

8.9 Polyreference Method in Second Order Form

Lembregts et al. (1986) address the shortcoming of the first-order formulation by using a second-order form of the equations. Applying again the derivative property of the one-sided Fourier transform (1.103) to equation (8.59), the Fourier transform of the second derivative of the impulse response is given by

$$\begin{aligned} \mathcal{F} \left\{ \frac{d^2 h(t)}{dt^2} \right\} &= -\omega^2 H(j\omega) - j\omega h(t)|_{t=0} - \dot{h}(t)|_{t=0} \\ &= -\omega^2 H(j\omega) - j\omega \Psi L^T - \Psi \Lambda L^T \end{aligned} \quad (8.74)$$

and noting that the Laplace transform of the second derivative of $h(t)$ is

$$\mathcal{F} \left\{ \frac{d^2 h(t)}{dt^2} \right\} = \mathcal{F} \left\{ \Psi \Lambda^2 e^{\Lambda t} L^T \right\} = \Psi \Lambda^2 [j\omega I - \Lambda]^{-1} L^T \quad (8.75)$$

The right hand sides of these two equations are thus equal:

$$-\omega^2 H(j\omega) - j\omega \Psi L^T - \Psi \Lambda L^T = \Psi \Lambda^2 [j\omega I - \Lambda]^{-1} L^T \quad (8.76)$$

This equation can be appended to equation (8.62) so that

$$\begin{aligned} \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) \\ -\omega^2 H(j\omega) \end{bmatrix} - \begin{bmatrix} 0 \\ \Psi L^T \\ -j\omega \Psi L^T - \Psi \Lambda L^T \end{bmatrix} &= \\ \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) - \Psi L^T \\ -\omega^2 H(j\omega) - j\omega \Psi L^T - \Psi \Lambda L^T \end{bmatrix} &= \begin{bmatrix} \Psi \\ \Psi \Lambda \\ \Psi \Lambda^2 \end{bmatrix} [j\omega I - \Lambda]^{-1} L^T \end{aligned} \quad (8.77)$$

Just as the right column of equation (8.62) was recognized as the first-order system eigenvalue problem in equation (8.63), here we can recognize that

$$I\Psi\Lambda^2 + A_1\Psi\Lambda + A_0\Psi = \mathbf{0} \quad (8.78)$$

is the second order eigenvalue problem as stated in equation (8.74) where $A_1 = M^{-1}C$ and $A_0 = M^{-1}K$. This can be rewritten as

$$\begin{bmatrix} A_0 & A_1 & I \end{bmatrix} \begin{bmatrix} \Psi \\ \Psi \Lambda \\ \Psi \Lambda^2 \end{bmatrix} = \mathbf{0} \quad (8.79)$$

Postmultiplying by $[j\omega I - \Lambda]^{-1} L^T$ then using equation (8.77) gives

$$\begin{aligned} [A_0 \quad A_1 \quad I] \begin{bmatrix} \Psi \\ \Psi\Lambda \\ \Psi\Lambda^2 \end{bmatrix} [j\omega I - \Lambda]^{-1} L^T &= \\ [A_0 \quad A_1 \quad I] \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) - \Psi L^T \\ -\omega^2 H(j\omega) - j\omega \Psi L^T - \Psi \Lambda L^T \end{bmatrix} &= \mathbf{0} \end{aligned} \quad (8.80)$$

Thus

$$[A_0 \quad A_1 \quad I] \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) \\ -\omega^2 H(j\omega) \end{bmatrix} = [A_1 \quad I] \begin{bmatrix} \Psi L^T \\ j\omega \Psi L^T + \Psi \Lambda L^T \end{bmatrix} \quad (8.81)$$

or

$$-\omega^2 H(j\omega) + A_1 j\omega H(j\omega) + A_0 H(j\omega) = B_0 + B_1(j\omega I) \quad (8.82)$$

where $B_0 = A_1 \Psi L^T + \Psi \Lambda L^T$ and $B_1 = \Psi L^T$. The coefficients for the homogeneous equation, A_0 and A_1 can be found using least squares and subsequently the eigenvalues and mode shapes can be solved using equation (8.78). Keep in mind that here $H(j\omega)$ is a matrix.

In block matrix form, this can be written as

$$[A_0 \quad A_1 \quad -B_0 \quad -B_1] \begin{bmatrix} H(j\omega) \\ j\omega H(j\omega) \\ I \\ j\omega I \end{bmatrix} = \omega^2 H(j\omega) \quad (8.83)$$

For the sake of convenience and consistency the transpose of the equation (not the Hermitian) is given by

$$[j\omega H^T(j\omega) \quad H^T(j\omega) \quad I \quad j\omega I] \begin{bmatrix} A_0^T \\ A_1^T \\ -B_0^T \\ -B_1^T \end{bmatrix} = \omega^2 H^T(j\omega) \quad (8.84)$$

Writing this for all frequencies for which frequency response data exists

$$[\mathcal{V}_{0-1,n_s} \quad \mathcal{U}_{0-1,n_s}] \begin{bmatrix} A_0^T \\ A_1^T \\ -B_0^T \\ -B_1^T \end{bmatrix} = -\mathcal{V}_{2,n_s} \quad (8.85)$$

which can be solved for the unknown coefficient matrices by

$$\begin{bmatrix} A_1^T \\ A_0^T \\ -B_0^T \\ -B_1^T \end{bmatrix} = [\mathcal{V}_{0-1,n_s} \quad \mathcal{U}_{0-1,n_s}]^\dagger (-\mathcal{V}_{2,n_s}) \quad (8.86)$$

Once A_0 and A_1 have been found, the mode shapes Ψ and eigenvalues Λ can be found from equation (8.78).

8.10 Free Response Methods

8.10.1 Ibrahim Time Domain Method

8.10.2 PolyMAX Method

8.11 Problems

Note: Data files can be found at <http://www.engineering.wright.edu/~jslater/VibrationTesting>.

1. The frequency response of a system is given in the file `mckiddat.mat`. Determine M , C , and K as well as the mode shapes, natural frequencies, and damping ratios.

9

Model Correction

9.1 Identification of Input forces from System Response

Allen and Carne (2006)Parloo et al. (2003)Fabunmi (1986)Friswell and Mottershead (1995)

9.2 Comparison of Finite Element Models and Identified Models

9.2.1 Finite Element Model Reduction Methods

System Equivalent Reduction Expansion Process (SEREP)

The System Equivalent Reduction Expansion Process (SEREP) is a modal truncation method which uses modal reassembly to generate a reduced order model in physical coordinates (O'Callahan 2000; O'Callahan et al. 1988). This is often an essential step in comparing finite element models to experimentally identified models because of the dramatic difference in the number of degrees of freedom between the two. It is also valuable when performing system simulations because maximum simulation time step is proportional to the period of the highest frequency of the model. Consider a finite element model defined by

$$M\ddot{\mathbf{x}} + C\dot{\mathbf{x}} + K\mathbf{x} = \mathbf{F} \quad (9.1)$$

The goal here is to reduce the size of the system so that the model only includes n modes or the n_r degrees of freedom that are necessary to the model. For the sake of simulation, we are often more concerned with the number of modes n which is prescribed to include all modes with frequencies less than approximately 120% of the bandwidth¹ of the excitation. In modal testing, however, the system will be reduced so that only nodes corresponding to the output sensor locations are retained ($n_r = p$) or a sum of the number of inputs and outputs if they are not *collocated* ($n_r = o + p$). In either case, the number of retained coordinates is constrained to be equal to the number of retained modes, and is chosen to be the larger of the two.

The coordinate vector \mathbf{x} is then partitioned into two parts, \mathbf{x}_r and \mathbf{x}_t , the retained and truncated coordinates, and the equations are reorganized to be in the form (considering only the undamped case

¹Maximum frequency

for now)

$$\begin{bmatrix} M_{rr} & M_{rt} \\ M_{tr} & M_{tt} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{x}}_r \\ \ddot{\mathbf{x}}_t \end{bmatrix} + \begin{bmatrix} K_{rr} & K_{rt} \\ K_{tr} & K_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{F}_r \\ \mathbf{0} \end{bmatrix} \quad (9.2)$$

Consider the eigensolution for the mass normalized eigenvectors, $\Psi = [\Psi_{ar} \quad \Psi_{at}]$ where r and t stand for retained and truncated respectively and a represents that all coordinates are retained in the vector. The modes to be retained, Φ_{ar} , is an $n \times n_r$ matrix, and the modes to be truncated, Φ_{at} , is an $n \times (n - n_r)$ matrix. Substituting

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_t \end{bmatrix} = \Psi \mathbf{r} = [\Psi_{ar} \quad \Psi_{at}] \begin{bmatrix} \mathbf{r}_r \\ \mathbf{r}_t \end{bmatrix} = \begin{bmatrix} \Psi_{rr} & \Psi_{rt} \\ \Psi_{tr} & \Psi_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{r}_r \\ \mathbf{r}_t \end{bmatrix}, \quad (9.3)$$

and pre-multiplying by

$$\Psi^T = [\Psi_{ar} \quad \Psi_{at}]^T$$

yields the equations of motion in modal coordinates

$$I\ddot{\mathbf{r}} + \Lambda\mathbf{r} = \Psi^T \mathbf{F} \quad (9.4)$$

with a transformed mass matrix of

$$\begin{bmatrix} \Psi_{ar}^T \\ \Psi_{at}^T \end{bmatrix} \begin{bmatrix} M_{rr} & M_{rt} \\ M_{tr} & M_{tt} \end{bmatrix} [\Psi_{ar} \Psi_{at}] = \begin{bmatrix} \Psi_{ar}^T M \Psi_{ar} & \Psi_{ar}^T M \Psi_{at} \\ \Psi_{at}^T M \Psi_{ar} & \Psi_{at}^T M \Psi_{at} \end{bmatrix} = \begin{bmatrix} I_{rr} & 0 \\ 0 & I_{tt} \end{bmatrix} \quad (9.5)$$

and a transformed stiffness matrix of

$$\begin{bmatrix} \Psi_{ar}^T \\ \Psi_{at}^T \end{bmatrix} \begin{bmatrix} K_{rr} & K_{rt} \\ K_{tr} & K_{tt} \end{bmatrix} [\Psi_{ar} \Psi_{at}] = \begin{bmatrix} \Psi_{ar}^T K \Psi_{ar} & \Psi_{ar}^T K \Psi_{at} \\ \Psi_{at}^T K \Psi_{ar} & \Psi_{at}^T K \Psi_{at} \end{bmatrix} = \begin{bmatrix} \Lambda_{rr} & 0 \\ 0 & \Lambda_{tt} \end{bmatrix} \quad (9.6)$$

We then truncate the modal vector $\mathbf{r} = [\mathbf{r}_r^T \quad \mathbf{r}_t^T]^T$ by assuming $\mathbf{r}_t = \mathbf{0}$. Equation (9.4) then reduces to

$$I_r \ddot{\mathbf{r}}_r + \Lambda_{rr} \mathbf{r}_r = \Psi_{ar}^T \mathbf{F}_r \quad (9.7)$$

Since now $\mathbf{x}_r = \Psi_{rr} \mathbf{r}_r$ from equation (9.3), we can substitute $\mathbf{r}_r = \Psi_{rr}^{-1} \mathbf{x}_r$ into equation (9.7) yielding

$$\Psi_{rr}^{-1 T} \Psi_{ar}^T M \Psi_{ar} \Psi_{rr}^{-1} \ddot{\mathbf{x}}_r + \Psi_{rr}^{-1 T} \Psi_{ar}^T K \Psi_{ar} \Psi_{rr}^{-1} \mathbf{x}_r = \Psi_{rr}^{-1 T} \Psi_{ar}^T \mathbf{F} \quad (9.8)$$

We can then define a coordinate transformation matrix T where

$$T = \Psi_{ar} \Psi_{rr}^{-1} = \begin{bmatrix} \Psi_{rr} \\ \Psi_{tr} \end{bmatrix} \Psi_{rr}^{-1} = \begin{bmatrix} I \\ \Psi_{tr} \Psi_{rr}^{-1} \end{bmatrix} \quad (9.9)$$

so that

$$\tilde{M} \ddot{\mathbf{x}}_r + \tilde{C} \dot{\mathbf{x}}_r + \tilde{K} \mathbf{x}_r = \tilde{\mathbf{F}} \quad (9.10)$$

where $\tilde{M} = T^T M T$, $\tilde{C} = T^T C T$, $\tilde{K} = T^T K T$ and $\tilde{\mathbf{F}} = T^T \mathbf{F}$ yields our reduced order model.

Guyan Reduction

Improved Reduction System

9.2.2 Normalized Cross Orthogonality

Avitabile and Pechinsky (1998) Avitabile (1999)

9.2.3 Modal Assurance Criterion (MAC)

Allemand (2003)

The *Modal Assurance Criteria (MAC)* or *Modal Shape Correlation Coefficient (MSCC)* is a measure of the correlation between two modes, whether they be two experimental modes, a physics derived mode and an experimental modes, or two physics derived modes and is given by

$$\text{MAC}(\psi_l, \psi'_m) = \frac{(\psi_l^H \psi'_m)^2}{(\psi_l^H \psi_l) (\psi'^H_m \psi'_m)} \quad (9.11)$$

Equation (9.11) can be compared to equation (2.121). Observing equations (2.116) and (2.6), the MAC is simply the normalized covariance between the two selected modes. In vibration testing, the MAC is

Consider an example of a cantilever beam in 3-D space before and after a point mass is added to the tip.

Table 9.1 Natural frequencies of cantilever beam before and after modification.

Mode number Natural Frequencies Before (Hz) Natural Frequencies After (Hz)

1	8.458	7.474
2	8.458	7.474
3	64.950	60.382
4	64.950	60.382
5	165.155	160.974
6	165.155	160.974
7	422.206	416.581
8	422.206	416.581
9	523.326	523.326
10	764.570	756.918
11	764.570	756.918
12	1017.303	949.384

9.2.4 Normalized Modal Difference

9.2.5 Coordinate Modal Assurance Criterion

Lieven and Ewins (1988)

9.2.6 Complex Mode Indicator Function

9.3 Model Correction

Ahmadian et al. (1998a,b, 2000); Friswell and Mottershead (1995, 1998); Friswell et al. (1998a,b, 2000a,b); Mares et al. (1999); Mottershead and Friswell (1998); Mottershead et al. (1999, 1996,?); Mottershead and Friswell (1993); Ouyang et al. (1998) Imregun et al. (1995) Schulz and Inman (1994)

Table 9.2 MAC values of cantilever beam relating modes before and after addition of tip mass.

	1	2	3	4	5	6	7	8	10	11
1	0.99	0.01	0.00	0.30	0.02	0.00	0.07	0.00	0.00	0.12
2	0.01	0.99	0.30	0.00	0.00	0.02	0.00	0.07	0.12	0.00
3	0.00	0.37	0.99	0.00	0.00	0.38	0.00	0.38	0.29	0.00
4	0.37	0.00	0.00	0.99	0.38	0.00	0.38	0.00	0.00	0.29
5	0.04	0.00	0.00	0.47	1.00	0.00	0.36	0.00	0.00	0.12
6	0.00	0.04	0.47	0.00	0.00	1.00	0.00	0.36	0.12	0.00
7	0.00	0.09	0.43	0.00	0.00	0.39	0.00	0.99	0.87	0.00
8	0.09	0.00	0.00	0.43	0.39	0.00	0.99	0.00	0.00	0.87
10	0.13	0.00	0.00	0.31	0.12	0.00	0.90	0.00	0.00	1.00
11	0.00	0.13	0.31	0.00	0.00	0.12	0.00	0.90	1.00	0.00

9.3.1 Parameter Corrections

9.3.2 Boundary Condition Errors

9.3.3 Correcting Theoretical Mode Shapes

9.3.4 Correcting the Theoretical Stiffness Matrix

9.3.5 Model Updating Using Modal Sensitivities

References

- ¹Clarke, D. W., "Generalized Least-Squares Estimation of the Parameters of a Dynamic Model," *Proceedings of the IFAC Symposium on Identification of Automatic Control Systems*, Paper 3.17, June 1967.
- ²Collins, J. D., Hart, G. C., Hasselman, T. K., and Kennedy, B., "Statistical Identification of Structures," *AIAA Journal*, Vol. 12, No. 2, 1974, pp. 185-190.
- ³Chen, Y. M., and Lin, Y., "An Iterative Algorithm for Solving Inverse Problems in Structural Dynamics," *International Journal for Numerical Methods in Engineering*, Vol. 19, June 1983, pp. 825-829.
- ⁴Berger, H., Chaquin, J. P., and Ohayon, R., "Finite Element Model Adjustment Using Experimental Data," *Proceedings of the 2nd International Modal Analysis Conference*, Society for Experimental Mechanics, Bethel, CT, 1984, pp. 638-642.
- ⁵Hoff, C. J., Bernitsas, M. M., Sandstrom, R. E., and Anderson, W. J., "Inverse Perturbation Method for Structural Redesign with Frequency and Mode Shape Constraints," *AIAA Journal*, Vol. 22, No. 9, 1984, pp. 1304-1309.
- ⁶Soeiro, F. J., and Hajela, P., "Damage Detection in Composite Materials Using Identification Techniques," *Proceedings of the 31st AIAA Structures, Structural Dynamics, and Materials Conference*, AIAA, Washington, DC, 1990, pp. 950-960.
- ⁷Hemez, F. M., and Farhat, C., "Correlating Finite Element Models to Modal Tests for Large Flexible Space Structures," *Proceedings of the 1993 CSMA/GAMNI/INRIA National Conference in Structural Design*, 1993, pp. 480-493.
- ⁸Maben, E., and Zimmerman, D. C., "Advances in Parameter Estimation Techniques Applied to Flexible Structures," NASA CP 3242, June 1994.
- ⁹Larson, C., and Zimmerman, D. C., "On the Use of Genetic Algorithms in Structural Damage Detection," *Proceedings of the 11th International Modal Analysis Conference*, Society for Experimental Mechanics, Bethel, CT, 1993, pp. 1095-1101.
- ¹⁰Fulcher, C. W., Marek, E. L., and Mayes, R. L., "Test/Analysis Correlation and Model Updating of the STARS I & II Missile Systems," *Structural Dynamic Modelling Test, Analysis and Correlation*, Bell and Bain, Glasgow, Scotland, U.K., 1993, pp. 433-446.
- ¹¹Carlin, R. A., and Garcia, E., "Parameter Optimization of a Genetic Algorithm for Structural Damage Detection," *Proceedings of the 14th International Modal Analysis Conference*, Society for Experimental Mechanics, Bethel, CT, 1996, pp. 1292-1298.
- ¹²Zimmerman, D. C., Yap, C. K., and Hasselman, T., "Evolutionary Approach for Model Correlation," *Mechanical Systems and Signal Processing*, Vol. 13, No. 4, 1999, pp. 609-625.
- ¹³Ruotolo, R., Surace, C., and Mares, C., "Damage Identification Using Simulated Annealing," *Proceedings of the 15th International Modal Analysis Conference*, Society for Experimental Mechanics, Bethel, CT, 1997, pp. 954-960.
- ¹⁴Holland, J. H., *Adaptation in Natural and Artificial Systems*, Univ. of Michigan Press, Ann Arbor, MI, 1975, Chaps. 1-6.
- ¹⁵Adamidis, P., "Parallel Evolutionary Algorithms: A Review," *Proceedings of the 4th Hellenic-European Conference on Computer Methods and Its Applications*, edited by E. A. Lipitakis, Dept. of Infomatics, Athens Univ. of Economics and Business, Athens, Greece, 1998, Session 12E.
- ¹⁶Cantu-Paz, E., "A Survey of Parallel Genetic Algorithms," *Calculateurs Paralleles, Reseaux et Systemes Repartis*, Vol. 10, No. 2, 1998, pp. 141-171.
- ¹⁷Rudic, Z., *UNA—Computer Program for Static and Dynamic Structural Analysis by Finite Element Method: User and Verification Manual*, Structural Analysis and Research Systems, Inc., 2001.
- ¹⁸Yap, K. C., and Zimmerman, D. C., "A Comparative Study of Structural Dynamics Modification and Sensitivity Method Approximation," *Mechanical Systems and Signal Processing*, Vol. 16, No. 4, 2002, pp. 585-597.
- ¹⁹Crawley, E. F., Van Schoor, M. C., and Bokhour, E. B., "The Mid-deck 0-Gravity Dynamics Experiment—Summary Report," NASA CR 4500, Jan. 1993.

10

Machine Health Monitoring

Adams (2001)Wowk (c1995)Wowk (1991)Wowk (c2000)Rao (2000)Rao (2000)Childs (1993)Vance (1988)Muszyńska (2005)

Contrary to prevailing perceptions, machine health monitoring dominates the field of vibration testing. A single failure in a production line can potentially cause catastrophic repercussions in terms of lost productivity during down time and repair of collateral damage that could have been avoided had imminent failure been identified and prevented. The techniques of machine health monitoring are partly based on the signal processing techniques of chapter 3, additional techniques, some that are hybrid, and experience with the machinery, specific, or as a class.

Common practice and advice by an experienced automobile mechanic is to drive their spouses vehicle once per week, but at least once per month. The purpose of this was to evaluate the vehicle via the multitude of audible, visual, and tactile indicators that can be used in assessing vehicle health. The trained ear of a mechanic can often identify a problem from very slight changes in the sound of the drive train.

Further, any competent driver is aware that a puddle of oil or other fluid beneath their vehicle is a potential sign of problems. A red color indicates a transmission problem, a yellow-green color indicates a coolant leak, and a black-brown color indicates an oil leak. An alert but less experienced driver may be concerned by a puddle of water underneath the vehicle on a dry day. However, the knowledgeable driver will recognize this as most likely the result of condensate from the air conditioning system.

Similarly, machine health monitoring is based on an understanding of the system and the experience of the assessor. Maintenance logs for the equipment must be kept, and any change in the requirements, whether fuel, oil, electrical power, hydraulic pressure, etc, are reasons for concern. Excessive debris material is also an obvious warning sign. Fastidious cleaning can preclude observation of debris if procedures aren't in place for alerting engineers when they increase. Thus it is vital that all users and personnel are educated to be aware of variations in the behavior of machinery. The best resource for identifying the onset of damage is often the accidentally-trained ear of the user of such equipment. The normal sounds and performance of the equipment will become exceptionally familiar to the user after an extended period of time, and an attentive user will be astutely aware of subtle changes.

This chapter addresses only the application of vibration testing techniques to use in machine health monitoring and should be used only as a part of a comprehensive maintenance program. Depending on the cost of machine failure, the use of vibration testing for health monitoring can be prohibitive.

10.1 Time Domain Analysis Methods

Time domain analysis techniques are not often employed in machine health monitoring for the simple reason that machinery running at a constant speed typically exhibits periodic behavior best represented by a Fourier series. The use of digital data acquisition equipment means that any Fourier analysis is performed using an FFT. However, some time domain signal processing can yield useful information, or better data for later frequency domain analysis.

10.1.1 Kurtosis

The simplest method for evaluating the health of a roller bearing system is the *kurtosis* given by equation (2.4) as

$$\hat{\kappa}_x = \frac{1}{T\sigma_x^4} \int_{-T/2}^{T/2} (x(t) - \hat{\mu}_x)^4 dt, \quad \text{or} \quad \frac{1}{N\sigma_x^4} \sum_{i=1}^N (x_i - \hat{\mu}_x)^4 \quad (10.1)$$

where $x(t)$ is typically an accelerometer response adjacent to the bearing (see Figure 10.1). Since the actual standard deviation is never available, the calculated standard deviation from equation (2.2) must be used instead. As described in chapter 2, the kurtosis value is a measure of how spiky a signal is.

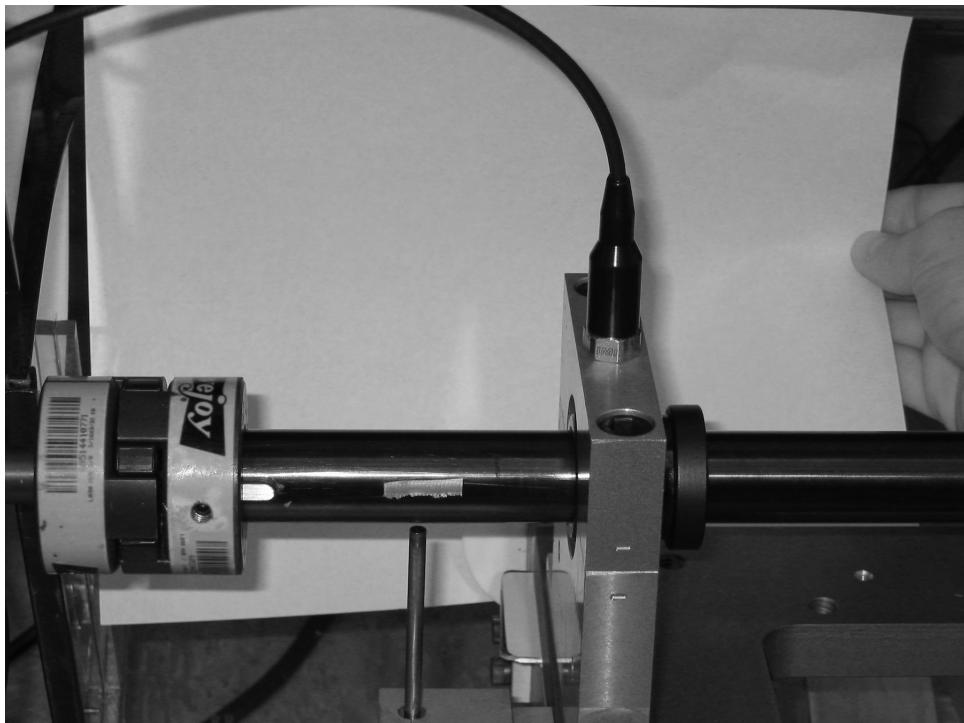


Figure 10.1 Accelerometer mounted adjacent to bearing

For a square wave the kurtosis value is 1 because all of the extrema at 1σ from the mean, while for a

sine the kurtosis increases to 1.5. Further, for a sawtooth wave such as in figure the kurtosis value is approximately 1.628. We take advantage of this in the case of damaged bearings by noting that failed bearings often generate spikes in the acceleration response at the moment the damaged part of the bearing comes into contact with another part of the bearing. As a result, the kurtosis value is increased by damage. Provided that the kurtosis value of the baseline healthy system is available, the kurtosis can provide a reasonable first measure that something is indeed failing. A repeated set of impulses such as those shown in figure 10.2 (and denoted by the \times points) will result in a kurtosis value of ∞ for the ideal theoretical case. However, when noise is added, the kurtosis will reduce to a value closer to 4 – 6, although values outside this range are not uncommon. The nominal healthy system should be free of

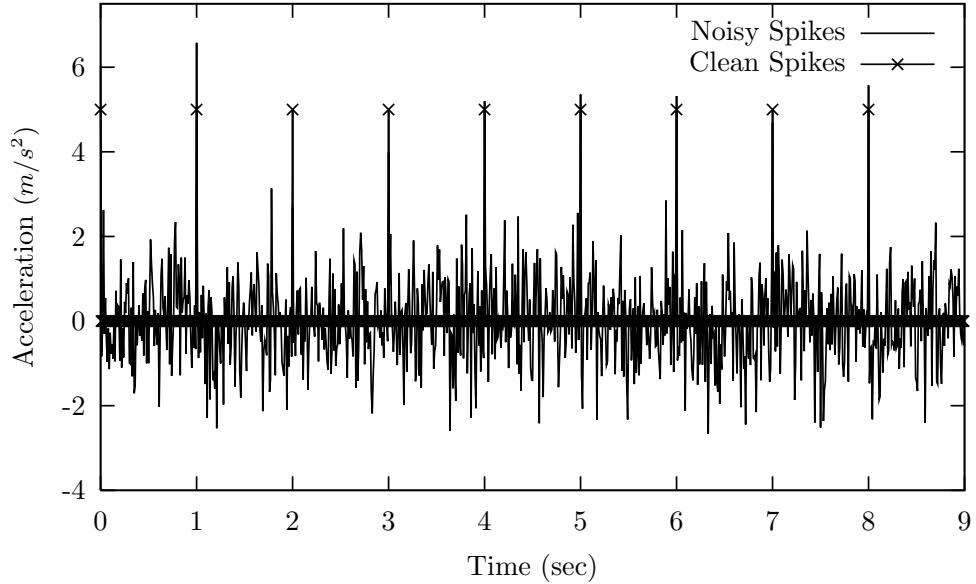


Figure 10.2 Artificial bearing fault data, with and without noise.

vibration and shocks when operating. As a result, sensor signals should ideally be comprised solely of Gaussian noise at low levels. Given that the nominal kurtosis value of a Gaussian distribution is 3¹ (Harris 1995), the *bias-adjusted kurtosis* value of

$$\hat{\kappa}_x = \frac{1}{T\sigma_x^4} \int_{-T/2}^{T/2} (x(t) - \hat{\mu}_x)^4 dt - 3, \quad \text{or} \quad \frac{1}{N\sigma_x^4} \sum_{i=1}^N (x_i - \hat{\mu}_x)^4 - 3 \quad (10.2)$$

is often used instead. Thus the baseline ideal system inducing no shock response should have a bias-adjusted kurtosis value of 0. However, confusion can occur when vibration of the system is at or near a resonance because the resulting bias-adjusted kurtosis value is -1.5 . A response near resonance will thus typically cause a bias-adjusted kurtosis value of between -1.5 and 0 due to the strong harmonic component at resonance combined with sensor noise. Other sources of harmonic response can also cause a negative shift in the bias-adjusted kurtosis. On the other hand, the sharp impulses caused by bearing or race defects impacting during rolling will cause bias-adjusted kurtosis values higher than

¹Also see problem 1 in chapter 2

0.0, often above 1.5. It is important to note that when both resonance and bearing faults occur the bias-adjusted kurtosis value can equal that of a healthy bearing (≈ 0). Further, as the bearing further degrades so that individual impulses are less common than continuous grinding the contribution of the bearing to the kurtosis value can disappear leading to a negative bias-adjusted kurtosis value. Thus it is important that kurtosis be used as an indicator when there is an understanding of the likely mode of failure and the system is not resonating. Zero or negative values of the bias-adjusted kurtosis not be taken as a guarantee of bearing health. Further, monitoring must be regular enough that failed bearings are not pulverized between a healthy reading and a subsequent reading. For the most important applications this may require constant monitoring. Kurtosis is best used as a simple indicator that something in the system has changed. It cannot assist in specifying what failure has occurred. Other techniques discussed in this chapter can be used to further identify what type of failure has occurred, including pulverization of a bearing.



Figure 10.3 Machine Fault Simulator. Photo courtesy SpectraQuest.[not yet!](#)

10.1.2 Time Domain/Ensemble Averaging

The natural reaction of the vibration analysis collecting machine vibration data is to convert the data to the frequency domain and commence averaging in order to reduce the impact of noise. While it is true that averaging in the frequency domain will smooth an auto-spectral density function and reduce variability due to noise, converting to the frequency domain removes one significant advantage that averaging in the time domain yields. A complex piece of machine equipment can have tens of elements each of which rotates at a unique rate. Further, events such as gear tooth meshing, bearing race point contact (the event of each bearing rolling over a given point in a race), and rotation of a blade on a disk

past a rub point each occur at unique frequencies. Machinery is typically a very noisy test environment, and frequencies corresponding to a multitude of faults can all appear in a single system. Further, when enough of these frequency components exist, the result may be the inability to uniquely identify the existence of any one of them individually.

A solution to this is *ensemble averaging*. Ensemble is the averaging of time histories across a number of individual records. The time records are recorded such that there start times correspond to an equivalent phase of the event of interest and through approximately the same phase of the event. Typically a record slightly longer than necessary is recorded so that cyclic behavior is easily observed.

For example, consider a shaft rotating at 3000 RPM² that is rubbing. A single revolution takes $\frac{60}{3000} = 0.02$ sec. In order to capture this event and obtain a time signature for it, data must be sampled for approximately 0.02 seconds per record. Ideally, a triggering event can be obtained using a tachometer or other device so that the start of each record will be as precisely as possible synchronized to the phase of the shaft. Figure ?? shows an optical probe being used to detect the passing of the keyway during each revolution. If such a trigger cannot be performed, then a long record can be broken into smaller records by using the sampling rate and shaft speed to calculate the periodicity of each single cycle data record. This is much more effective if the sampling rate of the data acquisition is an integer multiple of the speed of the shaft. Doing so ensures that each subsequent data cycle can start at a time corresponding to a data point, and not to a time between data points. The latter induces temporal smearing of the data and can result in loss of resolution in the data. A higher sampling rate, if high enough, can compensate sufficiently depending on the desired subsequent analysis needs.

The advantage of averaging in the time domain is that other events, such as gear meshing, bearing pass frequencies, etc. are often not synchronous with the data cycle. As a result these events appear to occur at random times within each data sample. Since they are not consistent in time, then averaging across time histories often effectively washes them out of the data so that the only remaining signal is the one of interest: the one synchronous with the trigger frequency.

Data collected via ensemble averaging can then be further processed using any of the techniques preceding or following this section, yielding more useful kurtosis information and auto spectrum densities containing only relevant frequency information.

10.2 Frequency Domain Analysis Methods

10.2.1 Auto Spectrum Density

The periodic impulsive signals such as bearing damage and damaged gear teeth can typically be observed better word better in the frequency domain. While the impulses in the artificial data of Figure 10.2 are easily observable, those in real data such as Figure ?? are not. Further, frequency domain analysis more readily identifies the frequencies of the impulse events, and further highlights other features of the events that are typically obscured by noise. Solution of problem 8 illustrates the frequency spectrum of an ideal repeated single-impulse event, while problem 9 illustrates rolling in and out of a crack in a race, or on and off a bearing flat. Figures ?? and ?? illustrate through simulated data what these can look like in more realistic data. Note that a frequency corresponding to the principle harmonic exists, along with higher harmonics due to the sharpness, or abruptness of the event. For the data with two repeated events, Frequencies of bearing responses.

²Rounds (or cycles) per minute.

Table 10.1 Response frequency multipliers for bearing faults. Ball bearing diameter, d , large diameter (between centers of ball bearings), D , number of ball bearing, N_b , and contact angle, θ . (Dimarogonas 1996)

Bearing Fault Type	Frequency Multiplier
Inner race fault	$\frac{N_b}{2} \left(1 + \frac{d}{D} \cos \theta\right)$
Outer race fault	$\frac{N_b}{2} \left(1 - \frac{d}{D} \cos \theta\right)$
Bearing fault	$\frac{D}{d} \left(1 - \left(\frac{d}{D}\right)^2 \cos^2 \theta\right)$
Cage (fundamental train)	$\frac{1}{2} \left(1 - \frac{d}{D} \cos \theta\right)$

10.2.2 Cepstrum

10.3 Joint Time Domain/Frequency Domain Analysis Methods

10.3.1 Spectral Maps (WaterFall Plots)

10.3.2 Campbell Diagram

10.3.3 Wavelets

10.4 Failure Identification

10.4.1 Unbalance

10.4.2 Rubbing

10.4.3 Misalignment

10.4.4 Bearing Faults

A

Decibels

The decibel is a measure of ‘power’ in a signal. Its origins are a desire to simplify the understanding and communication of power transmission, loss, and amplification in power supply networks. The decibel takes advantage of the property of \log_{10} such that

$$\log_{10}(xy) = \log_{10}(x) + \log_{10}(y) \quad (\text{A.1})$$

The definition of the \log_{10} of a signal is the Bell, directly or indirectly named after Alexander Graham Bell. The other unit (watts) is generally left off, but understood by context¹. Consider now a power supply of 11,000 watts (11 kW). Consider transmission through lines that result in a 90% power loss. This yields an available power at the other end of

$$11,000\text{kW} \times 0.90 = 9,900\text{kW} \quad (\text{A.2})$$

Consider this operation using Bells.

$$\begin{aligned} \log_{10}(11,000) + \log_{10}(0.90) &= \\ 4.04\text{Bells} - 0.05\text{Bells} &= 4.00\text{Bells} \end{aligned} \quad (\text{A.3})$$

One can easily see where if line losses were rated in Bells, then complex calculations of losses over a great number of lines is much more easily performed using the Bell. Now, note that for the example above, for a 10% power loss, one ends up working to two decimal places. The solution was to define the *decibel* as

$$S \text{ in dB} = 10 \times \log_{10}(S) \quad (\text{A.4})$$

Repeating our example, the available power is then

$$\begin{aligned} 10 \times \log_{10}(11,000) + 10 \times \log_{10}(0.90) &= \\ 40.4\text{Bells} - 0.4\text{dB} &= 40.0\text{dB} \end{aligned} \quad (\text{A.5})$$

Using only 1 decimal place results in an answer for this example with $10^{\frac{\pm 0.05}{10}} / 9,900 = 0.01\%$ error. This is due to the compaction effect of the log function. For example, 1 dB means a value of 1.259,

¹This is a poor practice in modern times because the use of the decibel (we’ll be getting there) is used with so many other units since its origination

where 100 dB means a value of 10,000,000,000. This compaction, however, doesn't destroy the ability to differentiate small values. For example, 0.001 W is -30 dB, while 0.002 W is -26.99 dB. In fact, a 50% difference between the power of two signals is always 3.0103 dB. For this reason, when considering auto-spectral densities of signals, or frequency response functions, the frequency at which the magnitude is 50% of the peak magnitude is commonly and interchangeably referred to as both the *half-power* point and -3 dB point. Referring to this value as the -3.0103 dB point only improves accuracy by 0.2%.

The definition of a decibel is often given in texts as $20 \times \log_{10}(X)$. This is correct if we are speaking about the magnitude of the Fourier transform (see equation (1.79)) of $x(t)$, or $\mathcal{F}(x(t)) = X(j\omega)$. The frequency response function of a system is defined in equation (3.3) as

$$H(j\omega) = \frac{X(j\omega)}{F(j\omega)} \quad (\text{A.6})$$

Rearranging gives

$$H(j\omega)F(j\omega) = X(j\omega) \quad (\text{A.7})$$

Multiplying equation (A.7) by its conjugate, and dividing by T gives

$$\begin{aligned} \bar{H}(j\omega)H(j\omega) \frac{1}{T} \bar{F}(j\omega)F(j\omega) &= \frac{1}{T} \bar{X}(j\omega)X(j\omega) \\ \bar{H}(j\omega)H(j\omega)S_{ff}(j\omega) &= S_{xx}(j\omega) \end{aligned} \quad (\text{A.8})$$

where $S_{ff}(j\omega)$ and $S_{xx}(j\omega)$ are defined by equations (3.15c) and (3.15b) respectively. In other applications, terms like this are referred to as *power spectral densities*, but the units are not those of power, so we will avoid that term. Nevertheless, writing equation (A.7) in terms of decibels,

$$10 \log_{10} (\bar{H}(j\omega)H(j\omega)) + 10 \log_{10} (S_{ff}(j\omega)) = 10 \log_{10} (S_{xx}(j\omega)) \quad (\text{A.9})$$

Noting that

$$\bar{H}(j\omega)H(j\omega) = |H(j\omega)|^2 \quad (\text{A.10})$$

we can rewrite equation (A.9) as

$$20 \log_{10} (|H(j\omega)|) + 10 \log_{10} (S_{ff}(j\omega)) = 10 \log_{10} (S_{xx}(j\omega)) \quad (\text{A.11})$$

Thus when describing the magnitude of a frequency response function, converting into dB is performed by applying

$$H(j\omega)\text{dB} = 20 \log_{10} (|H(j\omega)|). \quad (\text{A.12})$$

B

Linear Algebra

The following section is intended as a refresher and reference in linear algebra. Some topics, such as advanced eigenvalue issues, are not expected to be understood *a priori*. It is intended to show what is necessary to understand the mathematics being performed in this text without addressing the numerical issues of practical linear algebra computations. If the following sections are inadequate for reader interest, the reader should consult linear algebra texts by Horn and Johnson (1999) and Golub and Van Loan (1985). If this section moves too quickly for the reader, an introductory text such as that by Bellman (1960) will be more helpful.

B.1 Basics of Matrices and Vectors

For our purposes, we can consider matrices and vectors to be simply organized lists of numbers. The matrix A is represented by

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,M} \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots & a_{2,M} \\ a_{3,1} & a_{3,2} & a_{3,3} & \dots & a_{3,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N,1} & a_{N,2} & a_{N,3} & \dots & a_{N,M} \end{bmatrix} \quad (\text{B.1})$$

and its elements are represented by $a_{\ell,m}$. The size of the matrix is given as $N \times M$ where N is the number of rows and M is the number of columns. When referring to rows and columns of a matrix, or a specific element of a matrix, the row is always given before the column. For example, consider the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad (\text{B.2})$$

The matrix is 3×2 (3 rows and 2 columns) and $a_{3,1} = 5$.

A matrix is considered to be symmetric if its *transpose* is equal to itself. The transpose of a matrix is obtained by swapping corresponding rows and columns, or *flipping* the matrix about its diagonal. For example, consider taking the transpose of A as defined in equation (B.2)

$$A^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \quad (\text{B.3})$$

Clearly A is not symmetric because it is not equal to its transpose. Thus, a precondition for symmetry is that the matrix be *square*, having as many rows as columns.

The special case of a matrix with one column is a *vector*. A matrix with only a single row is called a *row vector*.

B.2 Multiplication of Vectors and Matrices

Consider the multiplication of a *matrix* A and a vector \mathbf{x} such that

$$\underbrace{A}_{N \times M} \underbrace{\mathbf{x}}_{M \times 1} = \underbrace{\mathbf{y}}_{N \times 1} \quad (\text{B.4})$$

where the respective sizes are shown in the underbraces. Note that in multiplication in linear algebra the sizes must always be consistent. The number of columns in the matrix A and the number of rows in the vector \mathbf{x} must be equal. This is because the rows of A are each in turn multiplied by the column that is \mathbf{x} . The result of the multiplication has a size corresponding to the number of rows of A and the number of columns of \mathbf{x} .

For example, using the matrix A of equation (B.2) and a vector \mathbf{x} defined by

$$\mathbf{x} = \begin{bmatrix} 7 \\ 8 \end{bmatrix} \quad (\text{B.5})$$

then

$$A\mathbf{x} = \underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}_{3 \times 2} \underbrace{\begin{bmatrix} 7 \\ 8 \end{bmatrix}}_{2 \times 1} = \underbrace{\begin{bmatrix} 1 \times 7 + 2 \times 8 \\ 3 \times 7 + 4 \times 8 \\ 5 \times 7 + 6 \times 8 \end{bmatrix}}_{3 \times 1} = \underbrace{\begin{bmatrix} 23 \\ 53 \\ 83 \end{bmatrix}}_{3 \times 1} = \underbrace{\mathbf{y}}_{3 \times 1} \quad (\text{B.6})$$

B.3 Multiplication of Matrices with Matrices

Multiplication of two matrices is a simple extension of multiplying a matrix and a vector.¹ Consider now multiplying the matrix A by a matrix B where

$$B = \begin{bmatrix} 7 & 9 \\ 8 & 10 \end{bmatrix} \quad (\text{B.7})$$

The matrix B can be considered to be two vectors, appended side by side, for the sake of multiplication. Note that the first column of B is the vector \mathbf{x} from the preceding section. The result is the same as the respective matrix/vector multiplications appended side by side, i.e.

$$AB = \underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}_{3 \times 2} \underbrace{\begin{bmatrix} 7 & 9 \\ 8 & 10 \end{bmatrix}}_{2 \times 2} = \underbrace{\begin{bmatrix} 1 \times 7 + 2 \times 8 & 1 \times 9 + 2 \times 10 \\ 3 \times 7 + 4 \times 8 & 3 \times 9 + 4 \times 10 \\ 5 \times 7 + 6 \times 8 & 5 \times 9 + 6 \times 10 \end{bmatrix}}_{3 \times 2} = \underbrace{\begin{bmatrix} 23 & 29 \\ 53 & 67 \\ 83 & 105 \end{bmatrix}}_{3 \times 2} \quad (\text{B.8})$$

The first column of the result in equation (B.8) can be recognized as the result given in equation (B.6). This is useful when considering multiplying matrices that are assembled from block matrices.

Further, observation of equation (B.8) illustrates the often used identity

$$(AB)^T = B^T A^T \quad (\text{B.9})$$

¹Actually, multiplying a matrix with a vector is a special case of multiplying two matrices.

which can be applied repeatedly to show that $(ABC)^T = C^T B^T A^T$, etc.

The *inverse!* matrix of a square matrix is defined as that matrix that when multiplied by the original matrix yields the identity matrix. For example, consider a matrix A . The matrix B is the inverse of A if

$$AB = I \quad (\text{B.10})$$

The inverse of a matrix is denoted by a superscript of -1 such that

$$B = A^{-1} \quad (\text{B.11})$$

An example method of obtaining the inverse of a matrix is given by equation (B.43). However, the inverse of a matrix is often not the desired result of a calculation, but instead an intermediate result. In these cases, using Gauss elimination if a much more prudent method Golub and Van Loan (1985); Horn and Johnson (1999); Meirovitch (1997). Programs such as MATLAB and Octave use the front divide by or back divide by to indicate use of Gauss elimination. For example, $A^{-1}B$ is coded as A while AB^{-1} is coded as A/B . Gauss elimination is not only faster than first performing the inverse, then the multiplication, but also more accurate. As illustrated by equation (B.43) in section B.8, a singular matrix cannot be inverted. Non-square matrices may be inverted using the pseudo-inverse as described in section B.10.

B.4 Multiplication of Vectors

Special cases of matrix multiplications are the *vector inner product*, also known as the *dot product*, and the *vector outer product*.

The vector inner product of a vector \mathbf{x} and a \mathbf{y} is defined as

$$\mathbf{x}^T \mathbf{y} = \sum_{i=1}^N x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_N y_N \quad (\text{B.12})$$

and is only permissible if \mathbf{x} and \mathbf{y} are $N \times 1$ and is a simple application of matrix multiplication.

The vector outer product is defined as

$$\mathbf{x}\mathbf{y}^T = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_N \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_N \\ \vdots & \vdots & \ddots & \vdots \\ x_N y_1 & x_N y_2 & \cdots & x_N y_N \end{bmatrix} \quad (\text{B.13})$$

Note that while the vector inner product is symmetric simply because the result is a scalar, the vector outer product is *not symmetric* unless $\mathbf{x} = \mathbf{y}$.

B.5 Linear Independence and Orthogonality of Vectors

A set of vectors

$$\mathbf{x}_i, \quad i = 1, 2, 3, \dots, n$$

in a linear space \mathcal{C}^{n^2} are linearly independent iff

$$\sum_{i=1}^n a_i \mathbf{x}_i = \mathbf{0}$$

cannot be satisfied without all $a_i = 0$ $i = 1, \dots, n$

²Complex Vector Space

B.5.1 Example

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad (B.14)$$

$$a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = \mathbf{0} \quad (B.15)$$

$$a_1 1 + a_2 0 = 0 \quad a_2 = 0 \quad (B.16a)$$

$$a_1 2 + a_2 2 = 0 \quad a_1 = 0 \quad (B.16b)$$

$\therefore \mathbf{x}_1$ and \mathbf{x}_2 are independent

B.5.2 Example

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \quad (B.17)$$

$$a_1 = 1, a_2 = 1, a_3 = -1$$

satisfies

$$a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 = \mathbf{0}$$

Thus \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 are not independent.

A more formal method of evaluating independence of vectors is through use of singular value decomposition as discussed in section B.9. For this problem, the assembled matrix $[\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3]$ has a single eigenvalue equal to zero indicating that only two of the vectors are independent. Thus, the matrix $[\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3]$ has a *rank* of 2. It is left to the reader to verify for this example that taking the transpose of the matrix does not effect the rank. Such a matrix is termed *rank deficient* because its rank (2) is less than its size (3).

B.6 Multiplication of Block Matrices

Matrices used in system identification are often assembled from smaller matrices. Such matrices are called *block matrices*. In light of this, consider the preceding example again. However, here we will define submatrices such that

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \quad (B.18)$$

and

$$B = \begin{bmatrix} 7 & 9 \\ 8 & 10 \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2] \quad (B.19)$$

Note for the multiplications to be feasible in block form, there must be matrix size consistency. e.g. the number of columns in the i th block column of A must be equal to the number of rows in the i th block row of B . The resulting multiplication of A and B can now be recognized as

$$AB = \begin{bmatrix} A_1 \mathbf{x}_1 & A_1 \mathbf{x}_2 \\ A_2 \mathbf{x}_1 & A_2 \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 23 \\ 53 \\ 83 \end{bmatrix} & \begin{bmatrix} 29 \\ 67 \\ 105 \end{bmatrix} \end{bmatrix} \quad (B.20)$$

Thus block matrices can be multiplied with the same rules as when considered in scalar form. In fact, non-block multiplications can be thought of as special cases of block matrices where the submatrices are all 1×1 .

B.7 Derivatives of Products of Matrices and Vectors

When writing functions of matrices and vectors it is not uncommon to need to take a derivative with respect to one of the vectors, often as part of a minimization problem, or in other cases to obtain a gradient. First consider

$$\mathbf{y} = A\mathbf{x} \quad (\text{B.21})$$

where

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \quad (\text{B.22})$$

and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (\text{B.23})$$

Then

$$\mathbf{y} = \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 \\ a_{2,1}x_1 + a_{2,2}x_2 \end{bmatrix} \quad (\text{B.24})$$

The derivative $\frac{d\mathbf{y}}{d\mathbf{x}}$, with both \mathbf{y} and \mathbf{x} having a length of 2, is defined as

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \begin{bmatrix} \frac{dy_1}{dx_1} & \frac{dy_1}{dx_2} \\ \frac{dy_2}{dx_1} & \frac{dy_2}{dx_2} \end{bmatrix} \quad (\text{B.25})$$

and is called the *Jacobian*. Observing equation (B.24),

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \frac{d(A\mathbf{x})}{d\mathbf{x}} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = A \quad (\text{B.26})$$

Consider next the common situation of a scalar y defined as

$$y = \frac{1}{2}\mathbf{x}^T A\mathbf{x} \quad (\text{B.27})$$

where A and \mathbf{x} are as defined above. Then

$$y = \frac{1}{2}(a_{1,1}x_1^2 + a_{1,2}x_1x_2 + a_{2,1}x_1x_2 + a_{2,2}x_2^2) \quad (\text{B.28})$$

The derivative $\frac{dy}{d\mathbf{x}}$ is defined as

$$\frac{dy}{d\mathbf{x}} = \begin{bmatrix} \frac{dy}{dx_1} & \frac{dy}{dx_2} \end{bmatrix} = \begin{bmatrix} a_{1,1}x_1 + \frac{1}{2}a_{1,2}x_2 + \frac{1}{2}a_{2,1}x_2 & \frac{1}{2}a_{1,2}x_1 + \frac{1}{2}a_{2,1}x_1 + a_{2,2}x_2 \end{bmatrix} \quad (\text{B.29})$$

If A is symmetric then $a_{1,2} = a_{2,1}$ and thus

$$\begin{aligned} \frac{dy}{d\mathbf{x}} &= \begin{bmatrix} a_{1,1}x_1 + \frac{1}{2}a_{1,2}x_2 + \frac{1}{2}a_{1,2}x_2 & \frac{1}{2}a_{1,2}x_1 + \frac{1}{2}a_{1,2}x_1 + a_{2,2}x_2 \end{bmatrix} \\ &= \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 & a_{1,2}x_1 + a_{2,2}x_2 \end{bmatrix} = \mathbf{x}^T A \end{aligned} \quad (\text{B.30})$$

If A is non-symmetric then

$$\frac{dy}{d\mathbf{x}} = \frac{1}{2}(\mathbf{x}^T A + \mathbf{x}^T A^T) \quad (\text{B.31})$$

B.8 The Eigenvalue Problem

The eigenvalue problem is obtained in a very wide array of mathematical problems.³ In vibration testing, it results from modal analysis and in calculation of the H_v frequency response function.

The general eigenvalue is one of finding λ and \mathbf{x} when

$$A\mathbf{x} = \lambda B\mathbf{x} \quad (\text{B.32})$$

and A and B are square $N \times N$ matrices. There are n linearly independent solutions \mathbf{x} called the *eigenvectors*. Each eigenvector has a corresponding *eigenvalue* λ . The eigenvalues λ are *usually* unique, but often are not. The more general form of the eigenvalue problem is

$$AX = BX\Lambda \quad (\text{B.33})$$

where $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, and Λ is a diagonal matrix of the eigenvalues corresponding to the eigenvectors assembled to form X .

Under special circumstances such as equation (1.160), the matrix B is the identity matrix and equation (B.33) becomes

$$AX = X\Lambda \quad (\text{B.34})$$

which is more familiar to most. Eigen-solvers are more often capable of handling the eigenvalue problem in the form of equation (B.34) than the form of equation (B.33). Many eigensolvers that can solve the eigenvalue problem form of equation (B.33) do so by transformation to the form of equation (B.34). This case can be obtained by pre-multiplying (B.33) by B^{-1} . For large matrices, Gauss elimination should be used to obtain $B^{-1}A$. Better yet, Cholesky decomposition (Meirovitch 1997) should be applied to B such that

$$B = B_c^T B_c \quad (\text{B.35})$$

where B_c is an upper triangular matrix⁴. The substitution

$$Y = B_c X \quad (\text{B.36})$$

can be made in equation (B.33), and then pre-multiplying by B_c^{T-1} gives

$$\left(B_c^{T-1} A B_c^{-1} \right) Y = Y\Lambda \quad (\text{B.37})$$

which is an equivalent form to equation (B.34). Once the solution is obtained, the eigenvectors X can be obtained from $X = B_c^{-1}Y$.

This methodology has significant advantages over simply inverting B . One is that it is more numerically robust, resulting in reduced propagation of numerical errors. A second is that when A and B are *Hermitian*⁵, the eigenvectors are orthogonal and the eigenvalues are real. These two pieces of information can be used by the algorithm programmer to improve computational efficiency, and can be used by the user to verify the quality of the results. These issues are discussed extensively in section 1.9.

Equation (B.34) is often written as a matrix decomposition form of

$$X^{-1}AX = \Lambda \quad (\text{B.38})$$

The rows of X^{-1} are referred to as the *left eigenvectors* because transposing equation (B.38) results in the alternate eigenvalue problem

$$X^T A^T X^{-1T} = \Lambda \quad (\text{B.39})$$

³The eigenvalue problem is also used for calculating principle stresses and strains, as well as principle axes and moments of inertia of rigid bodies

⁴All values below the diagonal are zero

⁵A Hermitian Matrix, A , is one where its conjugate transpose, represented by (\cdot^H) , is equal to itself.

For the sake of substitution and solution of linear algebra problems, a matrix is often decomposed using equation (??) such that

$$A = X\Lambda X^{-1} \quad (\text{B.40})$$

Further, if A is Hermitian, $X^{-1} = X^T$, and then

$$A = X\Lambda X^T \quad (\text{B.41})$$

The eigenvalue problem can thus be used to identify whether or not a matrix is *non-singular* or invertible.

Another useful form of (B.41) is to break it into a summation of *outer products* of the eigenvectors weighted by the corresponding eigenvalues. Accordingly, we can write

$$\begin{aligned} A &= X\Lambda X^T \\ &= [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} \\ &= \sum_{i=1}^n \mathbf{x}_i \lambda_i \mathbf{x}_i^T \end{aligned} \quad (\text{B.42})$$

If a matrix is invertible, it cannot have any eigenvalues equal to zero because

$$A^{-1} = (X^T)^{-1} \Lambda^{-1} X^{-1} \quad (\text{B.43})$$

Thus, since Λ is diagonal, and its inverse is

$$\Lambda^{-1} = \text{diag} [\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}] \quad (\text{B.44})$$

any zero eigenvalue will cause a singularity precluding an inverse.

B.9 Singular Value Decomposition - Principle Component Analysis

Singular value decomposition, or SVD, is an extension of the eigenvalue decomposition to non-square matrices. It is often used to identify, consolidate, and rank contributions of vectors to a large non-square matrix. The SVD of an $m \times n$ matrix A , where $m \geq n$ is defined by

$$\underbrace{A}_{m \times n} = \underbrace{U}_{m \times n} \underbrace{\Sigma}_{n \times n} \underbrace{V^T}_{n \times n} \quad (\text{B.45})$$

where

$$U^T U = I \quad (\text{B.46})$$

$$V V^T = I \quad (\text{B.47})$$

and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n \end{bmatrix} \quad (\text{B.48})$$

Some algorithms will alternatively return results satisfying

$$\underbrace{A}_{m \times n} = \underbrace{U}_{m \times m} \underbrace{\Sigma}_{m \times n} \underbrace{V^T}_{n \times n} \quad (B.49)$$

by appending zero rows to the matrix Σ and appending additional mutually orthogonal vectors to the matrix U . Since the nonzero appended columns of U correspond to zero-valued rows of Σ they have no impact on the matrix A when computed.

Just as in the eigenvalue problem, practical solution methodologies are extensive and sophisticated and will not be addressed here. The interested reader is referred to Golub and Van Loan (1985).

For illustration, consider obtaining the SVD of the matrix A defined by

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1.1 \end{bmatrix} \quad (B.50)$$

The problem of finding the singular values can quickly transformed into one of finding eigenvalues. Pre-multiplying (B.45) by its transpose yields

$$A^T A = V \Sigma U^T U \Sigma V^T \quad (B.51)$$

Applying equation (B.46), this simplifies to

$$A^T A = V \Sigma^2 V^T \quad (B.52)$$

Comparing this to equation (B.40) one should recognize the Σ^2 are the eigenvalues of the matrix $A^T A$. Further, the eigenvalue problem is a symmetric one because $A^T A$ is symmetric, which is easily proven by taking its transpose (likewise this can be done with the right size of equation (B.52)). The solution of the eigenvalue problem yields

$$\Sigma = \begin{bmatrix} 2.9 & 0.00 \\ 0.00 & 0.06 \end{bmatrix} \quad (B.53)$$

and

$$V^T = \begin{bmatrix} 0.70 & 0.72 \\ 0.72 & -0.70 \end{bmatrix} \quad (B.54)$$

The matrix U can then be found using (B.45) to be

$$U = A V \Sigma^{-1} = \begin{bmatrix} 0.49 & 0.30 \\ 0.49 & 0.30 \\ 0.49 & 0.30 \\ 0.52 & -0.85 \end{bmatrix} \quad (B.55)$$

where we have taken advantage of the fact that $V^T = V^{-1}$.

The numerical results of this example illustrate the application and behavior of the SVD analysis. Observing the original matrix A , ones first impression is that it is almost entirely constructed of values 1. In fact, only one value even slightly deviates from 1. This is illustrated in the resulting matrix U . The matrix U contains the *principle components*, or vectors, in the matrix A . The first column of A is a vector with nearly the same direction as a vector of ones but having been normalized. Notice then that the second column primarily shows a deviation of the fourth value. This is due to that small contribution of adding 0.1 to $A_{4,2}$. Importantly, Σ illustrates the degree of contribution of each of these vectors to the matrix and is almost always ordered from the highest to lowest value. Note that the first singular value is much greater than the second, indicating that the nearly constant vector is a much greater contributor to A . As can be observed in A , the second column of U has a relatively small contribution to A ,

and the second singular value highlights this. Further, $\sigma_i \geq 0$. The matrix V can then be recognized as an organizer. It determines how much of each σ -weighted column of U is needed to produce the corresponding column of A .

In practice, for very large matrices with redundant data, the higher-index values of σ tend towards 0, and for all practical purposes can be treated as zero. When this happens, the SVD problem is written as

$$\underbrace{A}_{n \times m} = \underbrace{U}_{m \times n} \underbrace{\begin{bmatrix} \Sigma_p & [0] \\ [0] & [0] \end{bmatrix}}_{p \times p} \underbrace{V^T}_{n \times n} \quad (B.56)$$

where the near-zero values of Σ have been set to zero leaving only p non-zero values. As a result, columns $p+1$ through n of U and rows $p+1$ through n of V^T can be discarded along with all of Σ outside of Σ_p . The resulting SVD problem statement then appears as

$$\underbrace{A}_{n \times m} = \underbrace{U}_{m \times p} \underbrace{\Sigma_p}_{p \times p} \underbrace{V^T}_{p \times n} \quad (B.57)$$

where $m \geq n \geq p$.

B.10 Pseudo-Inverse

The pseudo-inverse of a matrix is analogous to the inverse of a square matrix, with the exception that its used when a matrix is non-square. Consider obtaining the matrix B where the known matrix C is given by

$$\underbrace{A}_{m \times n} \underbrace{B}_{n \times q} = \underbrace{C}_{m \times q} \quad (B.58)$$

and none of the matrices are necessarily symmetric. Since A is not square, we cannot use a simple inverse. However, we can perform SVD using equation (B.57) on the matrix A giving

$$U \Sigma_p V^T B = C \quad (B.59)$$

Pre-multiplying both side by $V \Sigma_p^{-1} U^T$ gives

$$V \Sigma_p^{-1} U^T U \Sigma_p V^T B = \underbrace{I}_{n \times n} B = V \Sigma_p^{-1} U^T C \quad (B.60)$$

thus

$$B = V \Sigma_p^{-1} U^T C \quad (B.61)$$

Provided the SVD of A can be obtained, the only significant challenge that we could potentially face is inverting Σ_p . This is resolved by making sure to truncate those singular values below the point at which numerical errors become significant in order to improve the *condition number* of the inverse defined as

$$\text{Condition Number} = \frac{\sigma_p}{\sigma_1} \quad (B.62)$$

The lower the condition number, the less error prone the pseudo-inverse is. The minimum value that the condition number can have is 1 for an orthonormal matrix (such as V or equation (B.45)) and the maximum is unbounded (Crassidis and Junkins 2004; Golub and Van Loan 1985; Horn and Johnson 1999).

The pseudo-inverse of the matrix A is then defined as

$$A^\dagger = V \Sigma_p^{-1} U^T \quad (B.63)$$

where \cdot^\dagger designates a pseudo-inverse.

B.11 Linear Least Squares: Curve Fitting Polynomials

Consider two measured variables presumed to be related by

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots + a_{n_p} x^{n_p} \quad (\text{B.64})$$

where the coefficients a_i are desired. Further, presume data has taken such that for values of x_i , we have measured values of y, \hat{y}_i , instead of the true values. The values x_i are presumed to be perfectly correct. For each of the values x_i the presumed correct value of y_i is

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 + \dots + a_{n_p} x_i^{n_p} \quad (\text{B.65})$$

Combining these into matrix form gives

$$\mathbf{y} = \mathbf{X}\mathbf{a} \quad (\text{B.66})$$

where

$$\mathbf{X} = \begin{bmatrix} x_1^0 & x_1^1 & x_1^2 & x_1^3 & \cdots & x_1^{n_p} \\ x_2^0 & x_2^1 & x_2^2 & x_2^3 & \cdots & x_2^{n_p} \\ x_3^0 & x_3^1 & x_3^2 & x_3^3 & \cdots & x_3^{n_p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_N^0 & x_N^1 & x_N^2 & x_N^3 & \cdots & x_N^{n_p} \end{bmatrix} \quad (\text{B.67})$$

where \mathbf{X} is called the *Vandermonde matrix* in x where N must be greater than n_p . The squared error of the polynomial representation of the data is then

$$e^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \quad (\text{B.68})$$

The optimal values for \mathbf{a} are those that minimize this error, so taking the derivative of e^2 with respect to the coefficients, \mathbf{a} gives

$$\begin{aligned} \frac{de^2}{d\mathbf{a}} &= \frac{d}{d\mathbf{a}} \left((\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \right) \\ &= \frac{d}{d\mathbf{a}} \left(\mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} - 2\mathbf{a}^T \mathbf{X}^T \hat{\mathbf{y}} + \hat{\mathbf{y}}^T \hat{\mathbf{y}} \right) \\ &= 2\mathbf{X}^T \mathbf{X} \mathbf{a} - 2\mathbf{X}^T \hat{\mathbf{y}} = \mathbf{0} \end{aligned} \quad (\text{B.69})$$

thus

$$\mathbf{X}^T \mathbf{X} \mathbf{a} = \mathbf{X}^T \hat{\mathbf{y}} \quad (\text{B.70})$$

and the coefficients \mathbf{a} can be obtained by pre-multiplying both sides by $(\mathbf{X}^T \mathbf{X})^{-1}$ giving

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{y}} = \mathbf{X}^\dagger \hat{\mathbf{y}} \quad (\text{B.71})$$

where \cdot^\dagger means pseudo-inverse. Note that this procedure doesn't work if $\mathbf{X}^T \mathbf{X}$ is singular. This happens when either $N < n_p$ or one or more of the values x_i is equal to another value x_j . If some of the values x_i are relatively closely spaced, then $\mathbf{X}^T \mathbf{X}$ is near singular and numerical issues can make the results inaccurate. Section B.10 illustrates a more robust method for obtaining the pseudo-inverse, defined by equation (B.63), using singular value decomposition. Further, when n_p becomes large the solution can become very noise sensitive. In this case using Lagrange Polynomials can often be helpful (Harris 1995; Horn and Johnson 1999; Maia and E Silva 1998).

B.12 The Cayley-Hamilton Theorem

The Cayley-Hamilton theory states that a square matrix must satisfy its own characteristic equation. That is, for the $2n \times 2n$ matrix A , if we denote

$$\det(A - \lambda I) = 0 = \lambda^{2n} + a_1 \lambda^{2n-1} + a_2 \lambda^{2n-2} + \dots + a_{2n} \quad (B.72)$$

then

$$0 = A^{2n} + a_1 A^{2n-1} + a_2 A^{2n-2} + \dots + a_{2n} A^0 \quad (B.73)$$

This can be proven by diagonalizing A such that

$$A = \Phi \Lambda \Phi^{-1} \quad (B.74)$$

Substituting for A in equation (B.73) gives

$$\begin{aligned} 0 &= (\Phi \Lambda \Phi^{-1})^{2n} + a_1 (\Phi \Lambda \Phi^{-1})^{2n-1} + a_2 (\Phi \Lambda \Phi^{-1})^{2n-2} + \dots + a_{2n} (I) \\ &= \Phi \Lambda^{2n} \Phi^{-1} + a_1 \Phi \Lambda^{2n-1} \Phi^{-1} + a_2 \Phi \Lambda^{2n-2} \Phi^{-1} + \dots + a_{2n} \Phi \Lambda^0 \Phi^{-1} \end{aligned} \quad (B.75)$$

Premultiplying equation (B.75) by Φ^{-1} and post multiplying by Φ gives

$$0 = \Lambda^{2n} + a_1 \Lambda^{2n-1} + a_2 \Lambda^{2n-2} + \dots + a_{2n} I \quad (B.76)$$

which is just equation (B.72) written in matrix form for each of the individual eigenvalues, λ_i . Therefore equation (B.73) is true.

A further development of the Cayley-Hamilton theorem is that any matrix polynomial function can be represented as a polynomial of order $2n - 1$ or less. This results from equation (B.73). Solving for A^{2n}

$$A^{2n} = -a_1 A^{2n-1} - a_2 A^{2n-2} - \dots - a_{2n} A^0 \quad (B.77)$$

Thus, the term A^{2n} in a matrix polynomial equation can always be replaced by a polynomial of terms with power of $2n - 1$ and less. Higher powers of A can be obtained by multiplying equation (B.77) by A and successively substituting in lower power of A solutions.

Thus, any matrix function, $f(At)$, that can be represented as a power series can also be represented as

$$\begin{aligned} f(At) &= \sum_{i=0}^{\infty} \frac{1}{i!} \frac{d^i}{d\tau^i} f(\tau) \Big|_{\tau=t} t^i A^i \\ &= \sum_{i=0}^{2n-1} g_i(t) A^i \end{aligned} \quad (B.78)$$

where the functions $g_i(t)$ are obtained by algebraically combining the functions $\frac{1}{i!} \frac{d^i}{d\tau^i} f(\tau) \Big|_{\tau=t}$ and the a_i coefficients to the powers of A after application of equation (B.77) and its extension.

C

Variable Definitions

The conventions used for variable forms vary throughout the literature. Where they are consistent, the breadth of this book does not allow consistency with all fields. The following is the best attempt of the author to meet community standards, while still providing consistency throughout the text.

1. lowercase variables are scalars.
2. uppercase variables are matrices, *or* the Fourier/Laplace/Z-transforms of scalars, or the complex magnitudes of steady-state time histories. Context must be utilized to make a determination.
3. Calligraphic variables are always matrices. Lowercase calligraphic variables are time domain variables.
4. **bold** variables are vectors.
5. **bold** uppercase variables are vectors of transformed variables.
6. $\bar{\cdot}$ means complex conjugate.
7. dummy variables are clearly stated, and may have a defined meaning elsewhere in the text, but not in the context where they are used.
8. Functional dependence is necessary in some cases to explicitly determine the meaning of a variable.
9. \cdot^H means conjugate-transpose ($\bar{\cdot}^T$).
10. $\hat{\cdot}$ means measured quantity (with noise/error).
11. \cdot^\dagger means pseudo-inverse.
12. $[\cdot]$ is used occasionally used to emphasize that a variable is a matrix.

A	continuous time state matrix	B	control matrix	C	<i>second order systems:</i> damping matrix, <i>first order systems</i> measurement matrix, by context
${}_i A$	Modal constant/residue matrix of i^{th} mode	\tilde{B}	control matrix for second order system		
A_d	discrete time state matrix	B_d	discrete time control matrix	C_a	acceleration measurement matrix

C_d	displacement measurement matrix	K''	imaginary part of stiffness matrix	o	number of inputs (length of \mathbf{u})
C_v	velocity measurement matrix	K^*	complex stiffness matrix	ω	driving frequency
D	pass-through matrix	l	counter variable or arbitrary integer	$\omega_{(\cdot)}$	enumerated natural frequency
$\delta(\cdot)$	Dirac delta function			ω_d	damped natural frequency
$\delta_{(\cdot)}$	Kronecker delta function	L	Modal Participation Factors	ω_n	natural frequency
Δ	small discrete change	\mathcal{L}	Laplace transform	Ω	diagonal matrix of natural frequencies
e	$\ln(e) = 1$	λ	pole of system, root of characteristic equation	p	number of outputs (length of \mathbf{y})
η	loss factor				
f	applied force	Λ	<i>second order systems:</i> diagonal matrix of ω_n^2 , <i>first-order systems:</i> diagonal matrix of eigenvalues	P	Arbitrary state transformation matrix
\mathcal{F}	Fourier transform			ϕ	state-space mode shape
G	gyroscopic matrix			Φ	state-space mode shape matrix
h	impulse response function/frequency response function	m	scalar mass, counter variable, or arbitrary integer	ψ	physical-space mode shape
H	circulent matrix	M	mass matrix		
\mathcal{H}	Hankel matrix	\mathcal{M}	Markov parameters	Ψ	physical-space mode shape matrix
$H(j\omega)$	second order transfer function matrix	n	number of degrees of freedom or arbitrary integer	q	second-order orthogonal coordinate system
$H_{m,l}(\cdot)$	second order transfer function between input l and output m	n_m	number of modes	r	second-order modal coordinate vector
i	counter	n_p	number of terms in polynomial expansion	s	state-space modal coordinate vector
I	identity matrix			Σ	singular values
j	$\sqrt{-1}$	n_r	number of retained degrees of freedom in reduced order model, number of records for data	t	time
k	time or frequency index			T	period
K	stiffness matrix	n_s	number of spectral lines	$T(j\omega)$	transfer function matrix (state space)
K'	real part of stiffness matrix	N	Number of data points	v	complex magnitude of $q(t)$

\mathcal{V}	Augmented Vandermonde matrix	$\ddot{\mathbf{x}}$	acceleration vector	\mathbf{z}	$= [\mathbf{x}^T \quad \dot{\mathbf{x}}^T]^T$, state vector
\mathbf{x}	displacement vector	$\xi_{\mathbf{x}}(t)$	noise in the measured signal, $\hat{\mathbf{x}}$	ζ	damping ratio
$\dot{\mathbf{x}}$	velocity vector	\mathbf{y}	output vector		

D

Tables

D.1 Normal (Gauss) Distribution

The following table shows the positive side of the normal distribution PDF, and corresponding CDF. Here z is the normalized variable $z = \frac{x-\mu_x}{\sigma_x}$. Note that the PDF is symmetrical about $z = 0$, and that values of the CDF can be obtained using $\bar{P}(z) = 1 - P(-z)$ for $z < 0$.

z	PDF, $p_z(z)$	CDF, $P_z(z)$
0.00	0.398942280	0.500000000
0.05	0.398443914	0.519938806
0.10	0.396952547	0.539827837
0.15	0.394479331	0.559617692
0.20	0.391042694	0.579259709
0.25	0.386668117	0.598706326
0.30	0.381387815	0.617911422
0.35	0.375240347	0.636830651
0.40	0.368270140	0.655421742
0.45	0.360526962	0.673644780
0.50	0.352065327	0.691462461
0.55	0.342943855	0.708840313
0.60	0.333224603	0.725746882
0.65	0.322972360	0.742153889
0.70	0.312253933	0.758036348
0.75	0.301137432	0.773372648
0.80	0.289691553	0.788144601
0.85	0.277984886	0.802337457
0.90	0.266085250	0.815939875
0.95	0.254059056	0.828943874
1.00	0.241970725	0.841344746
1.05	0.229882141	0.853140944
1.10	0.217852177	0.864333939

Continued on next page

z	PDF, $p_z(z)$	CDF, $P_z(z)$
1.15	0.205936269	0.874928064
1.20	0.194186055	0.884930330
1.25	0.182649085	0.894350226
1.30	0.171368592	0.903199515
1.35	0.160383327	0.911492009
1.40	0.149727466	0.919243341
1.45	0.139430566	0.926470740
1.50	0.129517596	0.933192799
1.55	0.120009001	0.939429242
1.60	0.110920835	0.945200708
1.65	0.102264925	0.950528532
1.70	0.094049077	0.955434537
1.75	0.086277319	0.959940843
1.80	0.078950158	0.964069681
1.85	0.072064874	0.967843225
1.90	0.065615815	0.971283440
1.95	0.059594706	0.974411940
2.00	0.053990967	0.977249868
2.05	0.048792019	0.979817785
2.10	0.043983596	0.982135579
2.15	0.039550042	0.984222393
2.20	0.035474593	0.986096552
2.25	0.031739652	0.987775527
2.30	0.028327038	0.989275890
2.35	0.025218220	0.990613294
2.40	0.022394530	0.991802464
2.45	0.019837354	0.992857189
2.50	0.017528300	0.993790335
2.55	0.015449347	0.994613854
2.60	0.013582969	0.995338812
2.65	0.011912244	0.995975411
2.70	0.010420935	0.996533026
2.75	0.009093563	0.997020237
2.80	0.007915452	0.997444870
2.85	0.006872767	0.997814039
2.90	0.005952532	0.998134187
2.95	0.005142641	0.998411130
3.00	0.004431848	0.998650102

D.2 χ^2 Distribution (Chi-square)

Percentage points $\chi_{n,\alpha}^2$ of the Chi-square distribution for a selection of samples, n .

α	n						
	2	5	10	30	100	300	1000
0.001	0.00	0.21	1.48	11.59	61.92	229.96	867.48
0.005	0.01	0.41	2.16	13.79	67.33	240.66	888.56
0.010	0.02	0.55	2.56	14.95	70.06	245.97	898.91
0.025	0.05	0.83	3.25	16.79	74.22	253.91	914.26
0.050	0.10	1.15	3.94	18.49	77.93	260.88	927.59
0.100	0.21	1.61	4.87	20.60	82.36	269.07	943.13
0.200	0.45	2.34	6.18	23.36	87.95	279.21	962.18
0.500	1.39	4.35	9.34	29.34	99.33	299.33	999.33
0.800	3.22	7.29	13.44	36.25	111.67	320.40	1037.43
0.900	4.61	9.24	15.99	40.26	118.50	331.79	1057.72
0.950	5.99	11.07	18.31	43.77	124.34	341.40	1074.68
0.975	7.38	12.83	20.48	46.98	129.56	349.87	1089.53
0.990	9.21	15.09	23.21	50.89	135.81	359.91	1106.97
0.995	10.60	16.75	25.19	53.67	140.17	366.84	1118.95
0.999	13.82	20.52	29.59	59.70	149.45	381.43	1143.92

References

- Adams DE and Allemand RJ 2001 Residual frequency autocorrelation as an indicator of non-linearity. *International Journal of Non-Linear Mechanics* **36**(8), 1197–1211.
- Adams ML 2001 *Rotating Machinery Vibration*. Marcel Dekker, New York.
- Ahmadian H, Friswell M and Mottershead J 1998a Minimization of the discretization error in mass and stiffness formulations by an inverse method. *International Journal for Numerical Methods in Engineering* **41**, 371–387.
- Ahmadian H, Mottershead J and Friswell M 1998b Regularisation methods for finite element model updating. *Mechanical Systems and Signal Processing* **12**, 47–64.
- Ahmadian H, Mottershead J and Friswell M 2000 Damage location indicators from substructure mode shapes. *Inverse Problems in Engineering* **8**, 309–323.
- Allemand R 2003 The modal assurance criterion - twenty years of use and abuse. *Sound and Vibration* **37**(8), 14–23.
- Allemand RJ, Rost RW and Brown DL 1982 Dual input estimation of frequency response functions for experimental modal analysis of aircraft structures *1st International Modal Analysis Conference & Exhibit*, pp. 333–340 Union College, Schenectady, NY.
- Allemand RJ, Rost RW and Brown DL 1983 Multiple input estimation of frequency-response functions - excitation considerations *Mechanical Engineering*, vol. 105, pp. 90–90. ASME, 345 E 47TH ST, NEW YORK, NY 10017.
- Allen MS and Carne TG 2006 Comparison of inverse structural filter (isf) and sum of weighted accelerations technique (swat) time domain force identification methods In *47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference*. (ed. AIAA) number AIAA-2006-1884 in AIAA-2006 AIAA. AIAA.
- Anonymous 1981 *Webster's New Collegiate Dictionary*. G. & C. Merriam Co.
- Aoki M 1968 Control of large-scale dynamics systems by aggregation. *IEEE Transactions on Automatic Control* **AC-13**, 246–253.
- Avitabile P 1999 Why you can't ignore those vibration fixture resonances. *Sound and Vibration* **33**(3), 20–26.
- Avitabile P and Pechinsky F 1998 The coordinate orthogonality check (corthog). *Mechanical Systems and Signal Processing* **12**(3), 395–414.
- Bathe KJ and Wilson EL 1976 *Numerical Methods in Finite Element Analysis*. Prentice Hall, Inc., Englewood Cliffs.
- Bathe KJ and Wilson EL 1996 *Finite Element Procedures*. Prentice Hall, Inc., Englewood Cliffs.
- Bellman R 1960 *Introduction to Matrix Algebra*. McGraw-Hill Book Company, New York.
- Bendat JS and Piersol aG 2000 *Random Data Analysis and Measurement Procedures* 3rd edn. John Wiley & Sons, Inc., New York.
- Caughey TK and O'Kelly MEJ 1965 Classical normal modes in damped linear dynamic systems. *ASME Journal of Applied Mechanics* **32**, 583–588.
- Chen CT 1994 *System and Signals Analysis* 2nd edn. Saunders College Publishing, New York.

- Childs DW 1993 *Turbomachinery Rotordynamics: Phenomena, Modeling, and Analysis*. John Wiley & Sons, New York.
- Clough RW and Penzien J 2010 *Dynamics of structures* 2nd ed., rev edn. Computers and Structures, Berkeley, Calif.
- Cook RD, Malkus DS and Plesha ME 1989 *Concepts and Applications of Finite Element Analysis* 3rd edn. John Wiley & Sons, New York.
- Crassidis JL and Junkins JL 2004 *Optimal Estimation of Dynamic Systems*. Chapman & Hall.
- Dimarogonas AD 1996 *Vibration for Engineers* 2nd edn. Prentice-Hall, Inc., Upper Saddle River.
- Duhamel 1834 Mémoire sur les vibrations d'un système quelconque de points matériels. *Journal de l'École Polytech* **25**, 1–36.
- Eaton JW 2002 *GNU Octave*. Network Theory Ltd.
- (ed. Harris CM) 1995 *Shock and Vibration Handbook* 4th edn. McGraw Hill, New York.
- Ewins DJ 1992 *Modal Testing: Theory and Practice*. Research Studies Press Ltd., New York.
- Fabunmi Ja 1986 Effects of structural modes on vibratory force determination by the pseudoinverse technique. *AIAA Journal* **24**, 504–509.
- Friswell M and Mottershead J 1995 *Finite Element Model Updating in Structural Dynamics (Solid Mechanics and Its Applications)*. Kluwer Academic Publishers, Dordrecht.
- Friswell M and Mottershead J 1998 Special issue: International conference on identification in engineering systems. *Journal of Vibration and Control* **4**, 3–3.
- Friswell M, Inman D and Pilkey D 1998a Direct updating of damping and stiffness matrices. *AIAA Journal* **36**, 491–493.
- Friswell M, Mottershead J and Ahmadian H 1998b Combining subset selection and parameter constraints in model updating. *Journal of Vibration and Acoustics* **120**, 854–859.
- Friswell M, Mottershead J and Lees A 2000a Special issue - international conference on identification in engineering systems - university of wales swansea, uk - 29-31 march 1999. *Inverse Problems in Engineering* **8**, I–II.
- Friswell M, Mottershead J and Lees A 2000b Untitled. *Inverse Problems in Engineering* **8**, I–II.
- Golub GH and Van Loan CF 1985 *Matrix Computations*. Johns Hopkins University Press, Baltimore.
- Guttman I, Wilks SS and Hunter JS 1982 *Introductory Engineering Statistics* 3rd edn. Wiley, New York.
- Hamdan aMA and Nayfeh aH 1989a Errata: Measure of modal controllability and observability for first- and second-order linear systems. *AIAA Journal of Guidance, Control and Dynamics* **12**(5), –.
- Hamdan aMA and Nayfeh aH 1989b Measure of modal controllability and observability for first- and second-order linear systems. *AIAA Journal of Guidance, Control and Dynamics* **12**(3), 421–428.
- Hamming RW 1983 *Digital Filters* 2nd edn. Prentice-Hall, Inc., Englewood Cliffs.
- Ho BL and Kalman RE 1966 Effective construction of linear state variable models from input/output data *Proceedings of the 3rd Annual Allerton Conference on Circuit and System Theory*, pp. 449–459.
- Horn RA and Johnson CR 1999 *Matrix Analysis*. Cambridge University Press, New York.
- Imregun M, Visser WJ and Ewins DJ 1995 Finite element model updating using frequency response function data-i. theoy and initial investigation. *Mechanical Systems and Signal Processing* **9**(2), 187–202.
- Inman DJ 1989 *Vibration: With Control, Measurement, and Stability*. Prentice Hall, Inc., New York.
- Inman DJ 2000 *Engineering Vibration* 2nd edn. Prentice Hall, Inc., Upper Saddle River.
- Inman DJ 2006 *Vibration with control*. Wiley, Hoboken, NJ.
- Jackson LB 1989 *Digital Filters and Signal Processing* 2nd edn. Kluwer, Boston.
- Jeffrey A 2002 *Advanced Engineering Mathematics* Hartcourt/Academic Press chapter 11, pp. 665–671.
- Juang JN 1994 *Applied System Identification*. Prentice Hall, Inc.
- Juang JN and Pappa RS 1986 Effect of noise on modal parameters identified by the eigensystem realization algorithm. *Journal of Guidance, Control, and Dynamics* **9**(3), 294–303.
- Kailath T 1980 *Linear Systems*. Prentice Hall, New York.
- Kelly SG 1993 *Fundamentals of Mechanical Vibrations*. McGraw Hill, New York.
- Lembregts F, Leuridan JM, Zhang L and Kanda H 1986 Multiple input modal analysis of frequency response functions based on direct parameter identification *Proceedings of the 5th International Modal Analysis Conference and Exhibit*, pp. 589–598 Society for Experimental Mechanics.

- Leuridan J, De Vis D, Van der Auweraer H and Lembregts F 1986 A comparison of some frequency response function measurement techniques *4th International Modal Analysis Conference*, vol. II.
- Lieven NAJ and Ewins DJ 1988 Spatial correlation of modeshapes: the coordinate modal assurance criterion (comac) *Proceedings of the 6th International Modal Analysis Conference*, pp. 690–695.
- Maia NMM and E Silva JMM 1998 *Theoretical and Experimental Modal Analysis*. Research Studies Press, Ltd., Philadelphia.
- Mares C, Mottershead J and Friswell M 1999 Damage location in beams by using rigid-body constraints. *Damas 99: Damage Assessment of Structures* **167-1**, 381–390.
- Mcconnell KG 1995 *Vibration Testing: Theory and Practice*. John Wiley & Sons, New York.
- Meirovitch L 1997 *Principles and Techniques of Vibrations*. Prentice Hall, Inc., Englewood Cliffs.
- Miller I and Freund JE 1977 *Probability and Statistics for Engineers*. Prentice Hall of India, New Delhi.
- Mitchell LD 1982 Improved methods for the fast fourier-transform (FFT) calculation of the frequency-response function. *J MECH DES-T ASME* **104**(2), 277–279.
- Mottershead J and Friswell M 1998 Untitled. *Mechanical Systems and Signal Processing* **12**, 1–6.
- Mottershead J, Friswell M and Mares C 1999 A method for determining model-structure errors and for locating damage in vibrating systems. *Meccanica* **34**, 155–168.
- Mottershead J, Friswell M, Ng G and Brandon J 1996 Geometric parameters for finite element model updating of joints and constraints. *MECHANICAL SYSTEMS AND SIGNAL PROCESSING* **10**, 171–182.
- Mottershead JE 1998 On the zeros of structural frequency response functions and their sensitivities. *Mech. Syst. Signal Proc.* **12**(5), 591–597.
- Mottershead JE and Friswell MI 1993 Model updating in structural dynamics: A survey. *Journal of Sound and Vibration* **167**(2), 347–375.
- Muszyńska A 2005 *Rotordynamics*. Taylor & Francis, New York.
- Nashif AD, Jones DIG and Henderson JP 1985 *Vibration Damping*. John Wiley & Sons, New York.
- Newland DE 1993 *An Introduction to Random Vibrations, Spectral & Wavelet Analysis* 3rd edn. Longman Scientific and Technical, London.
- O'Callahan J 2000 Reduced model concepts. *Proceedings of the International Modal Analysis Conference - IMAC* **2**, 1528–1536.
- O'Callahan J, Avitabile P and Riemer R 1988 System equivalent reduction expansion process *7th International Modal Analysis Conference*.
- Otnes RK and Enochson L 1972 *Digital Time Series Analysis*. John Wiley and Sons, New York.
- Ouyang H, Mottershead J, Cartmell M and Friswell M 1998 Friction-induced parametric resonances in discs: Effect of a negative friction-velocity relationship. *Journal of Sound and Vibration* **209**, 251–264.
- Papoulis a and Pillai US 2002 *Probability, Random Variables, and Stochastic Processes* 4 edn. McGraw-Hill Higher Education, 1221 Avenue of Americas, New York, NY 10020.
- Parloo E, Verboven P, Guillaume P and Overmeire MV 2003 Force identification by means of in-operation modal models. *Journal of Sound and Vibration* **262**, 1610173.
- Parzen E 1972 *Modern Probability Theory and Its Applications*. Wiley easter Private Limited, New Delhi.
- Phillips CL and Nagle HT 1990 *Digital Control System Analysis and Design (3rd Edition)* 2nd edn. Prentice Hall, New York.
- Rao JS 2000 *Vibratory Condition Monitoring of Machines*. Narosa Publishing House, New Delhi.
- Rao SS 1999 *The Finite Element Method in Engineering* 3rd edn. Butterworth Heineman, Boston.
- Reid Jg 1983 *Linear System Fundamentals*. McGraw Hill, New York.
- Rocklin G, Crowley J and Vold H 1985 A comparision of h1, h2, and hv frequency response functions *A Comparision of H1, H2, and Hv Frequency Response Functions*, vol. 3rd International Modal Analysis Conference, pp. 272–278.
- Schulz MJ and Inman DJ 1994 Model updating using constrained eigenstructure assignment. *Journal of Sound and Vibration* **178**(1), 113–130.
- Skelton RE 1988 *Dynamic Systems Control: Linear Systems Analysis and Synthesis*. John Wiley & Sons, New York.

- Sólnes J 1997 *Stochastic Processes and Random Vibrations : Theory and Practice* vol. Preface. John Wiley & Sons. Includes bibliographical references (p. [415]-424) and index.
- Strum RD and Kirk DE 1989 *First Principles of Discrete Systems and Digital Signal Processing*. Addison Wesley, New York.
- Theude LLD c2005 Practical analog and digital filter design / les thede. Includes bibliographical references and index.
- Vance JM 1988 *Rotor Dynamics of Turbo-Machinery*. John Wiley and Sons, Inc.
- Vankarsen C and Allemand RJ 1984 Averaging for improved frequency-response functions. *Sound and Vibration* **18**(8), 18–24.
- Weaver W, Timoshenko SP and Young H 1990 *Vibration Problems in Engineering*. John Wiley & Sons, New York.
- Wicks AL and Mitchell LD 1987 Methods for the estimation of frequency-response functions in the presence of uncorrelated noise. *The International Journal of Analytical and Experimental Modal Analysis* **2**(3), 109–112.
- Wicks AL and Vold H 1986 The H_s frequency response function estimator *4th International Modal Analysis Conference*, vol. II, pp. 897–899.
- Wirsching PH, Paez TL and Ortiz K 1995 *Random Vibrations: Theory and Practice*. John Wiley & Sons, Inc., New York.
- Wowk V 1991 *Machinery Vibration : Measurement and Analysis / Victor Wowk*. McGraw-Hill., Includes bibliographical references (p. 347-348) and index.
- Wowk V c1995 *Machinery Vibration : Balancing / Victor Wowk*. McGraw-Hill., Includes bibliographical references (p. 311-313) and index.
- Wowk V c2000 *Machine Vibration. Alignment / Victor Wowk*. McGraw-Hill., Includes bibliographical references (p. 373-375) and index.
- Zhang L, Kanda H and Lembregts F 1986 Some applications of frequency domain polyreference modal parameter identification method *Proceedings of the 5th International Modal Analysis Conference and Exhibit*, pp. 1237–1245 Society for Experimental Mechanics.
- Zienkiewicz OC and Taylor RL 1989 *The Finite Element Method*. McGraw Hill, New York.

Index

- \cap , 56
- \mathcal{F} , 16
- \mathcal{L} , 10
- $\mathcal{S}(\tau)$, 124
- adjoint
 - classical, 36
- aggregation, 126
- aggregation matrix, 126
- aliasing, 93
- anti-resonance, 43
- ASD, 101
- augmented Vandermonde matrix, 148
- auto-correlation, 69
- auto-spectrum density
 - two-sided, 79
- averaging, 81, 94
 - ensemble, 172
 - time domain, 172
- balanced realization, 143
- block matrices, 180
- calculated mean, 49
- calculated variance, 49
- Cauchy-Schwartz inequality, 69
- Cayley-Hamilton Theorem, 186
- central limit theorem, 60, 65
- circulatory systems, 121
- cofactor, 43
- coherence, 82
- collocated, 163
- complex exponentials, 2
- complex modulus, 27
- complex stiffness, 28, 41
- compliance, 25, 156
- condition number, 185
- confidence interval, 63, 95
- continuous-time state-space model, 131
- continuum models, 112
- control matrix, 144
- controllability, 128
 - discrete, 135
 - index, 142
- controllability grammian, 128, 144
- controllability matrix
 - discrete, 135
 - generalized, 142
- controllable, 128, 135
- convolution integral, 31
- coordinate transformation
 - state-space, 126
- Correlation, 67
- correlation
 - auto, 69, 93
 - cross, 70, 93
- correlation coefficient, 58
- covariance, 57, 68
- critical damping, 6
- Cross Correlation, 67
- cross-correlation, 70
- damping ratio, 6
- dB, 175
- decibel, 175
- decimation, 93
- DFT, 88
- digital filters, 93
- Dirac delta function, 17, 28
- direct transmission, 131
- discrete controllability grammian, 136
- discrete controllability matrix, 135
- discrete Fourier transform, 88
- discrete observability grammian, 137
- discrete-time models, 133
- discrete-time state-space model, 131
- distribution
 - χ^2 , 51
 - χ^2 , 62, 194
 - t , 51
 - Chi-square, 51, 62, 194
 - Gauss, 193

- Gaussian, 51
- joint Gaussian, 58
- normal, 56, 193
- Rayleigh, 51, 57, 64, 65
- sine/cosine, 51
- truncated Gaussian, 51
- uniform, 51
- distribution function
 - joint, 56
- distributions
 - Gaussian, 56
- dot product, 179
- double least squares, 69
- drift
 - sensor, 94
- Duhamel's Integral, 31
- Duhamel's integral, 31
- effective inertia, 150
- effective stiffness, 150
- eigenvalue, 182
- eigenvalues
 - state-space, 123
 - second order system, 36
- eigenvectors, 182
- ensemble averaging, 95, 172, 173
- equivalent realization, 126
- ERA, 141
- ergodic signals, 67
- expected value, 49, 51
- fast Fourier transform, 88
- FFT, *see* fast Fourier transform
- filters
 - digital, 93
- first-order modal coordinates, 124
- Fourier series, 11, 170
- Fourier transform, 16, 23, 26
 - derivative, 20
 - integral, 21
 - linearity, 19
 - of $\sin(\Omega t)$, 18
 - of Dirac delta function, 17
 - of time convolution, 21
 - of time shifted function, 20
 - properties of, 19
- Fourier transform table, 23
- frequency resolution, 89
- frequency response function, 22, 77, 147
 - H_1 , 81
 - H_2 , 82
- H_3 , 84
- H_v , 83
- multiple input-single output, 84
- frequency resresponse function
 - H_v , 182
- FRF, *see* frequency response function
- Gaussian distribution, 56
- Gaussian random process, 56
- generalized controllability matrix, 142
- generalized observability matrix, 142
- Gibbs effect, 14, 105
- grammian
 - controllability, 144
 - observability, 144
- gyroscopic systems, 121
- half-power damping estimation, 112
- Hankel matrix, 141
- Hanning window, 97
- Heaviside function, 18, 21
- Hermitian, 156, 182
- Ho-Kalman theory, 141
- hysteresis, 28
- impulse response function, 26, 30, 130
 - from frequency response function, 22
- impulse response of a state space system, 130
- inertance, 27, 156
- input influence matrix, 123
- input matrix, 123
- inverse
 - matrix, 179
 - pseudo-, 152, 185
- inverse Fourier transform, 16
 - of frequency convolution, 22
- IRF, *see* impulse response function
- Jacobian, 54, 181
- joint cumulative distribution function, 56
- joint Gaussian distribution, 58
- joint normal distribution, 58
- Kronecker delta function, 38
- kurtosis, 47, 170
 - bias-adjusted, 171
- Laplace transform
 - properties of, 19
- Laplace transform, 10, 11, 52
 - one-sided, 10

- Laplace transform table, 11
leakage, 96
left eigenvectors, 182
linear least squares, 185
- MAC, 165
marginal, 57
Markov parameters, 131, 135, 138, 141
mass normalize, 37
mass orthonormal, 37
matrix, 178
 non-singular, 183
 singular, 183, 186
matrix exponential, 124
matrix inverse, 179
maximum observed frequency, 90
MDOF systems, *see* multiple degree of freedom systems, *see* multiple degree of freedom systems
mean, 47
 calculated, 60
minimum realization, 126
mobility, 27, 156
Modal Assurance Criteria (MAC), 165
modal constant, 42
modal coordinates, 127
 first-order, 124
 real, 127
modal equations, 39
modal frequency response functions, 41
modal matrix, 158
Modal Participation Factors, 131, 132
modal participation matrix, 158
Modal Shape Correlation Coefficient, 165
mode shapes, 132
models
 discrete time state-space, 133
moment
 calculating, 52
moment generating functions, 52, 53
MSCC, 165
multiple coherence, 87
multiple degree of freedom systems, 33, 121
- non-proportionally damped systems, 121
Normal distribution, 56
normalized covariance, 58
normalized random error, 95
Nyquist frequency, 90
Nyquist plot, 25
- observability, 128, 136
 discrete, 142
observability grammian, 130, 144
observability matrix, 130, 137
 generalized, 142
observable, 136
one-sided Laplace transform, 10
one-sided spectrum densities, 80
output equation, 131
output matrix, 131
- Parseval's formula, 22
partial coherence, 87
partial fraction, 153
PDF, *see* Probability Density Functions
poles, 132
Polyreference method, 158
power-spectrum density
 two-sided, 79
principle component analysis, 183
principle components, 184
principle of impulse and momentum, 29
Probability
 Density Functions, 50, 55
 Distribution Functions, 50
 Introduction to, 47
probability density function, 49
proportional damping, 41
pseudo-inverse, 148, 152, 185
- quadrature peak picking, 112
- random processes, 48
random signals, 48
rank, 180
rank deficient, 180
rational polynomial, 153
Rayleigh damping, 41
Rayleigh distribution, 57, 64, 65
real modal coordinates, 127
realization
 balanced, 143, 144
 equivalent, 126
 modal, 127
 real modal, 127
realization independent, 141
residuals, 148
residue, 26, 42, 152, 157
RMS value, 101
Roller Bearing Fault Frequencies, 174
Root Mean Square, 101

- s-domain, 10
- sample mean, 49
- sample variance, 49
- SDOF, *see* single degree of freedom system
- self-adjoint, 35
- self-adjoint systems, 35
- sensor drift, 94
- SEREP, 163
- Shannon's Sampling Theorem, 90
- signal bias, 94
- signals
 - ergodic, 67
 - finite time correlation, 70
 - random, 48
 - stationary, 65
- single degree of freedom system, 1
- Singular Value Decomposition, 156
- singular value decomposition, 142, 156, 183
- singular values, 142
- spectral density
 - power-, 176
- spectrum density, 77, *see* spectrum density
 - auto-, 77, 80
 - cross-, 77, 80
 - power-, 77
- Welch method, 96
- stability
 - state-space form, 123
- state controllability
 - discrete, 135
- state controllable, 135
- state matrix, 123, 159
- state observability, 136
- state space, 122
- state transition matrix, 124
- state-space, 122
 - discrete-time, 133
 - stability, 123
- state-space models, 123
- stationary signals, 65
- step function, 21
- stiffness orthogonal, 37
- strain gage, 72
- SVD, 183
- symmetric systems, 35
- System Equivalent Reduction Expansion
 - Process, 163
- system Markov parameters, 135
- systems
 - circulatory, 121
- gyroscopic, 121
- multiple degree of freedom, 33, 121
- non-proportionally damped, 121
- non-self-adjoint, 121
- self-adjoint, 35
- symmetric, 35
- table
 - Laplace transform, 11
 - time domain averaging, 172
 - time-domain identification, 135
 - transfer function, 11
 - transfer function matrix, 132
 - transpose, 177
 - trends
 - removing, 72
- Vandermonde matrix, 148, 152, 186
- Vandermonde matrix- augmented, 148
- variable definitions, 189
- variance, 47, 52
 - calculated, 64
 - sample, 64
- vector, 178
 - row, 178
- vector inner product, 179
- vector outer product, 179
- Welch method, 96
- window, 96
- windowing, 96
- windows
 - Blackman, 102
 - boxcar, 102
 - exponential, 102
 - Hamming, 102
 - Kaiser-Bessel, 102
 - Parzen, 102
- zero, 43
- zero padding, 93
- zero-order hold, 133