Ph.D. Qualifying Examination

Joseph Mariglio

December 24, 2013

Chapter 1

Miller Puckette

Abstract

If you were trying to measure how a solid object vibrates, the question might be broken down into two sub-questions: how does the object vibrate in general, and what is its specific vibrational state at a given moment in time. What models have you found that might be useful for describing the vibratory behavior in general, and what are the prospects for picking up specific vibrational state information? Is it possible to predict under what conditions it will be possible to make these measurements on real objects, and/or how the number of sensors available might affect the quality of measurements that are possible?

1.1 Derivation of the Non-Piezoelectric Wave Equation (no damping)

When an object vibrates, waves propagate through it. The term *wave* is used to describe the transfer of energy without the global translation of matter. To begin to talk about vibrations in finite substrates such as objects, we must first consider the case of a material without boundaries. The behaviors of finite objects will then precipitate from this more abstract case.

In fluids of sufficiently low viscosity, *stress* is the force normal to the surface per unit area:

$$T = \frac{F_n}{\text{Area}}$$

This is sometimes called the *nominal stress*, and it gives a reasonable approximation under certain conditions. In general cases, e.g. for solids or viscous liquids in 3 dimensions, stress is best described as a 2nd-rank tensor. Viewed in this sense, T describes the stress components on each face of an infinitesimally small cubic volume. We will refer to such an infinitesimal volume as a *particle*. Each component T_{ij} describes the stress exerted in a single dimension i, on a single surface j on the particle. [3, 18]

Stress on the object results in *strain*, a dimensionless proportion of displacements given by the ratio of the length of a stressed particle to an unstressed one Consider a lattice of particles with springs connecting them. Strain represents how much each spring has been deformed from its equilibrium point. The displacement, u, is a vector that represents the particle's position along the x, y, and z axes. Changes in these displacements will propagate throughout the object, from particle to particle. [3, 18]

Of particular interest are the local deformations of the object, rather than the overall translation of the object in space. These local deformations are given by the displacement gradient, ∇u, another 2nd-rank tensor[3, p. 12]:

$$\nabla u = \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{pmatrix}$$

This tensor describes local translations and rotations between neighboring particles. It was produced by taking the partial derivative of each component in vector u with respect to each spatial dimension.

The strain tensor S eliminates the contributions given by rotations, allowing us to describe only local rarefactions, compressions, or shear movements. The formula for S is:

$$S = \frac{\nabla u + (\nabla u)'}{2}$$

where x' denotes the *matrix transpose* of x, wherein the column and row indeces are exchanged, i.e., $x'_{i,j} = x_{j,i}$. [1]Alternatively, in terms of the components of S, we can write[3, p. 13]:

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{1.1.1}$$

[2]

In equation (1.1.1), each element of the strain matrix S_{ij} describes changes in displacement for combinations of two directions, which could be identical. The diagonal elements in strain matrix S_{ij}, with $i = j$, relate to rarefactions and compressions, where the two directions of changing displacement are the same (e.g. *longitudinal* waves). The off-diagonal components S_{ij} (with $i \neq j$) relate to shear motions, where the two directions of changing displacement are different (e.g. *transverse* waves). [3, 14]

We are now equipped to derive the equation of motion within an elastic solid. It is useful and common to assume the change in stress ΔT_i acts on a particle along the basis lengths Δx, Δy, and Δz uniformly. This approximation

---

[1]N.B. if u were a complex number (it can't be in this case), we would require the *adjoint* operator, which will be designated u^* henceforth, meaning we take the transpose and change the sign of the imaginary component. This may come in handy later.

[2]Readability demands we occasionally notate the x, y and z axes as x_1, x_2, and x_3, respectively. This notation underscores the similarities of the equations with respect to each axis, without requiring the development of matrices for each step. [3, p. 16] [18, p. 542]

is valid only because we will take the particle's volume to the infinitesimal limit later in the derivation. It is similar in spirit to the "Lumped Element Model" of Maxwell's Equations in electronics.

To find a force F_x acting in the x direction, for example, we multiply the appropriate stress component of tensor ΔT_x by the areas of each of the faces on the particle, i.e.[3, p. 15]:

$$
\begin{aligned}
F_x = &\left[(T_{xx} + \Delta T_{xx}) A_x - T_{xx} A_x \right] \\
&+ \left[(T_{xy} + \Delta T_{xy}) A_y - T_{xy} A_y \right] \\
&+ \left[(T_{xz} + \Delta T_{xz}) A_z - T_{xz} A_z \right]
\end{aligned} \tag{1.1.2}
$$

Newton's 2ⁿᵈ Law of Motion states that the force an object experiences equals its mass times its acceleration, or $F = m\ddot{u}$, where $\ddot{u} = \frac{\partial^2 u}{\partial t^2}$. The mass of the particle is its density ρ times its volume, or $\rho \Delta x \Delta y \Delta z$. Using equation (1.1.2), we can now find an equation that relates an acceleration in the x direction, $\ddot{u}_x$, to the resulting effect on the stress tensor, ΔT_{ij}[3, p. 16]:

$$
\begin{aligned}
\rho \Delta x \Delta y \Delta z \frac{\partial^2 u_x}{\partial t^2} = &\Delta T_{xx} \Delta y \Delta z \\
&+ \Delta T_{xy} \Delta x \Delta z \\
&+ \Delta T_{xz} \Delta x \Delta y
\end{aligned} \tag{1.1.3}
$$

Now we can simplify equation (1.1.3) and take the limit as the particle volume approaches zero to derive the *Equations of Motion* for a solid:

$$
\sum_{j=1}^{3} \frac{\partial T_{ij}}{\partial x_j} = \rho \frac{\partial^2 u_i}{\partial t^2} \tag{1.1.4}
$$

Above, i refers to the axis along which the motion is to be described.[3, p. 16] [18, p. 4]

Hooke's Law describes the behavior of an elastic material by relating the strain to the stress with a coefficient, c. It is a first-order approximation, which implies a linear relationship between these terms. The benefits of this approxi-

3

mation will be made clear when we discuss the *superposition principle* in section 1.2. In one dimension, Hooke's Law is $T = cS$. In the general case we are deriving, recall that stress, T_{ij}, is a 2nd-rank tensor, and so is strain, S_{ij}. To allow the coefficient c to map between these two 2nd-rank tensors, it is necessary to invoke it as a 4th-rank tensor c_{ijkl}. Stating Hooke's Law in 3 dimensions this way, and generalizing it for any stress vector, on each surface of the particle, T_{ij}, produces the *Elastic Constitutive Relation*[18, p. 542]:

$$T_{ij} = \sum_{k,l=1}^{3} c_{ijkl} S_{kl} \qquad (1.1.5)$$

It is desirable to reduce the number of unique components in equation (1.1.5). Recall that c_{ijkl} is a 4th-rank tensor—effectively a 4-dimensional matrix with 3 components in each dimension—with $3^4 = 81$ components total. Fortunately, not all of these components are unique. Exploiting the redundancies endemic to the stress and strain tensors, it is possible to rewrite equation (1.1.5), with T and S stated in terms of a single index I, and J, respectively. These new indeces will be denoted as a capitol letter. This is because $S_{kl} = S_{kl}$, as shown in equation (1.1.1). Reduced index I maps to standard tensor indeces ij, and J, analogously, to kl. Using this simplification, known as *reduced notation*, T_I and S_J can refer to any of the 6 unique components in T or S, by defining $S_J \equiv S_{kl} = S_{kl}$. The reduced notation form of S is shown below, expanded into matrix form[18, p. 542]:

$$S = \begin{bmatrix} S_1 & \frac{S_6}{2} & \frac{S_5}{2} \\ \frac{S_6}{2} & S_2 & \frac{S_4}{2} \\ \frac{S_5}{2} & \frac{S_4}{2} & S_3 \end{bmatrix} \qquad (1.1.6)$$

The reduced notation subscripts, used above, map to the standard tensor notation, developed in equation (1.1.5), as follows[3, p. 17] [18, p. 543]:

Reduced Index	Standard Tensor	Physical Meaning
1	xx	Longitudinal in x direction
2	yy	Longitudinal in y direction
3	zz	Longitudinal in z direction
4	yz = zy	Shear y - z
5	zx = xz	Shear z - x
6	xy = yx	Shear x - y

Similarly, the components in c are reduced to c_{IJ}, where capitol indeces IJ are an analogous reduction of the four indeces $ijkl$ to 36 unique components. This follows from the reduction to indeces I and J above. This 6 x 6 matrix can be further reduced, since $c_{IJ} = c_{JI}$. Most solids can be described with a stiffness matrix c_{IJ} consisting of only 21 components. Of these, 6 are diagonal—meaning they describe collinear relationships—and 15 are triangular—meaning they describe transverse relationships. The term *isotropic* denotes that the inter-particulate lattice behaves symmetrically in all directions. Isotropic crystals, for example, provide a helpful case to consider, as they require only 2 terms to describe fully. Combinations of these two elasticity coefficients, called the Lamé constants, are sufficient for describing all Hookean vibration in the material. The degree to which we can reduce a solid's stiffness, strain, and stress matrices c, S, & T, respectively, depends on the symmetries encoded in the material's inter-particulate relationships. [3] To elucidate especially complex behavior, we will occasionally assume isotropy in our treatment of waves.

For a general, non-isotropic solid, equation (1.1.5) can then be written, somewhat more humanely, as:

$$T_I = \sum_{J=1}^{6} c_{IJ} S_J \qquad (1.1.7)$$

Note that for plane waves, whose particle displacements act in a single direction, the same symmetries that permitted the simplifications from equation (1.1.5) to equation (1.1.7) also mean that $S_{kl} = \frac{\partial u_k}{\partial x_l}$. With that in mind, differentiating the above equation with respect to a single direction of particulate displacement

x_j gives the following[3]:

$$\sum_{j=1}^{3} \frac{\partial T_{ij}}{\partial x_j} = \sum_{j,k,l=1}^{3} c_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l} \qquad (1.1.8)$$

Finally, we can derive the *Non-piezoelectric Wave Equation* by setting the right side of the above equation equal to that of equation (1.1.4), i.e.:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \sum_{j,k,l=1}^{3} c_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l} \qquad (1.1.9)$$

Before moving on to solutions for this equation, we should explain what was meant at the outset of this document by "the transfer of energy without the global translation of matter." It should be clear from our definition of the strain tensor S that we are ignoring global translation and rotation of particles. Thus, all that remains is the definition of *energy* in this context. Energy in this system is defined in two components: *kinetic* and *potential*. The kinetic energy per unit volume is due to the particulate motion of the material, i.e.[18, p. 6],

$$W_k = \frac{\rho \dot{u}^2}{2} \qquad , \qquad (1.1.10)$$

where the density, ρ, is that of the unstressed material. The potential energy per unit volume, on the other hand, is due to the elasticity of the material, i.e., [18, p. 6]

$$W_p = \frac{TS}{2} \qquad . \qquad (1.1.11)$$

N.B. these are instantaneous values. In a lossless medium, these values will fluctuate together in antagonism. The total instantaneous energy of the system will be their sum. In a perfectly elastic medium, this value will not diminish, and we are in the possession of a perpetual motion machine of the 1st kind.

1.2 Solutions to the Wave Equation with Losses

The Non-piezoelectric Wave Equation (1.1.9) is in fact a system of 3 equations, for $i = x, y, z$. A possible solution to this equation could be a harmonic plane wave propagating forward in the positive x direction[3, p. 12]:

$$u(x, y, z, t) = \Re\{(u_x x + u_y y + u_z z)e^{i(\omega t - kx)}\} \qquad (1.2.1)$$

Harmonic solutions to the wave equation (1.1.9) consist of a complex exponential $e^{i(\omega t \pm kx_i)}$, wherein the term i denotes the imaginary number, $\sqrt{-1}$. This notation visually differentiates it from the index i. The term e denotes *Euler's number*, or $\lim_{x \to \infty}(1 + \frac{1}{x})^x$, whose y-intercept is 1.

Equation (1.2.1) uses the parameter ω to refer to the *angular frequency* of the wave. If f is the *frequency* of the wave in cycles / second, $\omega = 2\pi f$, so ω simply counts the rotations (over time) in terms of 2π. k is the *wave number*, essentially the spatial analogue of ω. By convention, for forward-traveling waves, the term kx is negative, whereas for backward-traveling waves, the sign is reversed. The $\Re\{x\}$ operator extracts the real component from a complex-valued x. [3, 18, 7]

Notice also how the wave function in equation (1.2.1) describes variations in the x, y and z components of the u vector, which represents particle displacement. These components u_x, u_y and u_z are amplitudes of the wave along each axis. Since the wave function maintains the same phase relationships across these axes, and the only variation is amplitude, the surfaces of constant phase are planes, orthogonal to the propagation vector, which is in this case a unit vector in the x direction.

Consider if the only non-zero component in the vector u were u_x. This would mean that the harmonic variation in particle displacement would only occur in the direction of the wave propagation. The wave motion would be expressed solely as a periodic compression and extension of the particle lattice. Such waves are called *longitudinal* waves. Conversely, if the only non-zero component in the vector u were either one of the components perpendicular to the direction of

wave propagation, i.e., u_y or u_z in the case of equation (1.2.1), the motion would be called a *shear* wave. This is analogous to the formalization posed in the discussions in 1.1 surrounding the definitions of stress and strain tensors. In the simplest case of a non-piezoelectric, isotropic solid, all waves can be considered a combination of longitudinal and shear plane waves. [18, p. 15]

Since, in equation (1.2.1), the values ω, t, k, and x, which multiply i, are real-valued, the exponential $e^{\mathrm{i}(\omega t \pm kx)}$ is complex. It can be stated in terms of its real and imaginary components separately as[7, p. 5]:

$$\Re\{e^{\mathrm{i}(\omega t \pm kx)}\} = \cos(\omega t \pm kx)$$
$$\Im\{e^{\mathrm{i}(\omega t \pm kx)}\} = \mathrm{i}\sin(\omega t \pm kx)$$

In order to make a physical measurement, we must ultimately extract the real component. To do this, we must subtract the imaginary component from the complex exponential function. It is often convenient to consider cases of wave motion wherein the material responds in a linear way to perturbations, e.g. for small values of S, T and u. This is precisely the approximation implied by Hooke's Law, as invoked in section 1.1. The *superposition principle* states that any solution to the wave equation, (1.1.9), can be described as a sum of harmonic solutions, of the form $e^{\mathrm{i}(\omega t \pm kx)}$. This means a thorough consideration of the harmonic case is sufficient to provide a generalized account, provided the restrictions imposed by linearity are obeyed.[7, p. 5] Thus, we must assume the real-valued measurements of wave behavior to be the linear superposition of two complex exponential functions, with their imaginary components inverted. The physical meaning of this is to form any solution to the wave equation as a superposition of pairs propagating sinusoids, which travel in opposite directions. This is known as *D'Alembert's solution* to the wave equation. We will use it many times.[7, p. 119]

Since $\frac{\partial^2 u_i}{\partial x^2} = -\omega^2 u_i$ and $\frac{\partial^2 u_i}{\partial t^2} = -k^2 u_i$, it is possible to rewrite equation

(1.2.1) into the *dispersion relation* for a longitudinal wave[3, p. 20]:

$$\rho\omega^2 = c_{IJ}k^2 \qquad (1.2.2)$$

[3] This relation allows us to calculate v: the speed at which a surface of constant phase propagates through the material. Recall, in the longitudinal plane wave case we began developing in equation (1.2.1), the surface is a plane, and the direction of particulate displacement of the wave is collinear to the vector of wave propagation. The term v, known as the *phase velocity* of a wave, describes the velocity with which those surfaces propagate through the material. The formula for v in any dimension x is $v_x = \sqrt{\frac{c_{IJ}}{\rho}}$. [3, p. 20]

When the elasticity coefficient c_{IJ} is complex, it can represent energy losses in the wave equation due to viscous forces between neighboring particles. This is a major source of attenuation in waves as they propagate through solids and liquids, although it is not the only one. The stresses due to viscosity T_{η_I} on a plane wave are[18, p. 8][3, p. 21]:

$$T_{\eta_I} = \sum_{J=1}^{6} \eta_{IJ}\dot{S}_J \qquad (1.2.3)$$

The term η has been introduced above to refer to the *coefficient of viscosity*.[4] N.B. η is a tensor with the same symmetries as c_{IJ}. The viscous stresses $T_{\eta_{IJ}}$ affect the total stresses T_I in the following way:

$$T_I = \sum_{J=1}^{6} c_{IJ}S_J + \eta_{IJ}\dot{S}_J \qquad (1.2.4)$$

Recall that a wave propagating in the positive direction has a negative displacement term. Therefore, η effectively scales a frictional force, in the direction opposing the particle velocity. In the case of a mechanical harmonic wave,

---

[3]N.B. the above equation uses reduced notation for components of stiffness matrix c.

[4]N.B. this is not the same η that will be used to describe the index of refraction in the context of EM theory

$\dot{S}_J = i\omega S_J$, and therefore the S_J term may be factored out of the sum like so:

$$T_I = \sum_{J=1}^{6} (c_{IJ} + i\omega\eta_{IJ})S_J \tag{1.2.5}$$

This permits us to describe both elasticity and damping with a single complex variable. We now apply this development from equation (1.2.5) to the Non-piezoelectric Wave Equation (1.1.9):

$$\rho\frac{\partial^2 u_i}{\partial t^2} = \sum_{j,k,l=1}^{3} (c_{ijkl} + i\omega\eta_{ijkl})\frac{\partial^2 u_k}{\partial x_j^2 \partial x_l^2} \tag{1.2.6}$$

Each component of this system can be independently found, using solutions of the form $u_i(x,t) = \Re\{Ae^{i(\omega t \pm kx_i)}e^{-\alpha x_i}\}$. What was once a sinusoid propagating in the x_i direction forever has now been "enveloped" by a falling exponential curve, whose level depends on the distance propagated by the wave and a scaling term on the function, α. In phenomenological terms, α determines how quickly the sinusoid decays, as a function of the distance it has propagated. It might help clarify the effects of damping to show this attenuation term as it relates to phase velocity v, angular frequency ω, density ρ, and the reactive portion of the elastic coefficient, η. Substituting the above solution for u_i and simplifying results in the following dispersion relation[24, p. 12][3, p. 22]:

$$\begin{aligned}
-\rho\omega^2 &= (c_{IJ} + i\omega\eta_{IJ})(\alpha + ik)^2 \\
&= (c_{IJ} + i\omega\eta_{IJ})(\alpha^2 + 2\alpha ik - k^2) \\
&= c_{IJ}(\alpha^2 - k^2) - 2\omega\eta_{IJ}k\alpha + 2\alpha ikc_{IJ} + i\omega\eta_{IJ}(\alpha^2 - k^2) \\
&= c_{IJ}(\alpha^2 - k^2) - 2\omega\eta_{IJ}k\alpha + i[2\alpha kc_{IJ} + \omega\eta_{IJ}(\alpha^2 - k^2)]
\end{aligned} \tag{1.2.7}$$

[5] Finally, we isolate the real and imaginary components:

$$-\rho\omega^2 = c_{IJ}(\alpha^2 - k^2) - 2\omega\eta_{IJ}k\alpha \tag{1.2.8}$$
$$0 = 2\alpha ikc_{IJ} + i\omega\eta_{IJ}(\alpha^2 - k^2) \tag{1.2.9}$$

---

[5]Notice that the k variable here refers to wave number and is not an index into c.

The attenuation factor α can be approximated by assuming $k \gg \alpha$, therefore the term $(\alpha^2 - k^2) \approx k^2$ in equation (1.2.9). After solving for α and simplifying, the result is[3, p. 22]:

$$\alpha = \frac{\omega^2 \eta_{IJ}}{2\rho v^3} \tag{1.2.10}$$

The viscous attenuation coefficient α is therefore proportional to the square of the frequency. This accounts for why, e.g., high frequencies get attenuated more than low frequencies, all other things considered. Furthermore, it has an inverse relationship to the cube of the phase velocity. Since shear waves typically travel slower than longitudinal waves, they often experience less attenuation per unit length.

Another way to look at complex coefficients of elasticity is in terms of energy. Recall in section 1.1 that we defined instantaneous values for the kinetic and potential energy per unit volume, in equations (1.1.10) and (1.1.11), respectively. For a periodic wave, the average kinetic energy is given by[18, p. 7]:

$$W_k = \frac{\Re\{\rho\dot{u}\dot{u}^*\}}{4} \quad , \tag{1.2.11}$$

and the average potential energy

$$W_p = \frac{\Re\{TS^*\}}{4} = \frac{\Re\{cSS^*\}}{4} \quad . \tag{1.2.12}$$

The total energy is still the sum of these, [18, p. 7]

$$W_a = \frac{\Re\{\rho\dot{u}\dot{u}^* + TS^*\}}{4} = \frac{\Re\{TS^*\}}{2} \quad , \tag{1.2.13}$$

and the complex power flow through an area, A, is

$$P_a = -\frac{\dot{u}^*TA}{2} \quad . \tag{1.2.14}$$

This setup supports the effort to make apparent what happens to the power flow, depending on c. If c is real, the propagation is lossless, and the power

P_a is real. If c is complex, then so is P_a, but of course we only take the real component. Thus power (and therefore energy) is conserved, albeit in a complex form.[18, p. 8] Yet another way to describe this loss is in terms of *impedance*, which will be explored in section 1.4.

1.3   Other Types of Waves

So far, we have developed an understanding of the propagation of longitudinal and shear waves. These are both categorized as *bulk waves*, because they require the material to extend for many periods of the wave. Momentarily, we take this opportunity to list some other simple types of vibrations that can travel through objects. Similarly to our treatments of the pure longitudinal and pure shear cases, these wave types will be shown propagating in a single direction. This is only possible if the wave never experiences a change the medium through which it travels. So, the waves also depend on the volumes through which they travel to be infinite and homogeneous in some direction.However these can also be described for wires, bars, or plates, and it will only take a minor conceptual leap to bring them into truly finite volumes. Rather than derive the wave equations for each of these types, a phenomenological account will be given.

Torsional waves are waves that propagate via torque. Torsional waves propagate with a phase velocity identical to that of transverse waves, and may even be considered a form of transverse wave for some volumes. This is because the coefficient of elasticity that governs the stress-strain relationship in torsional waves also does the same in shear waves for some materials. In both cases, the area of a cross-section perpendicular to the axis of propagation remains constant. [7, p. 90]

A wire, bar, or plate, whose thickness is small on the order of a wavelength of interest, may exhibit another type of wave motion, called the *quasi-longitudinal wave*.[7, p. 80] This type of wave is a form of longitudinal wave motion, except, unlike longitudinal waves, it propagates without bulk compression. Rather,

these waves are the propagation of tensile stress, and so while the field variables of this wave are oriented with the propagation vector, these waves occur in incompressible (or nearly incompressible) material. To make room for the longitudinal strain, these waves must also have a small transverse strain component. However, since the rod freely moves in the plane orthogonal to the propagation axis, there can be no stress wave traveling on this plane. The elasticity coefficient that relates stress to strain in quasi-longitudinal waves is *Young's modulus*, and is typically much smaller than that of the bulk longitudinal motion in the same material. [18, p. 14]

Flexural waves occur in a plate or wire when it bends. These are the waves most closely associated with the radiation of sound in air. Although they appear similar to transverse waves, the equations that govern their motion are completely unique. The description of these waves requires four variables, rather than the familiar two. These variables are the transverse velocity v_y, angular velocity w_z, the bending moment M_z, and the shear force F_y. These variables are related by a wave equation that looks utterly unlike that of the others—the derivative with respect to the spatial coordinate involves a fourth-order derivative, while the derivative with respect to time is of the usual second order. This means that in general, a linear, distortion-free propagation is impossible. Flexural waves are highly dispersive: their higher frequency components travel with faster phase velocities. [7, p. 95]

Two other important types of waves, *Plate* and *Surface* waves, require a more thorough understanding of boundary conditions. Although these wave motions happen freely as an object vibrates, they are most frequently encountered in waveguides. The mechanics behind such waves will be described briefly in 1.5, and their applications will be described in Chapter 2.

1.4 Impedance and Behavior at an Interface

In forward-moving plane waves, the *specific acoustic impedance* is defined as $Z = {-T}/{\dot{u}}$ and bears a relationship with the concept in electronics. In both fields, the term Z is a complex amplitude. In the case of alternating electromagnetic fields, objects like resistors, capacitors, and inductors are often modeled as complex impedances. This way, a simple electronic part can be described by its effects on the phase differences between the voltage V across it, and current I through it.[29, p. 168] In acoustics, the concept is analogous, and the two variables encompassed by complex acoustic impedance are velocity, $\dot{u} = \frac{\partial u}{\partial t}$, instead of current, and stress, $T = \frac{m\ddot{u}}{Area}$, instead of voltage.[7, p. 9] [6]

Just as with electromagnetic phenomena, when an acoustic plane wave meets a perpendicular planar change in impedance (e.g., a resistor, or in the case of light, a mirror), there is a component, T_{B1}, of the incident wave's stress, T_{F1}, that reflects back the way it came, and a component, T_{F2}, that transmits across the interface, continuing in the direction, T_{F1}, had been going. For acoustic waves meeting the interface obliquely, the story is much more complicated. These interactions differ considerably from the interactions exhibited by light at a specular interface, or electronic signals traveling down a transmission line. We will briefly explore the phenomena related to oblique acoustic reflection in section 1.5. Presently, we concern ourselves with the 1 dimensional case of acoustic impedance boundaries, which, when defined in terms of the stress, is very closely matched to the electromagnetic case, and much simpler than the general acoustic case.

In contrast to the specific impedance, which is an intrinsic property of the material in a resting state, the *characteristic impedance*, Z_0, is defined for a material through which a wave propagates. Characteristic impedance is depends on the wave propagating through the medium. For a forward-propagating plane wave with stress vector T_F, particle velocity $\dot{u}_F$, and phase velocity v_F, traveling through a material of unstressed density ρ, the characteristic impedance is

---

[6]N.B. be careful with analogies.

defined as

$$Z_0 = -\frac{T_F}{\dot{u}_F} = v_F \rho. \qquad (1.4.1)$$

To compare this characteristic impedance with that of a backward-propagating plane wave, i.e.,

$$Z_1 = -\frac{T_B}{\dot{u}_B} = -v_B \rho \qquad (1.4.2)$$

we see the sign has been reversed.

To develop the equations that determine the effect of characteristic impedance changes on wave propagation, consider a harmonic, longitudinal, plane wave, traveling in the $+x$ direction through a material of characteristic impedance Z_1, toward an infinitely long interface with a second material of characteristic impedance Z_2. At this interface, the values for the waves related to the stress, T, and particle velocity, $\dot{u}$, must be continuous. These boundary conditions must be obeyed at all times. Therefore, to find the coefficients of reflection and transmission, we first attend to the left side of the interface, summing the contributions from the incident and reflected waves. We try, [18, p. 10]

$$T_1 = T_{F1}e^{-ik_1 x} + T_{B1}e^{ik_1 x}$$
$$\dot{u}_1 = \dot{u}_{F1}e^{-ik_1 x} + \dot{u}_{B1}e^{ik_1 x} \qquad (1.4.3)$$

These are the two stress and particular velocity waves, traveling toward, and reflected from, the interface. The wave number k, and thus wavelength λ, is dependent on the characteristic impedance of the medium. Also notice in the above notation, we are not describing T as a tensor, but as a wave in the x direction. In the entirety of this example, the x directionality of the stress and particle velocity waves is assumed. Continuing the convention established in equations (1.4.1) and (1.4.2), the stresses T_{F1} & T_{B1} and particle velocities $\dot{u}_{F1}$ & $\dot{u}_{B1}$, each correspond to parameters of the forward- and backward-

propagating waves. We are now prepared to define the reflection coefficient Γ:

$$\Gamma = \frac{T_{B1}}{T_{F1}}, \qquad (1.4.4)$$

as the ratio of these two amplitudes. Writing the velocities in terms of stress from equations (1.4.1) and (1.4.2), we can describe the total stress and particle velocity on the left side of the interface, T_1 & $\dot{u}$, in terms of the incident wave's stress contribution and the reflection coefficient Γ, i.e.:

$$T_1 = T_{F1}\left(e^{-ik_1x} + \Gamma e^{ik_1x}\right)$$
$$\dot{u}_1 = -\frac{T_{F1}}{Z_1}\left(e^{-ik_1x} - \Gamma e^{ik_1x}\right) \qquad (1.4.5)$$

Meanwhile, to the right of the interface, a single transmitted wave propagates in the $+x$ direction:

$$T_2 = T_{F2}e^{-ik_2x}$$
$$\dot{u}_2 = -\frac{T_{F2}}{Z_2}e^{-ik_2x} \qquad (1.4.6)$$

Because this interface's boundary conditions require the stress T and velocity $\dot{u}$ to be continuous at the interface, we can rewrite the formula for the reflection coefficient Γ in terms of the two characteristic impedances:

$$\Gamma = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \qquad (1.4.7)$$

and we can define the stress transmission coefficient, $\mathcal{T}_T$, and power transmission coefficient, $\mathcal{T}_P$, as

$$\mathcal{T}_T \equiv \frac{T_{F2}}{T_{F1}} = 1 + \Gamma = \frac{2Z_2}{Z_2 + Z_1}$$
$$\mathcal{T}_P \equiv \frac{P_{F2}}{P_{F1}} = 1 + |\Gamma|^2 = \left(\frac{T_{F2}^2}{Z_2}\right) / \left(\frac{T_{F1}^2}{Z_1}\right). \qquad (1.4.8)$$

The above description, presented using longitudinal waves, is nearly analogous

in the case of shear waves. For a shear wave, the axis of particle velocity is orthogonal to the wave propagation axis, and there would typically be lower values for phase velocity v and impedance Z. However, other than these somewhat minor distinctions, as long as the plane wave's propagation vector is normal to an interface (alternatively, when the angle-of-incidence, $\theta_i = 0$), this simple behavior is observed.[18, pp. 9-14]

We can now test this model by demonstrating the effect of Z_1 versus Z_2 on the reflection and transmission of waves incident on a free boundary, i.e., when $Z_1 \gg Z_2$. Consider a rod extending in the $-x$ direction all the way to $-\infty$. At $x = 0$, this rod ends, and the boundary is completely free. This rod's width, z, and height, y, are small compared to the wavelengths of interest. Thus, wave propagation in this rod is restricted to a single dimension: its length. At the boundary, vibrations in this rod exhibit perfect lossless reflection. Incidentally, reverting to a single dimension also permits us to talk about force F, rather than stress T, which assumes an area. The interface is the outer surface of the object, so $Z_2 \approx 0$, and therefore $\Gamma \approx -1$. Using (1.4.5), we predict that the left side of the interface should have inverted stress waves traveling in opposite directions. Furthermore, again plugging into our formula (1.4.5), the model tells us that the left side of the interface should have non-inverted particle velocity waves, again traveling in opposite directions.

The boundary conditions, which must be satisfied at all times, are

$$x = 0 \qquad \& \qquad F = 0 \tag{1.4.9}$$

for this rod, because the end at $x = 0$ is free to move in a vacuum.

The rod is excited with a longitudinal sinusoidal *wave-train*: a wave that lasts long enough to be considered infinite. The function that describes this wave is $e^{i(\omega t - kx)}$. Since the wave is infinitely long, we don't have to worry about where the wave "starts:" it has no begining or end. Since we assume the system does not produce new frequencies, this wave can be represented as

a *phasor*, which means we will only write the terms inside the exponent. Thus, this wave can be written in terms of its force like so:

$$F_f = (\omega t - kx). \tag{1.4.10}$$

The above scenario seems innocuous enough. However, the boundary conditions have already been violated. At $x = 0$, the force is a sinusoid (like it is everywhere), whose amplitude is not zero at all times. To satisfy the boundary conditions in (1.4.9), we introduce a second wave-train, identical in all senses except for its propagation direction,

$$F_b = (\omega t + kx), \tag{1.4.11}$$

and so the total distribution of force along this fictitious rod, as a function of space and time is:

$$F(x,t) = F_f(\omega t - kx) + F_b(\omega t + kx). \tag{1.4.12}$$

To satisfy the boundary condition, we posit that

$$F_b(t,0) = -F_f(t,0), \tag{1.4.13}$$

which is consistent with our model's prediction above, in which the reflected force wave is inverted at a free end. [7, p. 116]

An interesting property here is that these two wave-trains would sum together such that, at all points, their phases would remain constant. Thus, the stress and velocity components of the disturbance would have phase velocities of zero, and the phenomenon that was once a traveling wave would no longer appear to be traveling at all: it would produce a *standing wave*. There would be points along the rod where the magnitude of the wave is always at a minimum (equilibrium). These points, which expand into lines when we talk about a two-dimensional surface, are called *nodes*. Conversely, there will be other points

along which the wave reaches a maximum once every time cycle. These regions are called *anti-nodes*.

Since the left boundary is at $-\infty$, this rod does not impart any particular restrictions on what wavelengths will satisfy the boundary conditions. As a result, we have to continue putting energy into it in order to see any of this behavior. [7, p.319] But what would happen if we gave the rod a finite length? So long as the relationship between the distance these waves travel and their fluctuation in time remains fixed (see equation (1.2.2)), we can imagine that a given amount of distance, with a given set of boundary conditions, would only allow for certain frequencies of wave phenomena to be reinforced in this way. We will return to this concept when we begin to talk about the vibratory *state* of an object, in section 1.6.

A soft barrier could be modeled as an interface with a material of characteristic impedance Z_2, still smaller than the initial impedance Z_1, but this time, $|\Gamma| < 1$. In this case, when the incident and reflected waves superpose, their wave velocities will not completely cancel, because the reflected wave will only be partially reflected. Thus there will be a traveling component and a standing component apparent within the rod, and only a traveling component on the outside. The traveling component—and therefore, the net energy—will propagate in the direction of the incident wave, with an amplitude equal to their difference. The resulting wave behavior within an object is sometimes referred to as a *partial* standing wave. [15, p. 275]

1.5 General Plane Wave Reflection and Refraction

To better demonstrate the general case of oblique plane waves on an impedance discontinuity, we must restrict our case to isotropic solids. Recall that, for isotropic solids, we can simplify the elasticity coefficient tensor c to just two components, called the *Lamé constants*, μ and λ. These constants are related

to the components of c as follows[18, p. 95]:

$$c_{11} + c_{22} + c_{33} = \lambda + 2\mu$$
$$c_{12} = c_{13} = c_{23} = c_{21} = c_{31} = c_{32} = 2\lambda \qquad (1.5.1)$$
$$c_{44} = c_{55} = c_{66} = \mu$$

In the general case, a pure longitudinal wave incident on a free interface at angle, θ_{li}, subtended from the normal, gives rise to two reflected waves: one longitudinal wave, propagating at an angle θ_{lr}, and one shear wave, propagating at an angle θ_{sr}. The propagation vectors of these three waves lie on a plane called the plane-of-incidence. In order to satisfy the boundary conditions for a free surface, namely that the normal stresses must be zero at the surface, the displacement component of this transverse wave must be orthogonal to the plane-of-incidence. Stress in any other direction would violate the boundary conditions.

In order to satisfy the boundary condition for all points on the surface plane, the phases of all incident waves must be synchronized along the surface. Since the planes of constant phase for the wave are no longer parallel to the surface plane, additional waves must be added to accomplish this.

Because the phase velocity—and therefore wavelength—of a shear wave in a given solid is typically less than that of a longitudinal wave, the angle a shear wave subtends from the normal will always be less than that of the two longitudinal waves, which will be identical. This suggests an acoustic version of Snell's law[18, p. 96] [7, p. 141]:

$$\frac{\sin(\theta_{sr})}{\sin(\theta_{li})} = \frac{k_l}{k_s} = \frac{v_s}{v_l} = \sqrt{\frac{\mu}{\lambda + 2\mu}}, \qquad (1.5.2)$$

where μ is the shear modulus— the elasticity constant for shear vibrations in an isotropic medium.

In optics, Snell's law describes the behavior of incident rays on an interface in terms of reflected and refracted components. However, the boundary conditions

are determined by the Fresnel Equations, and the system behaves differently. While for isotropic solids, shear and transverse polarizations travel at different speeds, there are no examples of such behavior in optics, barring the use of metamaterials. In air-borne acoustics, the medium cannot support shear waves, and does not have this either. Solids and viscous liquids seem to be unique in this.

The amplitudes of these waves are related by the parameters $\Gamma_{ll} = A_{lr}/A_{li}$, and $\Gamma_{sl} = A_{sr}/A_{si}$. They are: [18, p. 98]

$$\Gamma_{ll} = \frac{\sin(2\theta_s)\sin(2\theta_l) - (v_l/v_s)^2\cos(2\theta_s)^2}{\sin(2\theta_s)\sin(2\theta_l) + (v_l/v_s)^2\cos(2\theta_s)^2} \tag{1.5.3}$$

&

$$\Gamma_{sl} = \frac{2(v_l/v_s)\sin(2\theta_l)\cos(2\theta_s)}{\sin(2\theta_s)\sin(2\theta_l) + (v_l/v_s)^2\cos(2\theta_s)^2} \tag{1.5.4}$$

In the case of a shear plane wave incident on a free surface, familiar laws governing the amplitudes and angles subtended from the normal may be derived. Using equation (1.5.2) :

$$\frac{\sin(\theta_{lr})}{\sin(\theta_{si})} = \frac{k_s}{k_l} = \frac{v_l}{v_s} \tag{1.5.5}$$

Keeping in mind that $v_s < v_l$ for most solids, it becomes apparent that the angle θ_{lr} approaches ∞ as the angle θ_{si} vanishes. In real situations, there is a critical angle after which the reflected longitudinal wave disappears entirely. This pressure fluctuation, which must exist to satisfy the boundary conditions, is called a *surface wave*. The analogy to optics is invited once more, as these may be conceived of as subject to "total internal reflection." [7, p. 145]

Surface waves can propagate freely themselves, but their behavior is fairly distinct from other wave types. [7] [7, p. 150] *Rayleigh* waves are a type of free surface wave that exist in the surface of objects thicker than a wavelength of

---

[7] Technically, we have seen two surface waves already: the extensional and flexural waves discussed in section 1.3.

interest. These waves decay exponentially toward the interior of the object. However, these waves do not exhibit dispersion in an isotropic, homogenous solid.[7, p. 152][18, p. 113] Surface acoustic waves have practical applications, which will be discussed in section 2.2. The dynamics involved with these waves is too complex to derive here, but a phenomenological account may be given.

Recall that the boundary conditions at the surface of an object require the normal stress to be zero. The stress components in Rayleigh waves must sum together in such a way as to satisfy this condition. Furthermore, because the wave must have finite energy per unit propagation length, the wave's field variables must decrease in amplitude with the depth. [18, p. 110] The wave equation for such a disturbance, propagating for example in the positive z-direction, is [3, p. 70]

$$ u(x,y,z,t) = (u_x(y)e^{i\varphi_1}x + u_y(y)e^{i\varphi_2}y + u_z(y)e^{i\varphi_3}z)e^{i\omega}t^{-\gamma_z} \qquad , \qquad (1.5.6) $$

where γ is the complex propagation factor- essentially contributing an exponential decay through time. [18, p.112] These waves behave similarly to waves in liquid, because of their elliptical rotation.

1.6   Modes: An Introduction

We have barely dented the surface of the possibly infinite ways a general object can vibrate. Even for objects so simple they cannot be found in the real world, the vibrational energy flow through an object may take a staggering multiplicity of paths, as the above treatment has hopefully demonstrated. Even if we ignore the bizarre effects of the model described in section 1.5, the mere tracing of propagation vectors through an irregularly shaped object is an example of classical chaos. Compounded by this are the effects of polarization on a given vibration's wavelength, and damping. However, despite the complexity of the propagation mechanisms and the myriad ways in which a material can support the transfer of vibrational energy, most objects we encounter exhibit the highly improbable

behavior of accommodating the vibrations of certain frequencies over others.

In 2 dimensions, an analogous problem, frequently studied by mathematicians, is billiards. If a billiard ball—for our purposes a massless point-particle, experiencing lossless specular reflection at the boundaries—travels freely over an enclosed space, the behavior of the system is entirely determined by the shape of the space. If the space is an ideal circle, the paths traced by the billiard ball, i.e., the dynamics of the system, will be regular. After only a short number of reflections, a pattern can clearly be recognized in the propagation of the ball. Furthermore, the lengths of each path segment between collisions with the boundary are identical in this case. However, if the circle is flattened on one side, a completely different, frequently chaotic behavior emerges. [1]

Recall the discussion at the end of section 1.4, where a standing wave was first observed. We continue this simplified example to search for a way to describe those vibrations that are intrinsic to the structure through which they propagate. We can do this by closing the structure, so that its volume is finite. We then proceed by fitting a solution (or perhaps a set of solutions) to the wave equation as we have constructed it, along with the spatial constraints imposed by the boundary conditions, now that we have both of them.

While the previous one dimensional case was constructed with only a single, free, boundary at $x = 0$, a similar behavior could be expected from a finite object, such as a rod with one free end and one fixed end. Let the free end remain where it was in the previous example, and add a fixed end at length $x = l$. At a fixed end, the boundary conditions for the particulate velocity and the stress wave are the reverse of those of a free end, i.e.,

$$ x = l \qquad \text{and} \qquad \dot{u} = 0 \quad , $$

and therefore the force (or stress, if our rod's cross-sections had area) would vary maximally at $x = l$. In this system, one of the two boundaries rotates the wave's phase by π, so therefore, the wave-train requires 4 boundary reflections—two of

each—to return to its original location and phase.

Because this wave-train reaches a boundary, it must reflect back and prop-agate in the opposing direction. Thus, when the two wave-trains sum together, the system forms a standing-wave, all of whose field variables change periodically as a function of time and length.[7, p. 117] The time period of this particular standing-wave is:

$$p = 4l/v \qquad (1.6.1)$$

where v is the phase-velocity of the wave-train in the medium. This is the amount of time required for the wave to travel the length of the rod 4 times. We might be inclined look for more wave-trains which fit these propagation and boundary conditions.[7, p.117]

From the period, we find the frequency of the standing-wave to be $\frac{1}{p}$. Con-sider that the wave-train described above could consist of any wave satisfying the boundary conditions, so long as it still closes in on itself in phase. This system of self-reinforcing, periodic motion might very well support odd integer multiples of the spatial frequency, k, because those are the candidates likely to support the boundary conditions. Of course, whether this is actually true or not depends on the dispersion relation of the vibratory regime in the medium. Furthermore, damping effects might attenuate an otherwise valid frequency. However, at the very least, it might be reasonable to assume that these natural frequencies cluster into families, based on the above simplification.[7, p. 126]

Such a system of self-reinforcing periodic motion is known as a *normal mode*. A normal mode consists of both the wave-train's temporal frequency and its spatial path. Generally, the temporal frequency—which includes instantaneous complex amplitude—is called the *modal frequency*, or *eigenfrequency*. Similarly, the path traced by the closed loop—which also includes instantaneous complex amplitude—is called the *mode shape* or *eigenmode*. Since modes are encountered almost anywhere there is wave motion, they have many epithets. [7, p. 319]

Modes can be characterized by their shapes, which are complex-valued, and

always occur in complex-conjugate pairs. The mode shapes pair with their mode frequencies, which are also complex-valued, and occur in complex-conjugate pairs. This means that every mode requires two complex exponentials to describe it in abstract. The phasor description of a mode, therefore, requires an additional complex term to encode the location-dependency of the shape. To describe a particular state of the mode at a given time requires a third complex term, which describes the complex amplitude (or amplitude and phase-shift) of the system's state with respect to a single mode. Modes can thus describe temporal frequency and damping (or growth), spatial frequency, and complex amplitude. Thus, modes provide solutions to problems such as those discussed in 1.3.

Modes are often designated a number, which describes the amount of half-cycles taken up by the mode across the surface, in a single dimension. The mode whose period is described in equation (1.6.1) is of mode number zero, or, alternatively, a zeroth-order mode. This is the lowest frequency mode accommodated by the system described.

An important quality that modes have is their *completeness*, meaning that any behavior of a system, regardless of whether it is being perturbed by some outside force, can be expressed as a combination of that system's modes. Such a process of finding the coefficients that multiply the modes to result in the present behavior is called a *modal decomposition*. This regression holds true even if the system's behavior is overwhelmingly dominated by outside forces. For a proof of the completeness of modal decomposition, see [6, p. 281].

It is often said, in the literature, that modes are orthogonal. Careful attention must be paid to what is meant by this statement, because it implies that we could simply examine each mode shape, and ascertain that they are perpendicular, for example, by taking the dot product[7, p. 320]:

$$\int_S m'' \varphi_n(x,z) \varphi_m(x,z) \partial x \partial z = 0 \qquad \text{for} \qquad m \neq n, \qquad (1.6.2)$$

where m'' is the mass distribution (which likely involves y and another coordinate), and S is the (x, z) area of a face of the object. This demonstrates modal othogonality perfectly well in an abstract sense. Notice the dependence on the mass distribution, for example. However, an abstract understanding of this fact is not nearly as telling or profound as the physicality it points to. In order to demonstrate this, consider a system vibrating exclusively in some mode n. The total kinetic energy is [7, p. 320]

$$E_{kin} = \frac{1}{2} \int_S \dot{u}_n^2 m'' \varphi_n^2(x, z) \partial x \partial z. \qquad (1.6.3)$$

Meanwhile, the kinetic energy of the object vibrating in another mode m, such that, as in (1.6.2), $m \neq n$, is

$$E_{kin} = \frac{1}{2} \int_S \dot{u}_m^2 m'' \varphi_m^2(x, z) \partial x \partial z. \qquad (1.6.4)$$

If the two modes occur simultaneously, exactly as above in (1.6.3) and (1.6.4), and the modes are independent of one another, then, to avoid creating or destroying energy,

$$\frac{\frac{1}{2} \int_S \dot{u}_n^2 m'' \varphi_n^2(x, z) \partial x \partial z + \frac{1}{2} \int_S \dot{u}_m^2 m'' \varphi_m^2(x, z) \partial x \partial z}{\frac{1}{2} \int_S [\dot{u}_m m'' \varphi_m(x, z) + \dot{u}_n m'' \varphi_n(x, z)]^2 \partial x \partial z} \qquad (1.6.5)$$

which can only be true if (1.6.2) is true. [7, p. 322]

This demonstrates some of the conditions under which the orthogonality of the modes—either as given by variables in the model we have constructed, or as measured by instruments in the field—may not hold. If energy in this system is not conserved—i.e., if the system is not closed, then we cannot assume orthogonality. Frequently, we would like to measure the modal response of a system which is experiencing forced vibration, as in our example in section 1.4. Neither of these systems were closed, and therefore we cannot assume orthogonality of the modes. Furthermore, if the mass distribution is not known,

or is poorly approximated, we similarly cannot assume orthogonality.

1.7   Modal Analysis
via Partial Differential Equations

This technique begins with the assumption that the differential equation of motion in the object is known. Furthermore, the technique requires the assumption that the system is linear, or that the data has been conditioned such that it resembles an analysis of a linear system. The goal, as introduced qualitatively in the previous section, is to describe the behavior of the system as a linear combination of modes. The algorithm is presented without reference to dimensionality. It has been successfully applied to vectors of coefficients as well as matrices. [31, p. 317]

Another perspective: what we have been calling the "modes" are actually the *eigenvectors* of a matrix A, that describes the motion of the system as a *constant-coefficient linear differential equation*. Recall from linear algebra that *eigenvalues* are solutions λ, such that $Ax = \lambda A$, and the eigenvectors are x.[31, p. 283] If we think of a matrix A as a *linear transformation* of a vector, for example a filter, the eigenvectors are vectors that pass through the linear transformation unchanged except for a scale factor, λ, which is the eigenvalue. Our description of a mode shape is almost precisely this: a closed path, through which the eigenvalue rotates the vibration in phase. [23, p. 8] A constant-coefficient linear differential equation could be, e.g., the wave equation as derived in 1.1, however it need not merely be a 2nd-order differential equation. In fact, this equation could be of arbitrary order, up to the practical limitations of the abacus, and the amount of papyrus being used to implement the technique. This is because we assume, once again, the solutions to the equation of interest to be of the form $e^{-i\omega t}$, so differentiating is equivalent to multiplying by the factor $(-i\omega t)$. The modes are eigenvectors will form a basis which diagonalizes the single nth-order differential equation into a system of n 1st-order differential

equations. Such equations are of the general form[27, p. 16]

$$D^n x + a_{n-1} D^n - 1x + \cdots a_1 Dx + a_0 x = f, \qquad (1.7.1)$$

where the operator D is a differentiation with respect to time, x is the current displacement, and the values $a_0 \cdots a_{n-1}$ are the parameters of the differential equation, e.g., in the case of our wave equation, a_0 is the spring constant, a_1 is the frictional coefficient ("damping," in our terminology developed in section 1.2), and a_2 is the mass of the particle. The right-hand side, f, is the residual force. If $f = 0$, the system is not experiencing excitation from the outside. It is in a "homogeneous" state, reacting to the initial conditions $u(0)$. Let's say we wanted to know what this matrix would do after an arbitrarily long amount of time, in response to the initial conditions $u(0)$.

The first step is to convert this n$^{\text{th}}$-order equation into a system of first-order equations using matrices. The matrix to construct will be a *companion* matrix, one that applies the transition from x to $\dot{x}$, etc. Call it A in this general case. It looks like this:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_{n-1} & \cdots & \cdots & -a_1 & -a_0 \end{pmatrix} \qquad (1.7.2)$$

This matrix multiplies a column vector by the negative differential parameters and shifts the values up. In our vector u we will store previous increasing orders of derivatives of x. The entire equation in tableau becomes[27, p. 17]

28

$$D \begin{pmatrix} x \\ \dot{x} \\ \vdots \\ D^{n-1}x \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_{n-1} & \cdots & \cdots & -a_1 & -a_0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \vdots \\ D^{n-1}x \end{pmatrix} = 0 \quad (1.7.3)$$

or, more humanely, $Du = Au$. We can solve the response of the system as a sum of exponentials,

$$u_h(t) = e^{At}u(0). \tag{1.7.4}$$

If the matrix A has a set of unique eigenvalues, λ_n, the corresponding eigenvectors will be the modal vectors S_n. S is a matrix which "diagonalizes" A into Λ, a diagonal matrix with every λ down the middle. In this case, it would take a lot of graph paper to calculate the expansion of the matrix exponential, as its Taylor series involves taking arbitrarily high powers of a non-diagonal matrix. Fortunately, we can calculate the powers of a diagonal matrix very easily, and furthermore,[27, p. 20]

$$A^2 = (S\Lambda S^{-1})^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda\Lambda S^{-1} = S\Lambda^2 S^{-1},$$

so for our exponential function, we may instead write

$$Se^{\Lambda t}S^{-1} \tag{1.7.5}$$

Substituting this in for equation (1.7.4), we get the following:

$$u_h(t) = e^{At} = Se^{\Lambda t}S^{-1}u(0). \tag{1.7.6}$$

We then find the coefficients for the initial conditions using a regression of

the form

$$c = S^{-1}u(0), \qquad (1.7.7)$$

which is essentially a change of basis into the eigenvector basis: c tells us what combinations of S are needed to produce $u(0)$. Then we plug these coefficients back into our formula to find the response:

$$u_h(t) = \sum_{i=1}^{n} c_i e^{\lambda t}(S)_i \qquad (1.7.8)$$

When we examine the eigenvectors and eigenvalues of a constant-coefficient linear differential equation, many things become clear. We can even determine a solution for $t = \infty$. If we analyze the values in Λ, which will likely be complex, we will find that the mode shape in S with the largest corresponding $\Re\{\lambda_i\}$ will dominate, if there is a component of the initial condition c in this vector. If any $\Re\{\lambda_i\} = 0$, the system will tend towards an oscillation along the mode shape in the corresponding S. If any $\Re\{\lambda_i\} > 1$, the system is acausal. [31, p. 318]

Because we typically want to perform the algorithm for a mesh of mass-spring models in a network, we likely would form the vectors u into matrices, and adjust the dimensions accordingly.[27, p. 20]

The decomposition of the forcing function would reduce to:

$$Bu_f = S^{-1}e^{\Lambda t} \int_0^t e^{-D\tau} Sf(\tau)d\tau, \qquad (1.7.9)$$

where $\tau = 2\pi$. The total disturbance due to the forcing function and the homogeneous response is

$$u_f(t) = e^{At}v(0) + S^{-1}e^{\Lambda t} \int_0^t e^{-D\tau} Sf(\tau)d\tau \qquad (1.7.10)$$

This technique has the benefit of minimizing numerical error by avoiding discretization. Since the techniques for the decomposition of a system into its mode shapes, mode frequencies, and coefficients is a numerical process, we must

move from the continuous description of the previous sections, to a discrete one. There is an entire class of numerical techniques for making this leap, which is not the topic of interest presently. The curious or suspicious reader is referred to [31, p. 317] [26] [27] [6], and countless others. In fact, when numerically implemented, the technique described above must at some point discretize its output. The sacred compromise between accuracy and cost is a major factor in choosing one algorithm or another. However, hardware and algorithmic developments certainly make it difficult to predict whether a technique that has been formerly thought intractable will become available in the near future.

1.8   Other Modal Measurement Techniques

There are many numerical techniques for measuring modes in the wild. The above technique itself could made into one, although it is fairly ill-suited to the restrictions imposed by many real-time applications. Not all techniques are appropriate for field work. Many are extremely sensitive to environmental or sensor noise, and many are numerically unstable. The author's own technique, which will be presented and assessed at the close of this document, represents a first attempt at fitting the constraints in his own applications.

Furthermore, there is a distinct difference in constraints that apply to those algorithms designed for *modeling* a given phenomenon, versus those designed for *measuring* one. Up until this point, we have blithely assumed the information regarding spring constants, damping, and mass matrices would be readily available. This is a caveat not only of the method outlined in section 1.7, but others that will be named in this section. For some applications, this is fine, of course. However, the methods for extracting modal information that require some prior knowledge about the object to be measured, and specifically its governing equations, are at a distinct disadvantage in other applications. Similarly, those which require the object be excited by a known signal are unfortunately ill-suited for many applications. That being said, frequently the cost of these

techniques in terms of convenience is compensated by robustness, accuracy, or speed.

The first of these families of techniques we will assess are related to *Proper Orthogonal Decomposition.* [14] [17] [9] The basic process is not specific to the field of structure-borne sound, although each variant implies a slightly different implementation. Other epithets are: *Singular Value Decomposition* and *Principle Components Analysis.* These techniques all begin by arranging column vectors of zero-centered data into a matrix A, and creating a *Covariance matrix*, given as $R = 1/nAA'$, where n is the number of columns, i.e. the length of the recording in samples.

The process proceeds by finding the eigenvectors and eigenvalues of R. The eigenvectors of this matrix R are guaranteed to be orthogonal, because R is guaranteed to be real and symmetric. These orthogonal eigenvectors are called the *Proper Orthogonal Modes*, and the eigenvalues are the *Proper Orthogonal Values.* However, recall the earlier statements made in section 1.6 about the orthogonality of the mode shapes. The modes are (occasionally) orthogonal with respect to the mass matrix of the object. Actually, the modes are only orthogonal in terms of their energy. [7] *POD* attempts to fix this problem by multiplying R by the mass matrix.[14] This is a cold comfort though, because it requires knowledge about the mass matrix, careful placement of the sensors, and perfect energetic isolation of the object.

On the other hand, the Singular Value Decomposition factorizes the matrix as

$$A = U\Sigma W^*, \qquad (1.8.1)$$

where U are the eigenvectors of AA' , and W the eigenvectors of $A'A$. The squared eigenvalues are stored in Σ, and are called the Singular Values. Modern formulations of the Singular Value Decomposition find the eigenvalues of AA' without calculating it directly, and therefore could be used to implement POD (and other algorithms) extremely quickly. [33, p. 368] The SVD can be used

to compute a pseudoinverse when a matrix is singular (and thus no inverse exists), a useful application in *Least Squares* approximations (see below). PCA is essentially a POD, with the vectors ordered by the size of their corresponding λ. The first few principle components will normally contain the majority of the information in a dataset. [10, p. 4]

Another option for mode-solving techniques are those centered around Least Squares methods for estimation. Least Squares is a family of algorithms for regression and prediction, typically of large datasets. Regression and prediction are similar questions, but they are posed differently: regression takes data presumed to already have happened, and fits a line (usually in a fairly high dimensional dataset) that minimizes the mean squared orthogonal distance between the line and all of the points in the dataset. On the other hand, linear prediction first performs a regression and then predicts the next component. This is applied iteratively each time a new component comes in. *Fourier series* are in this family- they fit arbitrarily complicated functions with simpler functions, minimizing error in the least-squares sense. [31, p. 224]

By fitting the previous data vectors and predicting future data, these algorithms essentially do the inverse of the modeling process described in section 1.7. The goal is usually to construct a matrix which predicts the next component based on previous components. Once this is done, the task becomes an eigenvector search. In certain fields, these matrices are very large, and finding eigenvectors can become computationally intense. There have been a number of numerical techniques developed to assist with this complexity.

Krylov subspaces attempt to search for and approximate the eigenvectors of large matrices. These approximate eigenvectors and eigenvalues are called *Ritz* values and vectors. Krylov subspaces are spanned by the vectors formed by power iteration: a vector is chosen either at random or according to the application, and mapped onto matrix A, whose eigenvectors we are searching for. The process of multiplying the vector by increasingly high powers of the matrix, and storing the result at each iteration, continues until the vectors

33

converge within machine precision or some arbitrary number of iterations has been reached. The vectors span a space, called the Krylov subspace. [33, p. 428] However, this basis is not terribly useful because the vectors typically converge, so they are made orthonormal using *Gram-Schmidt*, which constructs an orthonormal basis by subtracting the components of vectors that project onto previously selected bases. [31, p. 234] In practice, doing this all at the end of the algorithm is probably a bad idea, because the numerical error accumulates both with the power iteration and Gram-Schmidt. [33, p. 439] The *Arnoldi* method applies a Gram-Schmidt-like orthonormalization step with each power iteration. It does not store the powers of the matrix at all. In fact, each successive step merely multiplies the previous vector by the matrix and then applies Gram-Schmidt. [33, p. 440]

Dynamic Mode Decomposition [5] is a relatively new technique which uses least-squares prediction to form a companion matrix, virtually identical to the one described in section 1.7. However, rather than placing the coefficients to a pre-determined polynomial in the coefficients block, DMD fills this block with predictor vectors which construct the data with minimum squared error. These are vectors of the form

$$c = (K'K)^{-1}K',\qquad(1.8.2)$$

where K is a matrix of the previous observations, arranged in columns (the example in section 1.7 used rows, which seems more common). This is a classic least-squares prediction formulation, and in cases where K has linearly independent columns the use of the expanded pseudoinverse is not necessary. The Ritz values and vectors (found using Arnoldi) of this companion matrix should correspond to the eigenfrequencies and eigenmodes of the system. [5]

Another approach is to find the SVD of K, and let K_* be K with the data shifted by one column. The diagonalization to solve is now

$$U^* K_* W \Sigma^{-1} = Y \Lambda Y^{-1},\qquad(1.8.3)$$

34

setting $V = UY$. Each column in V should then be scaled by each unique λ, with each subsequent component of the column in V taken to the next integer power of λ. [5]

This method is less sensitive to noise, but only works if the data matrix is full column rank. Another option is to implement what they call the "snapshots" method, which is used if the rows greatly outnumber the columns. This approach takes $K'K$ instead of K, which shrinks the dataset. The full factorization is $K'K = W\Sigma^2 W'$. Then it proceeds with (1.8.3), except that $U = KW\Sigma^{-1}$.[5]

In order to satisfy the boundary conditions, the data for these methods is usually mean-subracted. However, the authors showed that subtracting the mean from the data over the entire sample before taking the DMD is identical to taking the Discrete Fourier Trasform. The DFT is undesirable in this application for several reasons. First of all, it does not compute an error vector, which in this case would contain the forcing function and sensor noise. Second, the DFT cannot calculate drift on timescales larger than its window size. Third, the λs of the DFT have predetermined frequencies, located at the complex roots of unity. When the eigenfrequencies of the system land between these frequencies, the DFT gives misleading and erroneous data. Also, the DFT requires an integer number of cycles to be analyzed, and typically a windowing regime. [5]

There are a number of advantages to this method. It does not require that integer periods be analyzed, and requires no windowing. The algorithm is capable of predicting both linear and nonlinear behavior. It attempts to resolve true modal information, such as exact eigenfrequencies, damping / growth, and eigenmodes. It can resolve frequencies much lower than the block size. One thing that was noticeably missing from the article was a discussion of realtime considerations. [5]

1.9   A Proposed Mode-Solver

We now describe an algorithm for finding modes in multichannel datasets. The algorithm is intended to be maximally application-agnostic: no calibration, no *a priori* knowledge about the system, no precooked bases. Also, while this algorithm has not yet been implemented for realtime operation, it has been designed with this in mind. Of all the techniques for modal decomposition that have been described above, this algorithm bears the most similarity to DMD. However, the discovery that methods of this kind were even considered worthwhile came as a shock, after much of the development and testing had already occurred. Ultimately, it has been inspiring and encouraging to know that the pursuit is not a misguided waste of time.

This technique, like DMD and the PDE technique described in section 1.7, constructs a companion matrix, A, which multiplies the current audio block to predict what the next incoming sample will be on each input. Like both the PDE and DMD method, the matrix is largely a shifting submatrix, which is fed by a prediction submatrix. However, this method treats the data points from all locations on the surface as equally valid predictors of each new sample. This differs from both the PDE technique and from DMD, which make assessments of the (possibly many) channels as independent time series, only later finding eigenvectors that relate them. As a result, this technique leverages the many spatial samples in making a regression analysis, in addition to the time series.

Furthermore, this algorithm differs from DMD because it is a multirate process. The DMD behaves similarly to a DFT, where it is applied on a single block of data. This causes limitations in temporal accuracy, particularly if one wanted to calculate the residual, which would consist of the forcing function and noise. The proposed technique does not have these limitations.

For example, let X be an incoming matrix of data. This is a real-valued matrix with n rows and b columns, where b is the number of spatial measurement points, and n is the length in samples of the largest timescale we care to look

at, momentarily. Each column vector in X is a time series that progresses downward, so the last element in X is the most recent value observed at a given position. Each column vector in X is mean-subtracted. Let C be the covariance matrix defined as follows:

$$C_{i,j} = \sum_{i,j=1}^{(k+1)b} X([n-w-j:n-i],[1:b]) \cdot X([n-w-j:n-i],[1:b])' \quad (1.9.1)$$

This accumulates the variances over a period of w samples and forms a square, symmetric matrix of length $(k+1)b$, which we partition like so:

$$C = \begin{pmatrix} X_0 & X_1 \\ \hline X_1 & \end{pmatrix} \quad (1.9.2)$$

In C, then, the top $kb \times kb$ matrix is the covariance between the previous k windows of b channels. (Once again, the windows are accumulated over w samples.) Call this top-left submatrix X_0. The bottom-left submatrix has b rows and kb columns, and contains the covariance between the previous kb summations of length w, and the most recent values of b. Call this matrix X_1.

Another way to express these matrices, which might make them seem more familiar, is

$$
\begin{aligned}
X_0 &= K'K \\
X_1 &= K'u, \qquad \text{where} \qquad Kc = u.
\end{aligned}
$$
$$(1.9.3)$$

This permits the convenient formation of the normal equation

$$
\begin{aligned}
K'K\hat{c} &= K'u \\
\hat{c} &= (K'K)^{-1}K'u
\end{aligned}
$$

or, using our variables, $X_0^{-1}X_1 = P$, so named because it is our predictor submatrix. which constructs each of the values in x_1 as a linear combination of

all the previous values x_0. We cook this matrix P into a companion matrix of the form

$$A = \left(\begin{array}{c|c} 0 & I \\ \hline & P \end{array} \right) \qquad (1.9.4)$$

If we were performing a typical linear prediction, we would make b predictions, each of which uses the last $k - 1$ data points for each column vector time series. However, this algorithm makes use of the presumed linear dependence of the variables across space as well as time, so our b predictions will each use the last $(k - 1)wb$ data points. The hypothesis is that this would allow us to make a higher order prediction after less time has passed.

The eigendecomposition of this matrix $A = S\Lambda S^{-1}$ is an approximate modal decomposition of the system, in the least squares sense. In the current implementation, they are calculated using an Arnoldi implementation from the software package ARPACK. Regressions of the form described in equation (1.7.7) are possible, if the coefficients of another dataset are required. In this case, the new dataset could be modeled in terms of a combination of the homogeneous response + forcing function, as described in equation (1.7.10).

An initial set of experiments was undertaken to test the validity of the algorithm, and the conditions for convergence of the eigenfrequencies. Multichannel audio signals were synthesized using all-pole filters and white noise. The activations of the eigenfrequencies were randomly distributed across the channels for each signal. This was intended to simulate the coupling between mode shape and measurement location, which must be assumed to be unknown, therefore random. The signals could simulate a number of damping regimes by varying the distributions of the magnitudes of the poles.

A total of 20 parameter states were tested with 2 trials for each state. The stimuli were each 1000 samples long, with 4 channels of data. The two parameters that varied were the damping regimes and number of modes in the stimulus.

The stimuli were grouped by the magnitudes of their poles into 10 ranges

varying from 0-10%, 10-20%, 20-30%, and so on. The stimuli were further divided into 2 groups: one which exhibited 24 modes, the other which exhibited 28. The mode solver always attempted to find 24 eigenfrequencies.

The resulting eigenvalue sets were plotted against each other on the complex plane for each trial. Overall, the results suggest the mode solver performs better at finding modes which are less damped, which is fairly consistent with the findings of others. [5] [9] [17] [14]

The mode solver was also asked to plot the mode shapes as complex contours. Since the stimulus model did not explicitly specify these, they cannot be verified against any data. They just look pretty.

N.B., this algorithm does not suffer from the negative effects of mean-subtraction as referenced in [5], and above in section 1.8. Our method subtracts the mean of the entire data set from each point in the data set, then it breaks the entire data set into smaller matrices of length kb, spaced one sample apart, and proceeds with the DMD on each of those blocks. therefore, each submatrix is not re-centered, but rather permitted to drift on timescales much larger than k.

Therefore the frequency-domain distortions associated with the DFT, i.e. arranging the poles at evenly spaced intervals around the unit circle, will not occur. This technique's behavior, like that of DMD—and presumably other least-squares prediction methods—converges on DFT-like behavior as the time-series order, k, approaches the time-series length of dataset, n. Thus there appears to be a compromise between the number of modes this type of system can find, and the accuracy. Since we do not perform mean-subtraction across column vectors of X, this is perhaps another reason to use wider, rather than longer, datasets. However, this is just speculation at the moment.

Future tests will attempt to find the effects of increasing the number of channels. This will require a different implementation of the stimulus synthesizer. Other aspects of future work include verifying the forcing function estimation and mode shapes.