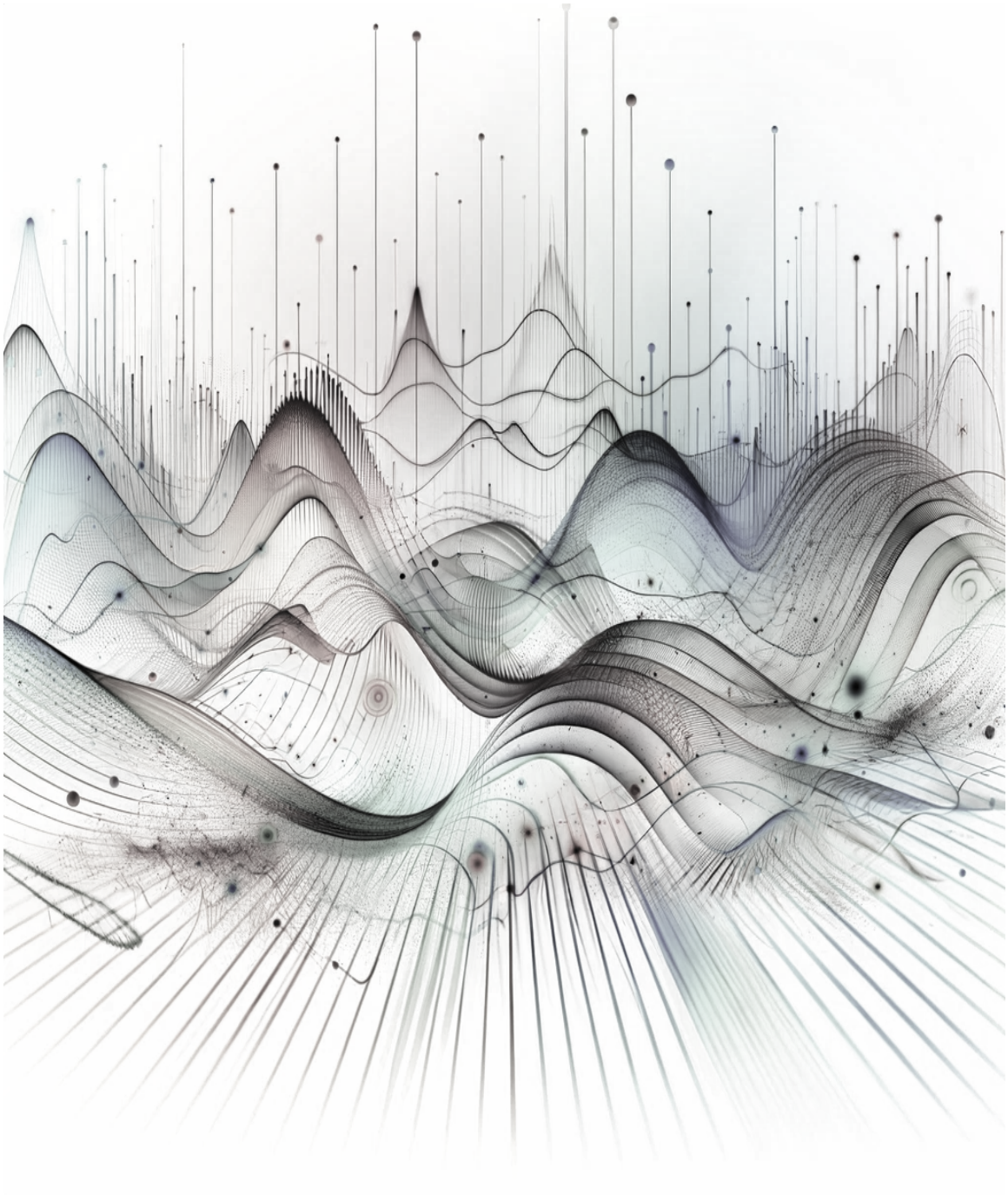
Lecture Notes on Vibrations

Bassam Bamieh

Department of Mechanical Engineering
University of California at Santa Barbara



Preface

These notes are an early draft. I have tried to be very careful with the writing and notation. However, there will be the inevitable typos and some errors. If you find any, please email the author at *bamieh@ucsb.edu*.

Contents

****

Part I

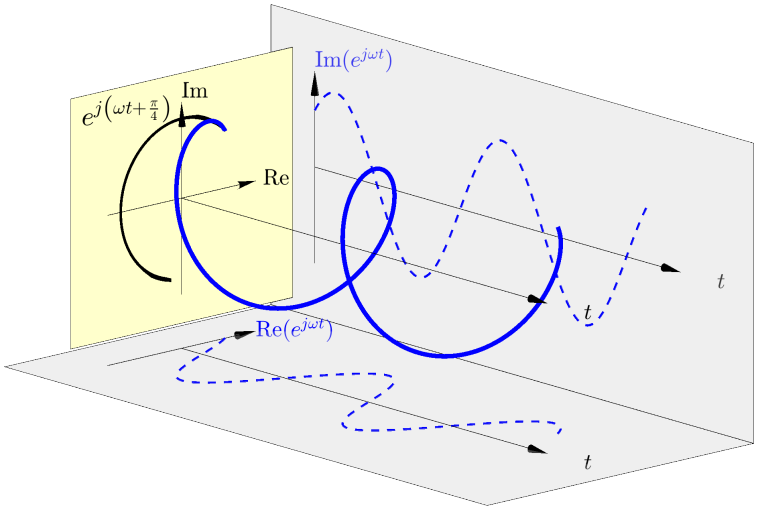
Modeling of Dynamical Systems and Signals

Modeling of Dynamical Systems and Signals

"Dynamical systems" are those whose configurations described by positions, velocities, voltages, currents etc. vary with time. These variations and oscillations are modeled using systems of differential equations, which constitute the dynamical model. These mathematical models predict the physical behavior of the system. Many mechanical systems' vibrations are very well modeled by systems of masses, springs and dampers. It is quite remarkable how sometimes complex mechanical systems can be modeled with just a few of those elements. The key point is that for engineering design, it is not the highest fidelity models that are useful, but rather the most parsimonious models that capture the phenomena and the design criteria.

For simple systems, models using Newton's laws of motion, together with characterization of spring and damper forces, suffice to write down the differential equations governing the system. For more complex systems with many mechanical degrees of freedom, especially those that move in 2D and 3D, Lagrangian Mechanics is a more convenient (and expedient) modeling framework.

"Signals" are the functions of time describing the system's configuration such as positions, velocities, etc. In vibrations analysis, these are typically sinusoidal signals of varying amplitudes and phases. Amplitude and phase relations between signals are crucial to understanding the underlying dynamic phenomena. They are best captured by "phasor analysis", which is basically the encoding of trigonometric identities using the arithmetic of complex numbers. Therefore a thorough understanding of and intuition for complex arithmetic is necessary for understanding these important amplitude and phase relationships.



Chapter 1

Dynamical Models: Masses, Springs and Dampers

The fundamental physical property that causes mechanical vibrations is some type of "restoring force" such as elasticity, or other forces that act similarly. It is surprising how many complex mechanical systems with a variety of component parts can be effectively modeled using an idealization of a mass connected to a fixed or moving structure via a simple spring and a damper. In this chapter, we will explore this type of modeling, which ultimately leads to differential equations whose solutions give the displacements, velocities and accelerations of mechanical vibrations. This chapter is concerned mainly with dynamical modeling*, i.e. given an arrangement of mechanical elements, derive the differential equations describing the dynamics. We do not discuss how to solve those equations in this chapter. Rather, subsequent chapters will explore solution properties of the types of equations we encounter here.*

1.1 Restoring Forces: Elastic Springs and Analogous Forces

The linear spring is the basic elastic element. The simplest version of a spring is characterized by Hooke's law[1] which states that the extension/compression Δ of a spring is proportional to the applied force F

$$\frac{F}{\Delta} \;=\; k \qquad \Leftrightarrow \qquad F \;=\; k\,\Delta \qquad \Leftrightarrow \qquad \frac{F}{k} \;=\; \Delta, \tag{1.1}$$

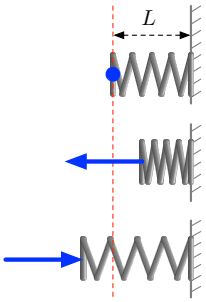
where k is the spring constant, which is also called the stiffness of the spring. The equation above justifies the term "stiffness" since a higher force is needed to achieve the same deflection Δ if k is increased.

The equation $F = k\Delta$ says that k is a conversion factor from displacement to force, and therefore k has units of force over distance. Values of F are typically given in Newtons/meter (N/m). Springs come in all shapes and sizes, and spring constants can vary widely from tens of thousands of N/m in vehicle suspensions to one-thousands N/μm in MEMS devices.

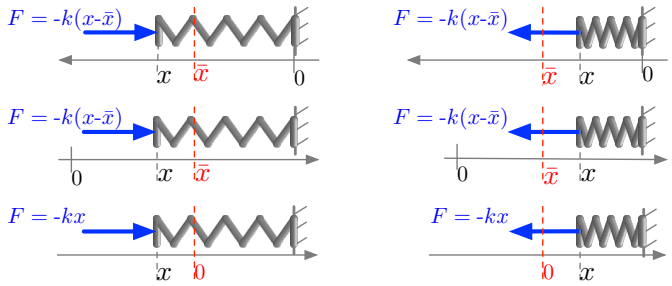
The relation (1.1) is incomplete since it does not specify the directionality of deflection relative to the forcing. The spring force always acts in a direction opposite to the spring's deflection as shown in Figure 1.1a. For a more precise description, we need to choose a coordinate system as shown in Figure 1.1b. This figure assumes that motion is confined to a single dimension. The variable x denotes the position of the movable end of the spring, while the other end is held in place. As the figure shows, Hooke's law states that the force exerted *by the spring* in this configuration is always given by the formula

$$F \;=\; -\,k\,(x - \bar{x})\,, \tag{1.2}$$

[1] Named after Robert Hooke, an English scientist who lived in the late 17th century and a contemporary of Newton.



(a) The force exerted by a spring on an object attached to it is shown in blue. It acts in a direction opposite to the spring's deflection away from its equilibrium (unstretched/uncompressed) length L. If the spring is compressed, the force is away from the spring (push). If it is extended, the force is towards the spring (pull).



(b) Various coordinate systems can be used to express a spring's deflection and the resulting force. Let $\bar{x}$ denotes the coordinate of the spring's movable end when uncompressed/unstretched, and x denote that coordinate in any other configuration. The spring's force is always given by $F = -k(x - \bar{x})$ regardless of the choice of coordinate system as shown above. However, depending on the direction of the coordinate axis, negative or positive values of $x - \bar{x}$ may imply either compression or extension. When possible, choosing the origin of the coordinate system so that $\bar{x} = 0$ (bottom figure) gives the simpler formula $F = -kx$.

Figure 1.1: A linear (Hookean) spring exerts a force that is proportional to its deflection from its *equilibrium* (uncompressed/unstretched) length. The constant of proportionality k is the "stiffness" of the spring. Various coordinate systems can be used to characterize the spring's forces. The reader should keep in mind that the forces depicted are exerted *by the spring*, not the forces exerted *on the spring* (which would be equal and opposite to those shown).

where $\bar{x}$ is the coordinate of the point of contact (between the spring and whatever it is acting on) when the spring is unstretched and uncompressed, i.e. in so-called *equilibrium*. The formula (1.2) holds regardless of the direction or the origin of the coordinate system. Three different choices are shown in Figure 1.1b, and we make the following observations for each.

- Figure 1.1b top: The origin is at the fixed end of the spring, and $\bar{x}$ is the position of the movable end *at equilibrium*. Thus $\bar{x}$ is actually the equilibrium length of the spring. The quantity $x - \bar{x}$ is the *spring's extension*, which is positive if the spring is extended, and negative if it is compressed as shown in the diagram.
- Figure 1.1b middle: The origin is chosen arbitrarily, but once that choice is made, $\bar{x}$ marks the coordinate of the movable end *at equilibrium*. With this choice of the coordinate axis direction, $x - \bar{x}$ is the *spring's compression*, i.e. positive if the spring is compressed and negative if it is extended.
- Figure 1.1b bottom: The origin is chosen so that $\bar{x} = 0$, i.e. at equilibrium the movable end is at $x = 0$, and therefore deviations from equilibrium are simply given by x. The force law then has the simple form $F(x) = -kx$.

The dynamics of bodies connected by springs can now be written down using Newton's second law, which leads to differential equations for positions and velocities. A mass is attached to a spring as shown in Figure 1.2. We choose a coordinate system, and select any point on the mass to mark its time-varying position by $x(t)$. The "reference position" $\bar{x}$ is the coordinate of that point when the spring is unstretched/uncompressed. The time-varying spring force is then $F(t) = -k\big(x(t) - \bar{x}\big)$, and Newton's second law states that

$$m\,\ddot{x}(t) \;=\; -\,k\,\big(x(t) - \bar{x}\big), \tag{1.3}$$

Note that while the position $x(t)$ is time-dependent, the reference position $\bar{x}$ is constant. This equation can be simplified by defining a new position variable $\tilde{x}(t)$

$$\tilde{x}(t) \;:=\; x(t) - \bar{x} \qquad \Rightarrow \qquad \ddot{\tilde{x}}(t) \;=\; \ddot{x}(t) \qquad \Rightarrow \qquad m\,\ddot{\tilde{x}}(t) \;=\; -\,k\,\tilde{x}(t) \tag{1.4}$$

This last differential equation is a little simpler to deal with than (1.3). Once this equation is solved, the solution to the original differential equation can be reconstructed from

$$x(t) \;=\; \tilde{x}(t) \;+\; \bar{x}.$$

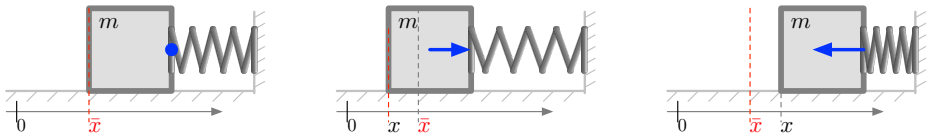


Figure 1.2: A mass is attached to a spring and the origin of the coordinate system is chosen arbitrarily. Any point on the rigid mass can be used to characterize its motion. Here the point chosen is the left end of the mass. That point is marked as $\bar{x}$ when the spring is in equilibrium and exerts no force on the mass (left figure). In any other configuration, the spring's force is given by $F = -k(x - \bar{x})$ since the difference $x - \bar{x}$ is also the compression of the spring.

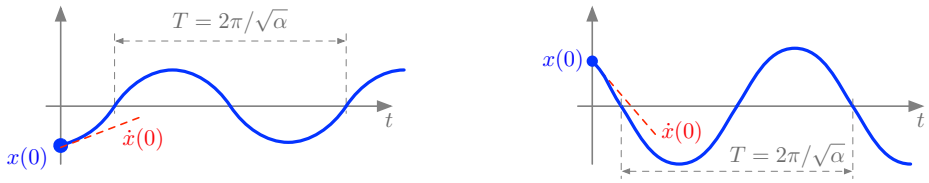


Figure 1.3: Solutions of a differential equation of the form $\ddot{x}(t) = -\alpha x(t)$ (where $\alpha > 0$) are oscillatory as a pure sinusoid with frequency $\sqrt{\alpha}$ rad/s. The amplitude and the phase of the sinusoid are determined from the initial conditions $x(0)$ and $\dot{x}(0$, but the "natural frequency" $\omega_n = \sqrt{\alpha}$ depends only on the parameter α in the differential equations and not on the initial conditions. The figure depicts two different solutions from two different initial conditions. The period of oscillation T is determined by the frequency as $T = 2\pi/\sqrt{\alpha} = 2\pi/\omega_n = 1/f$ where $f = \omega_n/2\pi$ is the frequency in cycles/second (i.e. Hz).

Alternatively, we can also choose the coordinate system from the start so that $\bar{x} = 0$, i.e. so that $x = 0$ corresponds to the spring being in equilibrium. In this case we can immediately write down an equation like (1.4), but for $x(t)$ rather than needing to define another variable

$$\begin{array}{rrcl} & m\,\ddot{x}(t) & = & -k\,x(t) \\ \Leftrightarrow & \ddot{x}(t) & = & -(k/m)\,x(t) \end{array} \qquad \Leftrightarrow \qquad m\,\ddot{x}(t) + k\,x(t) = 0. \tag{1.5}$$

The only difference between the solutions of (1.5) and those of (1.3) is the constant $\bar{x}$, which comes from shifting the origin of the coordinate system by $\bar{x}$. The choice of coordinate systems can significantly simplify the resulting differential equations for more complex systems with multiple masses. This is discussed in detail in Section 1.2.

Equation (1.5) is our first differential equation describing vibrations of a mass connected to a stationary anchor by a spring. You might recall from your differential equations courses that this equation has oscillatory solutions (when $m, k > 0$, which is the case here), with oscillation frequency of $\sqrt{k/m}$. We will study such solutions and their properties in detail in Chapter 3, but for now we state the general solution as

$$x(t) = A\,\cos\big(\omega_n t + \theta\big), \qquad \omega_n = \sqrt{\frac{k}{m}}, \tag{1.6}$$

where the oscillation frequency (in rad/s) ω_n is called the *natural frequency* of the system, A is the oscillation amplitude and θ is its phase. A and θ are determined by initial conditions (initial extension of the spring and initial velocity of the mass), while ω_n does not depend on initial conditions but only on the system parameters k and m.

Remark 1.1. In Chapter 3 we will derive the solution of any second order differential equation of the form

$$\ddot{x}(t) = -\alpha\,x(t), \tag{1.7}$$

where $\alpha > 0$ is any positive number. The solution of such an equation is always of the form

$$x(t) = A\,\cos\big(\omega_n t + \theta\big), \qquad \omega_n := \sqrt{\alpha}, \tag{1.8}$$

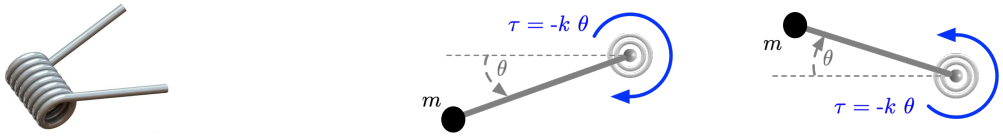


Figure 1.4: (*Left*) A picture of a torsional spring. (*Right*) Schematics of torsional springs. A torsional spring produces a *torque* $\mathcal{T}$ that's always opposite to its angle of deflection from equilibrium. A Hookean torsional spring produces a torque that is proportional to the angle of deflection $\mathcal{T} = -k\theta$. k is the stiffness of the spring and has units of torque/angle, e.g. Nm/deg. The dashed line above represents the *equilibrium angle* of the spring.

where A and θ, the so-called amplitude and phase respectively of the solution, depend on the particular initial conditions given for the problem. The solutions are always oscillatory with frequency $\omega_n = \sqrt{\alpha}$ *regardless of the initial conditions*. Two such solutions are shown in Figure 1.3. Note that the period of oscillation $T = 2\pi/\omega_n = 2\pi/\sqrt{\alpha}$, where ω_n is expressed in radians/second. Alternatively, $T = 1/f_n$, where $f_n = \omega_n/2\pi = \sqrt{\alpha}/2\pi$ is the frequency expressed in cycles/second (i.e. Hz).

1.1.1 Torsional Springs

When considering rotary motion, the equivalent of the spring element is the torsional spring depicted in Figure 1.4. For torsional springs, the "force law" is a relation between *torques* and *angles*. This is the counterpart of the relation between force and distance for standard springs. A "Hookean" torsional spring produces an opposing torque τ that is proportional to the angle θ of deflection from equilibrium

$$\tau \;=\; -\,k\,\theta. \tag{1.9}$$

The constant k is the "stiffness" of the torsional spring, and has units of torque per angle, such as Newton meters per degree for instance.

Recall that for rotational motion, Newton's second law relates angular acceleration $\ddot{\theta}$, mass moment of inertia J, and the applied torque τ

$$\mathrm{J}\,\ddot{\theta} \;=\; \tau.$$

Therefore if a structure with moment of inertia J is attached to a torsional spring with stiffness k, the differential equation that describes rotational motion is given by

$$\mathrm{J}\,\ddot{\theta}(t) \;=\; -\,k\,\theta(t) \qquad \Rightarrow \qquad \ddot{\theta}(t) \;=\; -\,(k/\mathrm{J})\,\theta(t). \tag{1.10}$$

This equation is of the form (1.7), and therefore its solutions are of the form (1.8)

$$\theta(t) \;=\; A\,\cos\big(\omega_n t + \theta\big), \qquad\qquad \omega_n \;=\; \sqrt{k/\mathrm{J}},$$

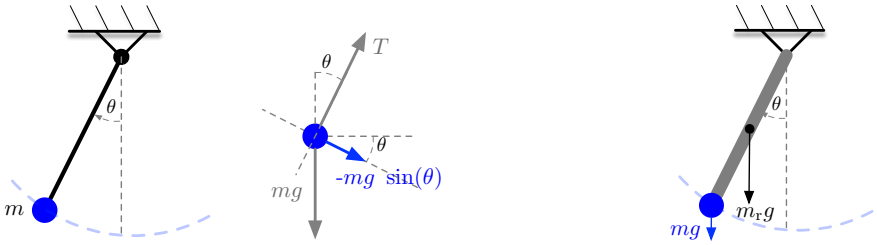
where the natural frequency ω_n is now determined by the ratio of the torsional spring stiffness to the body's moment of inertia about the rotation point. Again, the amplitude A and phase θ of the oscillation are determined by the initial conditions $\theta(0)$ and $\dot{\theta}(0)$.

Consider for example the configuration in Figure 1.4 where the rod connecting the mass m to the torsional spring is assumed massless, and the dimensions of the mass m itself are negligible. The moment of inertia then is simply $\mathrm{J} = ml^2$, where l is the length of the rod. The equation of motion is then

$$ml^2\,\ddot{\theta}(t) \;=\; -\,k\,\theta(t).$$

It should be noted that while a linear spring constant has units of force per linear length (e.g. N/m), a torsional spring constant has units of torque per angle (e.g. Nm/rad).

The reader should now compare Equations (1.10) and (1.5). They are very analogous mathematically. The angular position coordinate θ plays the same role as the linear position coordinate x, the mass moment of inertia J plays the same role as the mass m, and k indicates stiffness (though of two different types) in both cases. You would then expect the dynamical behavior of (1.10) to be exactly the same as (1.5), only the names have changed. The next examples gives yet more analogies.



(a) A mass m connected to a pivot via an inextensible string or a massless rod. The connecting element does not have its own dynamics since it is massless, but kinematically insures that the mass travels in a circular motion. Therefore the rod tension T cancels the radial component of gravity, and only the tangential component of gravity determines tangential acceleration and rotational motion as shown in the free-body diagram.

(b) If the connecting element is a rigid body with significant mass m_r, then its moment of inertia J as well as its gravity torque $-m_r g\ (l/2)\sin(\theta)$ must be taken into account to yield the force balance as expressed in Equation (1.12).

Figure 1.5: Two different models of the simple pendulum with the connecting element mass either ignored or accounted for.

1.1.2 Pendula

A common system in which oscillatory motion occurs is the free-swinging pendulum, which has no elastic forces, but gravity plays a role similar to spring forces. The pendulum is shown in Figure 1.5. A string or a rigid rod connects a mass m to a free, frictionless hinge attached to a frame. We present two different methods to derive the equations of motion depending on the assumptions about the connecting element.

1. If the connecting rod or string is assumed massless, we can use linear force balance as shown in Figure 1.5a. The rod/string is assumed inextensible. In this case, the motion of the mass is circular, and therefore the rod/string tension T and the radial component of gravity cancel out. The motion is then purely circular and completely determined by the tangential component of gravity which for any angle θ is given by

   $$-mg\ \sin(\theta)$$

   as shown in the free-body diagram. Notice the negative sign. Whichever convention we use for the angle orientation (i.e. positive for clockwise or positive for counter-clockwise), the tangential component of gravity acts in the opposite direction to the deflection angle.

   The equation of motion for this simple pendulum is then

   $$m\ \big(l\ddot{\theta}(t)\big)\ =\ -\,mg\ \sin\big(\theta(t)\big) \tag{1.11}$$

   where $l\ddot{\theta}$ is the tangential acceleration of the mass.

2. If the rod has non-negligible mass m_r, then we must use torque balance to account for the moment of inertia J of the rod. The equation of motion is then

   $$\underbrace{\big(ml^2+\mathrm{J}\big)}_{\substack{\text{combined moment of}\\ \text{inertia of } m \text{ and rod}}}\ \ddot{\theta}(t)\ =\ -\ \underbrace{mg\ l\ \sin\big(\theta(t)\big)}_{\substack{\text{gravity torque of}\\ \text{the mass } m}}\ -\ \underbrace{m_r g\ (l/2)\ \sin\big(\theta(t)\big)}_{\substack{\text{gravity torque of}\\ \text{the rod}}}$$
   $$=\ -\,gl\,(m+m_r/2)\ \ \sin\big(\theta(t)\big), \tag{1.12}$$

   where $l\sin\theta$ and $(l/2)\sin\theta$ are the moment arms of the mass m and rod m_r respectively. Note that the torques have opposite signs to the angular acceleration. Also note that if $m_r = 0$ and $\mathrm{J} = 0$ (massless rod assumption), Equation (1.12) reduces to Equation (1.11) after cancelling the l common factor from both sides.

Most pendula are operated in the regime of small oscillations, say $|\theta| \leq 10^\circ$. In this regime, the approximation $\sin(\theta) \approx \theta$ is quite good[2], and can be used to significantly simplify the analysis of the dynamics. For example, using this assumption, we can simplify (1.11) to give the *linearized* dynamics

$$\ddot{\theta}(t) \;=\; -\,(g/l)\;\theta(t). \qquad \text{(linearized dynamics for small } \theta) \tag{1.13}$$

Although the simple pendulum has no springs, its dynamical behavior is exactly like a Mass-Spring system.

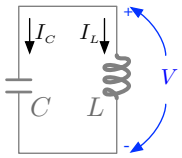
1.1.3 Mathematical Analogies

You should note now that Equation (1.13) "looks" exactly like (1.10) and (1.5). Let's compare the dynamics of all three systems, the linear spring, the torsional spring, and the pendulum together to see the analogy. Although each equation involves two constants, what really matters is the ratio of these two constants

$$\begin{aligned}
\ddot{x}(t) \;&=\; -\,(k/m)\;\;x(t), && \text{(mass/spring in linear motion)}\\
\ddot{\theta}(t) \;&=\; -\,(k/\mathrm{J})\;\;\theta(t), && \text{(mass/torsional-spring in rotational motion)}\\
\ddot{\theta}(t) \;&=\; -\,(g/l)\;\;\theta(t). && \text{(pendulum in rotational motion)}
\end{aligned}$$

Examine for instance the pendulum equation, where l acts like the mass m in the first equation, and gravitational acceleration g acts like the spring stiffness k. For small oscillations, the pendulum can be thought of as comparable to a mass-spring system. Rotational motion of the pendulum away from the downwards position elevates the mass, and gravity acts to restore the pendulum angle to zero (downwards position) just like a spring tries to restore a mass' position to its unextended length. The same pendulum taken from earth to the moon, where g is smaller, will oscillate like a mass connected to a weaker spring, and if taken to Jupiter, where g is much larger, will oscillate like a mass connected to a much stiffer spring. Finally, note that the length of the pendulum acts like an inertia term. The longer the pendulum arm, the more "apparent inertia" the pendulum has to rotational motion.

The analogies above are useful for applying dynamical intuition from spring-mass systems to other mechanical, or even electrical systems. To illustrate this last point consider the circuit shown here where an initial charge across the capacitor, or an initial current through the inductor ends up causing an oscillation. Denote the voltage across the capacitor and inductor by V (they are forced by the circuit wiring to have the same voltage across them), the current through capacitor as I_C, and the current through the inductor as I_L. Recall that the voltage-current relationships for capacitors and inductors are $C\dot{V} = I_C$ and $L\dot{I}_L = V$. The circuit wiring implies that $I_C = -I_L$. We can combine all three relations to get a differential equation for the voltage, or for either of the two currents as follows



$$\left.\begin{aligned} C\,\dot{V} &= I_C \\ L\,\dot{I}_L &= V \\ I_C &= -I_L \end{aligned}\right\} \;\Rightarrow\; \left\{\begin{aligned} &C\,\dot{V} = -I_L \;\;\Rightarrow\;\; C\,\ddot{V} = -\dot{I}_L \;\;\Rightarrow\;\; C\,\ddot{V} = -\tfrac{1}{L}\,V \;\;\Rightarrow\;\; \ddot{V} = -\tfrac{1}{LC}\,V \\ &L\,\dot{I}_L = V \;\;\Rightarrow\;\; L\,\ddot{I}_L = \dot{V} \;\;\Rightarrow\;\; L\,\ddot{I}_L = \tfrac{1}{C}\,I_C \;\;\Rightarrow\;\; \ddot{I}_L = -\tfrac{1}{LC}\,I_L \end{aligned}\right. \tag{1.14}$$

Note that since $I_C = -I_L$, the last equation also gives $\ddot{I}_C = -\frac{1}{LC}\,I_C$. Any of these three equations have oscillatory solutions with frequency $\sqrt{1/LC}$ whereby any initial charge across the capacitor gets discharged as a current through the inductor, which then recharges that capacitor in an infinite back-and-forth cycle in the absence of any resistance.

[2]Recall also that the Taylor series for $\sin(\theta)$ around $\theta = 0$ is $\sin(\theta) = \theta - \frac{1}{3}\theta^3 + \cdots$. Thus this approximation is just the first-order term in the Taylor series.

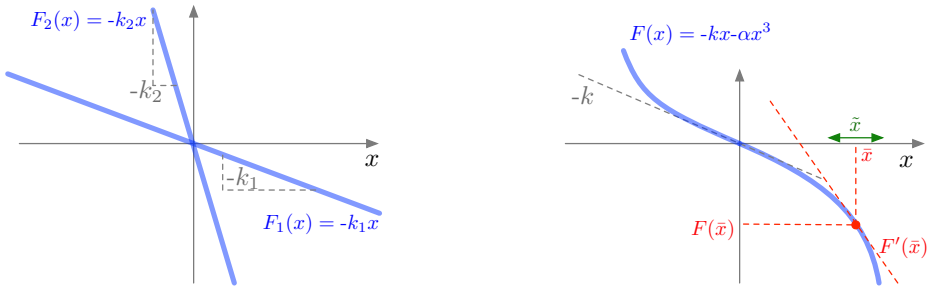


Figure 1.6: (*Left*) A linear (Hookean) spring has a force law $F(x) = -kx$, where x is the change of the spring's length from its equilibrium. The "stiffness" k is the slope of this graph. The stiffer the spring, the larger the slope, and consequently the higher the force produced by a certain deflection. Here $k_1 \ll k_2$ represents a "loose" spring compared to k_2 which represents a stiffer spring. (*Right*) A canonical example of a "stiffening" spring is mathematically modeled by the addition of a cubic term in the force law $F(x) = -kx-\alpha x^3$. For small deflections $x \approx 0$, the spring behaves like a linear spring with $F(x) \approx -kx$. For large deflections, the slope becomes larger, and thus the spring becomes stiffer. For small oscillations $\tilde{x} := x - \bar{x}$ around a mean deflection of $\bar{x}$, the oscillations of the force $F(x) - F(\bar{x}) \approx F'(\bar{x})\tilde{x}$ follow a linear spring, but with the higher stiffness $F'(\bar{x})$.

Equations (1.14) are exactly of the same form as the previous equations where the coefficient $1/LC$ plays the role of k/m in the linear mass-spring system for example. We will see later when we study electrical/mechanical analogies (Chapter **??**) that one analogy is to regard L as an inertia, and of $1/C$ as analogous to a spring's stiffness[3] so that $1/LC = (1/C)/L$. Alternatively, we can regard C as an inertia, and $1/L$ (sometimes called the inductor's "reluctance") as analogous to a spring's stiffness so that $1/LC = (1/L)/C$. Although these analogies are purely mathematical since they arise by examining the structure of differential equations, they are useful since they allow for applying mechanical intuition to electrical circuits and vice versa.

1.1.4 Nonlinear Springs, Forces and Linearizations

Hooke's law is usually only an approximation (although a very useful one) to elastic phenomena in general. In reality, the behavior of a spring or other elastic elements can be more complicated. Consider the graphical representation of Hooke's law in Figure 1.6 (left). The force $F(x)$ as a function of position x is a linear function

$$F(x) \;=\; -\,k\,x,$$

with slope $-k$. Stiffer springs with higher values of k correspond to linear graphs with larger slopes.

Graphs like those shown in the figure can actually be obtained experimentally. A load-cell (which measures force) can be attached to a spring, the spring gets pulled or pushed, the force recorded by the load cell, and the resulting deflection can be measured. When done at many values of force and deflection, one obtains the graph of force $F(x)$ versus deflection x. One common type of force law is that of a "stiffening spring". A simple mathematical model of such a spring is given by adding a cubic term to Hooke's law

$$F(x) \;=\; -\,k\,x \;-\; \alpha\,x^3, \tag{1.15}$$

where $\alpha > 0$ is some constant. A graph of such a force law is shown in Figure 1.6 (right). For small deflections $x \approx 0$, the spring acts like a Hookean spring with $F(x) \approx -kx$. This is because the linear term dominates the cubic term around $x \approx 0$. For larger deflection, the cubic term dominates, and higher forces are required to produce the same deflections, thus the term "stiffening" spring.

An insight into general force laws like that in (1.15) can be obtained considering small oscillations around some large extension or compression $\bar{x}$. Suppose a mass m attached to a nonlinear spring is subjected to an external, constant force $\bar{F}_c$. At equilibrium, the spring will have a deflection $\bar{x}$ such

[3] In fact, the inverse $1/C$ of capacitance is called "electrical elastance" precisely because of this analogy.

that the spring and constant external forces are balanced

$$F(\bar{x}) \;=\; \bar{F}_{\rm c}.$$

Now consider small oscillations around $\bar{x}$. Those oscillations are characterized by the "deviation from $\bar{x}$" defined as $\tilde{x} := x - \bar{x}$. Since $\ddot{x}(t) = \ddot{\tilde{x}}(t)$ we can write

$$m\,\ddot{\tilde{x}}(t) \;=\; F\big(x(t)\big) \;-\; \bar{F}_{\rm c}. \tag{1.16}$$

The right hand side can be approximated by using the first two terms in the Taylor series expansion of F around $\bar{x}$

$$\begin{aligned} F(x) \;&=\; F(\bar{x}) \;+\; F'(\bar{x})\;(x-\bar{x}) \;+\; O\big((x-\bar{x})^2\big) \\ \Rightarrow \qquad F(x) - F(\bar{x}) \;&\approx\; F'(\bar{x})\;\tilde{x}. \end{aligned}$$

With this approximation, and keeping in mind that $F(\bar{x}) = F_{\rm c}$, Equation (1.16) becomes

$$m\,\ddot{\tilde{x}}(t) \;=\; F'(\bar{x})\;\tilde{x}(t). \tag{1.17}$$

Note that the derivative $F'(\bar{x})$ acts like a "stiffness" in this equation. Thus in regions of the graph of $F(x)$ where small oscillations are encountered, the derivative $F'(x)$ acts like a "local stiffness", and the dynamic behavior (for small oscillations) is like a Hookean spring with stiffness $F'(\bar{x})$. This is illustrated in Figure 1.6 (right).

Linearization of General Nonlinear Forces

The linearization technique described for springs above is a general one. We now describe the linearization procedure in general for single-degree-of-freedom systems where Newtown's second law takes the general form

$$M\,\ddot{q}(t) \;=\; \mathcal{F}\big(q(t)\big), \tag{1.18}$$

where q is a "generalized coordinate" (e.g. a linear position coordinate such as x in (1.5), or an angular coordinate θ such as in the pendulum equation (1.11)). M is a "generalized mass" which might be an actual mass or a moment of inertia. $\mathcal{F}(.)$ represents the total sum of conservative forces, which must be a function of only the generalized coordinate q. For linear systems, $\mathcal{F}(.)$ is a linear function of q (such as in the Hookean spring), but $\mathcal{F}(.)$ may be more general as we saw in the "stiffening spring" example earlier. In the pendulum example (1.11), $q := \theta$, and $\mathcal{F}\big(\theta(t)\big) = -mg\sin\big(\theta(t)\big)$, which is clearly a nonlinear function of θ.

If there is a coordinate $\bar{q}$ such that the net force $\mathcal{F}(\bar{q}) = 0$, then this coordinate represents an *equilibrium* of the system. This means that if the system starts (say at time $t = 0$) at the equilibrium coordinate $q(0) = \bar{q}$, and with zero velocity $\dot{q}(0) = 0$, then the system will stay there so that $q(t) = \bar{q}$ for all $t \geq 0$. Now *small oscillations* around this equilibrium are described by the "shifted coordinate"

$$\tilde{q}(t) \;:=\; q(t) - \bar{q},$$

and their dynamics approximately obey the differential equation obtained from (1.18) by using the Taylor series expansion of $\mathcal{F}$ around $\bar{q}$ and ignoring all terms higher than first order

$$\begin{aligned} M\,\ddot{\tilde{q}}(t) \;&=\; M\left(\frac{d^2}{dt^2}\big(q(t)-\bar{q}\big)\right) \\ &=\; M\,\ddot{q}(t) && \text{(since } \bar{q} \text{ is constant in } t) \\ &=\; \mathcal{F}\big(q(t)\big) \\ &=\; \mathcal{F}\big(\bar{q}\big) \;+\; \mathcal{F}'(\bar{q})\,\big(q(t)-\bar{q}\big) \;+\; \text{higher-order-terms} \\ & && \text{(Taylor series expansion of } \mathcal{F} \text{ around } \bar{q}) \\ &=\; \cancelto{0}{\mathcal{F}(\bar{q})} \;+\; \mathcal{F}'(\bar{q})\;\tilde{q}(t) \;+\; \text{higher-order-terms in } \tilde{q} && (\textit{equilibrium} \Leftrightarrow \mathcal{F}(\bar{q}) = 0) \\ \Rightarrow \quad M\,\ddot{\tilde{q}}(t) \;&\approx\; \mathcal{F}'(\bar{q})\;\tilde{q}(t) && \text{(approximate equality since higher-order-terms ignored)} \end{aligned}$$

Thus for small oscillations, the approximate dynamics are like a Hookean spring with stiffness $\mathcal{F}'(\bar{q})$. This is precisely what we did with the stiffening spring (1.17), where the equilibrium is given by the static balance between the nonlinear spring force and the external force $\bar{F}_c$. This was also done in the pendulum example (1.13) where the equilibrium was due to having zero net forces on the pendulum when it is in the downward position $\bar{\theta} = 0$.

1.2 Coordinate Systems

Recall the change of variables from $x(t)$ to $\tilde{x}(t)$ in Equation (1.4) for a Mass-Spring system. This change of variables resulted in a simplified differential equation without constant terms. This is a specific instance of a general method where we always choose coordinate systems so that the differential equations are in their simplest form. In particular, we usually choose coordinate systems to so that constant forces, i.e. those that do not vary with time and thus have no effect on vibration analysis, are eliminated. Such constant forces may arise due to gravity for example, or due to pre-compression or pre-extension of springs. The following basic rule serves as a guideline for how to choose such convenient coordinate systems.

Rule 1.2. For any system of n masses connected by any number of springs, choose coordinate systems for each of the masses positions $x_1, \dots, x_n$ such that $x_1 = 0, \dots, \ x_n = 0$ correspond to static equilibrium.

This rule is illustrated in the next few examples. The reader should now note that we are using the term "equilibrium" in two different ways. The *equilibrium length* of a spring is the uncompressed/unstretched length, i.e. the length where it produces zero force. On the other hand when we say that a system of masses and springs is in *static equilibrium*, we mean that the sum of the forces on each mass is zero, so that the system starting from that configuration will not move.

Example 1.3. Consider a Mass-Spring system with gravity as shown in Figure 1.7a. L denotes the equilibrium length of the spring, and therefore its *extension* is $x(t) - L$. The system is in static equilibrium when the spring force is equal and opposite to gravity. If $\bar{x}$ denotes the mass's position *at static equilibrium*, then

$$k\,\big(\bar{x} - L\big) \;=\; mg \qquad \Rightarrow \qquad k\bar{x} \;=\; mg + kL. \qquad \text{(at static equilibrium)} \tag{1.19}$$

The dynamic equation is then

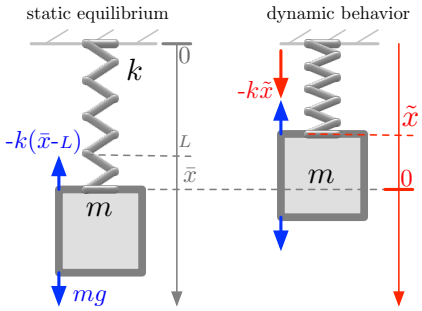
$$\begin{aligned} m\,\ddot{x}(t) \;&=\; -\,k\big(x(t) - L\big) \;+\; mg && \text{(Newton's 2nd law)} \\ &=\; -\,k\,x(t) \;+\; \big(kL + mg\big) && \text{(combining constants)} \\ &=\; -\,k\,x(t) \;+\; k\,\bar{x} && \text{(substituting from (1.19))} \\ &=\; -\,k\,\big(x(t) - \bar{x}\big) \end{aligned}$$

This equation looks exactly like (1.3), except that now $\bar{x}$ is determined by the balance between gravity and spring forces (1.19), as well as the choice of the coordinate system origin. We can again define a new coordinate system with $\tilde{x}(t) := x(t) - \bar{x}$, and the dynamical equation takes the simple form

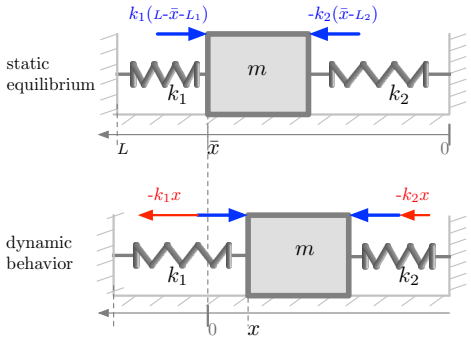
$$m\,\ddot{\tilde{x}}(t) \;=\; -\,k\,\tilde{x}(t). \tag{1.20}$$

As per Rule 1.2, if we are only interested in oscillations around static equilibrium, there is no need to find the static equilibrium explicitly. We simply define the origin to be the static equilibrium position (so that the static equilibrium $\bar{x}$ is at the origin, i.e. $\bar{x} = 0$) as shown in Figure 1.7a on the right, and immediately write down an equation equivalent to (1.20). Gravity plays no role in oscillations around static equilibrium in this case since it is a constant force[4].

[4]Warning: This is only true for linear dynamics. In certain nonlinear dynamics, constant forces such as gravity, or nonzero equilibrium extensions in nonlinear (stiffening) springs may actually alter the behavior of oscillations around static equilibrium.



(a) In static equilibrium, the spring is extended to balance gravity. In the coordinate system on the left, the balance is given by $k(\bar{x} - L) = mg$, where L is the unstretched length of the spring. The coordinate system on the right is chosen so that $\tilde{x} = 0$ corresponds to the static equilibrium point. In this coordinate system, the *net* force on m is then $-k\tilde{x}$.



(b) A mass between two compressed springs whose unstretched lengths are L_1 and L_2 respectively. The mass' width is negligible. In the coordinate system on top, the static equilibrium point is determined by the force balance $k_1(L - \bar{x} - L_1) - k_2(\bar{x} - L_1) = 0$ (compression means that $(x_2 - L_2) < 0$ and $(L - \bar{x} - L_1) < 0$). In the coordinate system at bottom where $x = 0$ implies static equilibrium, the net force on the mass is simply is $-(k_1 + k_2)x$.

Figure 1.7: Examples of systems with both constant and dynamic forces. The constant forces can be eliminated from the dynamical equations with the proper choices of coordinate systems.

Example 1.4. Consider the system shown in Figure 1.7b, which is a mass connected to two (possibly pre-compressed) springs. The width of the mass is considered negligible, and the equilibrium lengths of the springs are L_1 and L_2 respectively. In the coordinate system shown, the lengths of springs k_1 and k_2 at any time t are $(L - x(t))$ and $x(t)$ respectively. Their extensions are therefore

$$L - x(t) - L_1, \qquad x(t) - L_2$$

respectively. The dynamics of the mass are then given by

$$m\,\ddot{x}(t) \;=\; k_1\big(L - x(t) - L_1\big) \;-\; k_2\big(x(t) - L_2\big). \tag{1.21}$$

Note the signs on the forces. If k_1 is extended, i.e. $\big(L - x(t) - L_1\big) > 0$, its force acts in the positive x direction, thus the positive sign on k_1 above. If k_2 is extended, i.e. $\big(x(t) - L_2\big) > 0$, its force acts in the negative x direction, thus the negative sign on k_2.

We want to change the origin of the coordinate system in order to rewrite Equation (1.21) without the constant terms L, L_1 and L_2. This can be done by moving the origin to the point of *static equilibrium* $\bar{x}$, which is characterized by

$$k_1\big(L - \bar{x} - L_1\big) \;-\; k_2\big(\bar{x} - L_2\big) \;=\; 0. \tag{1.22}$$

Now define the deviation from static equilibrium $\tilde{x}(t) := x(t) - \bar{x}$, and the dynamics become

$$\begin{aligned} m\,\ddot{\tilde{x}}(t) \;&=\; k_1\big(L - x(t) - L_1\big) \;-\; k_2\big(x(t) - L_2\big) \\ &=\; k_1\big(L - (\tilde{x}(t) + \bar{x}) - L_1\big) \;-\; k_2\big((\tilde{x}(t) + \bar{x}) - L_2\big) \\ &=\; -\,k_1\tilde{x}(t) - k_2\tilde{x}(t) \;+\; \cancel{k_1\big(L - \bar{x} - L_1\big) - k_2\big(\bar{x} - L_2\big)} \qquad \text{(using (1.22))} \\ \Rightarrow \qquad m\,\ddot{\tilde{x}}(t) \;&=\; -\,(k_1 + k_2)\,\tilde{x}(t) \end{aligned} \tag{1.23}$$

Thus if we want to analyze the vibrations of this system, we simply analyze the differential equation (1.23) and its solutions. Note that this differential equation does not depend on the constant parameters L, L_1, L_2. If an expression for the solutions in the original coordinate system is needed, this is easy to do. Given any solution $\tilde{x}(t)$ of (1.23), the solution in the original coordinates is simply given from the definition $\tilde{x}(t) := x(t) - \bar{x}$ by

$$x(t) \;=\; \tilde{x}(t) + \bar{x}, \qquad \text{where } \bar{x} \;=\; \tfrac{k_1(L - L_1) + k_2 L_2}{k_1 + k_2}. \qquad \text{(solving for } \bar{x} \text{ from (1.22))}$$

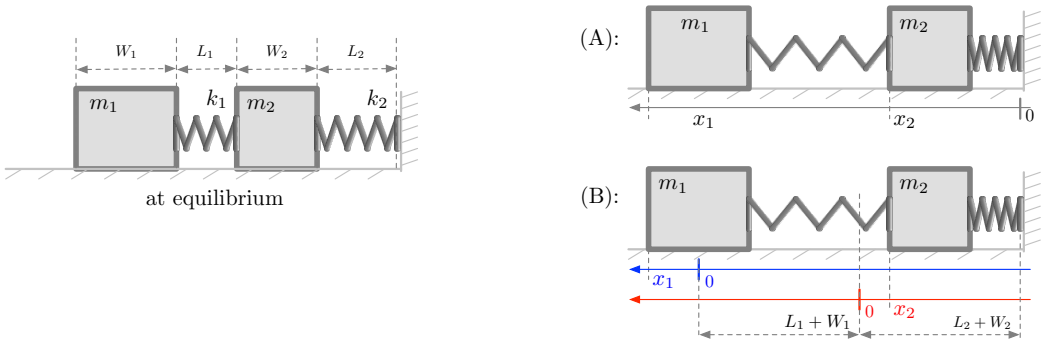


Figure 1.8: (*Left*) Two masses with two springs in equilibrium. L_1 is the unstretched length of spring k_1, W_1 is the width of m_1, and similarly for L_2 and W_2. (*Right, top*) If we use a single coordinate system for the positions of the two masses, we have to account for all those constant dimensions in the resulting differential equations (1.24) and (1.25). (*Right, bottom*) If we use different coordinate systems for each mass such that $x_1 = 0$ and $x_2 = 0$ correspond to static equilibrium, the resulting differential equations (1.27) have no constant terms.

Note that this amounts to only a constant shift by $\bar{x}$. The vibrational behavior of $x(t)$ and $\tilde{x}(t)$ is however the same.

Finally, it is interesting to compare the dynamics (1.23) to those of a single spring. Here both springs act "in concert", in that they both pull or push in the same direction at any one time. Therefore their combined effect is that of a single spring with stiffness $k_1 + k_2$. This is a specific example of a more general fact about interconnections of multiple springs that we will explore in detail in Section 1.5

1.2.1 Choice of Coordinates with Multiple Masses

Consider the system shown in Figure 1.8 where the equilibrium lengths L_1 and L_2 of the springs k_1 and k_2, as well as the widths W_1 and W_2 of the masses m_1 and m_2 respectively are explicitly shown. If we insist on a single coordinate system for both masses as shown in (A), we will have to explicitly account for the equilibrium lengths and masses' widths. This is a cumbersome exercise which we now illustrate, and then demonstrate how using two different coordinate systems in accordance with Rule 1.2 yields a much easier derivation of the equations of motion.

For the single coordinate system shown in Figure 1.8 (A), the equation of motion for m_1 is

$$m_1\, \ddot{x}_1(t) \;=\; -\,k_1\big(x_1(t) - x_2(t) - W_1 - L_1\big). \tag{1.24}$$

Note that $\big(x_1(t) - x_2(t) - W_1\big)$ is the instantaneous extension of spring k_1, and therefore its force $-k_1\big(x_1(t) - x_2(t) - W_1 - L_1\big)$ on m_1 is given by subtracting the equilibrium length L_1. As for m_2, it is subject to the same spring force as in (1.24), but in the opposite direction, as well as the spring force from k_2, which is the second term in the following equation

$$m_2\, \ddot{x}_2(t) \;=\; k_1\big(x_1(t) - x_2(t) - W_1 - L_1\big) \;-\; k_2\big(x_2(t) - W_2 - L_2\big). \tag{1.25}$$

Again, $\big(x_2(t) - W_2\big)$ is the instantaneous length of spring k_2, and therefore its *extension* is $\big(x_2(t) - W_2 - L_2\big)$, which produces a force $-k_2\big(x_2(t) - W_2 - L_2\big)$, i.e. in the rightwards direction when the extension is positive (and vice versa when the extension is negative, i.e. when spring is compressed).

Equations (1.24) and (1.25) together describe the dynamics of the two coupled masses. The constant terms L_1, L_2, W_1, W_2 make the equations a little cumbersome, and it would be nice to eliminate them. Again, we can do this in two different ways. The first is to define new coordinates by examining the equations, and seeing what is needed to "remove" the constant terms. Beginning with Equation (1.25), define $\tilde{x}_2 := x_2 - W_2 - L_2$. As before we have $\ddot{\tilde{x}}_2 = \ddot{x}_2$, and the second term in (1.25) will simply be $-k_2\tilde{x}_2$. To eliminate the constants in the first term (or equivalently in

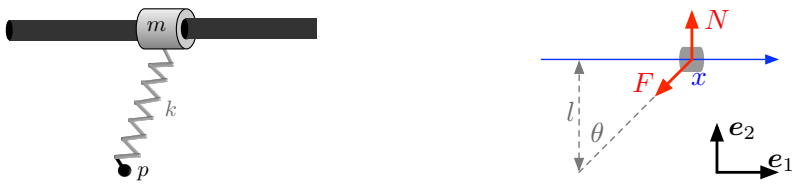


Figure 1.9: Example 1.5: A mass restricted to move without friction on a linear guide. The mass is connected with a spring to a fixed pivot point as shown. While the mass' motion is restricted to be one dimensional, the spring force is a vector in 2D whose direction varies with x. The motion along the guide is governed by the component of the spring force parallel to the guide.

Equation (1.24)), we have to define $\tilde{x}_1$ so that $\tilde{x}_1 - \tilde{x}_2$ becomes

$$\begin{aligned} \tilde{x}_1 - \tilde{x}_2 &= x_1 - x_2 - w_1 - L_1 \qquad \Rightarrow \qquad \tilde{x}_1 &= x_1 - x_2 - w_1 - L_1 + \tilde{x}_2 \\ & & &= x_1 - x_2 - w_1 - L_1 + \big(x_2 - w_2 - L_2\big) \\ & & &= x_1 - \big(w_1 + L_1 + w_2 + L_2\big). \end{aligned}$$

In these new coordinates, the combined equations of motions have the simpler form

$$\begin{aligned} m_1\, \ddot{\tilde{x}}_1(t) &= -\,k_1\big(\tilde{x}_1(t) - \tilde{x}_2(t)\big), \\ m_2\, \ddot{\tilde{x}}_2(t) &= \quad k_1\big(\tilde{x}_1(t) - \tilde{x}_2(t)\big) \;+\; k_2\, \tilde{x}_2. \end{aligned} \tag{1.26}$$

The masses' positions in original coordinates are obtained from

$$\begin{aligned} x_1(t) &= \tilde{x}_1(t) \;+\; \big(w_1 + L_1 + w_2 + L_2\big), \\ x_2(t) &= \tilde{x}_2(t) \;+\; \big(w_2 + L_2\big). \end{aligned}$$

Note that the constants L_1, L_2, w_1, w_2 have no effect on the equations (1.26) or their solutions. Thus knowledge of those constants is not needed to find the motion in the new coordinates. The constants only provide a "bias" to the solutions when expressed in the original coordinates.

Alternatively (and more easily), we can choose two different coordinate system origins for x_1 and x_2 as shown in (B). First choose the origin for x_2 so that $x_2 = 0$ corresponds to spring k_2 in equilibrium, and then choose the origin for x_1 such that $x_1 = 0$ and $x_2 = 0$ corresponds to spring k_1 in equilibrium. We can then write the equations directly as

$$\begin{aligned} m_1\, \ddot{x}_1 &= -\,k_1\big(x_1 - x_2\big) \\ m_2\, \ddot{x}_2 &= \quad k_1\big(x_1 - x_2\big) \;+\; k_2\, x_2 \end{aligned} \tag{1.27}$$

Note that these equations are identical to (1.26). Therefore with this scheme, we don't really need to know the equilibrium lengths of springs or widths of masses if we are only interested in understanding the oscillations. Those parameters are only needed to reconstruct the solutions in the original, common coordinate system.

1.2.2 Spring Forces in Higher Dimensions

When springs can extend in more than one dimension, more care is required to write down the effective equations of motion. Aside from the higher complexity of the equations, new features arise that are not there in purely one-dimensional problems. In particular, unlike the example of Figure 1.7b, any precompression of springs may actually change the nature of vibrations as the next example illustrates.

Example 1.5. Consider the Mass-Spring system shown in Figure 1.9. The mass is assumed to slide without friction on a linear guide and connected to a fixed pivot point p with a spring k. While the motion of the mass is in 1D, the spring force is a vector in 2D, and therefore care needs to be taken with the geometry.

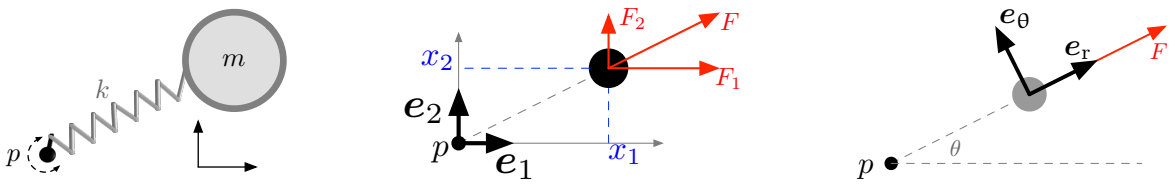


Figure 1.10: A mass moving in 2D while connected with a spring k to a fixed pivot point p. The motion can be analyzed in either a cartesian coordinate system (*center*), or the more convenient polar coordinate system (*right*). The polar coordinate system is more convenient in this case since the spring force is always in the radial $\boldsymbol{e}_{\mathrm{r}}$ direction.

Let L be the equilibrium length of the Spring (which may be different from the distance l shown). The spring's extension $\tilde{L}$ (the increase of the spring's length over its equilibrium length) is

$$\tilde{L} \;=\; \sqrt{x^2+l^2} - L.$$

This is the quantity that determines the magnitude of the spring force, while its direction is completely determined by the location x (see Figure 1.9 for the geometry)

$$\boldsymbol{F} \;=\; -\,k\left(\sqrt{x^2+l^2}-L\right)\;\left(\tfrac{x}{\sqrt{x^2+l^2}}\;\boldsymbol{e}_1 + \tfrac{l}{\sqrt{x^2+l^2}}\;\boldsymbol{e}_2\right),$$

where $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ are the unit coordinate vectors shown. The lateral motion of the mass is determined by the $\boldsymbol{e}_1$ component of the force, and therefore

$$m\,\ddot{x} \;=\; -\,k\left(\sqrt{x^2+l^2}-L\right)\;\tfrac{x}{\sqrt{x^2+l^2}} \;=\; -\,k\left(1-\tfrac{L}{\sqrt{x^2+l^2}}\right)\;x. \tag{1.28}$$

This is a nonlinear differential equation for $x(t)$. It can be linearized for small oscillations where $x \ll l$. The square root can then be approximated by $\sqrt{x^2+l^2} \approx l$. The differential equation becomes

$$\begin{aligned} & m\,\ddot{x}(t) \;=\; -\,k\,(1-L/l)\;\;x(t) \\ \Rightarrow\quad & m\,\ddot{x}(t) \;=\; -\,k_{\text{eff}}\;x(t), \qquad\qquad k_{\text{eff}} \;:=\; 1-L/l. \end{aligned}$$

Consider three cases depending on the relation between L and l.

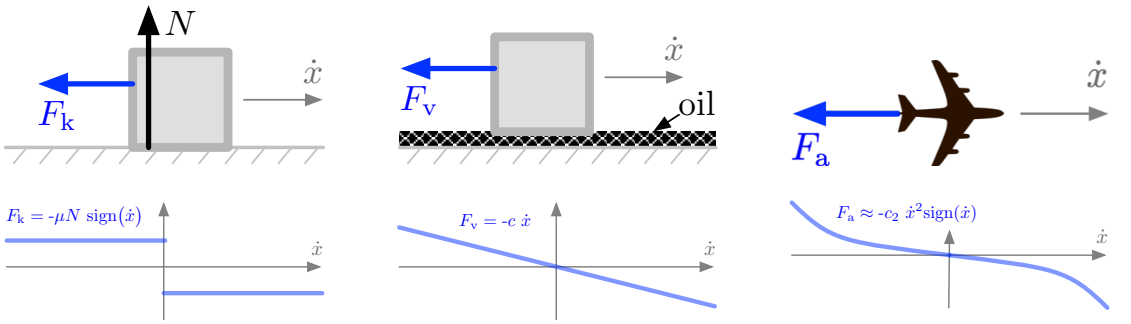
1. $l > L$: This means that the spring is stretched when $x = 0$. Since $L/l < 1$, the effective spring constant $k_{\text{eff}} < k$ is smaller than the actual spring constant. The larger l is (compared to L), the more the spring is prestretched, and K_{eff} is closer to the value of k.

2. $l = L$: In this case $k_{\text{eff}} = 0$. This means that up to first order in x, the spring exerts no force for small displacements x.

3. $l < L$: In this case $k_{\text{eff}} < 0$ is negative! What does does mean? If you think physically, the spring is actually compressed when $x = 0$. This is an example of an "unstable equilibrium"; an arbitrarily small initial change in x will cause the mass to slide either left or right (depending on the initial x), and it will not oscillate about the point $x = 0$. In this case we do not have small oscillations, but rather large ones, and small x assumption breaks down. In this case we have to reconsider the original differential equation (1.28).

Example 1.6. Consider a mass free to move in 2D while connected to a fixed pivot point by a spring k as shown in Figure 1.10. The motion can be analyzed in either a cartesian or a polar coordinate system. The equations are much simpler in the polar coordinate system since the spring force is always in the radial $\boldsymbol{e}_{\mathrm{r}}$ direction. In the polar system, the spring's extension is simply $(r-L)$ where L is the equilibrium length of the spring. Since there is no tangential $\boldsymbol{e}_\theta$ component of the spring force, the equations of motion are simply

$$\begin{aligned} m\,\ddot{r}(t) \;&=\; -\,k\;\Big(r(t)-L\Big), \\ m\,\ddot{\theta}(t) \;&=\; 0. \end{aligned} \tag{1.29}$$



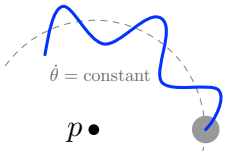
(a) Dry, kinetic friction drag. N is the surface-normal force, and the friction force is $F_k = -\mu N \text{ sign}(\dot{x})$, where μ is the kinetic friction coefficient. Although the magnitude $|F_k|$ is not dependent on velocity, the direction is.

(b) Viscous (Stokes) drag is prominent at low Reynolds numbers, i.e. either low velocities or high viscosities such as in motion lubricated with oil. It is linearly proportional to velocity $F_v = -c\,\dot{x}$.

(c) Aerodynamic drag is prominent at high Reynolds numbers. It is approximated as being proportional to the square of the speed, with direction opposite to velocity $F_a \approx -c_2\,\dot{x}^2 \text{ sign}(\dot{x})$.

Figure 1.11: The three main types of dissipative drag forces. All are a function of velocity, and their direction is always opposite to velocity.

Thus the mass will oscillated like a 1D Mass-Spring system in the radial direction, while it may be stationary or rotate with constant angular velocity in the tangential direction as illustrated in the diagram on the right.



For comparison, we redo the analysis in cartesian coordinates. The spring's extension is given by

$$\tilde{L} \;=\; \sqrt{x_1^2 + x_2^2} - L.$$

The vector of the spring force is then

$$\boldsymbol{F} \;=\; -\,k\left(\sqrt{x_1^2 + x_2^2} - L\right)\;\left(\frac{x_1}{\sqrt{x_1^2+x_2^2}}\;\boldsymbol{e}_1 + \frac{x_2}{\sqrt{x_1^2+x_2^2}}\;\boldsymbol{e}_2\right).$$

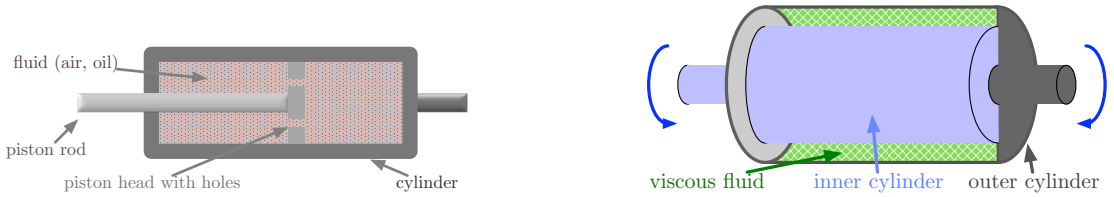
The differential equations for the cartesian coordinates $x_1(t)$ and $x_2(t)$ are then

$$\begin{aligned} m\,\ddot{x}_1 \;&=\; -\,k\left(\sqrt{x_1^2 + x_2^2} - L\right)\frac{x_1}{\sqrt{x_1^2+x_2^2}} \;=\; -\,k\left(1 - L\,\frac{1}{\sqrt{x_1^2+x_2^2}}\right)x_1, \\ m\,\ddot{x}_2 \;&=\; -\,k\left(\sqrt{x_1^2 + x_2^2} - L\right)\frac{x_2}{\sqrt{x_1^2+x_2^2}} \;=\; -\,k\left(1 - L\,\frac{x_2}{\sqrt{x_1^2+x_2^2}}\right)x_2, \end{aligned} \tag{1.30}$$

where the argument t has been suppressed for notational simplicity. Note that these are two coupled, nonlinear differential equations which are much more difficult to analyze than the equations (1.29) written in polar coordinates. The term *coupled* here means that the equation for $x_1(t)$ depends on $x_2(t)$ and vice versa, thus the two equations have to solved *together*, unlike (1.29) where the two equations are independent of each other. It is not even clear how one should linearize (1.30) or what "small oscillations" could mean. Thus in some problems, a judicious choice of coordinate systems may significantly simplify the analysis.

1.3 Damping Forces

So far we have only considered conservative forces due to elasticity or gravity. In vibrations analysis, it is also important to consider dissipative forces such as the various forms of drag and friction. There are many types of dissipative forces in mechanics, but the three most important ones are (a) Kinetic (Dry) Friction, (b) Viscous (Stokes) Drag, and (c) Aerodynamic Drag. They are described



Linear Motion Damper

Rotational Motion Damper

Symbol:

Symbol:

Figure 1.12: (*Left*) Schematic of a linear-motion viscous damper. Two rods are connected to a sealed cylinder and a movable piston head respectively. The motion is linear. The piston head has small holes through which a viscous fluid can pass back and forth between the two compartments. This creates resistance to motion, and the resulting force between the two rods is proportional and opposite to their relative velocity. (*Right*) A similar concept creates a rotary damper. A sealed cylinder contains a smaller, solid cylinder, and the gap between between the two is filled with thin layer of a viscous fluid. The motion is rotational, and the differential torque between the two connecting rods is proportional and opposite to the relative rotational velocity.

mathematically in terms of velocity $\dot{x}$ as follows[5]

$$\begin{aligned} F_{\text{k}} &= -\mu N \,\text{sign}(\dot{x}), && \text{(dry, kinetic friction)} \\ F_{\text{v}} &= -c\,\dot{x}, && \text{(viscous drag)} \\ F_{\text{a}} &= -c_1\dot{x} - c_2\,\dot{x}^2\,\text{sign}(\dot{x}) \;\approx\; -c_2\,\dot{x}^2\,\text{sign}(\dot{x}), && \text{(aerodynamic drag)} \end{aligned}$$

where for kinetic friction, N is the force normal to a surface, and μ is the kinetic friction coefficient with that surface. The three different types of drag force laws are illustrated in Figure 1.11. The viscous drag force law, also known as Stokes' drag, is prominent at low Reynolds numbers, i.e. at low velocities or high viscosities. It is sometimes referred to as viscous friction as might occur between two surfaces with a viscous lubricant in between. In contrast, kinetic friction is sometimes called "dry friction". The aerodynamic drag force law is valid at high Reynolds numbers (e.g. cars and planes in air), where the quadratic term dominates the linear term. The latter is often ignored in that expression.

For vibration analysis of mechanical systems, the main drag force of concern is viscous drag, and to a lesser extent dry friction drag. In fact, viscous drag elements called *dampers* are deliberately introduced in mechanism to achieve certain vibration control specifications. Think about dampers on doors or drawers as well as in vehicle suspension systems. The construction of a basic viscous damper is illustrated in Figure 1.12. It is usually a sealed cylinder with a movable piston, and a fluid (such as air, or some type of oil) that flows between two compartments through small holes. Unlike a spring, this mechanism exerts no forces when stationary regardless of extension or compression. The force it generates is well approximated by linear proportionality to velocity and opposite to its direction

$$F_{\text{v}} = -c\,\dot{x}, \tag{1.31}$$

where c is referred to as the *damping coefficient*. Since c converts velocities to forces, it has units of force per velocity, e.g. N/m/s. Typical vehicle shock absorber dampers have damping coefficients in the range of 1000s N/m/s. A shock absorber is a combination of a spring and damper as shown in Figure 1.13.

Figure 1.12 also shows a rotational damper that has a torque law which is similarly proportional, but to rotational velocity

$$\tau = -c\,\omega,$$

[5]The function sign(α) is defined as $+1$ if $\alpha > 0$, -1 if $\alpha < 0$, and 0 if $\alpha = 0$.

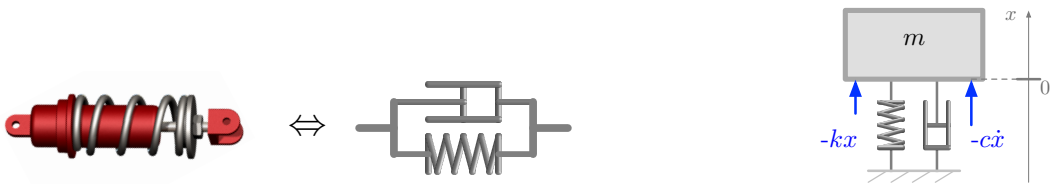


Figure 1.13: (*Left*) A typical shock absorber is made up of a viscous-damper cylinder with a spring coiled around it. It is represented schematically as a parallel connection of a spring and a damper. (*Right*) A mass m connected to a fixed frame with both a spring and damper experiences a spring force $-kx$ proportional to displacement, and a damper force $-c\dot{x}$ proportional to its velocity.

where ω is the relative rotational velocity between the two ends. Figure 1.12 also shows schematic symbols for dampers. The one for the linear motion damper is widely used. On the other hand, there is no universally accepted schematic symbol for the rotational damper. The one shown is the one chosen for this book.

Consider the mass m in Figure 1.13 connected to a fixed frame by a spring k and a damper with coefficient c. Choose the origin $x = 0$ of the coordinate system such that the spring is in equilibrium. The spring's force is then simply $-kx$. The choice of origin does not effect the expression for the damper's force, which is $-c\dot{x}$, and velocity $\dot{x}$ is independent of where the coordinate system's origin is. The equation of motion is therefore

$$m\,\ddot{x}(t) \;=\; -\,k\,x(t) \;-\; c\,\dot{x}(t) \qquad \Leftrightarrow \qquad \boxed{m\,\ddot{x}(t) \;+\; c\,\dot{x}(t) \;+\; k\,x(t) \;=\; 0.} \tag{1.32}$$

Note that the two forces are opposite to spring deflection and velocity respectively.

Comparing the differential equation (1.32) to (1.5) for a system with only a spring, we see that it is also a 2nd order differential equation, but now we have a term that depends on the first time derivative as well. We will study such equations in detail in Chapter 3. Unlike Equation (1.5) which has only oscillatory solutions, Equation (1.32) may have non-oscillatory behavior in the so-called "highly-damped" regime, or decaying (damped) oscillations in the so-called "lightly damped" case. The relative values of the constants m, k, and c determine which of those two regimes the system is in.

In addition to spring and damper forces, we often have externally generated and possibly time-varying forces. Consider the system illustrated in Figure 1.14, where a mass m is connected to a fixed frame through a spring and a damper. In addition, the mass is subject to a possibly time-varying external force $f(t)$. Such forces can be generated by controlled actuator elements which can generate forces whose magnitudes are controlled by some external mechanism (either electronic or mechanical). $f(t)$ can also model uncontrolled external forces such as the effect of wind gusts, ground vibrations, and the like. The main new ingredient is that $f(t)$ can be time varying rather than the constant forces we encountered earlier (e.g. gravity). To write down the equation of motion for this system, we simply apply Newton's 2nd law again with $f(t)$ added to the sum of the forces

$$m\,\ddot{x}(t) \;=\; -\,k\,x(t) \;-\; c\,\dot{x}(t) \;+\; f(t) \qquad \Leftrightarrow \qquad \boxed{m\,\ddot{x}(t) \;+\; c\,\dot{x}(t) \;+\; k\,x(t) \;=\; f(t),} \tag{1.33}$$

where we have assumed the mass to slide without friction with the supporting surface. Note that this equations is valid provided the origin of the coordinate system is chosen so that $x = 0$ corresponds to the spring in equilibrium. This will be our practice from now on without explicitly and repeatedly stating it.

It is instructive to compare Equation (1.33) to Equation (1.32) which has no external forcing. The latter is a *homogenous* 2nd order differential equation, while the former is the same differential equation, but with a forcing term, i.e. an inhomogenous differential equation. In the context of vibration analysis, the preferred terms are "free vibrations" (for (1.32)) and "forced vibrations" for (1.33). More generally, we will refer to systems like (1.33) as a system with "an input", since forcing terms in the differential equation may be more general than forces, such as other displacements, velocities or electronic signals. Differential equations like (1.33) will be studied in detail in Chapters 4 and 6 for various types of inputs.

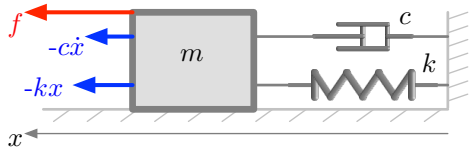


Figure 1.14: A Mass-Spring-Damper system subject to an external, time varying force $f(t)$. While the spring and damper forces are determined by the mass' configuration (position and velocity), the force $f(t)$ is determined "externally", i.e. independently of the mass' configuration. Such forces can model various phenomena, e.g. an electronically controlled actuator which produces a force whose value is determined by an electronic command, or the effect of external mechanical forces such as wind gusts or ground vibrations.

1.4 Mechanical Degrees of Freedom

Any mechanical system composed of an assembly of rigid bodies connected with either flexible or rigid linkages has a certain number of "mechanical degrees of freedom". Roughly speaking, the number of degrees of freedom is the number of *coordinates* needed to uniquely describe the configuration of the system. If the system has a single mass constrained to move in only one dimension (e.g. Figures 1.2 or 1.7), then a single coordinate x (or θ in the case of the torsional spring of Figure 1.4 or the pendulum) can be used to describe the motion of that mass. Note that this is the case regardless of the number of springs or dampers the mass is attached to. Springs and dampers are typically assumed massless[6], and therefore do not add any additional degrees of freedom. For example, in Figure 1.7b, a single mass is connected by two springs to two rigid anchors, and therefore this system has a single degree of freedom, namely the position of the mass.

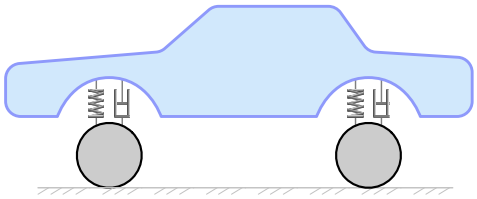
A system with two masses such as that of Figure 1.8 has two degrees of freedom, namely the positions of each mass. Note that here, each mass is constrained to move in only one dimension (since only lateral motion occurs in this example). Recall that when deriving the differential equations (1.26) for this system, we ended up with *two, coupled, second order* differential equations for the whole system. This is a pattern we will see regularly. A system with n-Degrees-Of-Freedom (n-DOF) will always be modeled using n coupled differential equations, each of second order. The analysis of such systems is best done using matrix methods, which we will begin developing in Chapter 7.

A single mass can also have multiple degrees of freedom if it can move in more than one coordinate. A nice example of this is the modeling of automotive suspension systems illustrated in Figure 1.15. A vehicle is typically considered as a single rigid mass. A hierarchy of dynamical models can be developed as follows. In Figure 1.15b, the car is modeled as a single mass interacting with the ground via a spring and damper, but the motion is just linear vertical motion described by a single coordinate x. The differential equation describing the dynamics is the single second-order equation (1.32) for a Mass-Spring-Damper system.

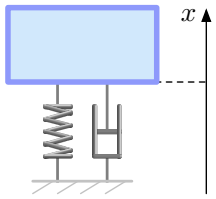
A more realistic suspension model is shown in Figure 1.15c sometimes referred to as the "half-car model"[7]. Here the car is assumed to be rigid body in the plane, which requires three coordinates (x, y, θ) to completely describe its motion (e.g. (x, y) for the location of the center of mass, and θ for the rotation of the body about the out-of-plane axis). In suspension models, one is mainly interested in vertical motion, and horizontal motion is assumed steady, therefore not contributing to vertical vibrations. Thus the half-car model has two degrees of freedom (x, θ) describing the vertical displacement of the center of mass (we relabeled vertical displacement as x rather than y for later notational consistency) and the rotation angle respectively. Alternatively (and equivalently) we can use the vertical positions x_1, x_2 of the two contact points between the rigid body and the support structure. Those represent another equivalent 2-DOF model. Whichever choice we make for the coordinates of the half-car model, we would then expect two, coupled, second-order differential equations to model this system. Those are derived in Section 1.7.2 for illustration.

An even more realistic suspension model is shown in Figure 1.15d. The car is still one single rigid

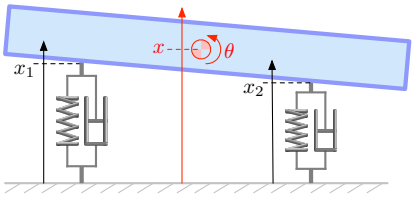
[6] or at least of much smaller masses than those of the rigid bodies they are attached to.
[7] This is also called the "sprung beam" model if the mass is thought of as a rigid beam rather than a vehicle.



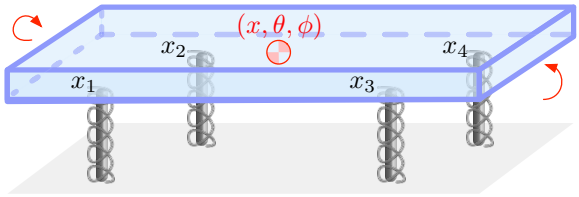
(a) A typical automotive suspension model assumes the shock absorber and suspension assembly at each tire to be modeled by a spring and a damper. The car body is assumed to be a rigid body.



(b) A "quarter car" model is a 1 degree of freedom model where the car is assumed to be s single mass that can only move (not rotate) vertically with vertical coordinate x.



(c) A "half car" model is a 2 degree of freedom model for motion of a supported rigid body in the plane. The coordinate choices can be either the vertical displacements (x_1, x_2) of the two support points, or alternatively (x, θ), where x is the vertical position of the center of mass, and θ is the angle of rotation of the rigid body in the plane.



(d) A "full car" model is a 3 degree of freedom model of the car as a rigid body in three dimensions. One choice of coordinates is (x, θ, ϕ) for the vertical displacement of center of mass, the azimuthal θ and polar ϕ angles of rotation respectively. Alternatively, the vertical coordinates (x_1, x_2, x_3, x_4) of the four support points can be used. However, those four variables have a single constraint, which is consistent with the system having only 3 degrees of freedom.

Figure 1.15: A hierarchy of automotive suspension models with one, two and three mechanical degrees of freedom. A common simplifying assumption in such vibration analysis is that all displacements are purely vertical, and horizontal motion is steady, and therefore does not contribute to vertical vibrations.

body, but it can rotate in three dimensions. Again, our main interest is in vertical displacement and body orientation, so one choice of coordinates would be (x, θ, ϕ) describing vertical displacement of the center of mass, the azimuthal θ and polar ϕ orientation of the body. Equivalently, we can use the vertical displacements (x_1, x_2, x_3, x_4) of the four support points. However, because the body is assumed rigid, there is actually a kinematic constraint on those four coordinates, which reduces the number of degrees of freedom to 3 rather than 4. The details of this model, which is more involved to derive, will be presented later.

In general, if we have n masses, where each can move in m coordinates, then the total mechanical degrees of freedom of the entire system will be the product nm, and the model will consequently require nm, coupled, second-order differential equations.

1.5 Connections of Multiple Springs and Dampers

When multiple springs and dampers are connected together, it is possible to aggregate their effects into a smaller number of elements with different effective stiffness and damping coefficients. The simplest such case is when they are connected in parallel as shown in Figure 1.16. Assuming that both springs have the same equilibrium lengths, then regardless of the extension/compression of the springs, both of their forces always act in this same direction. Similarly both damper forces act in the same direction regardless of the mass' velocity direction. Each set of forces then add up as

$$-k_1 x - k_2 x \;=\; -(k_1 + k_2)\, x \;=\; k_{\text{eff}}\, x \qquad \Rightarrow \qquad \boxed{k_{\text{eff}} \;=\; k_1 + k_2,} \tag{1.34}$$

$$-c_1 \dot{x} - c_2 \dot{x} \;=\; -(c_1 + c_2)\, \dot{x} \;=\; c_{\text{eff}}\, \dot{x} \qquad \Rightarrow \qquad \boxed{c_{\text{eff}} \;=\; c_1 + c_2.} \tag{1.35}$$

Thus the two parallel springs can be replaced by a single spring with stiffness $k_{\text{eff}} = k_1 + k_2$, and the two parallel dampers can be replaced by a single damper with coefficient $c_{\text{eff}} = c_1 + c_2$. We again emphasize that the spring forces have the expression above only if the origin of the coordinate

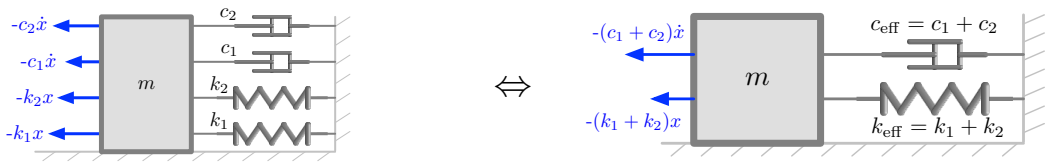


Figure 1.16: When springs and dampers are connected in parallel, their resulting forces simply add up. This results in an effective stiffness of $k_{\text{eff}} = k_1 + k_2$, and an effective damping coefficient of $c_{\text{eff}} = c_1 + c_2$.

system $x = 0$ is such that both springs are in equilibrium. If that is not the case, the conditions of static equilibrium would have to be derived, and the origin of the coordinate system shifted so that the equations above hold. This procedure is explored in Exercise 1.1.

A different configuration in which forces add up in the same manner is shown in Figure 1.7b, where the two springs are on *opposing sides* of the mass. Although they are on opposite sides, when one spring is in effective compression, the other one is in effective extension, and both forces act in the same direction. Thus the effect of the forces add up as demonstrated by Equation (1.23), and in this case we can also replace the two springs by an equivalent $k_{\text{eff}} = k_1 + k_2$. A similar argument applies to two dampers connected on opposite sides of a mass.

An alternative way to connect spring and damper elements is in series, and it turns out that the effective stiffness and damping coefficients behave differently in this case. To analyze this situation, let's first consider all the forces in a Mass-Spring-Damper system as shown in Figure 1.17a where the "internal forces" on the spring and damper are shown. For simplicity, springs and damper elements are assumed massless. A massless element simply "transmits" forces from one end to another as shown in the diagram. To see this, suppose the element had a small mass ϵ, and $\tilde{L}(t)$ refers to the deviation of its length from equilibrium. Let F_{l} and F_{r} be the two forces acting on it at the left and right edges respectively. Newton's second law says

$$\epsilon \, \ddot{\tilde{L}}(t) \;=\; F_{\text{l}} \;+\; F_{\text{r}}.$$

If we let $\epsilon \to 0$ to model the case of negligible mass, the left hand side is zero, and we get that $F_{\text{l}} = -F_{\text{r}}$, i.e. the forces on each side of a massless element are equal and opposite as claimed in the diagrams of Figure 1.17a.

Now consider the two springs shown in Figure 1.17b. The two springs have internal forces of $k_1 L_1$ and $k_2 L_2$ respectively, where L_1 and L_2 are the deviations of the springs' lengths from their equilibrium[8]. Since the contact point between the two springs is also massless, those two forces must be equal in magnitude, i.e. $k_1 L_2 = k_2 L_2$. This equality implies that the extensions of the two springs can not be arbitrary, but must be related by

$$\frac{k_1}{k_2} \;=\; \frac{L_2}{L_1}. \tag{1.36}$$

This means that if for example k_1 is stiffer than k_2, then L_2 will be longer than L_1 since k_2 is a softer spring, and therefore will extend more than k_1 when subjected to the same force.

The effective stiffness k_{eff} is determined by the requirement that the effective spring force $k_{\text{eff}}(L_1 + L_2)$ is equal to the individual spring forces $k_1 L_2 = k_2 L_2$. Using $k_1 L_1$ for instance gives

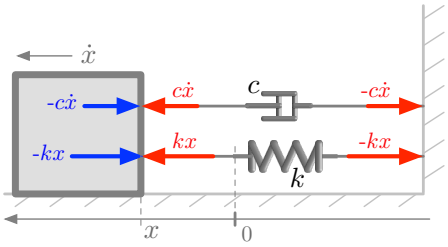
$$k_{\text{eff}}(L_1 + L_2) = k_1 L_1 \qquad \Rightarrow \qquad k_{\text{eff}} \;=\; \frac{k_1 L_1}{L_1 + L_2} = \frac{k_1}{1 + L_2/L_1}$$

(dividing top and bottom by L_1)

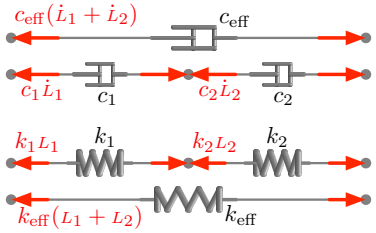
$$= \frac{k_1}{1 + k_1/k_2} \qquad \text{(using (1.36))}$$

$$\Rightarrow \qquad \boxed{k_{\text{eff}} \;=\; \frac{k_1 k_2}{k_2 + k_1}.} \tag{1.37}$$

[8] In this analysis we only keep track of force magnitudes, as their directions are explicitly shown in Figure 1.17b.



(a) Depicting all the internal forces in a Mass-Spring-Damper system. Here the spring is in extension, and moving with a positive velocity. Therefore both damper and spring forces on the mass (in blue) are negative. On the left side of the spring and damper, the forces of the mass on the spring and damper are equal and opposite respectively. The spring and damper elements are usually assumed massless, and therefore they simply transmit forces from one end to the other as shown. On the right side are the forces of the stationary frame on the spring and damper elements, which are equal and opposite to the forces on those elements at the other end.



(b) To find the effective damping and stiffness for two elements connected in series, we must analyze the internal reaction forces. Massless elements (as the damper and spring are assumed to be) transmit forces as shown. For the dampers diagram, all forces shown are equal in magnitude, and similarly for the springs diagram. L_1 and L_2 are the deviations of the lengths of the respective elements from their equilibrium values.

Figure 1.17: Analysis to reveal the effective stiffness k_{eff} and damping c_{eff} when elements are connected in series.

For dampers, an exactly parallel analysis leads to a similar formula

$$c_{\text{eff}}(\dot{L}_1+\dot{L}_2)=c_1\dot{L}_1 \quad \text{and} \quad \frac{c_1}{c_2}=\frac{\dot{L}_2}{\dot{L}_1} \qquad \Rightarrow \qquad \boxed{c_{\text{eff}} = \frac{c_1c_2}{c_2+c_1}.} \tag{1.38}$$

The expressions (1.37) and (1.38) may look familiar. If we rewrite them as

$$k_{\text{eff}}=\frac{k_1k_2}{k_2+k_1} \quad \Leftrightarrow \quad \frac{1}{k_{\text{eff}}}=\frac{1}{k_1}+\frac{1}{k_2}, \qquad c_{\text{eff}}=\frac{c_1c_2}{c_2+c_1} \quad \Leftrightarrow \quad \frac{1}{c_{\text{eff}}}=\frac{1}{c_1}+\frac{1}{c_2},$$

we see a familiar expression from electrical circuits. In fact, springs and dampers add up in series and in parallel exactly like capacitors and inductors add up in series and parallel connections[9]. Note that resistors on the other hand add up in the opposite manner. These analogies with electrical circuits will become clearer when we discuss Electrical-Mechanical analogies later on.

It is useful to remember some basic rules about parallel and series connections of springs and dampers. A parallel connection (1.34) of two springs gives a k_{eff} which is *stiffer* than either spring. On the other hand, a series connection (1.38) of two springs gives a k_{eff} which is *softer* than either spring. This is useful for system design. If a connection needs to be stiffened, a spring must be added in parallel. It it needs to be softened, a spring must be added in series. The exact constants for the added springs can be calculated from the relations (1.34) or (1.37), along with the design specifications. Similar statements apply to damping elements.

redo this in terms of compliance

A simple example of a series connection of two springs is shown in Figure 1.18 as a combined model of an automotive suspension spring and tire stiffness. The interaction of a fully inflated tire with a roadway can be modeled as a stiff spring k_{t}. A typical suspension system has spring constants in the range $k_{\text{s}} \approx$ 20-30 KN/m (Kilo-Newtons/meter). The stiffness of a fully-inflated tire interacting with a road surface is much higher, on the order of 200 KN/m. Ignoring the mass of the tire, the two springs are essentially connected in series (as shown in Figure 1.18). The effective spring constant is then given by (1.37) as

$$k_{\text{eff}} = \frac{k_{\text{s}}k_{\text{t}}}{k_{\text{s}}+k_{\text{s}}} = \frac{20 \text{ KN/m } 200 \text{ KN/m}}{20 \text{ KN/m} + 200 \text{ KN/m}} \approx 18 \text{ KN/m}.$$

Thus the effective stiffness is not much different from that of the softer spring k_{s}. This is because the tire spring is about an order of magnitude stiffer than the suspension spring, and therefore their

[9] For example, two capacitors C_1 and C_2 in parallel have a net capacitance of C_1+C_2, but when connected in series, their net capacitance is $C_1C_2/(C_1+C_2)$.

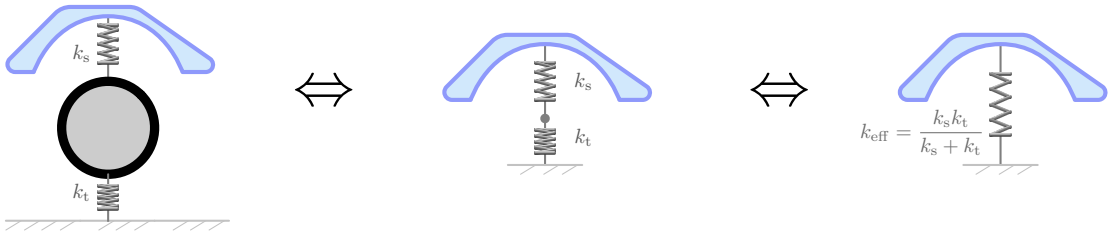


Figure 1.18: A combined model of a suspension system stiffness k_s, and the tire/road interaction which is modeled as a much stiffer spring k_t for a fully inflated tire. The two springs are in series, and because $k_t \gg k_s$, the series connection has an effective stiffness k_{eff} which only slightly lower than k_s, the softer of the two springs.

series connection is only slightly softer than the softer of the two springs. This is similar to the connection of two resistors *in parallel* where one resistance is much higher than the other. The effective resistance is then slightly lower than the low resistance path. Whether the effects of tire elasticity can be ignored in vibration analysis depends on the purpose and accuracy of the analysis. In either case, incorporating tire stiffness is a simple matter of altering the effective spring constant.

Finally, it might be natural to ask the question of what happens if two different types of elements, like a spring and a damper are connected in series? This one is a little tricky and requires a finer analysis as explored in Exercise 1.2.

1.6 The Energy Method

For systems with a single degree of freedom, we can find the natural frequency more quickly using energy calculations. Let $q(t)$ be the "generalized" position coordinate of such a system (e.g. a linear position or angular position coordinate), the kinetic energy of such a system is usually of the form

$$\text{KE} \;=\; \tfrac{1}{2} c_k \, (\dot{q})^2.$$

For example, for the Mass-Spring system (1.5) and the pendulum (1.11), their respective kinetic energies are

$$\begin{aligned} \text{KE} &= \tfrac{1}{2} m \, \dot{x}^2 && \text{linear Mass-Spring system} \\ \text{KE} &= \tfrac{1}{2} m \, (l\dot{\theta})^2 \;=\; \tfrac{1}{2} m l^2 \, \dot{\theta}^2 && \text{pendulum} \end{aligned}$$

Thus the coefficient c_k is m in the first case and ml^2 for the second.

For systems where conservative forces are linear functions of position coordinates, the potential energy (of the conservative forces) will be quadratic in the position coordinate[10], and thus of the form

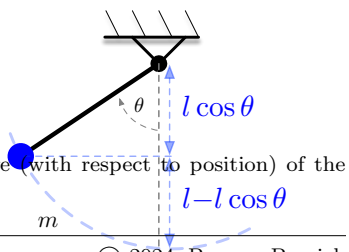
$$\text{PE} \;=\; \tfrac{1}{2} c_p \, q^2.$$

For example, for the Mass-Spring system, the spring potential energy is given by

$$\text{PE} \;=\; \tfrac{1}{2} k \, x^2.$$

On the other hand, for the pendulum system, the gravitational potential energy written in terms of the angular coordinate θ is (see figure)

$$\text{PE} \;=\; mg \, l\big(1 - \cos(\theta)\big).$$



[10] Recall that for conservative forces, the force is the negative of the derivative (with respect to position) of the potential energy function. The derivative of a quadratic is linear.

This is not quadratic in θ, but if we expand the Taylor series for $\cos(\theta)$ around $\theta = 0$ we get

$$\begin{aligned} 1 - \cos(\theta) &= 1 - \left(1 - \tfrac{1}{2!}\theta^2 + \tfrac{1}{4!}\theta^4 - \cdots\right) \\ &= \tfrac{1}{2!}\theta^2 - \tfrac{1}{4!}\theta^4 + \cdots \end{aligned} \tag{1.39}$$

Now similar to what we did with linearizing forces to model small oscillations, this corresponds to retaining up to quadratic terms in potential energy, we therefore approximate this potential energy with its quadratic term

$$\text{PE} \;\approx\; \tfrac{1}{2} mgl\; \theta^2.$$

Thus the coefficient c_{p} for the pendulum is mgl.

Newtown's second law for systems that have quadratic kinetic and potential energies of the form

$$\text{E} \;=\; \tfrac{1}{2} c_{\text{k}}\; \dot{q}^2 \;+\; \tfrac{1}{2} c_{\text{p}}\; q^2$$

is written as (this also follows immediately from the Euler-Lagrange equations)

$$c_{\text{k}}\; \ddot{q}(t) \;=\; -\, c_{\text{p}}\; q(t).$$

The natural frequency of such a system is then given by the ratio

$$\omega_{\text{n}} \;=\; \sqrt{c_{\text{p}}/c_{\text{k}}}. \tag{1.40}$$

We can think of c_{k} as an effective inertia and c_{p} as an effective stiffness. For the Mass-Spring system, the formula gives the (by now) familiar $\omega_{\text{n}} \;=\; \sqrt{k/m}$. For the pendulum, it gives

$$\omega_{\text{n}} \;=\; \sqrt{c_{\text{p}} c_{\text{k}}} \;=\; \sqrt{mgl/ml^2} \;=\; \sqrt{g/l},$$

which is the result arrived at from the earlier calculation in (1.13). The formula (1.40) simplifies some calculations as the next example will show.

A Metronome

A musical metronome is basically a pendulum designed so that it has a "tunable" oscillation frequency that can be adjusted over a the range of frequencies required to "keep the beat" for any particular musical practice performance. The main design requirement is that its frequency should be "tunable", and that its size not be too big (e.g. to fit nicely onto a desktop). If we attempt to use a simple pendulum as a metronome, its natural frequency is $\omega_{\text{n}} = \sqrt{g/l}$ as analyzed earlier. Clearly g is not a design parameter, but the mass can be placed on a slider to make l a tunable parameter. We will see in the calculations at the end of this example that this is not a practical design for the required range of frequencies. The pendulum length would be impractically long.

The metronome ingeniously solves this design problem. It has two masses rather than one, one of which is on a slider, and both are rigidly connected, but free to rotate about a common pivot point. See Figure 1.19 for a diagram. We will now analyze this system to show that it has a much more tunable range of frequencies (with a small size) than the standard pendulum.

Although this system has two masses, it is really a single degree of freedom system. There is only one position coordinate, namely θ since the two masses are connected by a rigid rod that pivots around a point in between the two masses. Using the energy method, it is not too hard to compute the natural frequency of this system as a function of the masses and the two lengths. Assuming the connecting rod to be massless, the kinetic energy is

$$\text{KE} \;=\; \tfrac{1}{2}\left(m_2(l_2\dot{\theta})^2 \;+\; m_1(l_1\dot{\theta})^2\right) \;=\; \tfrac{1}{2}\left(m_2 l_2^2 + m_1 l_1^2\right)\; \dot{\theta}^2 \;=:\; \tfrac{1}{2} c_{\text{k}}\; \dot{\theta}^2.$$

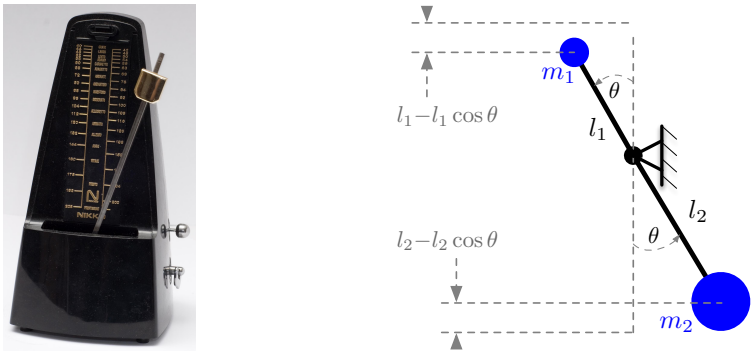


Figure 1.19: (*Left*) A typical wind-up metronome (bottom mass not visible) and, (*Right*) a simplified mechanical model. Although this system has two masses, both masses are connected by a massless rigid rod, and therefore it has only one mechanical degree of freedom which is the (angular) position coordinate θ.

The potential energy is the sum of the two gravitational potential energies of the masses using the quadratic approximation (1.39) of the cosine (see also Figure 1.19 for the geometry)

$$\begin{aligned} \text{PE} &= m_2 g\, l_2\big(1-\cos(\theta)\big) \;-\; m_1 g\, l_1\big(1-\cos(\theta)\big) \\ &\approx \tfrac{1}{2} m_2 g\, l_2 \theta^2 \;-\; \tfrac{1}{2} m_1 g\, l_1 \theta^2 \;=\; \tfrac{1}{2} g\,(m_2 l_2 - m_1 l_1)\,\theta^2 \;=:\; \tfrac{1}{2} c_{\text{p}}\,\theta^2. \end{aligned}$$

Note that each potential energy term is chosen with a constant offset so that each is zero at $\theta = 0$.

The formula (1.40) now gives the oscillation frequency of the metronome as

$$\omega_{\text{n}} \;=\; \sqrt{c_{\text{p}}/c_{\text{k}}} \;=\; \sqrt{g\,\frac{m_2 l_2 - m_1 l_1}{m_2 l_2^2 + m_1 l_1^2}}. \tag{1.41}$$

It should be noted that this formula would have required more work to derive using free body diagrams and force balances.

Now analyze (1.41) to see how it can be useful for design. First note that if $m_2 l_2 < m_1 l_2$, then the quantity under the square root is negative. This actually means that this system would not oscillate under this condition. This is physically intuitive since this condition means that the metronome system is "top heavy", and the top mass will simply fall all the way to the bottom instead of oscillating. Now in the metronome shown in Figure 1.19, m_1, m_2, l_2 are fixed, and l_1 is variable using the slider. Let's make a simplifying assumption that $m_2 = 10 m_1$ and write l_1 as a multiple of l_2

$$\begin{aligned} m_2 &= 10 m_1 \\ l_1 &= \alpha l_2 \end{aligned} \qquad \Rightarrow \qquad \omega_{\text{n}} \;=\; \sqrt{g\,\frac{m_2 l_2 - m_1 l_1}{m_2 l_2^2 + m_1 l_1^2}} \;=\; \sqrt{g\,\frac{10 m_1 l_2 - \alpha\, m_1 l_2}{10 m_1 l_2^2 + \alpha^2\, m_1 l_2^2}} \;=\; \sqrt{\frac{g}{l_2}}\,\sqrt{\frac{10-\alpha}{10+\alpha^2}}.$$

The first factor $\sqrt{g/l_2}$ is constant, and the second factor $\sqrt{(10-\alpha)/(1+\alpha^2)}$ is tunable. If α is just below 10, the natural frequency is very low $\omega_{\text{n}} \approx 0$. This corresponds to the tunable mass being at the topmost position in the metronome of Figure 1.19. Note that the scale on that metronome is labeled in "Beats Per Minute" (BPM) (so 60 BMP corresponds to 1Hz). The other extreme is when α is decreased to zero, which then means that $\sqrt{(10-\alpha)/(10+\alpha^2)} \to 1$, and ω_{n} approaches the maximum frequency possible which is $\sqrt{g/l_2}$. If say $l_2 = 2\text{cm}$, then $\sqrt{9.8\ \text{m/s}^2 / 0.02\ \text{m}} \approx 22\ \text{rad/s} = 3.5\ \text{Hz} = 210\ \text{BPM}$.

Finally, we should compare the system of Figure 1.19, which is a pendulum with two unbalanced masses, to the simple pendulum with one mass. The simple pendulum has a natural frequency of $\omega_{\text{n}} = \sqrt{g/l}$. To get a frequency range of say

$$\omega_{\text{n}} \;=\; (40-200)\,\text{BPM} \;\approx\; (2/3 - 3.3)\,\text{Hz} \;\approx\; (4.2 - 21)\,\text{rad/s},$$

we would need a length range of

$$l \;=\; \frac{g}{\omega_n^2} \;\approx\; \frac{9.8\ \mathrm{m/s^2}}{(4.2-21)^2\ \mathrm{rad^2/s^2}} \;\approx\; (0.56-0.02)\,\mathrm{m}.$$

The difficulty is at the lower end of the desired frequency range. If you want 40 BPM, then you would need a simple pendulum of almost half a meter length, which is unwieldy, and certainly much smaller than a typical metronome. With the unbalanced double mass design, one can reach the lower end of the frequency with a device that essentially can be made as small as desired.

The Rayleigh-Ritz Method

The Energy Method described above is often referred to as the Rayleigh-Ritz Method in more general systems. It is exact for single degree of freedom systems, but only approximate for system with higher degrees of freedom, for which the method provides approximation to the "eigenvalues" of n-DOF systems, and those eigenvalues in turn determine the several natural frequencies of vibrations. These approximations can be surprisingly good especially for deformable bodies for which exact calculations are often quite involved. This topic will be discussed in Chapter 10.

1.7 Further Examples

In this section we present further example that each demonstrate a physical system or an analysis technique that will be covered in more depth in a later chapter. The first case is a simple 1-DOF approximation to a clamped beam, which in reality is a continuum system of the type studied in Chapters 9 and 10. The second is the so-called sprung-beam, or equivalently half-car model discussed earlier. This a 2-DOF system that is best studied using matrix methods as developed starting in Chapter 7.

1.7.1 Approximation of Cantilevered Beams

Vibration analysis of flexible and deformable bodies such as beams and plates is the subject of later chapters (Chapters 9 and 10). Unlike mathematical models of rigid bodies which involve Ordinary Differential Equations (ODEs), the analysis of flexible bodies require the use of Partial Differential Equations (PDEs). However, in this section we present a simple rigid-body approximation of a clamped-free beam as a single degree of freedom system with a torsional spring.

Consider the clamped-free beam shown in Figure 1.20. This beam has many "modes" (i.e. shapes) of vibrations. In fact, it has an infinite number of such modes. Again, this will be studied in detail in later chapters. Figure 1.20 (*Left*) shows the so-called "first mode" of vibration. We will try to model this mode of vibration using the highly simplified model shown in Figure 1.20 (*Middle*), where the clamped-free beam is replaced by an equivalent (in mass and geometry) rigid beam pivoting around a torsional spring. According to (1.10), the dynamics of such a system are given by

$$\mathrm{J}\ \ddot{\theta}(t) \;=\; -\,k_{\mathrm{eff}}\ \theta(t), \tag{1.42}$$

where $\mathrm{J} = mL^2/3$ is the moment of inertia around the end point (m is the mass and L is the length of the rigid-beam respectively), and k_{eff} is the (yet to be determined) effective torsional spring constant.

How is k_{eff} in (1.42) determined? One approximation is to think of the static deflection scenario shown in Figure 1.20 (*Right*). In the static case, a force F applied at the free end causes a vertical deflection δy at that end given by the equation[11]

$$\delta y \;=\; \frac{1}{EI}\ \frac{L^3}{3}\ F, \tag{1.43}$$

[11] This comes from solving the static beam equation $EI\ y^{(4)}(x) = 0$, where $y(x)$ is the vertical beam deflection at location x. The clamped-end boundary conditions are $y(0) = 0$, $y^{(1)}(0) = 0$, and the free-end boundary conditions are $y^{(2)}(L) = 0$, $y^{(3)}(L) = -\frac{1}{EI}F$.

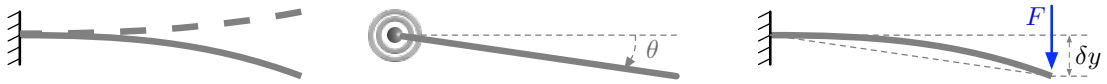


Figure 1.20: (*Left*) A full model of the vibrations of a clamped-free beam requires using a partial differential equation to solve for the "modes" of vibration. Here, the first such mode is depicted. (*Middle*) A simplified approximate model for a clamped-free beam is as a rigid beam pivoting around a torsional spring. (*Right*) To determine the effective torsional spring constant, we can use the static deflection equation for a clamped-free beam which states that the deflection $\delta y = \frac{L^3}{3EI}\, F$, where F is the force applied at the free end. The effective torsional spring constant is then obtained from the ratio of applied static moment FL to the resultant static angular deflection $\delta\theta \approx \delta y/L$, i.e. $k_{\text{eff}} = \frac{FL}{\delta\theta} \approx \frac{FL}{\delta y/L} = 3EI/L$.

where E is Young's modulus and I is the beams cross-sectional "area moment of inertia" (not to be confused with J above). Thus k_{eff} can be determined by comparing the static deflection of a torsional spring model to that of the flexible beam

$$\begin{aligned} k_{\text{eff}} \;=\; \frac{\text{applied static torque}}{\text{static angular deflection}} \;&=\; \frac{F\ L}{\delta\theta} && \text{(for the torsional spring model)} \\ &\approx\; \frac{F\ L}{\delta y/L} && \text{(using } \delta\theta \approx \sin(\delta\theta) = \delta y/L \text{ for small } \delta\theta\text{)} \\ &=\; \frac{F\ L^2}{\frac{1}{EI}\ \frac{L^3}{3}\ F} && \text{(using (1.43))} \\ &=\; 3EI/L. \end{aligned}$$

Now using this (together with $\mathrm{J} = mL^2/3$) in (1.42) gives

$$\frac{mL^2}{3}\,\ddot{\theta}(t) \;=\; -\,\frac{3EI}{L}\,\theta(t) \qquad \Leftrightarrow \qquad \ddot{\theta}(t) \;=\; -\,\frac{9EI}{mL^3}\,\theta(t). \tag{1.44}$$

Equation (1.44) implies that the natural frequency of oscillation of the torsional beam model is

$$\omega_{\text{n}} \;=\; 3\sqrt{\frac{EI}{mL^3}}\ \text{rad/s}.$$

This is not too far off from the actual frequency for a clamped-free beam (as will be calculated in Chapter 10) of $3.52\sqrt{EI/mL^3}$ rad/s. Depending on the application, this may or may not be a reasonable approximation.

1.7.2 2-DOF Example: Sprung Beam or Half-Car Model

In this section we detail the modeling of the half-car model discussed earlier in Figure 1.15c. This system has two mechanical degrees of freedom, and the brief analysis we present here will serve as a preview of the matrix methods developed later for vibrations analysis of n-Degree Of Freedom (n-DOF) systems.

Two coordinate systems for the half-car model (also sometimes referred to as the (rigid) sprung beam) are depicted in Figure 7.1. Let x be the vertical displacement of the center of mass, and x_1 and x_2 be the vertical displacements of the support points. Assume the origin of the coordinate systems for x_1, x_2 and θ are chosen so that $x_1 = 0$, $x_2 = 0$ and $\theta = 0$ correspond to the two springs being at equilibrium. Let m be the mass of the rigid body, then Newton's second law for vertical and rotational motions are

$$\begin{aligned} m\,\ddot{x} \;&=\; -\,k_1x_1 - k_2x_2 - c_1\dot{x}_1 - c_2\dot{x}_2, \\ \mathrm{J}\,\ddot{\theta} \;&=\; l_1\,(k_1x_1 + c_1\dot{x}_1) \;-\; l_2\,(k_2x_2 + c_2\dot{x}_2)\,, \end{aligned} \tag{1.45}$$

where J is the moment of inertia about the center of mass. Note the signs on the torques $l_1\,(k_1x_1 + c_1\dot{x}_1)$ and $l_2\,(k_2x_2 + c_2\dot{x}_2)$ which are due to the sign convention on θ being measured positively in the

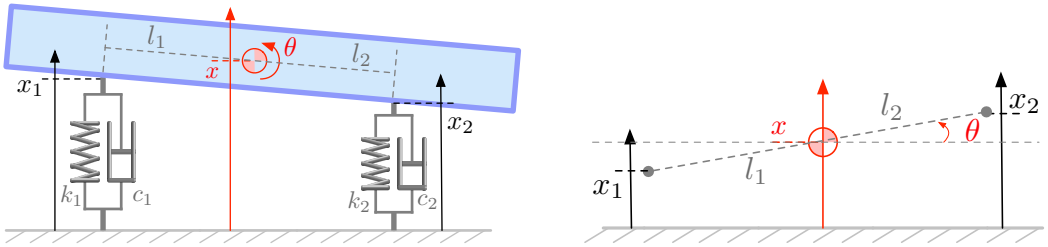


Figure 1.21: (*Left*) Schematic of the "sprung-beam" or "half-car" model, which has two degrees of freedom. The coordinates are either (x_1, x_2), the vertical displacements of the support points, or equivalently (x, θ) the vertical displacement of the center of mass and the rotation angle of the mass respectively. (*Right*) The kinematics used to relate the two coordinate systems (x, θ) and (x_1, x_2).

counter-clockwise direction. For example, when x_1 is positive, spring k_1 exerts a downwards force on its support point, which is a positive torque on the mass. On the other hand, when x_2 is positive, spring k_2 exerts a downwards force, which results in a clockwise (negative) torque on the mass.

We need to rewrite the equations above in terms of only one set of coordinates, either (x, θ) or (x_1, x_2), but not both. The first option is to use the kinematics to express (x_1, x_2) in terms of (x, θ). To do this, we make the simplifying assumption (reasonable for small oscillations in θ) that the springs and dampers move only vertically[12] (see Figure 7.1 (*Right*))

$$\left.\begin{aligned} x_1 &\approx x - l_1 \sin\theta \approx x - l_1\theta \\ x_2 &\approx x + l_2 \sin\theta \approx x + l_2\theta \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} m\,\ddot{x} &= -(k_1+k_2)\,x + (k_1l_1 - k_2l_2)\,\theta \\ &\quad -(c_1+c_2)\,\dot{x} + (c_1l_1 - c_2l_2)\,\dot{\theta} \\ \mathrm{J}\,\ddot{\theta} &= (k_1l_1 - k_2l_2)\,x - (k_1l_1^2 + k_2l_2^2)\,\theta \\ &\quad (c_1l_1 - c_2l_2)\,\dot{x} - (c_1l_1^2 + c_2l_2^2)\,\dot{\theta} \end{aligned}\right. \tag{1.46}$$

Note an important feature of these two equations. The first one is for the derivative $\ddot{x}$, but it depends on both x and θ and their derivatives. Similarly, the second equation is for $\ddot{\theta}$, but it also depends on both x and θ and their derivatives. The equations are therefore *coupled*. Each equation cannot be solved separately, they have to be solved together. In the language of dynamics, the equations for $\ddot{x}$ and $\ddot{\theta}$ describe vertical and rotational dynamics respectively. The fact that their equations are coupled means that vertical vibrations effect rotational vibrations and vice versa. A very useful technique that helps reveal the underlying structure of complex equations like (7.2) is to rewrite them as a *single matrix differential equation* as follows.

$$\begin{bmatrix} m\,\ddot{x} + (c_1+c_2)\,\dot{x} - (c_1l_1 - c_2l_2)\,\dot{\theta} + (k_1+k_2)\,x - (k_1l_1 - k_2l_2)\,\theta \\ \mathrm{J}\,\ddot{\theta} - (c_1l_1 - c_2l_2)\,\dot{x} + (c_1l_1^2 + c_2l_2^2)\,\dot{\theta} - (k_1l_1 - k_2l_2)\,x + (k_1l_1^2 + k_2l_2^2)\,\theta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{1.47}$$

$$\Leftrightarrow \quad \begin{bmatrix} m & 0 \\ 0 & \mathrm{J} \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} c_1+c_2 & c_2l_2 - c_1l_1 \\ c_2l_2 - c_1l_1 & c_1l_1^2 + c_2l_2^2 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} k_1+k_2 & k_2l_2 - k_1l_1 \\ k_2l_2 - k_1l_1 & k_1l_1^2 + k_2l_2^2 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{1.48}$$

The equation (7.3) is exactly the two equations (7.2) written as the two components of a vector. This vector equation can be further rearranged into the matrix-vector equation (7.4) in which each term represents all derivatives of a given order. Here the three terms collect derivatives of second, first and zeroth orders respectively.

An alternate model is to rewrite equations (7.1) using the coordinates (x_1, x_2) by substituting for (x, θ) in terms of (x_1, x_2). First note that the mapping $(x, \theta) \mapsto (x_1, x_2)$ in (7.2) can be written in matrix-vector form and inverted as follows

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -l_1 \\ 1 & l_2 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} x \\ \theta \end{bmatrix} = \frac{1}{l_1+l_2} \begin{bmatrix} l_2 & l_1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (l_2x_1 + l_1x_2)/(l_1+l_2) \\ (x_2 - x_1)/(l_1+l_2) \end{bmatrix}$$

[12] This is clearly not the case if the beam is assumed to be a rigid body. If $x_1 \neq x_2$, then the support points will have to move laterally. This motion can be assumed negligible for small oscillations θ. An alternative way to view this is that if we write the full two dimensional model that allows for lateral motion, then linearizing around $\theta \approx 0$ will give the model above that ignores lateral motion.

Substituting these expressions for (x, θ) in (7.1) gives

$$\begin{aligned} m\ (l_2\ddot{x}_1 + l_1\ddot{x}_2)/(l_1+l_2) &= -k_1x_1 - k_2x_2 - c_1\dot{x}_1 - c_2\dot{x}_2 \\ \mathrm{J}\ (-\ddot{x}_1 + \ddot{x}_2)/(l_1+l_2) &= l_1\,(k_1x_1 + c_1\dot{x}_1) \ - \ l_2\,(k_2x_2 + c_2\dot{x}_2) \end{aligned}$$

These equations can be reorganized into a matrix-vector form similar to (7.4) as follows

$$\frac{1}{l_1+l_2}\begin{bmatrix} ml_2 & ml_1 \\ -\mathrm{J} & \mathrm{J} \end{bmatrix}\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} + \begin{bmatrix} c_1 & c_2 \\ c_1l_1 & -c_2l_2 \end{bmatrix}\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} + \begin{bmatrix} k_1 & k_2 \\ k_1l_1 & -k_2l_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0. \tag{1.49}$$

n-DOF Mass-Spring-Damper System in Matrix Form

Note the similarity in the structures of the equations (7.4) and (7.5). They are both of the following form

$$\begin{bmatrix} \quad \\ M \\ \quad \end{bmatrix}\begin{bmatrix} \quad \\ \ddot{\boldsymbol{x}}(t) \\ \quad \end{bmatrix} + \begin{bmatrix} \quad \\ C \\ \quad \end{bmatrix}\begin{bmatrix} \quad \\ \dot{\boldsymbol{x}}(t) \\ \quad \end{bmatrix} + \begin{bmatrix} \quad \\ K \\ \quad \end{bmatrix}\begin{bmatrix} \quad \\ \boldsymbol{x}(t) \\ \quad \end{bmatrix} = \begin{bmatrix} \quad \\ 0 \\ \quad \end{bmatrix} \tag{1.50}$$

$$\Leftrightarrow \qquad M\ddot{\boldsymbol{x}}(t) \ + \ C\dot{\boldsymbol{x}}(t) \ + \ K\boldsymbol{x}(t) \ = \ 0, \tag{1.51}$$

where $\boldsymbol{x}(t)$ is a *vector* of coordinates, and M, C, K are *matrices* of system's coefficients. Such equations are the n-DOF generalizations of the single DOF equation (1.32). M is referred to as the "mass matrix", and similarly, C and K are called the damping and stiffness matrices respectively. Equations (1.50) and (7.6) are exactly the same. The first equation (1.50) graphically emphasizes the dimensions of the matrices and vectors for illustration. The second equation is written in the much more compact matrix-vector form, but the reader should always keep in mind that despite the simplicity of writing it in that form, a lot of information is "coded into" the entries of the matrices M, C, and K.

We will see in Chapter 7 how to write all n-DOF systems in the standard form (7.6), which will then enable a unified method of analysis for all such systems.

Decoupling and Normal Modes

When analyzing n-DOF systems written in a matrix form like (7.6), certain coordinates are much better than others. Take for example the equations (7.4) written in (x, θ) coordinates, and assume a special case where $k_1 = k_2 = k$, $c_1 = c_2 = c$, and $l_1 = l_2 = l$. The equations then simplify to

$$\begin{bmatrix} m & 0 \\ 0 & \mathrm{J} \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} 2c & 0 \\ 0 & 2cl^2 \end{bmatrix}\begin{bmatrix} \dot{x} \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 2k & 0 \\ 0 & 2kl^2 \end{bmatrix}\begin{bmatrix} x \\ \theta \end{bmatrix} = 0. \tag{1.52}$$

Note that the matrices M, C and K are now *diagonal* (i.e. all off-diagonal entries are zero). This matrix equation is therefore *decoupled* into two scalar equations

$$m\ \ddot{x}(t) \ + \ 2c\ \dot{x}(t) \ + \ 2k\ x(t) \ = \ 0, \tag{1.53}$$

$$\mathrm{J}\ \ddot{\theta}(t) \ + \ 2cl^2\ \dot{\theta}(t) \ + \ 2kl^2\ \theta(t) \ = \ 0. \tag{1.54}$$

Note that the differential equation for $x(t)$ does not involve $\theta(t)$, and similarly, the differential equation for $\theta(t)$ does not involve $x(t)$. This means each differential equation can be solved independently of the other. In this case we say that the *dynamics of x and θ are decoupled* in the sense that they do not influence each other. We call the motion of $x(t)$ the *vertical mode*, and that of θ the *rotational mode* of vibration respectively. Those are the *normal modes* of vibration of this 2-DOF system. In general, an n-DOF system will have n different normal modes of vibration.

Let's make more simplifying assumptions to get some intuition for the distinction between vertical and rotational vibrations. Assume no damping $c = 0$, and that the support points of the sprung beam in Figure 7.1 are at each end. Additionally, assume the beam is slender so that its moment of inertia (about the center of mass) is that of a bar of uniform length $2l$

$$\mathrm{J} \ = \ m(2l)^2/12 \ = \ ml^2/3$$

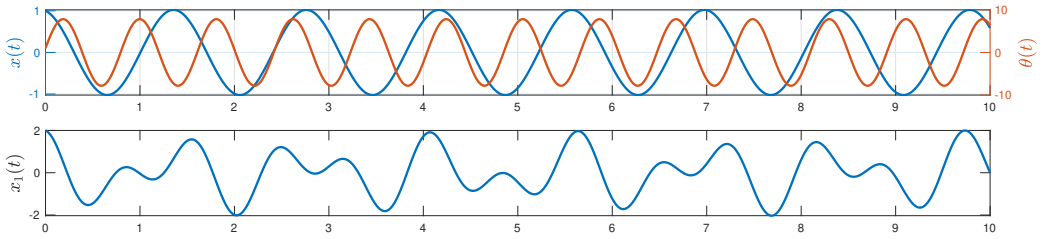


Figure 1.22: Vibrations of the sprung beam model in the special case where vertical and rotational dynamics (7.8)-(7.9) are decoupled. (*Top*) Both rotational $\theta(t)$ and vertical $x(t)$ vibrations are pure sinusoids, but with different frequencies. The frequency of rotational vibrations $\theta(t)$ is higher than that of vertical vibrations $x(t)$ as predicted by the formulas (7.11) for their respective natural frequencies. (*Bottom*) The vibration of the support point $x_1(t)$ is a superposition of the two modes (vertical and rotational) of vibrations.

With these assumptions, the rotational motion equation (7.9) becomes

$$ml^2/3\ \ddot{\theta}\ +\ 2kl^2\ \theta\ =\ 0 \qquad \Rightarrow \qquad \ddot{\theta}\ +\ \frac{6k}{m}\ \theta\ =\ 0. \tag{1.55}$$

Now comparing the vertical motion equation (7.8) (with $c = 0$) and the rotational motion equation (7.10), we see that they are both of the Mass-Spring type, but with two different natural frequencies

$$\begin{aligned} \text{vertical vibrations:} \quad & \omega_{\text{n}} \ = \ \sqrt{2k/m} \\ \text{rotational vibrations:} \quad & \omega_{\text{n}} \ = \ \sqrt{6k/m} \ = \ \sqrt{3}\ \sqrt{2k/m}. \end{aligned} \tag{1.56}$$

Thus rotational vibrations have a natural frequency that is $\sqrt{3}$ times higher than that of vertical vibrations.

Any free vibrations of this 2-DOF system will be a *superposition* of those two modes of vibrations with those two different frequencies. For example, the motion of one of the support points, say $x_1(t)$ is linear combination of both $x(t)$ and $\theta(t)$ since (recall (7.2))

$$x_1(t)\ =\ x(t)\ -\ l\ \theta(t).$$

Figure 7.2 shows an example of the vibration of such a system. Note how vertical motion $x(t)$ and rotational motion $\theta(t)$ are both pure sinusoids, but with different natural frequencies. The vibration of $x_1(t)$ appears to be some superposition (a linear combination) of the two motions.

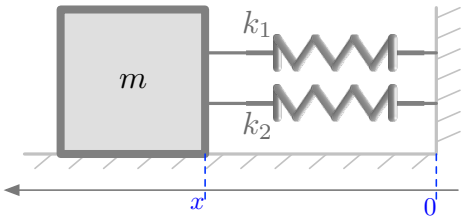
The fact that the matrix equation (7.7) involved only *diagonal* matrices is what led to the decoupling of the dynamics of x and θ as described by the two mutually-independent equations (7.8) and (7.9). Although this looks like a "lucky accident" due to the symmetry of the problem, there is a general method to take any coupled matrix problem representing an n-DOF system, and find the fundamental *normal modes* of vibration in a similar manner to what was done above. The mathematical technique is based on matrix diagonalization, which in turn is based on finding eigenvalues of eigenvectors of matrices. This is the topic of normal mode analysis developed in Chapter 7.2.

Exercises

Exercise 1.1

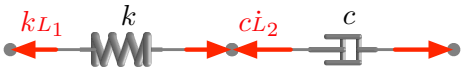
Consider the diagram below where springs k_1, k_2 have equilibrium lengths L_1, L_2 respectively.

(a) Find an expression for the mass' position x at static equilibrium. (b) What is the natural frequency of oscillation of this system, and how does it compare to the case when the two springs have equal equilibrium lengths (i.e. when $L_1 = L_2$)?



Exercise 1.2

A spring and a damper in series. TBC



Exercise 1.3

Coordinate systems TBC

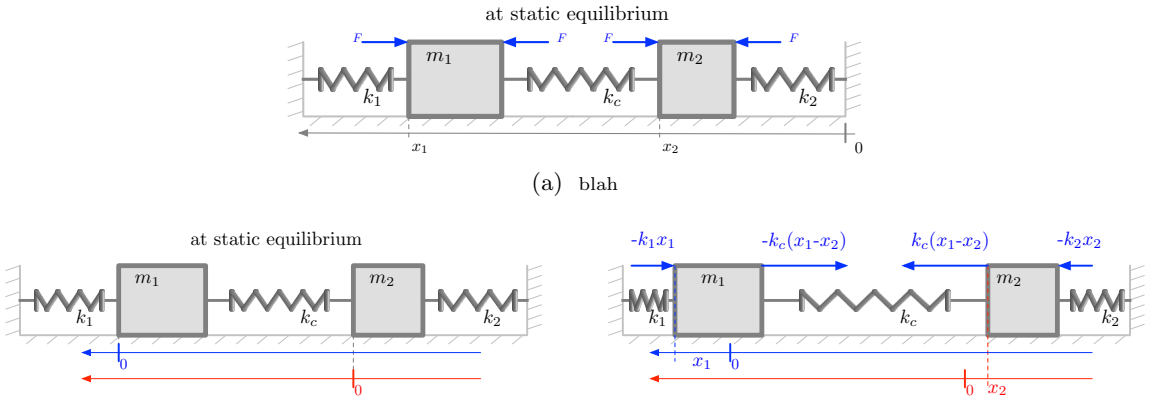
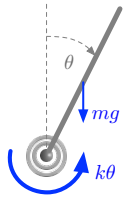


Figure 1.23: blah

Exercise 1.4

Consider the rigid beam on a torsional spring shown here. The beam has uniform density, mass m, length l, and torsional spring constant k. Gravity acts downwards as shown. You may assume the small oscillations approximation of $\sin(\theta) \approx \theta$.



1. Derive a formula for the natural frequency of oscillation ω_n of this beam in terms of m, k, l, and g. Is the frequency higher or lower than that when ignoring gravity?
2. What is the minimum stiffness k required so that the beam does not "fall over"?

Chapter 2

Complex Numbers, Functions and Phasors

Vibration analysis of harmonically forced systems involves additions, differentiation and integration of sinusoidal signals. Phasor analysis greatly simplifies these mathematical relations by representing sinusoidal signals that have a common frequency as vectors in the complex plane. The amplitude and phase of a signal is encoded as the magnitude and phase of the complex number, i.e. the phasor, representing it. Many trigonometric identities are encoded in the arithmetic of complex numbers in this manner, and the use of phasor analysis obviates the need for the complicated manipulation of trigonometric identities. Familiarity with complex numbers is thus essential for vibration analysis, and this chapter provides a quick introduction and review of complex arithmetic.

Introduction and Motivation

Under harmonic excitation, and after the effects of initial conditions have died down (i.e. asymptotically as $t \to \infty$), mechanical vibrations as well as voltage and current oscillations in AC circuits are sinusoids[1] (sine and cosine functions, and combinations thereof). All such functions can be represented as follows

$$x(t) \;=\; \mathsf{x}\,\cos\big(\omega\, t \;+\; \theta\big), \tag{2.1}$$

where x is the *amplitude*, ω is the *frequency* (in rad/s), and θ is the *phase shift* (which could be positive or negative).

To analyze mechanical vibrations or electrical oscillations, we often have to add and subtract sinusoids, as well as differentiate and integrate them. If you remember your trigonometric identities, you will recall that adding two sinusoids gives another sinusoid. For example, we can use the identity $\cos(\alpha) + \cos(\beta) \;=\; 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)$ to add two cosines with equal amplitudes but different phases

$$\begin{aligned} x_1(t) + x_2(t) \;&=\; \mathsf{x}\cos\big(\omega t \;+\; \theta_1\big) \;+\; \mathsf{x}\cos\big(\omega t \;+\; \theta_2\big) && (2.2)\\ &=\; 2\,\mathsf{x}\;\cos\big(\omega t + \tfrac{\theta_1+\theta_2}{2}\big)\;\;\cos\big(\tfrac{\theta_1-\theta_2}{2}\big) && (2.3)\\ &=\; \Big(2\mathsf{x}\cos\big(\tfrac{\theta_1-\theta_2}{2}\big)\Big)\;\;\cos\big(\omega t + \tfrac{\theta_1+\theta_2}{2}\big) && (2.4)\\ &=:\; \mathsf{y}\;\cos\left(\omega t + \phi\right) \;\;=:\; y(t). && (2.5) \end{aligned}$$

We see that those two cosines add up to another cosine but with amplitude $\mathsf{y} := 2\mathsf{x}\cos\big(\frac{\theta_1-\theta_2}{2}\big)$ and phase $\phi := (\theta_1+\theta_2)/2$. Note that if the amplitudes of the two original sinusoids were not equal, the trigonometric identities needed would have been much more complicated. In general, the amplitude and phase of the sum $y(t) = x_1(t) + x_2(t)$ will depend on amplitudes and phases of both x_1 and x_2.

Many other formulas for combinations of different sinusoids can be derived using trigonometric identities, but the procedures can quickly become a cumbersome algebraic nightmare! Here comes

[1] This will be demonstrated in Chapter 4.

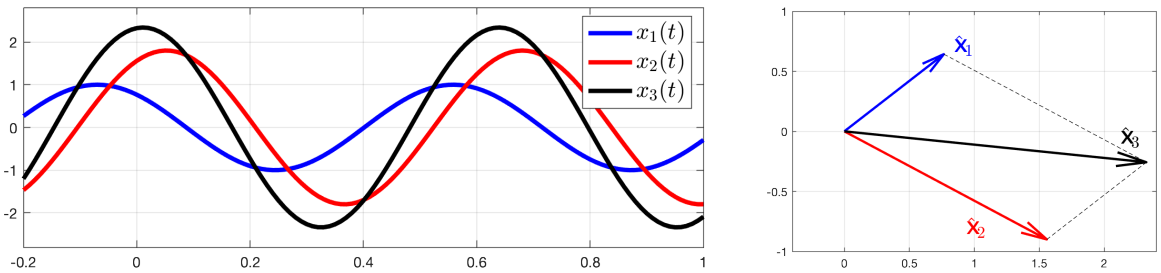


Figure 2.1: *(Left)* Two sinusoids of the same frequency, but different amplitudes and phases, $x_1(t) = \cos(10t + 40^\circ)$ and $x_2(t) = 1.8\ \cos(10t - 30^\circ)$ are added together to form a third sinusoid $x_3(t) = x_1(t) + x_2(t)$. It is complicated to derive an expression for $x_3(t)$ using trigonometric identities. The *phasor representation* considerably simplifies this. Signals are represented as "phasors", which are vectors in the complex plane where the complex number $\hat{x}_i$ "represents" the time function $x_i(t)$. (*Right*) The magnitude and phase of the phasor $\hat{x}_3$ (shown in black) representing $x_3(t)$ is simply obtained by the vector addition in the complex plane of the phasors $\hat{x}_1$ (blue) and $\hat{x}_2$ (red) representing $x_1(t)$ and $x_(2)$ respectively. The proof of this statement is in Section 2.3

in the concept of *phasors*, which allow for a very compact representation of sinusoids, and *encode all additive trigonometric identities simply in the arithmetic of complex numbers*! We will develop this "phasor calculus" in this chapter, but as a preview, Figure 2.1 illustrates how two sinusoids can be added *as functions* by simply adding their representing phasors *as vectors* in the complex plane. Assuming the frequency ω is fixed, each sinusoid of the form (2.1) is represented by a complex number (its *phasor representation*), which is also a 2-vector in the plane (Figure 2.1, right). The magnitude of the phasor is the length of the vector, and it is the amplitude x of the corresponding sinusoid. The phase of the phasor as a complex number is the phase θ of the sinusoid. The addition of two sinusoids of the same frequency is another sinusoid with its own amplitude and phase, the phasor representing that addition is just the addition of the two original phasors as complex numbers (or equivalently as 2-vectors in the plane). In this way, phasors give a very convenient way of adding sinusoids of the same frequency without invoking any trigonometric identities. We will also shortly see that the phasor representation allows for very simple calculations involving differentiation and integration of sinusoids. These nice differentiation and additive properties of phasors allow us to easily analyze relationships between currents and voltages in AC circuits with RLC components, as well as vibrations in complex mass-spring-damper systems.

To get good proficiency for this "phasor anlaysis" it is important to review and be fully proficient with complex number arithmetic as done in Section 2.2.

2.1 Sinusoids, Phase and Time Shifts

Sinusoidal signals are specified by their frequencies or periods, amplitudes and phase or time shifts. All sinusoids can be represented in either the "sine" or "cosine" form. They are periodic functions, which means that the signal over all time is a repetition of some function defined over a fundamental period. Time shifts of functions represent either time delays or advances, and these shifts can be described as either time or phase shifts.

The "cosine" (or "sine") Representation

We use the term *sinusoid* to refer to sine and cosine functions, and combinations thereof. All such functions can be represented in the "cosine form"

$$u(t) \ = \ \mathsf{u}\ \cos\big(\omega t + \theta\big), \tag{2.6}$$

where t is the independent variable (usually interpreted as time). The sinusoid is completely specified by its *parameters*, which are the constants u, ω and θ. u is the *amplitude*, ω is the *frequency* (in

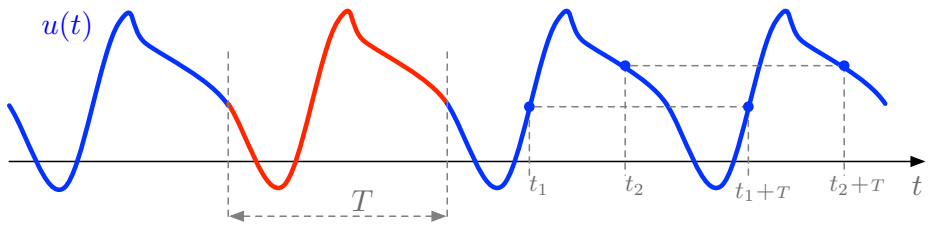


Figure 2.2: A signal is *periodic* with period T (T-periodic) if it is uniquely determined by its values over a time interval of length T. This also means that a T-periodic signal is generated from a function defined over an interval of width T (shown in red), and then repeating it an infinite number of times with shifts of multiples of T. A T-periodic signal u has the property that for any t, the values of u at times that are "T apart" are equal, i.e. for any t, $u(t+T) = u(t)$. The is illustrated above for two different times t_1 and t_2.

rad/s or deg/s), and θ is the *phase shift* (in rad or deg), which could be positive or negative. For example, if we need to represent a sine, the identity

$$\sin\big(\omega t\big) \;=\; \cos\big(\omega t - 90^\circ\big) \;=\; \cos\big(\omega t - \pi/2\big)$$

means that a sine is just a cosine with a -90° phase shift (called a 90° "phase delay" or "phase lag" because it is negative). This also means that we could represent any sinusoid such as (2.6) in the alternate "sine form"

$$u(t) \;=\; \mathsf{u}\,\sin\big(\omega t + \phi\big),$$

where $\phi = \theta - \pi/2$. The manipulations of signals (and their phasors) are exactly the same regardless of whether one chooses the sine or cosine forms, but the choice must be made upfront before calculations are done. We will mostly use the cosine form.

Periodic Signals

The signal in (2.6) is a special case of a periodic signal. More generally, a signal $u(t)$ is called periodic (or more precisely T-periodic) if it has the property

$$\text{for all } t, \qquad u(t) \;=\; u(t+T),$$

where the constant T is called the *period*. This is illustrated in Figure 2.2. A T-periodic signal is completely determined by its values over any time interval of length T. In Figure 2.2 a particular choice is made for this interval (shown in red), but any other interval can be taken provided it is of length T.

For a sinusoidal signal of the form (2.6), its period is determined by its frequency ω. To find the relationship between ω and T, recall that the cosine function is itself periodic with period 2π (or 360°) since for all α, $\sin(\alpha + 2\pi) = \sin(\alpha)$. Therefore ωt has to advance by 2π for $\cos(\omega t + \theta)$ to repeat, and we conclude that the period T of $\cos(\omega t + \theta)$ is given by

$$\begin{aligned} T \;&=\; 2\pi/\omega && \text{if } \omega \text{ is given in rad/s,} \\ T \;&=\; 360^\circ/\omega && \text{if } \omega \text{ is given in deg/s.} \end{aligned}$$

Note that in order to make sense of $\cos(\omega t + \theta)$, the quantity ωt must have units of radians or degrees. Therefore ω must have units of radians/time or degrees/time.

Another unit for measuring frequencies is Hz (Hertz), which are *cycles/second*. Cycle is another name for the period T, so we can say that Hz = *periods/second*. Thus if f is measured in Hz, then $1/f$ is the period expressed as seconds per period (equivalently seconds per cycle)

$$T \text{ (in seconds/cycle)} \;=\; \frac{1}{f \text{ in (cycles/second)}}.$$

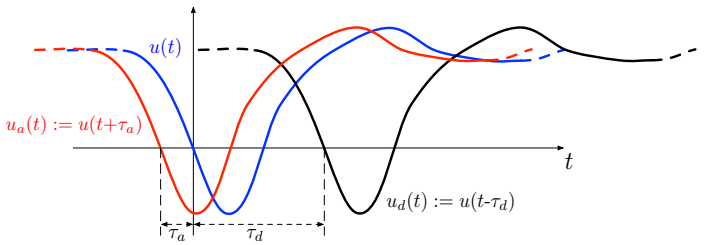


Figure 2.3: Illustration of time delays and time advances of signals, and the corresponding shifts of of their graphs. The signal $u(t)$ delayed by τ_d is the signal $u_d(t) := u(t - \tau_d)$. The graph of the time-delayed signal u_d is exactly the graph of the original signal $u(t)$ but *right shifted* by the amount of time delay τ_d. The signal $u_a(t) := u(t + \tau_a)$ is the original signal $u(t)$ advanced by τ_a time units. Its graph is a *left-shifted* version of the graph of $u(t)$.

The relations between T, f in Hz and ω in rad/s are thus

$$T \;=\; \frac{2\pi}{\omega} \;=\; \frac{1}{f}, \qquad\qquad \omega \;=\; 2\pi\, f,$$

with a similar relationship if ω is given in deg/s. A simple way to remember that last formula is as a unit conversion. There are 2π radians per period, so

$$f\,\left(\frac{\cancel{\text{cycle}}}{\text{s}}\right)\quad 2\pi\,\left(\frac{\text{rad}}{\cancel{\text{cycle}}}\right) \;=\; \omega\,\left(\frac{\text{rad}}{\text{s}}\right).$$

Time and Phase Shifts

To understand phase shifts, it is important to understand time delays and time advances. Let $u(t)$ be a signal[2]. The signal "delayed" by $\tau_d \geq 0$ time units is $u_d(t) := u(t - \tau_d)$, because e.g. the value of the signal u at time 0 occurs in the signal u_d at the later time τ_1. This is illustrated in Figure 2.3. Note also that in terms of graphs of signals, the graph of u_d is the same as the graph of u, but *right shifted* by τ_d. Thus time delays correspond to right shifts of the signal graph. On the other hand, the signal "advanced" in time by $\tau_a \geq 0$ time units is $u_a(t) := u(t + \tau_a)$. Note again for example that the value of u at time $t = 0$ occurs in u_a at time $t = -\tau_a$, i.e. before $t = 0$, and we can say that u_a is a time-advanced version of u. Figure 2.3 shows that a time advance corresponds to a *left shift* of the signal's graph.

When a signal is periodic, we can describe time shifts in terms of phase shifts. The phase shift is defined as the time shift *as a fraction of the period* expressed in degrees or radians, i.e. for any periodic signal

$$\text{phase shift} \;:=\; \begin{cases} \dfrac{\text{time shift}}{\text{period}}\; 360^\circ, & \text{if expressed in degrees,} \\ \dfrac{\text{time shift}}{\text{period}}\; 2\pi, & \text{if expressed in radians.} \end{cases}$$

Conversely, if a phase shift θ of a T-periodic signal is given, the the corresponding time shift τ is

$$\tau \;=\; \frac{\theta}{360^o}\; T \;=\; \frac{\theta}{2\pi}\; T, \tag{2.7}$$

depending on the units in which θ is given. Note that both time and phase shifts can be either positive or negative. Now we consider a pure sinusoid and a "phase shifted" version of it (here assumed given in radians)

$$\begin{aligned} v_1(t) &= A\cos\left(\omega t\right) \\ v_2(t) &= A\cos\left(\omega t + \theta\right) \;=\; A\cos\left(\omega\left(t + \frac{\theta}{\omega}\right)\right) \;=\; A\cos\left(\omega\left(t + \frac{\theta}{2\pi}T\right)\right) \;=\; v_1\left(t + \frac{\theta}{2\pi}T\right). \end{aligned}$$

Thus v_2 is v_1 but with a time shift of $\frac{\theta}{2\pi}T$. This is exactly (2.7) in this special case. This is illustrated for the signal $\sin\left(2\pi 10t\right)$ in Figure 2.4.

[2]The term *signal* is used to denote a function where the independent variable is *time*.

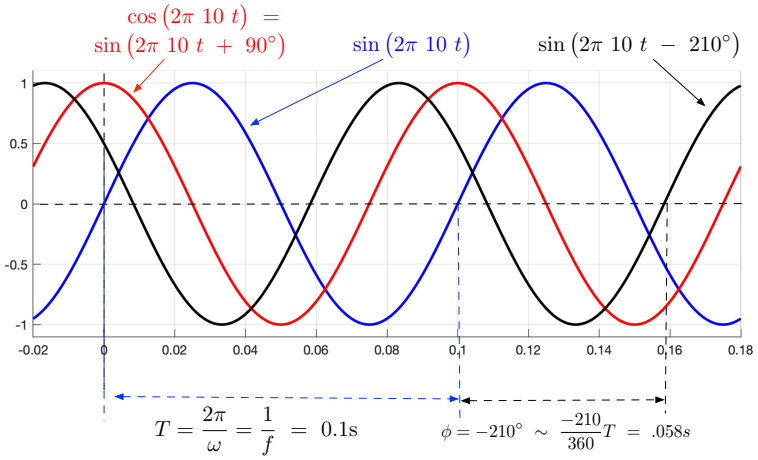


Figure 2.4: The blue signal is $\sin(2\pi\ 10\ t)$. Its period can be measured as the time between *upwards zero crossings*, which here appears to be 0.1s, and agrees with the formula $T = \frac{1}{f} = \frac{1}{10}$. The red signal is $\cos(2\pi\ 10\ t)$ which can be written as $\sin(2\pi\ 10\ t\ +\ 90^\circ)$, so it has a *phase lead* of 90° over the blue signal. A phase lead is equivalent to a *time advance* of $\bar{t} = \frac{90^\circ}{360^\circ} T = \frac{1}{4}\ 0.1s\ =\ 0.025s$, which is what the plot shows. A time advance corresponds to a *left shift* of the signal's plot. The signal in black is $\sin(2\pi\ 10\ t\ -\ 210^\circ)$, thus it has a *phase lag* of 210° over the blue signal. A phase lag corresponds to a *time delay*, and in this case it is $\bar{t} = \frac{210^\circ}{360^\circ} T\ =\ 0.058s$. A time delay corresponds to a *right shift* of the signal's plot.

2.2 Background: Arithmetic of complex numbers

A complex number z is a tuple of real numbers $z = (\alpha, \beta)$, where α and β are the real and imaginary parts respectively. We normally write

$$z\ =\ \alpha + j\beta,$$

where $j^2 := -1$. We will use the notation $\mathbb{R}(z)$ and $\mathbb{I}(z)$ to refer to the real and imaginary parts of a complex number, i.e.

$$z\ =\ \alpha + j\beta \qquad \Leftrightarrow \qquad \mathbb{R}(z) = \alpha,\ \ \mathbb{I}(z) = \beta.$$

The definition $j^2 := -1$ specifies the rules for adding and multiplying complex numbers as

$$z_1 + z_2 = (\alpha_1 + j\beta_1)\ +\ (\alpha_2 + j\beta_2)\ =\ (\alpha_1 + \alpha_2) + j(\beta_1 + \beta_2) \tag{2.8}$$

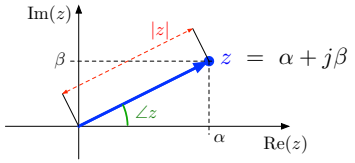
$$\begin{aligned} z_1 z_2 = (\alpha_1 + j\beta_1)(\alpha_2 + j\beta_2) &= \alpha_1\alpha_2 + j^2\beta_1\beta_2 + j\alpha_1\beta_2 + j\alpha_2\beta_1 \\ &= (\alpha_1\alpha_2 - \beta_1\beta_2) + j(\alpha_1\beta_2 + \alpha_2\beta_1) \end{aligned} \tag{2.9}$$

Thus addition of two complex numbers is done by adding their real and imaginary parts respectively. Multiplication however "mixes" the real and imaginary parts together. Both addition and multiplication operations can be viewed geometrically as shown next.

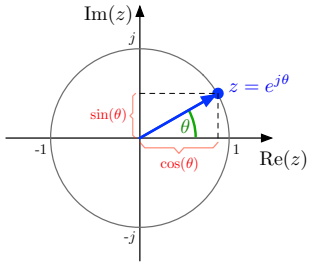
A complex number is best visualized as a vector in two dimensions with cartesian components (α, β). The *magnitude* of z (written $|z|$) is the length of this vector, and its *phase* (written $\angle z$) is the angle the vector makes with the horizontal axis measured counter-clockwise. See figure 2.5a. From basic trigonometry, the relations for the magnitude and phase of $z = \alpha + j\beta$ are

$$\begin{aligned} |z| &= \sqrt{\alpha^2 + \beta^2}, \qquad \angle z\ =\ \tan^{-1}\frac{\beta}{\alpha} \\ \alpha &= |z|\cos(\angle z), \qquad \beta\ =\ |z|\sin(\angle z) \end{aligned}$$

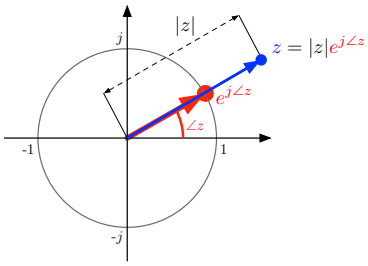
Another convenient form with which to write complex numbers is the *polar form*. It is derived using Euler's formula $e^{j\theta} = \cos(\theta) + j\sin(\theta)$, which can be thought of as describing complex numbers



(a) A complex number $z = \alpha + j\beta$ can be considered as a 2-vector in the plane with cartesian coordinates (α, β). The length of the vector is the *magnitude* $|z|$ of the complex number, and its *phase* $\angle z$ is the angle it makes with the positive real axis (counter-clockwise).

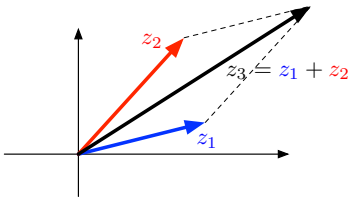


(b) The *unit circle* of the complex plane is the set of all complex numbers with magnitude 1. Euler's formula says that all complex numbers of the form $e^{j\theta} = \cos(\theta) + j\sin(\theta)$ lie on the unit circle as shown. The real number θ is the angle the vector $e^{j\theta}$ makes with the real axis.

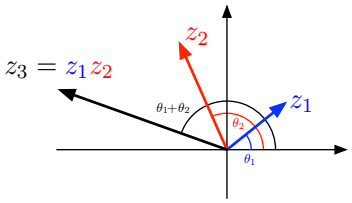


(c) The polar form $z = |z|e^{j\angle z}$ of a complex number can be interpreted as follows. The unit-length vector $e^{j\angle z}$ specifies z's direction, and the real scalar $|z|$ specifies its length. z is then the product of those two. Note that for any complex vector z, $e^{j\angle z}$ is another complex vector of the *same phase* as z, but with *unit* magnitude.

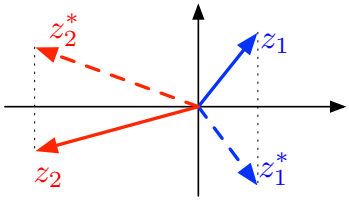
Figure 2.5: Complex numbers are identified with 2-vectors in the plane, which is why it is called the "complex plane". Complex numbers can be represented in either cartesian or polar forms. The polar form is best appreciated if one understands Euler's formula and its relation to the unit circle of the complex plane.



(a) Complex numbers add just like 2-vectors in the plane. The real and imaginary parts of $z_1 + z_2$ are the sum of the real and imaginary parts of z_1 and z_2 respectively.



(b) Multiplication of complex numbers is best visualized in their polar form. The phases add up, i.e. $\angle(z_1 z_2) = \angle z_1 + \angle z_2$, and the magnitudes (length of vectors) multiply $|z_1 z_2| = |z_1||z_2|$.



(c) Conjugation of a complex number flips the sign of its imaginary part. This can be nicely visualized as a reflection about the real axis.

Figure 2.6: Geometric views of the addition, multiplication and conjugation of complex numbers.

on the *unit circle* of the complex plane (this is all complex numbers with magnitude of 1 as shown in Figure 2.5b.) We can thus write any complex number z as

$$z \;=\; \alpha + j\beta \;=\; |z|\cos(\angle z) + j|z|\sin(\angle z) \;=\; |z|\,\big(\cos(\angle z) + j\sin(\angle z)\big) \;=\; |z|\,e^{j\angle z},$$

where again $|z|$ is the magnitude and $\angle z$ is the phase. This is called the "polar" representation of the complex number and is illustrated in Figure 2.5c.

In (2.8) we saw that the sum of two complex numbers is easy to define with cartesian coordinates. The product (2.9) however was a little more complicated. The product is easier to write in the polar representation since magnitudes multiply and phases add

$$z_1 z_2 \;=\; |z_1|\,e^{j\angle z_1}\,|z_2|\,e^{j\angle z_2} \;=\; |z_1|\,|z_2|\,e^{j(\angle z_1 + \angle z_2)}\,.$$

Note that it is easiest to add complex numbers in the real/imaginary (cartesian) representation, while it is easiest to multiply them in the polar representation. See Figure 2.6 for a geometrical view of these operations.

Some other useful relations which can be demonstrated using the polar representation are (try

these)

$$\begin{aligned}
|z_1 z_2| &= |z_1||z_2| && \textit{(the magnitude of the product is the product of the magnitudes)}\\
\angle(z_1 z_2) &= \angle z_1 + \angle z_2 && \textit{(the phase of the product is the sum of the phases)}\\
\left|\frac{z_1}{z_2}\right| &= \frac{|z_1|}{|z_2|} && \textit{(the magnitude of the ratio is the ratio of the magnitudes)}\\
\angle\left(\frac{z_1}{z_2}\right) &= \angle z_1 - \angle z_2 && \textit{(the phase of the ratio is the difference of the phases)}
\end{aligned} \tag{2.10}$$

For example $\angle\left(\frac{1}{j}\right) = \angle(1) - \angle(j) = 0^o - 90^o = -90^o$. Note that $\frac{1}{j} = -j$.

Given a complex number $z = \alpha + j\beta$ its *complex conjugate* z^* is defined by

$$z^* := \alpha - j\beta.$$

Euler's formula immediately implies that in polar form, conjugation simply negates the phase, i.e.

$$\left(A\, e^{j\theta}\right)^* = A\, e^{-j\theta}.$$

This has the geometric interpretation shown in Figure 2.6c. Conjugation *reflects* the complex number about the real axis. It doesn't change the length of the vector, but reverses its phase. This is illustrated in Figure 2.6c. It is easy to show that conjugation obeys the following rules

$$\begin{aligned}
\left(z_1 + z_2\right)^* &= z_1^* + z_2^*\\
\left(z_1 z_2\right)^* &= z_1^* z_2^*,
\end{aligned}$$

and that we can recover the real and imaginary parts from the complex number and its conjugate

$$\begin{aligned}
z + z^* &= 2\,\mathbb{R}(z) && \left(\text{since } z + z^* = (\alpha + j\beta) + (\alpha - j\beta) = 2\alpha\right)\\
z - z^* &= 2j\,\mathbb{I}(z) && \left(\text{since } z - z^* = (\alpha + j\beta) - (\alpha - j\beta) = 2j\beta\right),
\end{aligned}$$

2.3 The Phasor Representation of Sinusoidal Functions

The basic idea behind phasor representations is to represent a sinusoidal signal of a given frequency using a complex number, i.e. to *encode* the amplitude and phase of the sinusoid in the magnitude and phase of the complex number respectively. First, recall Euler's formula which implies that a cosine can be written as the real part of an exponential with an imaginary argument

$$e^{j\phi} = \cos(\phi) + j\sin(\phi) \qquad \Rightarrow \qquad \cos(\phi) = \mathbb{R}\left(e^{j\phi}\right) \qquad \Rightarrow \qquad \cos\left(\omega t\right) = \mathbb{R}\left(e^{j\omega t}\right).$$

The last equation shows that we can use Euler's formula to write a *cosine time function* as the real part of a *complex-valued function of time* $e^{j\omega t}$. "Complex-valued" means that for each t, the function $e^{j\omega t}$ has its value as a complex number. Since for each t, the quantity $e^{j\omega t}$ is complex, three dimensions are required to visualize the graph of such a function, one dimension for t, and two dimensions for the (time-varying) complex number $e^{j\omega t}$. This visualization is illustrated in Figure 2.7.

Alternatively, another geometrically intuitive way to visualize this function is as a *curve in the complex plane* parameterized by t. The function $e^{j\omega t}$ in particular traces a circular curve in the complex plane that goes around the unit circle counterclockwise starting from 1. This is illustrated in Figure 2.8. Since $e^{j\omega t}$ is a point on the unit circle at angle ωt radians (counter-clockwise from 1), the curve goes around the unit circle once every $\frac{2\pi}{\omega}$ time units. Thus the larger ω is, the faster this

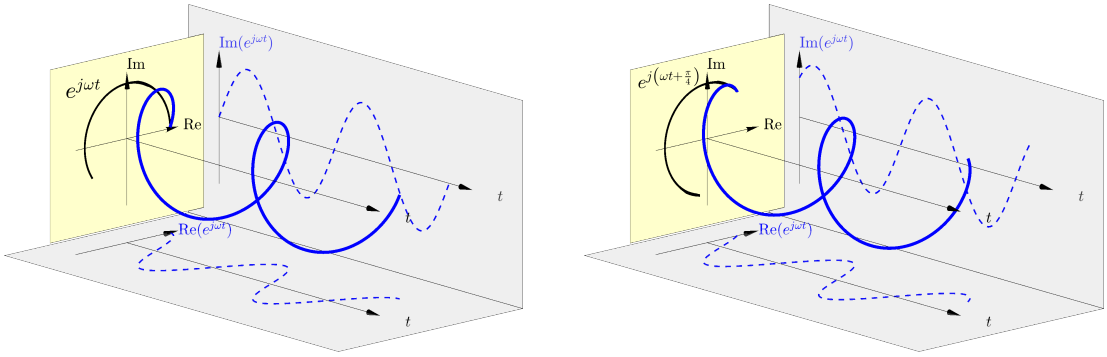


Figure 2.7: The complex-valued function of time $e^{j\omega t}$ visualized as a helical curve in three dimensions. Two dimensions are needed for the complex plane (shown here as the yellow plane), and the third dimension for t. At $t = 0$, the function $e^{j\omega t} = e^0 = 1$ starts at 1, and then rotates counterclockwise along the unit circle (depicted as the solid black curve in the complex plane, with only the initial part of the curve shown for clarity). The full graph of $e^{j\omega t}$ as a function of t is the blue helical curve. Euler's formula states that $e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$. The dashed curves show the real and imaginary parts as the projections of the helical curve onto the planes $(\mathbb{I}(e^{j\omega t}), t)$ and $(\mathbb{R}(e^{j\omega t}), t)$, which show the $\sin(\omega t)$ and $\cos(\omega t)$ functions respectively (in the left figure). The left figure depicts the function $e^{j\omega t}$, while the right figure depicts $e^{j(\omega t+\pi/4)}$ which starts at $e^{j\pi/4}$ at time $t = 0$. The imaginary and real parts of the right figure show a sine and a cosine, but with phase shifts.

curve goes around the unit circle. The *projection* of this curve onto the real axis gives the real-valued time function $\mathbb{R}(e^{j\omega t}) = \cos(\omega t)$.

Now fix a frequency ω, and consider a sinusoid of that frequency, but arbitrary amplitude x and phase θ

$$x(t) \;=\; \mathsf{x}\,\cos\big(\omega t+\theta\big) = \mathsf{x}\,\mathbb{R}\Big(e^{j(\omega t+\theta)}\Big) \;=\; \underbrace{\mathbb{R}\big(\mathsf{x}e^{j\theta}\; e^{j\omega t}\big)}_{\substack{\text{separate the time-dependent}\\ \text{from the time-constant factors}}} \;=\; \mathbb{R}\big(\hat{\mathsf{x}}\; e^{j\omega t}\big)\,, \qquad \text{where } \hat{\mathsf{x}} \;:=\; \mathsf{x}\,e^{j\theta}. \tag{2.11}$$

There are two factors in $\mathbb{R}\big(\hat{\mathsf{x}}\; e^{j\omega t}\big)$, the time-function $e^{j\omega t}$, which involves the fixed frequency ω, and thus is the same for all sinusoids of that frequency. The complex number $\hat{\mathsf{x}} = \mathsf{x}\; e^{j\theta}$ now *encodes* the amplitude and phase of the signal $\mathsf{x}\cos\big(\omega t + \theta\big)$ and uniquely represents it. This is illustrated in Figure 2.8 where three different examples are shown. The reader should examine those figures while imagining the evolution of the respective curves in the complex plane and their projections onto the real axis which produce the signals as a time functions. We now make a formal definition.

Definition 2.1. *Given a fixed frequency ω, the* phasor representation *of a time function of the form $x(t) = \mathsf{x}\;\cos(\omega t + \theta)$ is the complex number $\hat{\mathsf{x}} := \mathsf{x}e^{j\theta}$. We write this representation as*

$$x(t) \;=\; \mathsf{x}\;\cos\big(\omega t + \theta\big) \;=\; \mathbb{R}\big(\hat{\mathsf{x}}\; e^{j\omega t}\big) \qquad \longleftrightarrow \qquad \hat{\mathsf{x}} \;=\; \mathsf{x}\; e^{j\theta} \;=\; |\hat{\mathsf{x}}|\; e^{j\angle\hat{\mathsf{x}}}. \tag{2.12}$$

A word of caution is needed here. When we say for example that a displacement or a voltage signal $x(t) = \mathsf{x}\;\cos(\omega t + \theta)$ is represented by the complex number $\hat{\mathsf{x}} = \mathsf{x}e^{j\theta}$, *it does not mean that the voltage $x(t)$ is complex.* The voltage $x(t)$ is always a real number! We are simply using the phasor representation because we can add sinusoids by adding their phasor representation using complex arithmetic, and therefore avoid the use of messy trigonometric identities as shown next.

Remark 2.2. [on notation] We will adhere to the following standardized notation. A sinusoidal signal (a real-valued function of time) is always denoted by italic small font, e.g. $x(t) = \mathsf{x}\;\cos(\omega t + \theta)$. Its amplitude (which is a positive real number) is always denoted by a sans-serif font x. The phasor representing $x(t)$ is a complex number which is denoted by a sans-serif font with a "hat", e.g. $\hat{\mathsf{x}} = \mathsf{x}e^{j\theta}$. Therefore the font choices for another signal denoted by $u(t)$ would be

$$u(t) \;=\; \mathsf{u}\;\cos\big(\omega t + \phi\big) \qquad \longleftrightarrow \qquad \hat{\mathsf{u}} \;=\; \mathsf{u}\; e^{j\phi}.$$

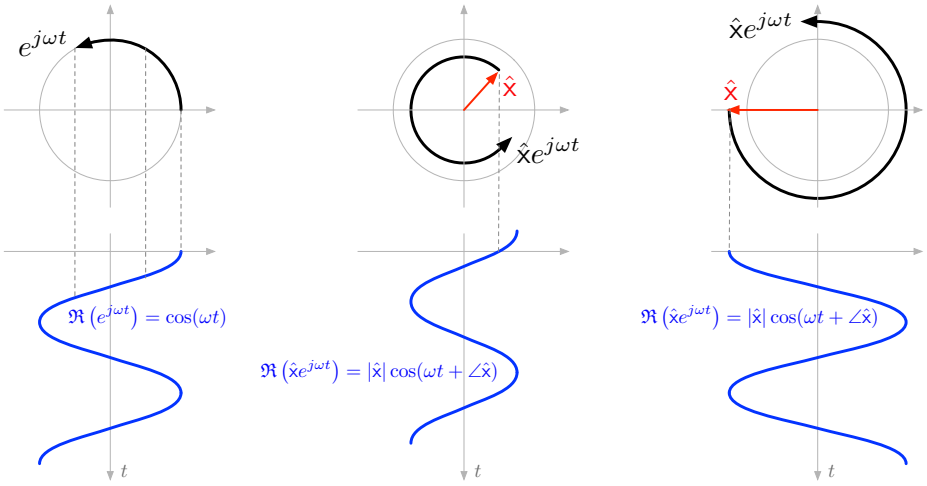


Figure 2.8: (*Left*) The *complex-valued function of time* $e^{j\omega t}$ is a circular curve in the complex plane that goes around the unit circle counterclockwise starting from 1. Its real part is the function $\mathbb{R}\left(e^{j\omega t}\right) = \cos(\omega t)$. (*Middle and Right*) When multiplied by a complex number $\hat{\mathrm{x}}$, the function $\hat{\mathrm{x}}e^{j\omega t}$ also traces a circular curve in the complex plane, but starting at $\hat{\mathrm{x}}$ rather than 1, with a circle of radius $|\hat{\mathrm{x}}|$. The real part of this function is $\mathbb{R}\left(\hat{\mathrm{x}}e^{j\omega t}\right) = |\hat{\mathrm{x}}|\cos\left(\omega t + \angle\hat{\mathrm{x}}\right)$, where $|\hat{\mathrm{x}}|$ and $\angle\hat{\mathrm{x}}$ are the magnitude and phase of the complex number $\hat{\mathrm{x}}$ respectively. Thus for a fixed frequency ω, the time function $\mathrm{x}\cos\left(\omega t + \theta\right)$ is completely determined by the complex number $\hat{\mathrm{x}} = \mathrm{x}e^{j\theta}$, and thus we call $\hat{\mathrm{x}}$ its *phasor representation*.

Now consider two sinusoids of the same frequency, but different amplitudes and phases, and use the complex represenation (2.12) for each of them

$$\begin{aligned}
x_3(t) \;=\; x_1(t) \;+\; x_2(t) &= \mathrm{x}_1\cos\left(\omega t + \theta_1\right) \;+\; \mathrm{x}_2\cos\left(\omega t + \theta_2\right) \\
&= \mathbb{R}\Big(\mathrm{x}_1\; e^{j(\omega t+\theta_1)}\Big) \;+\; \mathbb{R}\Big(\mathrm{x}_2\; e^{j(\omega t+\theta_2)}\Big)\,, \\
&= \mathbb{R}\big(\hat{\mathrm{x}}_1\; e^{j\omega t}\big) \;+\; \mathbb{R}\big(\hat{\mathrm{x}}_2\; e^{j\omega t}\big)\,, \qquad (\text{where } \hat{\mathrm{x}}_1 = \mathrm{x}_1 e^{j\theta_1},\ \hat{\mathrm{x}}_2 = \mathrm{x}_2 e^{j\theta_2}) \\
&= \mathbb{R}\big(\hat{\mathrm{x}}_1\; e^{j\omega t} \;+\; \hat{\mathrm{x}}_2\; e^{j\omega t}\big) \qquad (\text{since the real parts add up}) \\
&= \mathbb{R}\big((\hat{\mathrm{x}}_1 + \hat{\mathrm{x}}_2)\; e^{j\omega t}\big) \;=\; \mathbb{R}\big(\hat{\mathrm{x}}_3\; e^{j\omega t}\big) \qquad (\text{where } \hat{\mathrm{x}}_3 = \hat{\mathrm{x}}_1 + \hat{\mathrm{x}}_2)
\end{aligned}$$

Thus the phasor $\hat{\mathrm{x}}_3$ representing $x_3(t) = x_1(t) + x_2(t)$ is simply the sum (as a complex number) of the phasors $\hat{\mathrm{x}}_1$ and $\hat{\mathrm{x}}_2$ representing $x_1(t)$ and $x_2(t)$ respectively. This is the concept that was illustrated in the introduction in Figure 2.1, and we now state it as a theorem.

Theorem 2.3. *Given a frequency ω, the addition of two sinusoids of this frequency can be expressed using their phasor representations as follows*

$$\begin{aligned}
x_1(t) &= \mathrm{x}_1\cos\left(\omega t + \theta_1\right) \quad &\longleftrightarrow \quad \hat{\mathrm{x}}_1 &= \mathrm{x}_1 e^{j\theta_1}, \\
x_2(t) &= \mathrm{x}_2\cos\left(\omega t + \theta_2\right) \quad &\longleftrightarrow \quad \hat{\mathrm{x}}_2 &= \mathrm{x}_2 e^{j\theta_2}, \\
x_3(t) \;=\; x_1(t) + x_2(t) \;&= \mathrm{x}_3\cos\left(\omega t + \theta_3\right) \quad &\longleftrightarrow \quad \hat{\mathrm{x}}_1 + \hat{\mathrm{x}}_2 = \hat{\mathrm{x}}_3 &= \mathrm{x}_3 e^{j\theta_3},
\end{aligned}$$

i.e. the amplitude x_3 and phase θ_3 are obtained from the complex number addition

$$\hat{\mathrm{x}}_3 \;=\; \hat{\mathrm{x}}_1 \;+\; \hat{\mathrm{x}}_2 \qquad \Leftrightarrow \qquad \mathrm{x}_3\; e^{j\theta_3} \;=\; \mathrm{x}_1\; e^{j\theta_1} \;+\; \mathrm{x}_2\; e^{j\theta_2}$$

Note that to arrive at x_3 and θ_3 given $\mathrm{x}_1, \mathrm{x}_2, \theta_1, \theta_2$, one typically has to convert $\mathrm{x}_1 e^{j\theta_1}$ and $\mathrm{x}_2 e^{j\theta_2}$ to cartesian form since complex addition is most easily done in that form, and then convert the result back to polar form to obtain x_3 and θ_3. Note that usually, these calculations are done by calculator or software, which have complex arithmetic built in, but usually don't have trigonometric identities built in.

The phasor representation is not only useful for adding sinusoidal functions, but also for differentiating and integrating them. This makes it particularly suited to treating certain types of

differential equations that arise in harmonically-forced vibrations. This is the topic of Chapter 4, but we give next a brief glimpse of this technique.

2.3.1 Phasor Representation of Sinusoidal Derivatives

If $y(t)$ is a sinusoids with a given phasor representation, what is the phasor representation of its derivative $\dot{y}(t)$? There is a simple answer to this question which will greatly simplify the analysis of constant-coefficient ODEs with sinusoidal forcing. One way to calculate is as follows

$$\begin{aligned}
& y(t) = \mathsf{y}\,\cos(\omega t+\theta) \qquad \longleftrightarrow \qquad \hat{\mathsf{y}} = \mathsf{y}\,e^{j\theta} \\
\Rightarrow \quad & \dot{y}(t) = -\omega\mathsf{y}\,\sin(\omega t+\theta) \\
& \qquad = -\omega\mathsf{y}\,\cos(\omega t+\theta-90^\circ) && \text{(since } \sin(\phi)=\cos(\phi-90^\circ) \text{ for any } \phi\text{)} \\
& \qquad = \omega\mathsf{y}\,\cos(\omega t+\theta-90^\circ+180^\circ) && \text{(since } \cos(\phi)=-\cos(\phi+180^\circ) \text{ for any } \phi\text{)} \\
& \qquad = \omega\mathsf{y}\,\cos(\omega t+\theta+90^\circ).
\end{aligned}$$

Thus the amplitude of $\dot{y}$ is the amplitude of y multiplied by ω, and the phase of $\dot{y}$ is a 90° *advance* over the phase of y. We can summarize this by the following relations

$$\begin{aligned}
& y(t) \qquad \longleftrightarrow \qquad \mathsf{y}\,e^{j\theta} \\
\Rightarrow \quad & \dot{y}(t) \qquad \longleftrightarrow \qquad \omega\mathsf{y}\,e^{j(\theta+90^\circ)}.
\end{aligned} \tag{2.13}$$

Another way to arrive at the derivative relations is first to obtain the derivative of the complex-valued exponential function $e^{j\omega t}$. The time derivative of any complex-valued function is defined using the derivatives of the real and imaginary parts as follows. First, let u be a complex-valued function of the time variable t

$$u(t) := u_{\mathrm{r}}(t) + j\,u_{\mathrm{i}}(t),$$

where $u_{\mathrm{r}}(t)$ and $u_{\mathrm{i}}(t)$ are the (real-valued) functions representing the real and imaginary parts of $u(t)$ respectively. We define the time derivative of such a function as

$$\frac{d}{dt}\,u(t) := \left(\frac{d}{dt}u_{\mathrm{r}}(t)\right) + j\,\left(\frac{d}{dt}u_{\mathrm{i}}(t)\right),$$

i.e. by differentiating the real and imaginary parts individually. Now we apply this definition to find the time derivative of $e^{j\omega t}$ by using Euler's formula

$$\begin{aligned}
\frac{d}{dt}\,e^{j\omega t} = \frac{d}{dt}\,\big(\cos(\omega t)+j\sin(\omega t)\big) &= -\omega\sin(\omega t) + j\,\omega\cos(\omega t) \\
&= j\omega\,\big(\cos(\omega t) + j\sin(\omega t)\big) = j\omega\,e^{j\omega t} && (2.14)\\
&= \omega\,e^{j(\omega t+\pi/2)}, && (2.15)
\end{aligned}$$

where we used the fact that $j = e^{j\pi/2}$ in the last equality. We can use this formula to take a time derivative of the more general function $\mathsf{u}\,e^{j(\omega t+\theta)}$ by

$$\begin{aligned}
\frac{d}{dt}\,\left(\mathsf{u}\,e^{j(\omega t+\theta)}\right) = \mathsf{u}\frac{d}{dt}\,\left(e^{j\omega t}\,e^{j\theta}\right) = \mathsf{u}\,e^{j\theta}\,\frac{d}{dt}\,\left(e^{j\omega t}\right) &= \mathsf{u}\,e^{j\theta}\,j\omega\,e^{j\omega t} = \mathsf{u}\,e^{j\theta}\,e^{j\pi/2}\,\omega\,e^{j\omega t} \\
\Rightarrow \qquad \frac{d}{dt}\,\left(\mathsf{u}\,e^{j(\omega t+\theta)}\right) &= \omega\mathsf{u}\,e^{j(\omega t+\theta+\pi/2)}
\end{aligned} \tag{2.16}$$

Thus differentiation amplifies the amplitude by a factor of ω, and increases the phase by 90°. Taking the real parts of both sides of this last formula gives the same answer as was obtained earlier in (2.13).

Examine the formula (2.14), and observe that it is the same as the classic formula $\frac{d}{dt}e^{\alpha t} = \alpha e^{\alpha t}$ when α is real. The formula is exactly the same when α is complex (in this case $\alpha = j\omega$). We demonstrate this for the function e^{zt} where $z = (\alpha + j\beta) \in \mathbb{C}$ is any complex number

$$\begin{aligned}
\frac{d}{dt}\, e^{zt} \;&=\; \frac{d}{dt}\, e^{(\alpha+j\beta)t} \;=\; \frac{d}{dt}\, \big(e^{\alpha t}e^{j\beta t}\big) \;=\; \frac{d}{dt}\, \Big(e^{\alpha t}\big(\cos(\beta t) + j\sin(\beta t)\big)\Big) \\
&\overset{1}{=}\; \alpha e^{\alpha t}\cos(\beta t) - \beta e^{\alpha t}\sin(\beta t) \;+\; j\,\big(\alpha e^{\alpha t}\sin(\beta t) + \beta e^{\alpha t}\cos(\beta t)\big) \\
&\overset{2}{=}\; e^{\alpha t}\big(\alpha\cos(\beta t) - \beta\sin(\beta t) \;+\; j\,\big(\alpha\sin(\beta t) + \beta\cos(\beta t)\big)\big) \\
&\overset{3}{=}\; e^{\alpha t}\,(\alpha + j\beta)\,\big(\cos(\beta t) + j\sin(\beta t)\big) \\
&\overset{4}{=}\; e^{\alpha t}\,(\alpha + j\beta)\,e^{j\beta t} \;=\; z\; e^{zt},
\end{aligned}$$

where $\overset{1}{=}$ follows from the product rule of differentiation of real functions, $\overset{2}{=}$ is a rearrangement, $\overset{3}{=}$ follows from the product formula (2.9) for two complex numbers, and $\overset{4}{=}$ is another use of Euler's formula.

In Chapter 4 we will see how these concepts can be used to greatly simplify the analysis of arbitrary constant-coefficient differential equations with sinusoidal forcing terms. This will provide us with a powerful technique to convert *differential equations to algebraic equations*. Those algebraic equations will give the relations between the respective amplitudes and phases of forcing and response terms, and define the so-called *frequency response* of any system. Concepts such as resonance, and vibration isolation and suppression can be completely characterized in terms of the frequency response.

Part II

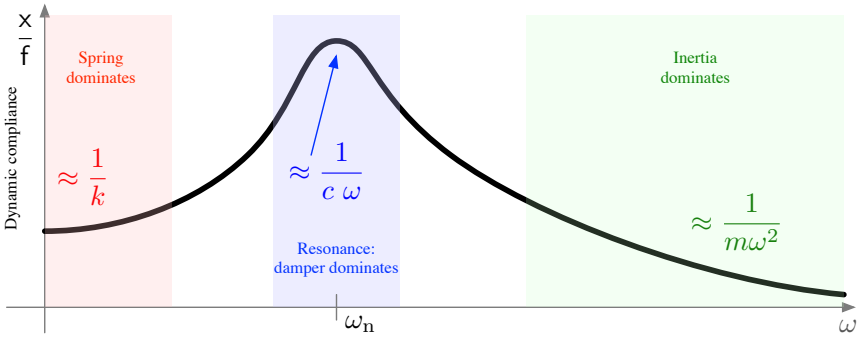
Single Degree of Freedom Systems

Single Degree of Freedom Systems

This part concerns analysis of systems with a single mechanical degree of freedom. The main mathematical tools used are solutions of scalar second-order differential equations. Unforced equations model so-called "free vibrations", which arise when systems are in an initial disequilibrium, and exhibit vibrations as they return to equilibrium (rest). Such vibrations are characterized by the system's "natural frequency", which is determined by its mechanical parameters.

Harmonically "forced vibrations" arise when systems are subjected to continuous oscillatory forcing. This forcing can be actual mechanical forces due to wind, ground vibrations, actuation, etc. Forcing can also occur due to one part of a mechanical system acting on another part through displacements, velocities, etc. The key tool in analyzing forced vibrations is the "frequency response" derived from phasor representations of oscillatory signals, or from "transfer function" descriptions of the dynamics. These give a full understanding of the important phenomenon of resonance, which occurs due to matching of the forcing frequency and the system's natural frequency.

Single degree-of-freedom systems are highly idealized, and most mechanical system are more complex with many degrees of freedom. None the less, the analysis of these simple systems are extremely useful for many engineering design problems. Those include the design of vibration isolation/attenuation systems, vibrometers and accelerometers.



Chapter 3

Free Vibrations

The term "free vibrations" refers to the setting when no external oscillatory forces are acting on a system. In this case, the vibrations are entirely due to initial extensions of springs or initial velocities of masses, i.e. due to non-equilibrium initial conditions. The motion in these cases is obtained from solutions of homogenous constant-coefficient differential equations. For single degree of freedom systems, the solutions are characterized by the locations of two possibly complex roots of a characteristic equation. Underdamped oscillations occur when roots are complex-conjugate pairs, and overdamped motion occurs when the the roots are real. A complete characterization of the oscillation frequency, its amplitude and phase, as well as oscillation decay is possible in terms of system parameters. A particularly important non-dimensional parameter, the "damping ratio", characterizes similar oscillations in a wide variety of mechanical systems. After a quantitative analysis of free vibrations, it is possible to infer some system parameters such as damping/mass and spring-stiffness/mass ratios from time-resolved experimental measurements of positions or velocities.

Background from Constant-Coefficient Differential Equations

We begin by recalling the main technique for solving constant-coefficient, homogenous (i.e. unforced) Ordinary Differential Equations (ODEs) of the form

$$\alpha_n\, x^{(n)}(t) \;+\; \alpha_{n\text{-}1}\, x^{(n\text{-}1)} \;+\cdots+\; \alpha_1\, x^{(1)}(t) \;+\; \alpha_0\, x(t) \;=\; 0. \qquad t \geq 0. \tag{3.1}$$

Let's parse this notation carefully. A term like $x^{(k)}(t)$ denotes the k'th derivative of the function $x(t)$. The $n+1$ numbers $\alpha_0, \ldots, \alpha_n$ are the *coefficients* of this equation. Note the indexing of the coefficients in comparison to the order of derivative in each term. Equation (3.1) is called an *n'th order* differential equation since the highest derivative that occurs in this equation is $x^{(n)}(t)$. This equation has $n+1$ terms unless some of the coefficients $\{\alpha_k\}_{k=0}^{n}$ are zero, in which case it will have less. For example, consider our Mass-Spring equation

$$m\,\ddot{x}(t) \;+\; k\,x(t) \;=\; 0. \tag{3.2}$$

This equation is of the form (3.1). It is 2nd order since the highest derivative that appears in it is $\ddot{x}(t)$. Note that for this equation

$$\begin{aligned} m\ddot{x}(t) + kx(t) = 0 \qquad &\Leftrightarrow \qquad \alpha_2 x^{(2)}(t) + \alpha_1 x^{(1)}(t) + \alpha_0 x(t) = 0 \\ &\Rightarrow \qquad \alpha_0 = k,\ \alpha_1 = 0,\ \alpha_2 = m, \end{aligned}$$

where $\alpha_1 = 0$ since there is no first order derivative term.

Now the steps for solving an equation of the form (3.1) are as follows

1. Form the *characteristic polynomial* of this equation from its coefficients $\alpha_0, \ldots, \alpha_n$

$$P(s) \;:=\; \alpha_n s^2 + \alpha_{n\text{-}1} s^{n\text{-}1} + \cdots + \alpha_1 s + s. \tag{3.3}$$

   This is just a polynomial in the variable s.

2. Find the n *roots* of this polynomial, and call them $\bar{s}_1, \ldots, \bar{s}_n$. Note that some of them may be complex, and since the coefficients $\{\alpha_k\}$ are real, any complex roots must come in complex conjugate pairs.

3. Any solution of (3.1) must be a linear combination of the (possibly complex) exponential functions $\{e^{\bar{s}_k t}\}$, i.e.

$$x(t) \;=\; \mathsf{a}_1\, e^{\bar{s}_1 t} \;+\cdots+\; \mathsf{a}_n\, e^{\bar{s}_n t}, \tag{3.4}$$

   where $\bar{s}_1, \ldots, \bar{s}_n$ are the roots found earlier, and $\mathsf{a}_1, \ldots, \mathsf{a}_n$ are constants that depend on the particular initial conditions. If there are any complex roots, they must be complex conjugates, e.g. $\bar{s}_i = \bar{s}_j^*$. The corresponding constants a_i and a_j are generally complex, and must also be complex conjugates, i.e. $\bar{s}_i = \bar{s}_j^* \quad \Rightarrow \quad \mathsf{a}_i = \mathsf{a}_j^*$.

4. Note that since we have n unknown constants, we need n initial conditions (on $x(0)$ and its derivatives up to the $n-1$ derivative) to uniquely solve for the unknown coefficients. The constants $\mathsf{a}_1, \ldots, \mathsf{a}_n$ can be obtained from the initial conditions by solving a system of n linear equations.

In this chapter we apply the above procedure for single-degree-of-freedom systems with and without damping. Single degree of freedom means that the differential equation is of 2nd order, and therefore there are only two roots of the characteristic polynomial. The analysis largely depends on whether the roots are complex conjugates or both real. The analysis also consists of two main steps outlined below.

1. Complex conjugate roots of the characteristic polynomial correspond to *undamped* or *underdamped* vibrations. The locations in the complex plane of those roots determine the vibration frequency as well as the decay rate of vibrations. Alternatively, when both roots are real, the observed behavior is a decaying response with no oscillations. This analysis applies to any set of initial conditions, i.e. the conclusions here are independent of the initial conditions.

2. To determine exactly the amplitude and phase of vibrations, one needs to solve the differential equations for given initial conditions. We will derive explicit formulas for this. The derivations can be a little involved, but the final answers are useful for solving certain problems.

It is worth emphasizing the step 1 above is more important than step 2, especially for design problems, where the frequency and decay rate of vibrations is to be "tuned". The exact solutions in step 2 can always be obtained from numerical simulations, although the formulas we will derive can be useful for some simple design problems as well.

3.1 Undamped Free Vibrations

The free vibrations of a Mass-Spring system are mathematically described by the solutions of Equation (3.2) from given initial conditions. Since this is a 2nd order differential equation, finding the solution requires two initial conditions $x(0)$ and $\dot{x}(0)$, i.e. the initial position and velocity of the mass. We now determine the general form of a solution in terms of the parameters m and k, as well as the initial conditions using the general procedure for solving ODEs outlined in the introduction.

Although there are two coefficients m and k in the the differential equation (3.2), it is only the ratio of those two coefficients that matters to the solution since we can divide through by m

$$m\ddot{x}(t) + kx(t) = 0 \qquad \Leftrightarrow \qquad \ddot{x}(t) \;+ \tfrac{k}{m} x(t) = 0 \tag{3.5}$$

to obtain an equivalent differential equation depending on only k/m. The characteristic polynomial of this equation is easy to factor as follows

$$\bar{s}^2 + \tfrac{k}{m} = 0 \qquad \Leftrightarrow \qquad \bar{s}^2 = -\tfrac{k}{m} \qquad \Leftrightarrow \qquad \bar{s}_{1,2} = \pm j\sqrt{\tfrac{k}{m}}. \tag{3.6}$$

The coefficients m and k represent the two physical properties of mass and stiffness, both of which by definition are always positive. The ratio $-k/m$ is therefore a negative number, and thus we have two *purely imaginary* roots of the characteristic polynomial which are complex conjugates.

According to the general form (3.4), solutions of Equation (3.5) are of the form

$$x(t) \;=\; \mathsf{a}\, e^{j\omega_n t} \;+\; \mathsf{a}^*\, e^{-j\omega_n t}, \qquad\qquad \omega_n := \sqrt{\tfrac{k}{m}}. \tag{3.7}$$

They are purely oscillatory solutions with frequency $\omega_n := \sqrt{k/m}$ which is called the *natural frequency* of the system. The constants a and a^* are complex conjugate numbers (recall that the notation a^* stands for the complex conjugate of a), which are determined by the initial conditions $x(0)$ and $\dot{x}(0)$.

The solution (3.7) is written as the sum of two complex-conjugate functions of t, and is therefore a real function of t, which can be rewritten in two equivalent ways. The first follows from recalling the fact that for any complex number z, $z + z^* = 2\mathbb{R}(z)$, and therefore

$$x(t) \;=\; \mathsf{a}\, e^{j\omega_n t} + \mathsf{a}^* e^{-j\omega_n t} \;=\; \mathsf{a}e^{j\omega_n t} + \big(\mathsf{a}e^{j\omega_n t}\big)^* \;=\; 2\,\mathbb{R}\big(\mathsf{a}\, e^{j\omega_n t}\big) \;=\; 2\,|\mathsf{a}|\,\cos\big(\omega_n t + \angle\mathsf{a}\big), \tag{3.8}$$

where $|\mathsf{a}|$ and $\angle\mathsf{a}$ are the magnitude and phase of the complex number a. Now we see perhaps more clearly that the solution is purely oscillatory with a frequency of $\omega_n = \sqrt{k/m}$ rad/s.

A slightly different form of (3.8) can be written down to conform with phasor representations of sinusoids by absorbing the factor of 2 into the constants as follows

$$x(t) \;=\; 2\,\mathbb{R}\big(\mathsf{a}\, e^{j\omega_n t}\big) \;=\; 2\,|\mathsf{a}|\,\cos\big(\omega_n t + \angle\mathsf{a}\big) \;=\; |\hat{\mathsf{x}}|\,\cos\big(\omega_n t + \angle\hat{\mathsf{x}}\big) \;=\; \mathbb{R}\big(\hat{\mathsf{x}}\, e^{j\omega_n t}\big) \tag{3.9}$$

where we simply redefined the complex number $\hat{\mathsf{x}} = 2\mathsf{a}$. Note that a and $\hat{\mathsf{x}}$ have the same phase.

Another way of writing the solution (3.9) in terms of only real functions is obtained by expressing the complex number $\hat{\mathsf{x}} = \alpha + j\beta$ in terms of its real and imaginary parts α and β respectively, and using Euler's formula as follows

$$\begin{aligned} x(t) \;=\; \mathbb{R}\big(\hat{\mathsf{x}}\, e^{j\omega_n t}\big) \;&=\; \mathbb{R}\big((\alpha + j\beta)\,\big(\cos(\omega_n t) + j\sin(\omega_n t)\big)\big) \\ &=\; \alpha\,\cos(\omega_n t) \;-\; \beta\,\sin(\omega_n t). \end{aligned} \tag{3.10}$$

This solution is a sum of two sinusoids, both of the same frequency ω_n. Recall that such a sum is always another sinusoid of the same frequency, but different magnitude and phase, which is precisely the form (3.9).

We have so far obtained the solution in three different forms which we now summarize

$$x(t) = \begin{cases} \frac{1}{2}\left(\hat{\mathsf{x}}\, e^{j\omega_n t} \;+\; \hat{\mathsf{x}}^*\, e^{-j\omega_n t}\right) \\ |\hat{\mathsf{x}}|\,\cos\left(\omega_n t + \angle\hat{\mathsf{x}}\right) \\ \alpha\,\cos(\omega_n t) \;-\; \beta\,\sin(\omega_n t) \end{cases}, \quad \text{where} \quad \begin{cases} \omega_n = \sqrt{k/m}\;\left(\begin{array}{l}\text{determined by}\\ \text{system parameters}\end{array}\right) \\ \hat{\mathsf{x}} = \alpha + j\beta\;\left(\begin{array}{l}\text{determined by}\\ \text{initial conditions}\end{array}\right) \end{cases} \tag{3.11}$$

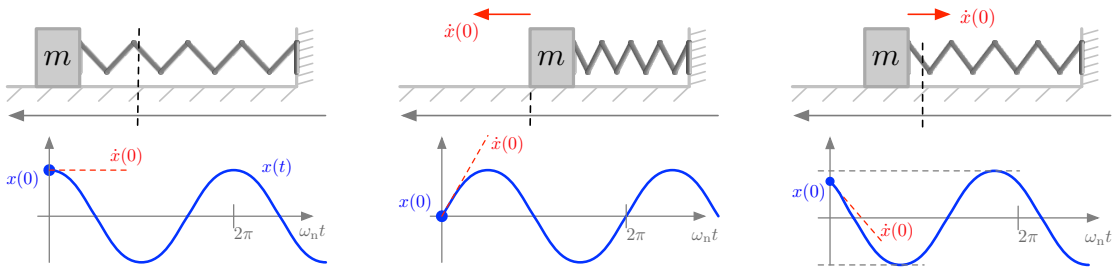
Which of those three equivalent forms we end up using is a matter of convenience, and is usually determined by what further analysis is to be done.

The constants $\hat{\mathsf{x}} = \alpha + j\beta$ in (3.11) are determined from the initial conditions $x(0)$ and $\dot{x}(0)$ by solving a system of 2 linear equations. We now derive that relation in order to rewrite the solutions directly in terms of initial conditions. First, write the initial conditions in terms of the complex constant $\hat{\mathsf{x}}$

$$\begin{aligned} x(0) &= \tfrac{1}{2}\left(\hat{\mathsf{x}}\, e^{j\omega_n t} \;+\; \hat{\mathsf{x}}^* e^{-j\omega_n t}\right)\Big|_{t=0} = \left(\hat{\mathsf{x}} + \hat{\mathsf{x}}^*\right)/2 \\ \dot{x}(0) &= \tfrac{1}{2}\,\frac{d}{dt}\left(\hat{\mathsf{x}} e^{j\omega_n t} + \hat{\mathsf{x}}^* e^{-j\omega_n t}\right)\Big|_{t=0} = \tfrac{1}{2}\left(\hat{\mathsf{x}}\, j\omega_n e^{j\omega_n t} - \hat{\mathsf{x}}^*\, j\omega_n e^{-j\omega_n t}\right)\Big|_{t=0} = \left(\hat{\mathsf{x}}\, j\omega_n - \hat{\mathsf{x}}^*\, j\omega_n\right)/2 \end{aligned}$$

This system of equations can be written in matrix-vector form and solved as follows

$$\begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix} = \tfrac{1}{2}\begin{bmatrix} 1 & 1 \\ j\omega_n & -j\omega_n \end{bmatrix}\begin{bmatrix} \hat{\mathsf{x}} \\ \hat{\mathsf{x}}^* \end{bmatrix} \quad\Rightarrow\quad \begin{bmatrix} \hat{\mathsf{x}} \\ \hat{\mathsf{x}}^* \end{bmatrix} = \frac{-1}{j\omega_n}\begin{bmatrix} -j\omega_n & -1 \\ -j\omega_n & 1 \end{bmatrix}\begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix} \quad\Rightarrow\quad \begin{aligned} \hat{\mathsf{x}} &= x(0) \;+\; \tfrac{1}{j\omega_n}\,\dot{x}(0) \\ &= x(0) \;-\; j\tfrac{1}{\omega_n}\,\dot{x}(0) \end{aligned}.$$



(a) A mass starting from an initial positive extension $x(0)$, and at rest ($\dot{x}(0) = 0$). The motion is $x(t) = x(0)\cos(\omega_n t)$, i.e. the phase is zero, and the amplitude of oscillation is exactly the magnitude of the initial extension.

(b) A mass starting with no initial extension ($x(0) = 0$), but with positive initial velocity $\dot{x}(0)$. The motion is $x(t) = A\sin(\omega_n t) = A\cos(\omega_n t - 90^\circ)$. The amplitude of oscillation is $A = \dot{x}(0)/\omega_n = \dot{x}(0)\sqrt{m/k}$. Thus smaller inertia or stiffer spring imply a smaller oscillation amplitude for a given initial velocity.

(c) A mass with non-zero initial extension and velocity. The oscillation $x(t) = A\cos(\omega_n t + \theta)$ has amplitude A and phase θ that depend on values of both $x(0)$ and $\dot{x}(0)$ according to (3.14). Note that since the initial spring force and velocity are in the same direction, the oscillation amplitude is larger than the initial extension.

Figure 3.1: Illustration of the free vibrations of a Mass-Spring system starting from three different initial conditions. The mass' position is given by Theorem 3.1 as $x(t) = A\cos(\omega_n t + \theta)$, where the amplitude A and phase θ depend on initial conditions, but solutions always oscillate with the *natural frequency* $\omega_n = \sqrt{k/m}$ rad/s, which is independent of the initial conditions. Note the scaling of the time axes. The equilibrium position of the mass is indicated by the vertical dashed line.

We have therefore obtained the real and imaginary parts of $\hat{\mathrm{x}}$ in terms of the initial conditions. Since α and β in (3.11) are those parts respectively

$$\hat{\mathrm{x}} \;=\; \alpha + j\beta \;=\; x(0) \;-\; j\tfrac{1}{\omega_n}\,\dot{x}(0) \qquad\Rightarrow\qquad \alpha = x(0), \qquad \beta = -\tfrac{\dot{x}(0)}{\omega_n},$$

we have therefore obtained explicit expressions for the solution in all its forms. We summarize this in the next statement.

Theorem 3.1. *Consider a differential equation of the form*

$$\ddot{x}(t) \;=\; -\,\omega_n^2\; x(t), \tag{3.12}$$

with initial conditions $x(0)$ and $\dot{x}(0)$. The solution is given in either of the following two equivalent forms

$$x(t) \;=\; x(0)\;\cos(\omega_n t) + \tfrac{\dot{x}(0)}{\omega_n}\;\sin(\omega_n t), \tag{3.13}$$

$$\;=\; \mathrm{x}\;\cos\big(\omega_n t + \theta\big), \qquad \mathrm{x} = \sqrt{x^2(0) + \tfrac{\dot{x}^2(0)}{\omega_n^2}}, \qquad \theta = -\tan^{-1}\tfrac{\dot{x}(0)}{\omega_n\; x(0)}. \tag{3.14}$$

Remark 3.2. We can apply this theorem to problems other than the Mass-Spring system. Recall the mathematical analogies from Chapter 1 where differential equations of three different systems were compared

$$\ddot{x}(t) \;=\; -\,(k/m)\;\; x(t), \qquad \text{(mass/spring in linear motion)}$$

$$\ddot{\theta}(t) \;=\; -\,(k/\mathrm{J})\;\; \theta(t), \qquad \text{(mass/torsional-spring in rotational motion)}$$

$$\ddot{\theta}(t) \;=\; -\,(g/l)\;\; \theta(t). \qquad \text{(pendulum in rotational motion)}$$

All three equations are of the form (3.12) where in each case ω_n is equal to $\sqrt{k/m}$, $\sqrt{k/\mathrm{J}}$ and $\sqrt{g/l}$ respectively. More generally, the theorem applies to any second order equation that can be rearranged in the form $\ddot{x}(t) = -\alpha\, x(t)$, where $\alpha > 0$ maybe some combination of coefficients. In this case ω_n in (3.12) would just be $\sqrt{\alpha}$.

Figure 3.1 illustrates the vibrational motion of a Mass-Spring system starting from three different initial conditions. Each case demonstrates some property of the solutions as follows.

- Figure 3.1a: When a mass is released from rest (i.e. $\dot{x}(0) = 0$) from an initial extension of $x(0)$, the solution is a pure cosine with zero phase

$$x(t) \;=\; x(0)\;\cos\big(\omega_{\rm n} t\big).$$

  In this case, the amplitude of oscillation is $|x(0)|$, which is exactly the magnitude of the initial extension or compression.

- Figure 3.1b: When a mass has positive initial velocity $\dot{x}(0) > 0$ starting at the equilibrium location of the spring (i.e. $x(0) = 0$). The motion is given by

$$x(t) \;=\; \tfrac{\dot{x}(0)}{\omega_{\rm n}}\;\sin\big(\omega_{\rm n} t\big) \;=\; \dot{x}(0)\sqrt{\tfrac{m}{k}}\;\cos\big(\omega_{\rm n} t - 90^\circ\big).$$

  Note that the mass starts from zero extension, and then oscillates with an amplitude that depends not only on the initial velocity, but also on m and k. For a given initial velocity, the larger the mass' inertia m, the larger the amplitude, and the higher the stiffness k, the smaller is the resulting amplitude. This makes physical sense since an initial velocity with a large m means a large initial kinetic energy, and we expect the oscillations to be bigger. This will be made precise with the energy analysis below.

- When a mass has both non-zero initial extension and velocity, the amplitude and phase are given by (3.14). The amplitude in particular will always be larger than the initial extension. This follows from the formula for x in (3.14) which implies that x is always larger than $x(0)$ if $\dot{x}(0) \neq 0$. It can be clearly seen in Figure 3.1c.

Velocity Amplitude and Phase

Consider the general solution (3.14) for the mass' position, from which we get the velocity by simply differentiating that formula

$$x(t) = \mathsf{x}\;\cos\big(\omega_{\rm n} t + \theta\big), \qquad \mathsf{x} = \sqrt{x^2(0) + \tfrac{\dot{x}^2(0)}{\omega_{\rm n}^2}}, \qquad \theta = -\tan^{-1}\tfrac{\dot{x}(0)}{\omega_{\rm n}\;x(0)},$$

$$\begin{aligned}
\dot{x}(t) &= -\mathsf{x}\;\omega_{\rm n}\;\sin\big(\omega_{\rm n} t + \theta\big) \\
&= \;\mathsf{x}\;\omega_{\rm n}\;\sin\big(\omega_{\rm n} t + \theta + 180^\circ\big) && \text{(because } -\sin(x) = \sin\left(x \pm 180^\circ\right)\text{)} \\
&= \;\mathsf{x}\;\omega_{\rm n}\;\cos\big(\omega_{\rm n} t + \theta + 180^\circ - 90^\circ\big) \;=\; \mathsf{x}\;\omega_{\rm n}\;\cos\big(\omega_{\rm n} t + \theta + 90^\circ\big) && (\sin(x) = \cos\left(x - 90^\circ\right))
\end{aligned}$$

Two things are worth noting here. First, the the velocity amplitude is the position amplitude x multiplied by the frequency $\omega_{\rm n}$. This is physically intuitive, if the mass is oscillating back and forth very rapidly (with high frequency $\omega_{\rm n}$), then even for small position oscillations with amplitude x, the velocity amplitude $\mathsf{x}\;\omega_{\rm n}$ can be quite large, and vice versa if $\omega_{\rm n}$ is small (i.e. slow oscillations mean relatively low velocities). The ratio of displacement to velocity amplitudes is

$$\frac{\text{velocity amplitude}}{\text{displacement amplitdue}} \;=\; \frac{\omega_{\rm n}\mathsf{x}}{\mathsf{x}} \;=\; \omega_{\rm n}$$

The stiffer a system is (larger $\omega_{\rm n}$ is referred to as "stiffer"), the larger the velocity amplitudes are relative to displacement.

The other important observation is that position and velocity are always 90° out of phase. In particular, velocity has a 90° *phase lead* over position (recall that phase leads are equivalent to time advances). This is just an instance of the fact that the derivative of any sinusoid always has a 90° phase lead over it. This 90° phase difference is clearly seen in Figure 3.2. An intuitive interpretation to keep in mind is that since velocity has a phase lead over position, then velocity "anticipates" position, i.e. can predict where the mass will be 90° later, which is equivalent to $T/4$ in time units, where T is the period.

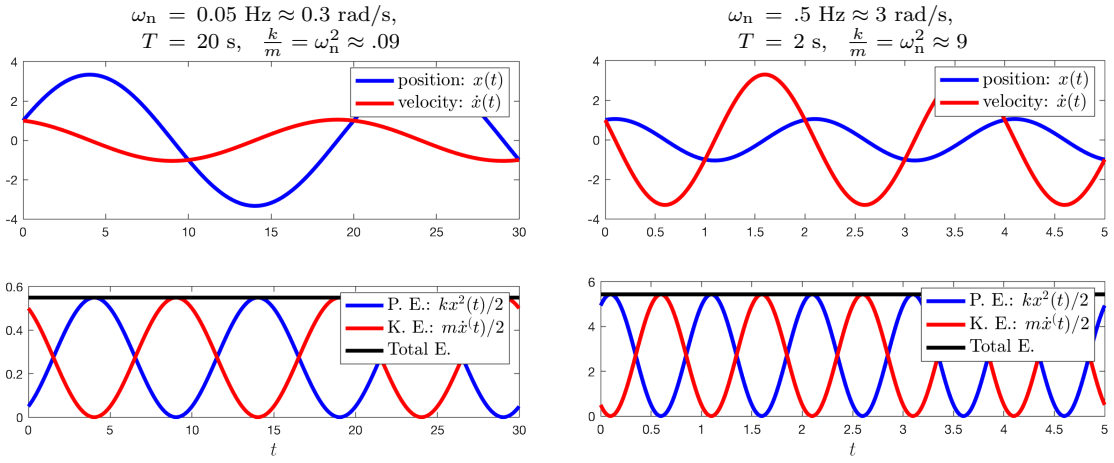


Figure 3.2: A comparison of the position and velocity oscillations for two systems with a low natural frequency ω_n (*left*), and a high ω_n (*right*). Note that the velocity signal always "leads" the position signal by a 90° phase shift. The bottom plots show the kinetic and potential energy oscillations over time. The sum of the two energies is always constant (in time) since this is a conservative system with no damping.

Energy

The total mechanical energy of the Mass-Spring system is the sum of the spring potential energy and the mass' kinetic energy

$$E(t) \;:=\; V(t) \;+\; T(t) \;:=\; \tfrac{1}{2}k\, x^2(t) \;+\; \tfrac{1}{2}m\, \dot{x}^2(t), \tag{3.15}$$

thus it depends on both position and velocity. Equation (3.14) shows that oscillation amplitudes depend on both initial position $x(0)$ and initial velocity $\dot{x}(0)$. This expression can be rewritten in terms of the initial total energy $E(0)$ which is the sum of the initial kinetic energy $T(0) := \frac{1}{2}m\, \dot{x}^2(0)$ and initial spring potential energy $V(0) := \frac{1}{2}k\, x^2(0)$. Starting from (3.14)

$$\begin{aligned} \mathsf{x}^2 \;&=\; x^2(0) + \tfrac{\dot{x}^2(0)}{\omega_n^2} \;=\; x^2(0) + \tfrac{\dot{x}^2(0)}{k/m} \;=\; \tfrac{2}{k}\left(\tfrac{1}{2}k\, x^2(0) + \tfrac{1}{2}m\, \dot{x}^2(0)\right) \\ &=\; \tfrac{2}{k}\Big(V(0) + T(0)\Big) \;=\; \tfrac{2}{k}\, E(0). \end{aligned} \tag{3.16}$$

Note that the factor $2/k$ does not depend on the initial conditions. Thus the amplitude x depends only on the total energy $E(0)$ of the initial conditions rather than the details of velocity versus position. This explains why the peak extension/compression in Figure 3.1c (which is equal in magnitude to the amplitude x) is larger than the initial extension. Since the initial velocity is non-zero, the total initial energy $E(0)$ is larger than the initial spring potential energy $V(0)$, the latter being determined by the initial spring extension.

The Spring-Mass system is subject to only the spring force, which is a conservative force. We therefore expect total energy to be conserved. We can verify that directly without actually solving the differential equation as follows. If $E(t)$ is conserved, then it must be constant in time, i.e. $E(t) = E(0)$. To show that, simply take the time derivative

$$\begin{aligned} \tfrac{d}{dt}E(t) \;&=\; \tfrac{d}{dt}\left(\tfrac{1}{2}k\, x^2(t) + \tfrac{1}{2}m\, \dot{x}^2(t)\right) && (3.17) \\ &=\; k\, x(t)\, \dot{x}(t) + m\, \dot{x}(t)\, \ddot{x}(t) && \text{(using the chain rule)} \\ &=\; k\, x(t)\, \dot{x}(t) - k\, \dot{x}(t)\, x(t) \;=\; 0. && \text{(substituting from the diff. eq. } m\ddot{x} = -kx\text{)} \end{aligned}$$

Therefore $E(t)$ is constant in time. The calculation we just went through is a very important one. In fact, it is such an important technique that it bears repeating that we *did not need to solve the differential equation* $m\ddot{x} + kx = 0$ in order to conclude that $E(t)$ is conserved. We just needed

the differential equation as well as the definition of energy. We will encounter such calculations again in Section 3.4 on *state space representations*, where we will see that they are useful even for non-conservative systems.

3.2 Damped, Free Vibrations

In this section we examine the free vibrations of a system with damping. As the name suggests, the effect of damping will make any free oscillations decay. We will identify two regimes called the *underdamped* and the *overdamped* cases respectively. In the underdamped regime, oscillations decays exponentially with decay being faster the higher the damping. In the overdamped regime, the responses also decay exponentially, but with no oscillations due to the higher value of damping. The boundary between the two regimes is referred to as the *critically damped* case. Those regimes will be identified using the value of a non-dimensional parameter ζ called the *damping ratio* .

Recall the differential equation for a single mass connected to a spring and damper. Although it has three parameters m, c and k, only two ratios matter since we can divide any two of the parameters by a third. The typical choice is to divide by the mass m

$$m\,\ddot{x}(t) \;+\; c\,\dot{x}(t) \;+\; k\,x(t) \;=\; 0 \qquad \Leftrightarrow \qquad \ddot{x}(t) \;+\frac{c}{m}\,\dot{x}(t) \;+\frac{k}{m}\,x(t) \;=\; 0. \tag{3.18}$$

The characteristic polynomial of this equation is a little more complicated than that of (3.5). Its roots can be found using the quadratic formula

$$\bar{s}^2 + \tfrac{c}{m}\bar{s} + \tfrac{k}{m} = 0 \qquad \Leftrightarrow \qquad \bar{s}_{1,2} = -\tfrac{c}{2m} \pm \sqrt{\tfrac{c^2}{4m^2} - \tfrac{k}{m}}. \tag{3.19}$$

Any solution is thus a sum of two, possibly complex, exponentials

$$x(t) \;=\; \mathsf{a}_1\, e^{\bar{s}_1 t} \;+\; \mathsf{a}_2\, e^{\bar{s}_2 t},$$

where the coefficients a_1 and a_2 are determined from initial conditions. Note however that the *exponents* in $e^{\bar{s}_1 t}$ and $e^{\bar{s}_2 t}$ are determined by the system parameters m, c, and k, and are thus independent of initial conditions. We will typically be interested characterizing the vibrations of a system from any initial conditions, and therefore the behavior of the functions $e^{\bar{s}_1 t}$ and $e^{\bar{s}_2 t}$ will deserve a detailed analysis.

Although it is easy to write down formula (3.19), it is a little difficult to see how the roots change as we change the three parameters m, c, and k. It turns out that there is another set of equivalent parameters that more easily characterize the behavior of the roots. First consider some special cases.

1. When there is no damping $c = 0$, we have two purely imaginary roots

   $$c \;=\; 0 \qquad \Rightarrow \qquad \bar{s}_{1,2} \;=\; \pm\, j\sqrt{k/m} \;=\; \pm\, j\omega_{\mathrm{n}}$$

   as we saw earlier in (3.6).

2. When c is small, the discriminant in (3.19) is negative, which gives *two complex-conjugate roots*.

3. When c is large, the discriminant in (3.19) is positive, and therefore there are *two real roots*.

Given observations 2 and 3, there must be a special intermediate value of c for which there are two repeated real roots. This value is referred to as "critical damping" c_{c} and is characterized by the discriminant being 0

$$\frac{c_{\mathrm{c}}^2}{4m^2} - \frac{k}{m} \;=\; 0 \qquad \Rightarrow \qquad c_{\mathrm{c}} \;=\; 2\sqrt{mk}.$$

When the damping is critical, the roots are repeated and real, and are given by

$$c = c_{\mathrm{c}} \qquad \Rightarrow \qquad \bar{s}_{1,2} \;=\; -\,\frac{c}{2m} \;=\; -\,\frac{2\sqrt{mk}}{2m} \;=\; -\sqrt{\tfrac{k}{m}} \;=\; -\,\omega_{\mathrm{n}}.$$

Notice the reappearance of the *natural frequency* ω_n in the special case of critical damping. The behavior of the roots appears to be largely dependent on whether the damping is above or below the critical value. It is very useful to define a non-dimensional parameter that captures this fact. This parameter is called the *damping ratio*, and it is simply the ratio of actual damping c to the critical damping value c_c

$$\zeta := \frac{c}{c_c} \quad \Rightarrow \quad \zeta = \frac{c}{2\sqrt{mk}}. \tag{3.20}$$

Note that ζ is "unitless" since it is the ratio of two damping parameters that have the same units. If $\zeta < 1$, the damping is below c_c. If $\zeta = 1$, the then damping is critical $c = c_c$, and if $\zeta > 1$, the damping is above the critical value $c > c_c$.

We now rewrite the formula (3.19) for the characteristic roots in terms of the parameters ζ and ω_n. First note that (3.20) gives the ratio c/m as

$$\zeta = \frac{c}{2\sqrt{mk}} \quad \Rightarrow \quad \frac{c}{m} = 2\zeta\sqrt{\tfrac{k}{m}} = 2\zeta\omega_n. \tag{3.21}$$

Using this, the differential equation (3.18) can be rewritten as

$$\ddot{x}(t) + \tfrac{c}{m}\,\dot{x}(t) + \tfrac{k}{m}\,x(t) = 0 \quad \Leftrightarrow \quad \ddot{x}(t) + (2\zeta\omega_n)\,\dot{x}(t) + \omega_n^2\,x(t) = 0, \tag{3.22}$$

and the characteristic roots rewritten as

$$\begin{aligned} \bar{s}_{1,2} = -\tfrac{c}{2m} \pm \sqrt{\tfrac{c^2}{4m^2} - \tfrac{k}{m}} \quad &\Leftrightarrow \quad \bar{s}_{1,2} = -\zeta\omega_n \pm \sqrt{\zeta^2\omega_n^2 - \omega_n^2} \\ &\Leftrightarrow \quad \boxed{\bar{s}_{1,2} = \left(-\zeta \pm \sqrt{\zeta^2 - 1}\right)\omega_n,} \end{aligned} \tag{3.23}$$

where the roots are now expressed in terms of only two parameters, the damping ratio ζ, and the natural frequency ω_n. With this last expression, we see more clearly the effect of ζ on the roots. Figure 3.3 shows the locations of the roots in the complex plane as ζ ranges from 0 to ∞. There are two main regimes, the so-called "overdamped" (when the roots are real), and the "underdamped" (when the roots are complex), with the boundary between them as the critically damped case. We now investigate the details of the three cases.

- Overdamped: $\zeta > 1 \quad \Rightarrow \quad \zeta^2 - 1 > 0 \quad \Rightarrow \quad \sqrt{\zeta^2 - 1}$ is a real number. In this case, the two characteristic roots in (3.23) are both negative real numbers[1], and any solution is therefore a sum of two *decaying exponentials*

$$\boxed{x(t) = \mathsf{a}_1\, e^{\bar{s}_1 t} + \mathsf{a}_2\, e^{\bar{s}_2 t},} \qquad \bar{s}_{1,2} = \left(-\zeta \pm \sqrt{\zeta^2 - 1}\right)\omega_n, \tag{3.24}$$

  The two exponentials $e^{\bar{s}_1 t}$ and $e^{\bar{s}_2 t}$ are called the *modes* of the system. The exact linear combination of these two modes in any particular response depends on the initial conditions.

  There is an important relation between the two roots in that their product is independent of ζ as can be seen from

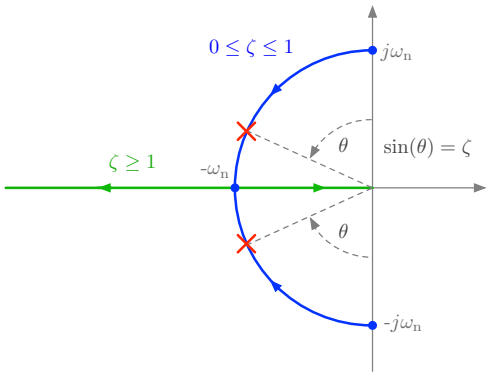
$$\bar{s}_1\bar{s}_2 = \left(-\zeta + \sqrt{\zeta^2 - 1}\right)\left(-\zeta - \sqrt{\zeta^2 - 1}\right)\omega_n^2 = \left(\zeta^2 - \left(\zeta^2 - 1\right)\right)\omega_n^2 = \omega_n^2.$$

  This means in particular that there is a reciprocal relation between them

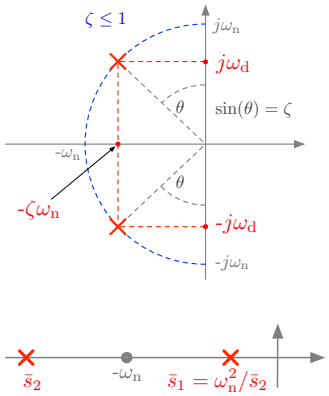
$$\bar{s}_2 = \omega_n^2/\bar{s}_1.$$

  Now suppose we keep ω_n fixed, and increase ζ from 1 to ∞. The expression (3.23) implies that $\bar{s}_2 = -\zeta - \sqrt{\zeta^2 - 1} \to -\infty$, and the reciprocal relation therefore implies $\bar{s}_1 \to 0$. In this case we call $e^{\bar{s}_2 t}$ the *fast mode* (since it decays quickly), and we call $e^{\bar{s}_1 t}$ the *slow mode* (since it decays slowly). Therefore, for heavily damped systems, the response is always a combination of a slow and fast mode. Usually, it is the slow mode that "dominates" the response since the fast mode decays quickly enough that its contribution to the overall response is negligible after a short time. Such a response and its two slow and fast components is shown in the last panel of Figure 3.4.

[1] Note that $\bar{s}_1$ is clearly a negative number. $\bar{s}_2$ is negative since $\sqrt{\zeta^2 - 1} < \zeta$ when $\zeta > 1$.



(a) A depiction of how the roots $\bar{s}_{1,2}$ "move" in the complex plane as ζ changes from 0 to ∞. The *underdamped* case corresponds to $0 \le \zeta \le 1$, and the roots are complex conjugates that lie on a semi-circle of radius ω_n. The *critically damped* case is when $\zeta = 1$, which corresponds to a double root on the real axis at $-\omega_n$. The *over-damped* case is when $\zeta > 1$, and the roots are both real, and as $\zeta \to \infty$, one roots limits to $-\infty$, while the other limits to 0, as shown by the green arrows.



(b) (*Top*): In the underdamped case, the imaginary part ω_d of the roots gives the *oscillation frequency*, while the real part $-\zeta\omega_n$ gives the decay exponent. (*Bottom*): In the overdamped case, the two real roots have a reciprocal relation $\bar{s}_2 = \omega_n^2/\bar{s}_1$. This implies that as $\zeta \to \infty$, $\bar{s}_1 \to -\infty$, while $\bar{s}_2 \to 0$ as shown in (a) by the green curves.

Figure 3.3: The locations in the complex plane of the characteristic roots $\bar{s}_{1,2}$ (denotes by the red crosses) of damped vibrations for various values of the damping ratio $\zeta = c/c_c$ (where c_c is the critical damping value), while the undamped natural frequency ω_n is held constant.

- Critically Damped: $\zeta = 1$, and therefore $\bar{s}_{1,2} = -\omega_n$, i.e. a double, negative real root. The general form of the solution corresponding to a double root at $\bar{s}_{1,2} = -\omega_n$ is

$$\boxed{x(t) \;=\; \mathsf{a}_1\, e^{-\omega_n t} \;+\; \mathsf{a}_2\, te^{-\omega_n t},} \tag{3.25}$$

  where the coefficients a_1 and a_2 are determined by initial conditions. In this case, the *modes* are really a single exponential mode $e^{-\omega_n t}$, but with an additional term of $te^{-\omega_n t}$. Both of those terms decay exponentially with exponent $-\omega_n$.

- Underdamped: $\zeta < 1 \quad \Rightarrow \quad \zeta^2 - 1 < 0 \quad \Rightarrow \quad \sqrt{\zeta^2 - 1}$ is imaginary. Therefore $\bar{s}_{1,2}$ are two complex-conjugate roots. Figure 3.3b illustrates this case. Let's see what the real and imaginary parts are in terms of the parameters ζ and ω_n

$$\begin{aligned} \bar{s}_{1,2} \;&=\; \left(-\zeta \pm j\sqrt{1-\zeta^2}\right)\omega_n. \;=\; -\zeta\omega_n \;\pm\; j\,\sqrt{1-\zeta^2}\,\omega_n \\ &=:\; -\zeta\omega_n \;\pm\; j\,\omega_d \qquad \text{(defining the } \textit{damped frequency } \omega_d := \omega_n\sqrt{1-\zeta^2}\text{)} \end{aligned} \tag{3.26}$$

  The imaginary part $\omega_d := \omega_n\sqrt{1-\zeta^2}$, which determines the oscillation frequency of the solution, is naturally called the "damped" oscillation frequency[2]. It is always smaller than the natural frequency ω_n (the frequency of oscillation if there were no damping) as seen in Figure 3.3b. The real part $-\zeta\omega_n$ determines an "exponentially decaying envelope" for the oscillations. To see both of the above facts, start from the expression $\bar{s}_{1,2} = -\zeta\omega_n \pm j\omega_d$ for the characteristic roots, and

[2]The oscillation frequency ω_d is sometimes referred to as the "damped natural frequency", or the "damped circular frequency".

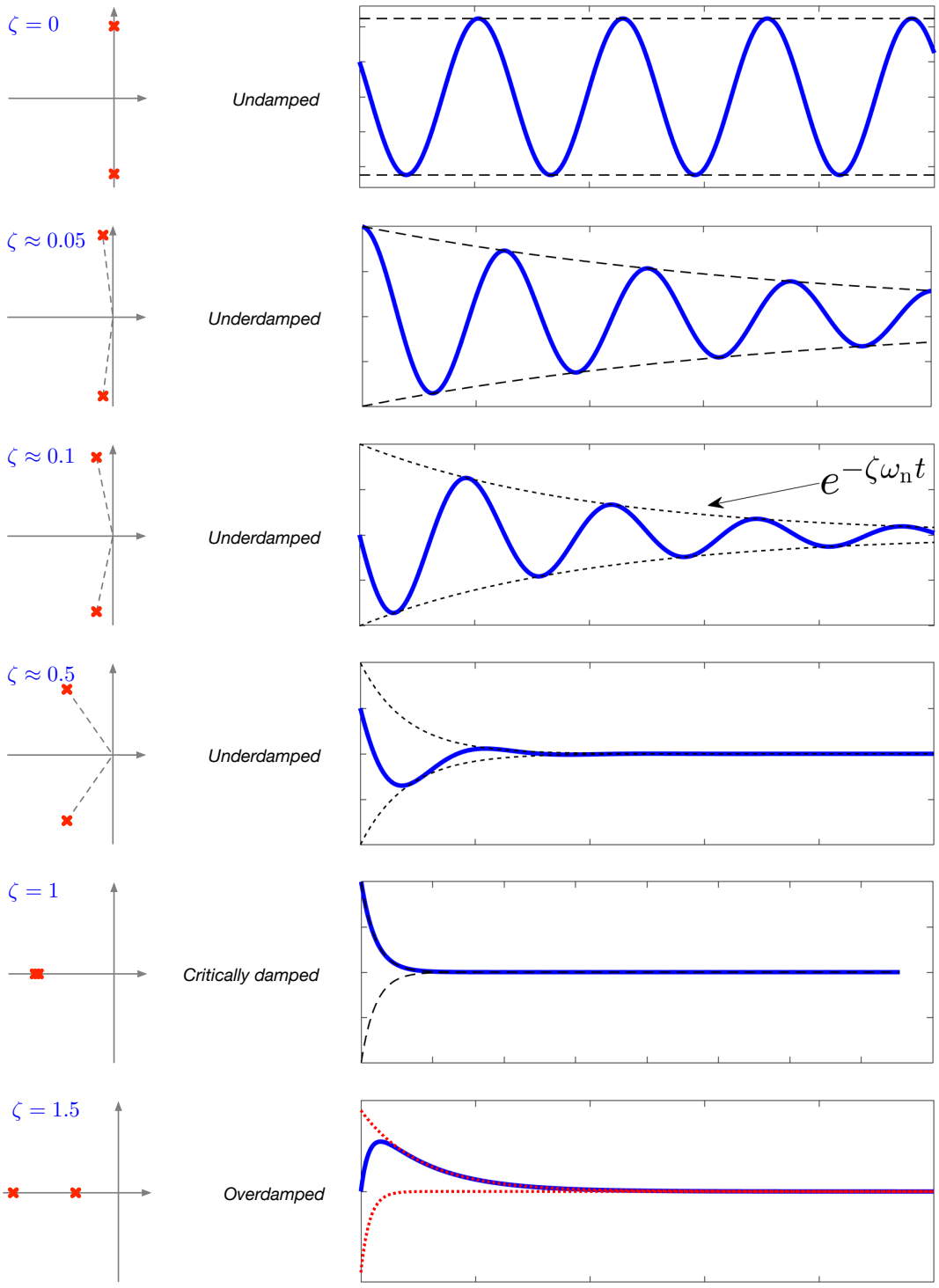


Figure 3.4: The effect of the damping ratio ζ on the qualitative shape of the response. The undamped case ($\zeta = 0$), three underdamped cases ($\zeta < 1$), the critically damped ($\zeta = 1$), and an overdamped ($\zeta > 1$) case are shown. Both the characteristic roots locations in the complex plane (left), and the time responses (*right*) are shown. The "decay envelops" of the underdamped cases are shown. The overdamped case shows the slow/fast component modes (dashed red lines). The time axis has the same limits in all of the above cases for the sake of comparison.

the general form of the solution

$$\begin{aligned} x(t) &= \mathsf{a}\, e^{(-\zeta\omega_n + j\omega_d)t} + \mathsf{a}^*\, e^{(-\zeta\omega_n - j\omega_d)t} \\ &= e^{-\zeta\omega_n t}\left(\mathsf{a}\, e^{j\omega_d t} + \mathsf{a}^*\, e^{-j\omega_d t}\right) \\ &= e^{-\zeta\omega_n t}\; 2\mathbb{R}\left(\mathsf{a}\, e^{j\omega_d t}\right) \\ &= e^{-\zeta\omega_n t}\; 2|\mathsf{a}|\cos\left(\omega_d t + \angle\mathsf{a}\right) \end{aligned} \tag{3.27}$$

$$\Rightarrow \quad \boxed{x(t) = \underbrace{\mathsf{x}\, e^{-\zeta\omega_n t}}_{\substack{\text{exponentially} \\ \text{decaying envelope}}} \underbrace{\cos\left(\omega_d t + \theta\right)}_{\substack{\text{sinusoid with frequency } \omega_d \\ \text{and phase } \theta}}}, \qquad \mathsf{x} := 2|\mathsf{a}|, \quad \theta := \angle\mathsf{a}, \tag{3.28}$$

where again the constants $\mathsf{x} = 2|\mathsf{a}|$ and $\theta = \angle\mathsf{a}$ are determined by initial conditions.

As already stated, the rate of exponential decay is determined by the real part $-\zeta\omega_n$ of the roots. If ω_n is held constant, then as ζ increases from 0 to 1, the response decays faster. This is illustrated in the various panels of Figure 3.4 which the reader should now contemplate. The oscillation frequency ω_d is always smaller than the natural frequency ω_n as can be seen from the relation

$$\omega_d = \omega_n \sqrt{1-\zeta^2}. \tag{3.29}$$

However, for typical values of ζ which range between 0 and say 0.3, the relative difference is difficult to perceive as you can see from Figure 3.4. For example, at $\zeta = 0.3$, the quantity $\sqrt{1-\zeta^2} \approx 0.95$, and thus the two frequencies differ by only about 5%. Even in the case of $\zeta = 0.5$ shown in Figure 3.4, it is difficult to distinguish the oscillation frequency from that of the case $\zeta = 0.01$ visually.

There is an interesting geometry in the underdamped case. As already mentioned, the roots lie on a semicircle of radius ω_n. To see that, examine their magnitude (as complex numbers)

$$|\bar{s}_{1,2}| = \left|\left(-\zeta \pm j\sqrt{1-\zeta^2}\right)\omega_n\right| = \sqrt{\zeta^2 + (1-\zeta^2)}\;\; \omega_n = \omega_n.$$

Therefore, as ζ changes from 0 to 1, the two complex conjugate roots move along this semicircle until they "meet" at $-\omega_n$ at the critical damping values $\zeta = 1$. This "path" is illustrated in Figure 3.3a. A particularly nice geometric interpretation of ζ is that it is the sine of the angle the roots make with the imaginary axis as shown in Figure 3.3b and follows from

$$\begin{aligned} \sin(\theta) &= \frac{\zeta\omega_n}{\omega_n} = \zeta \\ \cos(\theta) &= \frac{\omega_d}{\omega_n} = \frac{\omega_n\sqrt{1-\zeta^2}}{\omega_n} = \sqrt{1-\zeta^2} \end{aligned}$$

Remark 3.3. There is an important property of the damping ratio ζ that needs to be emphasized. Since this quantity is dimensionless, it allows for comparing vibration behavior across a large variety of mechanical systems as shown in the table of Figure 3.5. In some sense, ζ is the equivalent of the "Reynolds number" for vibration analysis. Just like the Reynolds number in fluid flow allows for comparison of flow behavior over varying spatial scales, viscosities and velocities, the damping ratio ζ characterizes vibrational behavior across systems with differing materials, construction and sizes as illustrated by the examples in the table.

Remark 3.4. For a system with any damping $c > 0$, the characteristic roots always have *negative real part*, or equivalently they lie strictly in the *left half of the complex plane*. This can be seen from Figure 3.3a, as well as the expressions (3.23) and (3.26) in the overdamped and underdamped cases respectively. This means that any response (from any initial condition) will always decay exponentially.

Mechanical System	Typical Damping Ratio ζ
Metals (in elastic range)	< 0.01
Continuous Metal Structures	$0.02 - 0.04$
Metal Structures with Joints	$0.03 - 0.07$
Aluminum/Steel Transmission Lines	≈ 0.0004
Small Diameter Piping Systems	$0.01 - 0.02$
Large Diameter Piping Systems	$0.02 - 0.03$
Automotive suspension (sedans)	$0.2 - 0.25$
Automotive suspension (sports cars)	$0.3 - 0.25$
Rubber	0.05
Large Buildings during Earthquakes	$0.01 - 0.05$

Figure 3.5: Typical damping ratio values for a variety of mechanical systems. Since the damping ratio ζ is a non-dimensional number, it can serve as a metric for comparison across a wide variety of mechanical systems. Data from *Orban, F. "Damping of materials and members in structures." Journal of Physics: Conference Series. Vol. 268. No. 1. IOP Publishing, 2011.*

This property is referred to as "stability" in dynamical systems. For the Mass-Spring-Damper system, it holds when $m, c, k > 0$ as is always the case. Note that the case of undamped vibrations $c = 0$ is a mathematical idealization. All physical systems have some amount of damping, even if microscopic. For extremely lightly damped systems where $\zeta \approx 0$, the exponential decay can be so slow so as not to be significant over the time length of the experiment. In this case, the mathematical idealization of $c = 0$ is useful and convenient.

Solutions in terms of Initial Conditions

Equations (3.24), (3.25), and (3.28) give the form of the solutions for the three cases respectively. The important qualities of these solutions such as oscillation frequencies and decay rates are given in terms of only ζ and ω_n, which depend on the system parameters, but not on initial conditions. It is less important, but convenient to derive explicit expressions for the coefficients that appear in those formulas as a function of initial conditions $x(0)$ and $\dot{x}(0)$. For the sake of completeness, formulas for these solutions are included in Appendix 3.A.

3.3 Inferring Parameters from Experimental Data

When dealing with a complex mechanical system, it is usually difficult to obtain a detailed description of every part's mechanical properties such as dimensions, masses, spring constants and the like. For example, suppose we have an automotive suspension system which we model as a Mass-Spring-Damper system. The details of the system are unknown but we can conduct an experiment where the system is given an initial displacement and the subsequent position (or velocities or accelerations) are measured as a function of time. Given this data, is it possible to find the mass, spring constant, and damping coefficient of the system?

To answer the above question, consider a typical set of experimental data as shown in Figure 3.6 which shows the position of a mass (versus time) that is oscillating due to some non-equilibrium initial condition. This data could for example be collected from a position sensor (e.g. strain gauge, optical encoder, etc.) over time. The trajectory looks like a typical Mass-Spring-Damper system in the under-damped regime, i.e. it is a signal of the form (3.28)

$$x(t) \;=\; \mathsf{x}\, e^{-\zeta\omega_n\, t}\, \cos\big(\omega_d t + \theta\big), \qquad\qquad \omega_d = 2\pi/T. \tag{3.30}$$

From the data, we can easily infer the oscillation period T by either measuring the time between zero crossings (provided we remove any constant offset in the signal), or the time between peaks.

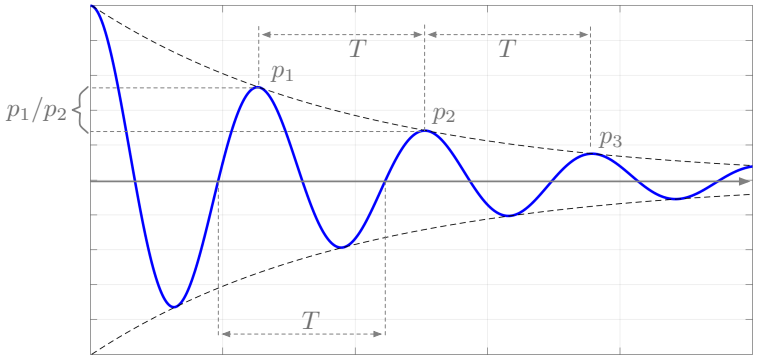


Figure 3.6: An example of the use of experimental data to infer system parameters. The signal shown could be recorded position measurements of a Mass-Spring-Damper system's free vibrations from a non-equilibrium initial condition (velocity or acceleration measurements also yield similar signals). From the data, one can readily obtain the period of oscillation T as well as the *logarithmic decrement* δ defined as the logarithm of the "ratio of successive peaks" $\delta := \ln(p_k/p_{k+1})$. The system's damping ratio ζ and its undamped natural frequency ω_n can be obtained from the simplified relations $\zeta \approx \delta/2\pi$ and $\omega_n \approx 2\pi/T$, which are reasonable approximations for the range $\zeta \leq 0.3$.

In Figure 3.6, it is assumed that the constant offset has been removed and we simply measure the time between zero crossings, which then gives the signal's period T. From the period we can find the damped oscillation frequency by

$$\omega_d = 2\pi/T \qquad \text{(because } \cos(\omega_d t + \theta) \text{ is periodic with period } T = 2\pi/\omega_d)$$

We can also find the decay rate of the envelope from measuring the "decay" of successive peaks in the signal. Those peaks are labeled $p_1, p_2, p_3, \ldots$ in Figure 3.6. Since the envelope is of the form $Ae^{-\zeta\omega_n t}$, we can get information about ζ and ω_n from the ratio of any two *successive* peaks (which are a period T apart in time) as follows

$$\frac{p_k}{p_{k+1}} = \frac{\times\, e^{-\zeta\omega_n \bar{t}}}{\times\, e^{-\zeta\omega_n(\bar{t}+T)}} = \frac{\times\, e^{-\zeta\omega_n \bar{t}}}{\times\, e^{-\zeta\omega_n \bar{t}}\, e^{-\zeta\omega_n T}} = e^{\zeta\omega_n T} = e^{2\pi\zeta\omega_n/\omega_d} \qquad \text{(since } T = 2\pi/\omega_d)$$

$$= e^{2\pi\zeta/\sqrt{1-\zeta^2}} \qquad \text{(since } \tfrac{\omega_n}{\omega_d} = \tfrac{1}{\sqrt{1-\zeta^2}} \text{ by (3.29))}$$

Note that the ratio p_k/p_{k+1} of two successive peaks is independent of which two successive peaks we choose (i.e. it is the same for any k). This leads us to define the *logarithmic decrement* δ, which is a number readily obtained from experimental data by simply measuring the ratio of any two successive peaks

$$\delta := \ln\frac{p_k}{p_{k+1}} \quad \Leftrightarrow \quad \frac{p_k}{p_{k+1}} = e^{\delta} \quad \Rightarrow \quad \delta = 2\pi\,\zeta\,\frac{1}{\sqrt{1-\zeta^2}}.$$

Thus the logarithmic decrement gives the damping ratio ζ by solving for it from the above relation

$$\boxed{\zeta = \frac{\delta}{\sqrt{\delta^2 + 4\pi^2}}}. \tag{3.31}$$

Note that both δ and ζ are non-dimensional. Note also he limiting behaviors at the two extremes

$$\zeta \to 0 \quad \Rightarrow \quad \delta \to 0 \quad \Rightarrow \quad p_k/p_{k+1} \to 1 \quad \Rightarrow \quad \text{peaks don't decay}$$
$$\zeta \to 1 \quad \Rightarrow \quad \delta \to \infty \quad \Rightarrow \quad p_k/p_{k+1} \to \infty \quad \Rightarrow \quad \text{peaks decay rapidly}$$

Once we find δ from the data, ζ can be obtained from (3.31). Next, the undamped natural frequency ω_n can be found from the damped frequency ω_d (which is determined from the period T)

by using (3.29) again

$$\omega_{\rm n} \;=\; \frac{\omega_{\rm d}}{\sqrt{1-\zeta^2}} \;=\; \frac{2\pi}{T}\;\frac{1}{\sqrt{1-\zeta^2}}. \tag{3.32}$$

The relations (3.31) and (3.32) can be simplified by observing that for $\zeta \leq 0.3$ (which is a typical range for many mechanical systems), the approximation $\sqrt{1-\zeta^2} \approx 1$ is reasonable[3]. The relations (3.31) and (3.32) then simplify to

$$\boxed{\zeta \;\approx\; \delta/2\pi, \qquad \omega_{\rm n} \;\approx\; \omega_{\rm d} \;=\; 2\pi/T.} \qquad \text{(an approximation for the range } \zeta \leq 0.3) \tag{3.33}$$

Thus we can directly go from (δ, T) to $(\zeta, \omega_{\rm n})$ using those simplified relations.

Is it possible to determine m, c, and k from ζ and $\omega_{\rm n}$? Clearly not, since again it is only the ratios c/m and k/m that determine dynamic behavior

$$m\ddot{x}(t)+c\dot{x}(t)+kx(t)=0 \qquad \Leftrightarrow \qquad \ddot{x}(t)+\tfrac{c}{m}\dot{x}(t)+\tfrac{k}{m}x(t)=0 \qquad \Leftrightarrow \qquad \ddot{x}(t)+(2\zeta\omega_{\rm n})\dot{x}(t)+\omega_{\rm n}^2 x(t)=0.$$

Thus the ratios can be obtained from the definitions of ζ and $\omega_{\rm n}$ as

$$\frac{k}{m}=\omega_{\rm n}^2, \qquad \frac{c}{m}=2\zeta\omega_{\rm n}.$$

Remark 3.5. The formulas given above for the calculation of system parameters from experimental data are approximate, but useful none the less. When more accurate estimation of system parameters is desired, the better approach is to do a regression fit of the experimental data to the system model. This type of regression is commonly used in engineering design, and goes by the name of "system identification".

Remark 3.6. The same technique presented above can be used if alternatively velocity or acceleration measurements are available. Indeed, if the position trajectory is of the form (3.30), then velocity is obtained by differentiation using the product rule

$$\begin{aligned} & x(t) \;=\; \mathsf{x}\, e^{-\zeta\omega_{\rm n}\, t}\; \cos\big(\omega_{\rm d}t+\theta\big) \\ \Rightarrow \qquad & \dot{x}(t) \;=\; (-\mathsf{x}\zeta\omega_{\rm n})\, e^{-\zeta\omega_{\rm n}t}\; \cos\big(\omega_{\rm d}t+\theta\big) - (\mathsf{x}\omega_{\rm d})\, e^{-\zeta\omega_{\rm n}t}\; \sin\big(\omega_{\rm d}t+\theta\big) \\ & \qquad\quad \;=\; -\,\mathsf{x}\, e^{-\zeta\omega_{\rm n}t}\Big(\zeta\omega_{\rm n}\; \cos\big(\omega_{\rm d}t+\theta\big) + \omega_{\rm d}\; \sin\big(\omega_{\rm d}t+\theta\big)\Big) \\ & \qquad\quad \;=\; (-\mathsf{x}\omega_{\rm n})\, e^{-\zeta\omega_{\rm n}t}\; \cos\big(\omega_{\rm d}t+\phi\big), \end{aligned} \tag{3.34}$$

where ϕ is some phase angle determined by the other parameters[4]. Thus velocity is of the same form as position; it is a decaying sinusoid with the same frequency and decay envelope as position, but with a different amplitude and phase. Since only the envelope decay rate and the oscillation frequency determine ζ and $\omega_{\rm n}$, then we could use a velocity measurement to determine system parameters using exactly the same method described earlier for position measurements.

Finally, if we differentiate velocity to get acceleration, the derivative of (3.34) will be of the form

$$\ddot{x}(t) \;=\; \mathsf{x}\,\omega_{\rm n}^2\, e^{-\zeta\omega_{\rm n}t}\; \cos\big(\omega_{\rm d}t+\psi\big),$$

for some phase ψ. This is again of the same form as the position signal, with yet another amplitude and phase, but with the same oscillation frequency and decay envelope. Therefore, acceleration measurements (which are typically the easiest to obtain with modern accelerometers) can alternatively be used to determine system parameters.

[3] For example, if $\zeta = 0.3$, then $\sqrt{1-\zeta^2} = 0.95$, thus the approximation has a 5% error, which is likely similar to the experimental error in measuring the peak ratios p_k/p_{k+1}.

[4] Note that $\cos(\omega_{\rm d}t+\theta)$ and $\sin(\omega_{\rm d}t+\theta)$ are two sinusoids of the same frequency, thus any linear combination of them is a sinusoid of the same frequency and some phase ϕ. Since they are exactly 90° out of phase, the amplitude of their linear combination is easy to calculate by Pythagoras as $\sqrt{\zeta^2\omega_{\rm n}^2+\omega_{\rm d}^2} = \sqrt{\omega_{\rm n}^2}$ where the last equality follows from the relation (3.29) which implies $\omega_{\rm d}^2 = \omega_{\rm n}^2(1-\zeta^2)$. The phase ϕ of the sum can also be calculated, but is irrelevant to the current argument.

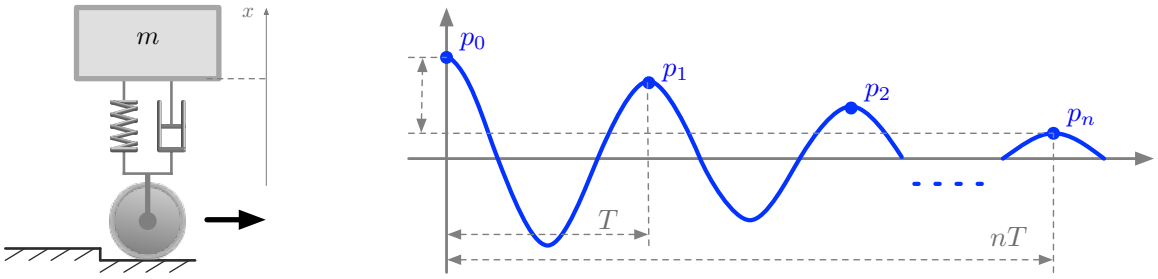


Figure 3.7: The setting of Example 3.7. A suspension system starts from a non-equilibrium position at zero vertical velocity (e.g. after having encountered a bump, and reaching its maximum extension at zero velocity, which is marked by the peak p_0). After n cycles, the peak ratio $p_0/p_n = e^{n\delta}$ is given by the logarithmic decrement δ as

The logarithmic decrement is also useful for quick calculations even when the system parameters are known as the next example demonstrates.

Example 3.7. Consider an idealized suspension system as shown in Figure 3.7 which is starting from a non-zero deflection with zero velocity. This initial deflection is marked as p_0 in the Figure. The system parameters ω_n and ζ are known. One question that can be asked is *how many periods of oscillation does it take for the peak deflection to be reduced to less than* 1% *of the initial deflection?* In the figure, this is labeled as peak p_n where n is unknown and is to be determined. Using the logarithmic decrement, we can express the ratio of p_0 to p_n as

$$\frac{p_0}{p_n} = \frac{p_0}{p_1}\frac{p_1}{p_2} \cdots \frac{p_{n-2}}{p_{n-1}}\frac{p_{n-1}}{p_n} = \underbrace{e^{\delta} \cdots e^{\delta}}_{n \text{ times}} = e^{n\delta}.$$

Note that there are n oscillation periods between peak p_0 and peak p_n (see Figure 3.7).

If we demand that p_n should be 1% of p_0, then by taking $\ln$ of both sides of the above

$$\ln\left(e^{n\delta}\right) = \ln\left(\frac{p_0}{p_n}\right) \quad \Rightarrow \quad n = \frac{1}{\delta}\ln\left(\frac{p_0}{p_n}\right) = \frac{\ln(100)}{\delta} \approx \frac{\ln(100)}{2\pi\zeta} \approx \frac{0.73}{\zeta} \quad (\text{since } \delta \approx 2\pi\zeta)$$

If for example $\zeta = 0.2$ (which would be a rather "soft" automotive suspension system), then

$$n \approx \frac{0.73}{0.2} = 3.65.$$

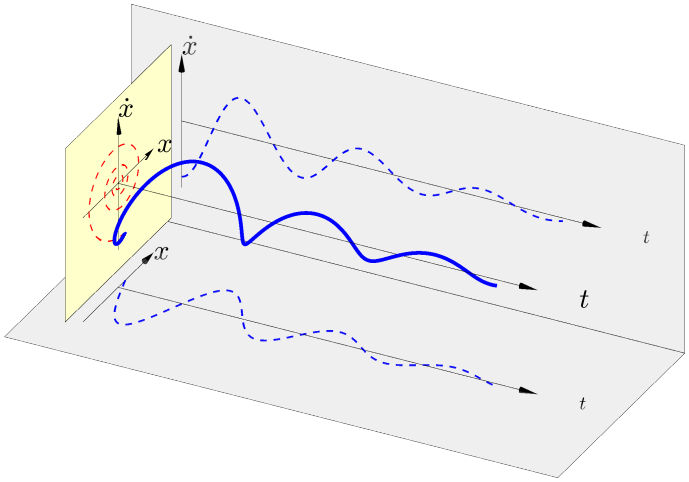
Thus it would take 4 oscillation periods to suppress the initial deflection to less than 1% of its value.

Note that we did not need knowledge of ω_n for the calculation above. That is because the question is asked in terms of "number of oscillation periods" as the unit. If the question were "how much time does it take ...", then knowledge of ω_n would be necessary.

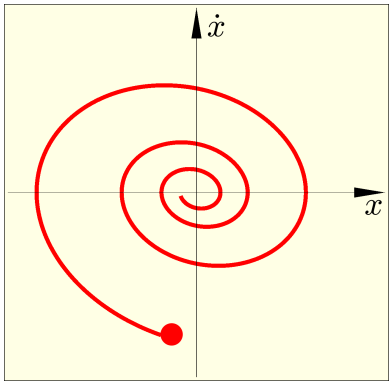
3.4 Analysis in State Space

So far we have viewed position and velocity solutions of the MSD system as functions of time as shown for example in Figure 3.2. There are other ways to view the motion of this system which reveal additional information not easily visible in time plots. This brings us to the concept of "state space". We begin first with some plots to develop intuition.

Figure 3.8 shows a plot of the solutions of a typical MSD system. The full solution is given for both position and velocity $\big(x(t), \dot{x}(t)\big)$ as a function of time. Since there are 3 variables x, $\dot{x}$ and t, a plot of this solution is a *curve in 3 dimensional space* as shown in Figure 3.8a. Note the helical shape, which a signature of position and velocity oscillating alternately, as well as the decay of the helix towards the time axis, which is a signature of the decay of oscillations due to damping. This plot has all the information about the systems solutions. Projection of the helix onto the (x, t) plane



(a) The full picture of the solutions requires a 3D plot. The coordinates here are $x(t)$, $\dot{x}(t)$ and time t. A damped system will typically produce a helical curve that decays towards the time axis. Projections of this curve onto the (x,t) plane give the standard position vs. time plot, and projections onto the $(\dot{x},t)$ plane give velocity versus time. Projections onto the $(x,\dot{x})$ plane give the state space trajectories plot shown in part (b).



(b) A plot of the "trajectories" of the MSD system in state space. This is a plot of velocity $\dot{x}(t)$ versus position $x(t)$, which are the coordinates in the 2-dimensional state space. The time information is lost in this plot. A damped system will typically have trajectories that "spiral inwards" from the initial condition (red dot) towards the origin.

Figure 3.8: Different ways of visualizing the solutions of a Mass-Spring-Damper (MSD) system either in the 2-dimensional state space of velocity versus position, or in 3D using $(t,x,\dot{x})$ as coordinates. The 3D plot contains all the information, and the three different projections of it give the standard position vs. time, velocity vs. time, as well as the state-space trajectories plot.

gives the position-vs-time plot, while projection onto the $(\dot{x},t)$ plane gives the velocity-vs-time plot. In addition, projection of the helix onto the $(x,\dot{x})$ plane gives a new kind of plot of velocity-vs-position as shown in Figure 3.8b. Such a plot shows features that are not readily visible in plots versus time such as those in Figure 3.2. In particular, we see more clearly how velocity and position "alternate in amplitude", i.e. as velocity increases, position gets close to zero and vice versa. This is of course due to the repeated exchange of kinetic and potential energies. The fact that the curve in Figure 3.8b "spirals inwards" towards the zero point $(0,0)$ reflects the decay of trajectories due to damping. While the time information is lost in this plot, the other features just mentioned are more clearly visible when solutions are represented in the $(x,\dot{x})$ plane.

The 2-dimensional plane $(x,\dot{x})$ is called the *state space* of the MSD system[5], and curves $\big(x(t),\dot{x}(t)\big)$ in this plane representing solutions from any particular initial condition are called *trajectories* in this state space. Figure 3.8b shows one such trajectory from one particular initial condition.

Given any dynamical system described by differential equations, a very useful concept is that of the state of that system. The state at any time is informally defined as the "minimal quantity" needed to predict the evolution of the system going forward in time without knowledge of the past history of the system. For example, consider a second-order differential equation of the form

$$m\,\ddot{x}(t) + c\,\dot{x}(t) + k\,x(t) \;=\; 0. \tag{3.35}$$

If we know the position and velocity $\big(x(\bar{t}),\dot{x}(\bar{t})\big)$ at any time $\bar{t}$, then we can solve the equation forward and compute positions and velocities $\big(x(t),\dot{x}(t)\big)$ for all future time $t \geq \bar{t}$, i.e. we can regard $\big(x(\bar{t}),\dot{x}(\bar{t})\big)$ as initial conditions ans obtain the solutions forward in time for $\bar{t} \leq t < \infty$. Note that we do not need to know the past of the trajectory (history) $\big\{\big(x(t),\dot{x}(t)\big),\ t<\bar{t}\big\}$, we only need to know the 2-vector $\big(x(\bar{t}),\dot{x}(\bar{t})\big)$ at time $\bar{t}$.

[5]In the physics literature, the state space is alternately termed the "phase space", and when it's two dimensional, it is called the "phase plane". This use of the term "phase" should not be confused with the phase of sinusoids or the phase of a complex number, which are unrelated. In this book we only use the terms "state" and "state space" for this concept.

With the above informal definition of the state, we see that the state of (3.35) at time t must be the vector $\big(x(t), \dot{x}(t)\big)$ of position and velocity. The second-order differential equation (3.35) can then be rewritten as a *vector, first-order* differential equation as follows. We first define $z_1(t) := x(t)$ and $z_2(t) := \dot{x}(t)$ as our "state variables". This means that $\dot{z}_1(t) = z_2(t)$ by definition, and the original system differential equation (3.35) gives another differential equation

$$\left.\begin{array}{l} z_1(t) := x(t) \\ z_2(t) := \dot{x}(t) \end{array}\right\} \qquad \Rightarrow \qquad \begin{bmatrix} \dot{z}_1(t) \\ \dot{z}_2(t) \end{bmatrix} = \begin{bmatrix} z_2(t) \\ -\frac{k}{m} z_1(t) - \frac{c}{m} z_2(t) \end{bmatrix}, \tag{3.36}$$

where we have written two scalar differential equations as a single *vector differential equation*. The second equation for $\dot{z}_2$ in (3.36) is a rewriting of the original equation (3.35) using the new variables z_1 and z_2

$$\ddot{x}(t) \; = -\tfrac{k}{m} x(t) - \tfrac{c}{m} \dot{x}(t) \qquad \Rightarrow \qquad \ddot{x}(t) \; = \; \dot{z}_2(t) = -\tfrac{k}{m} z_1(t) - \tfrac{c}{m} z_2(t).$$

An important feature of (3.36) is that it is a "first-order" equation, i.e. no variables are differentiated more than once, while the original equation (3.35) is second order. The conversion of a scalar, 2nd-order equation to a vector, 1st-order equation is achieved at the expense of using two variables $z_1(t), z_2(t)$ rather than the single original variable $x(t)$. Note that the It is a fact that all state-space models are of the form of vector, 1st-order differential equations as we will see in later chapters when we discuss state-space models for N-DOF systems.

When the dynamics are linear, i.e. when the differential equations describing the dynamics involve only linear combinations of variables and their derivatives, the state-space model (3.36) can be written in a nice matrix-vector form as follows

$$\tfrac{d}{dt} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} \tag{3.37}$$

The reader should now convince themselves that this matrix-vector equation is exactly the same as the vector differential equation in (3.36). The state-space model (3.37) is for a single MSD system. It is a special case of more general state-space models for linear dynamical systems, and in particular for larger N-DOF mechanical systems which we will study in later chapters. The most general models take the form

$$\begin{bmatrix} \dot{z}_1(t) \\ \vdots \\ \dot{z}_n(t) \end{bmatrix} = \begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} z_1(t) \\ \vdots \\ z_n(t) \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{bmatrix} \\ \dot{\boldsymbol{z}}(t) \\ \\ \end{bmatrix} = \begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{z}(t) \\ \\ \end{bmatrix}, \tag{3.38}$$

where $\boldsymbol{z}(t)$ is an n-dimensional "state vector" (typically containing all positions and velocities in a mechanical system), and A is matrix determined by the parameters of the system (such as damping coefficients, masses, spring constants, etc.). We will see later that all dynamical properties of such systems are determined by linear-algebraic properties of the matrix A such as its eigenvalues and eigenvectors. A short glimpse of this general analysis will be given at the end of this section for the special case of the single MSD system (3.37).

State space models are also useful for understanding the evolution (in time) of energy in mechanical systems as we explore next.

Energy

Recall the definition of mechanical energy from (3.15) earlier as the sum of potential and kinetic energies

$$E(t) \; := \; V(t) + T(t) \; := \; \tfrac{1}{2} k \; x^2(t) + \tfrac{1}{2} m \; \dot{x}^2(t). \tag{3.39}$$

Note that as a function of current position $x(t)$ and current velocity $\dot{x}(t)$, total energy is a *quadratic function* Q of position and velocity

$$Q\big(x, \dot{x}\big) := \tfrac{1}{2} \left(k x^2 + m \dot{x}^2\right) \qquad \Rightarrow \qquad E(t) = Q\big(x(t), \dot{x}(t)\big).$$

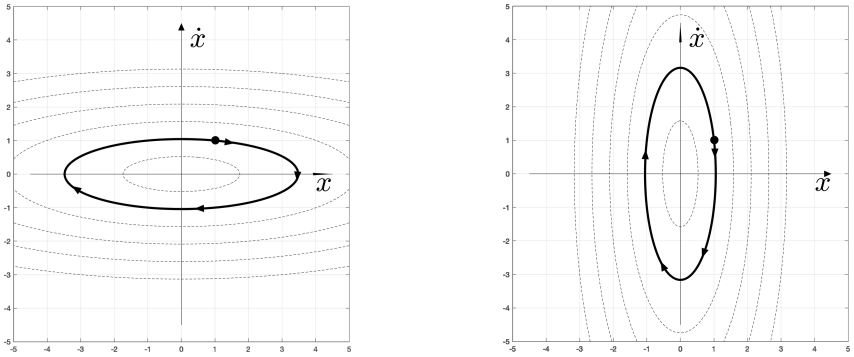


Figure 3.9: Depicting the position and velocity trajectories of an undamped MSD system in the state space $(x, \dot{x})$. The dashed curves are the level sets of the energy function which are ellipses. The solid curves are the trajectories starting from the initial condition $\big(x(0), \dot{x}(0)\big) = (1, 1)$ (shown as the black dot). Trajectories for an undamped system always trace out a level set (i.e. an ellipse) of the energy function. The trajectories in the above cases evolve clockwise. According to (3.40), eccentricity of the ellipses is determined by $\omega_{\rm n}$. On the left is the case $\omega_{\rm n} = .3$, while the right figure shows the case $\omega_{\rm n} = 3$.

For an MSD system with no damping, energy is conserved (as shown in (3.17)), and therefore any trajectory $\big(x(t), \dot{x}(t)\big)$ in state space must evolve on *level sets* of the quadratic energy function $Q(.,.)$

$$E(t) = Q\big(x(t), \dot{x}(t)\big) \;=\; \text{constant} \;=\; E(0).$$

Level sets of quadratic functions in the plane are ellipses, and we therefore expect these trajectories to trace out ellipses in state space as shown in Figure 3.9, where level sets of Q are shown as dashed curves, and trajectories from a specific initial conditions are shown solid curves.

Trajectory ellipses can be characterized in terms of system parameters by rewriting Q in terms the natural frequency $\omega_{\rm n}$ as follows

$$Q(x, \dot{x}) \;=\; \tfrac{1}{2}\Big(k\, x^2 + m\, \dot{x}^2\Big) \;=\; \tfrac{m}{2}\Big(\tfrac{k}{m}\, x^2 + \dot{x}^2\Big) \;=\; \tfrac{m}{2}\Big(\omega_{\rm n}^2\, x^2 + \dot{x}^2\Big).$$

Thus the level sets of Q are characterized by

$$Q(x, \dot{x}) = \text{constant} \qquad \Leftrightarrow \qquad \omega_{\rm n}^2\, x^2 + \dot{x}^2 = \text{constant}. \tag{3.40}$$

This is the equation of an ellipse whose eccentricity is determined by the coefficient $\omega_{\rm n}$, where if $\omega_{\rm n} < 1$, the major axis of the ellipse is horizontal, while if $\omega_{\rm n} > 1$, the major axis is vertical. This is confirmed in Figure 3.9 for the two cases $\omega_{\rm n} = .3$ and $\omega_{\rm n} = 3$.

We can also study the evolution of energy in damped MSD systems. The damper's force is non-conservative and actually dissipates mechanical energy. Over time, the mechanical energy (sum of spring potential and mass kinetic energies) will be converted to heat (due to viscous friction) inside the damper. Recall that in (3.17) we showed energy conservation by taking the time-derivative of the energy function and showed that derivative to be zero. In the damped case, we can go through a similar exercise and calculate exactly at what rate mechanical energy is being lost. Starting again from the definition of mechanical energy (3.39), we can take its derivative with respect to time and use the system dynamics to substitute

$$\begin{aligned}
\tfrac{d}{dt}E(t) \;&=\; \tfrac{d}{dt}\Big(\tfrac{1}{2}k\, x^2(t) + \tfrac{1}{2}m\, \dot{x}^2(t)\Big) \\
&=\; k\, x(t)\, \dot{x}(t) + m\, \dot{x}(t)\, \ddot{x}(t) && \text{(using the chain rule)} \\
&=\; k\, x(t)\, \dot{x}(t) - \dot{x}(t)\;\big(kx(t) + c\dot{x}(t)\big) && \text{(substituting from the diff. eq. } m\ddot{x} = -(kx + c\dot{x})\text{)} \\
&=\; -\,c\, \dot{x}^2(t). && (3.41)
\end{aligned}$$

For a damped system, the constant $c > 0$ is always positive, and the term $\dot{x}^2(t) \geq 0$. Therefore the derivative of energy is always negative (or if velocity is zero). The rate of energy dissipation is

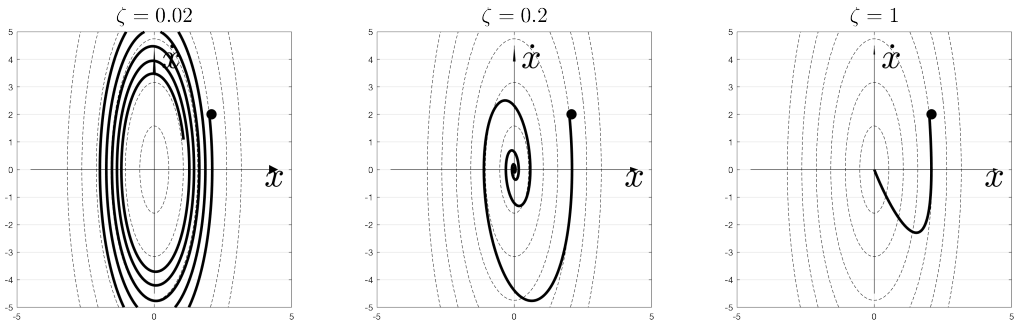


Figure 3.10: Examples of trajectories in state space $(x, \dot{x})$ for a damped system. The dashed curves are the level sets of the mechanical energy function (the sum of spring-potential energy and the mass' kinetic energy). In contrast to the undamped cases of Figure 3.9, the trajectories of a damped system no longer trace the energy level set, but rather "descend" along those level sets since energy is being continuously dissipated as heat in the damper. The larger the damping ratio ζ, the faster the trajectories limit towards the zero state $(x, \dot{x}) = (0, 0)$. The "inward spiraling" of the trajectories indicate decaying oscillations of both position and velocity.

proportional to the product $c\dot{x}^2(t)$ of the damping coefficient c and the square of the velocity, the higher this product, the faster mechanical energy is converted into heat.

Figure 3.10 illustrates this energy dissipation for three different values of ζ (which trends in the same manner as c if we keep m and k fixed). The level sets of mechanical energy are also plotted for comparison. We see that the trajectories do indeed "descend" on the energy level sets, with the descent being faster the larger ζ is. Observe also that the trajectories "spiral inwards" towards the zero state. This inwards spiraling reflects decaying oscillations along the position and velocity coordinates.

In the study of stability and control design, this kind of "dissipation analysis" we just went through is generalized through the concept of so-called "Lyapunov functions". The key point is that we did not have to actually solve the differential equations of the system in order to obtain the dissipation rate (3.41) above, we simply substituted the form of the differential equations without solving them.

Linear Algebra Methods for State-Space Models

We close this section with an illustration of how linear algebra methods, namely using eigenvalues and eigenvectors can be used to obtain (or characterize) solutions to state-space models like (3.37). This can be thought of as a "geometric view" of solutions of differential equations, and in essence provides a geometric method for solving differential equations qualitatively. This discussion can also be considered a brief preview of the matrix methods of normal mode analysis used for n-DOF systems in Chapter 7, but is not necessary for the remainder of the single DOF analysis chapters.

Consider again the MSD system of (3.37), and rewrite its "A-matrix" in terms of ω_n and ζ

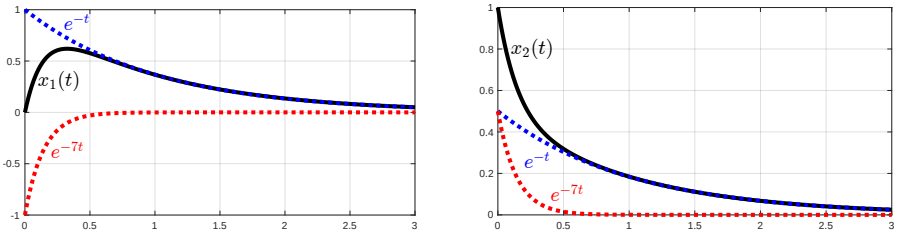
$$A \;:=\; \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{bmatrix} \;=\; \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{bmatrix}.$$

The eigenvalues of A are obtained from the "characteristic polynomial" of the matrix A

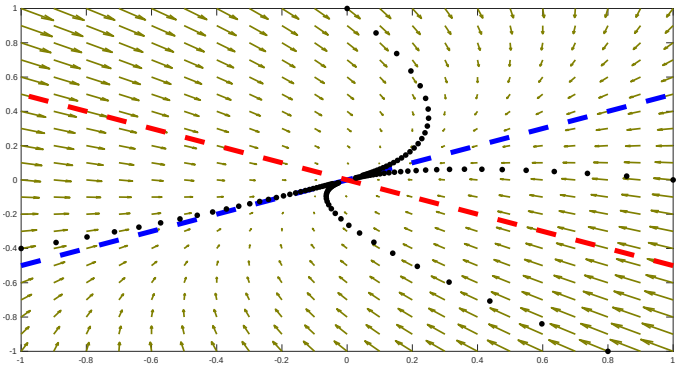
$$\begin{aligned} \det(sI - A) \;=\; \det \begin{bmatrix} s & -1 \\ \omega_n^2 & s + 2\zeta\omega_n \end{bmatrix} \;&=\; s\,(s + 2\zeta\omega_n) + \omega_n^2 \;=\; s^2 + 2\zeta\omega_n\, s + \omega_n^2 \\ &=\; (s - \bar{s}_1)(s - \bar{s}_2) \;=\; s^2 - (\bar{s}_1 + \bar{s}_2)\, s + \bar{s}_1\bar{s}_2, \end{aligned}$$

where $\bar{s}_{1,2}$ are the roots of this polynomial obtained from the quadratic formula as

$$\bar{s}_{1,2} = \left(-\zeta \pm \sqrt{\zeta^2 - 1}\right)\omega_n, \qquad \text{also note that} \qquad \bar{s}_1\bar{s}_2 = \omega_n^2, \qquad \bar{s}_1 + \bar{s}_2 = -2\zeta\omega_n.$$



(a) The time-evolution of a two dimensional system from some initial condition. This system has two real eigenvalues at -1 and -7. The trajectories x_1 and x_2 are linear combinations $x_l(t) = \alpha_{l,1}e^{-t} + \alpha_{l,2}e^{-7t}$ of the two system modes. Those components are shown (labeled by e^{-t} and e^{-7t} although they are $\alpha_{l,1}e^{-t}$ and $\alpha_{l,2}e^{-7t}$ respectively, and therefore should add up to the corresponding state as shown). The components e^{-7t} and e^{-t} are referred to as the "fast" and "slow" modes respectively. The fast mode dominates the response for initial times, after which the slow mode dominates.



(b) The vector field of a 2-dimensional linear system $\dot{x} = Ax$ in which A has two negative-real eigenvalues. The red and blue dashed lines show the one-dimensional subspaces spanned by the two *eigenvectors* respectively. Note how the vector field points straight in the direction of zero along those subspaces. Thus any initial condition lying in either subspace will evolve in that subspace for all time as per (??). Those two subspaces are referred as the *invariant subspaces*. The dotted black curves represent trajectories from various initial conditions, with dots representing the trajectories' positions at a uniform grid in time. Note that large spacing between successive dots represents a "fast moving" trajectory.

Figure 3.11: Modal and state-space analysis of an overdamped system.

Note that these are exactly what we called the "characteristic roots" of the 2nd order differential equation derived earlier (3.23)[6].

The roles played by the eigenvalues and eigenvectors is best illustrated by the overdamped example of Figure 3.11. In the overdamped case $\zeta > 1$, and the roots $\bar{s}_{1,2}$ are two (negative) real roots. In this example, they are at -1 and -7, and therefore we know from (3.24) that all solutions must be of the form

$$x(t) \;=\; \mathsf{a}_1\, e^{-t} + \mathsf{a}_2\, e^{-7t}.$$

The function e^{-t}, e^{-7t} are called the "modes" (or "pure modes") of the system, and this equation says that all solutions are linear combinations of the pure modes of motion. The exact values of the linear combination coefficients $\mathsf{a}_1, \mathsf{a}_2$ depend on the initial conditions. Figure 3.11a illustrates this for two different sets of initial conditions. Since the e^{-7t} mode decays much faster than the e^{-t} mode, they are referred to as the "fast" and "slow" modes respectively. Beyond an initial transient period after which the effects of the fast mode e^{-7t} become negligible, the solution $x(t)$ is dominated by by the slow mode e^{-t}.

Evolution of the state as a *vector* in $\mathbb{R}^n$: Another insight can be obtained by plotting the trajectories of $x(t)$ as curves in the 2-dimensional state space of $\dot{x}(t)$ versus $x(t)$. This is shown in Figure 3.11b.

[6]Note that we used the term "characteristic polynomial" for an ODE (3.3), and also for a *matrix* here. It can be shown in general that the matrix A of a state-space model of an ODE has exactly the same characteristic polynomial as the ODE it came from, thus justifying this terminology.

Relabel $x(t) =: x_1(t)$ and $\dot{x}(t) =: x_2(t)$, and write the state-space model as

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{bmatrix} \dot{\boldsymbol{x}}(t) \end{bmatrix} = \begin{bmatrix} \quad A \quad \end{bmatrix} \begin{bmatrix} \boldsymbol{x}(t) \end{bmatrix}. \tag{3.42}$$

This vector differential equation is fully described by the *vector field* shown in Figure 3.11b. At each point $\boldsymbol{x}$ in the plane, a vector equal to $A\boldsymbol{x}$ is drawn. Geometrically, for any trajectory $\boldsymbol{x}(t)$, the vector $\dot{\boldsymbol{x}}(t)$ is the *tangent vector* to the trajectory at the point $\boldsymbol{x}(t)$. The geometric meaning of the differential equation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t)$ is that any trajectory must *follow the vector field*, i.e. at any point $\boldsymbol{x}(t)$, its tangent vector $\dot{\boldsymbol{x}}(t)$ must equal the vector $A\boldsymbol{x}(t)$. This gives a "graphical method" of solving the differential equation. Starting from any initial condition $\boldsymbol{x}(0)$, simply follow the vector field to obtain the trajectory for $t \geq 0$. Several such trajectories are shown by the dotted curves in Figure 3.11b.

The eigenvectors of A play a special role in the state-space picture. Denote the eigenvectors of the eigenvalues $\bar{s}_1 = -1$, $\bar{s}_2 = -7$ by $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ respectively[7]. Those two eigenvectors (more precisely, "eigen-directions") are shown as the dotted blue and red lines in the figure. It can be shown that if we use those two eigenvectors as new coordinates, then any trajectory can be written in the new coordinate system as

$$\begin{bmatrix} \boldsymbol{x}(t) \end{bmatrix} \;=\; \alpha_1\, e^{-t} \begin{bmatrix} \boldsymbol{v}_1 \end{bmatrix} + \alpha_2\, e^{-7t} \begin{bmatrix} \boldsymbol{v}_2 \end{bmatrix}. \tag{3.43}$$

The coefficients α_1, α_2 depend on the initial conditions. This form of the solution shows an important property. If for example the initial condition $\boldsymbol{x}(0)$ lies anywhere on the line spanned by $\boldsymbol{v}_1$, then $\alpha_2 = 0$, and the trajectory will remain along that line for all $t \geq 0$, i.e.

$$\begin{bmatrix} \boldsymbol{x}(0) \end{bmatrix} = \alpha_1 \begin{bmatrix} \boldsymbol{v}_1 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} \boldsymbol{x}(t) \end{bmatrix} = \alpha_1\, e^{-t} \begin{bmatrix} \boldsymbol{v}_1 \end{bmatrix}, \quad t \geq 0.$$

Similarly, if the initial condition lies on the line spanned by $\boldsymbol{v}_2$, then the trajectory stays on that line, but will decay towards zero like e^{-7t} instead. This can also be seen in the figure by observing how the vector field directions do not change as you move along each of those lines. Along those lines, all the vectors point directly at the origin.

Thus the two eigenvectors of A represent very special directions in state space where trajectories starting along those lines will always remain along those lines. All other trajectories will decay along curves towards the origin. This phenomenon justifies naming the eigenvalues/vectors pairs $(\bar{s}_1, \boldsymbol{v}_1)$ and $(\bar{s}_2, \boldsymbol{v}_2)$ as "pure modes" of motion (also called "normal modes"). If the initial condition starts purely in one of those two directions, it will stay in that direction for all time. All other initial conditions have solutions which are "mixtures" of those two pure modes as in (3.43). We will see in later chapters how similar decompositions of motions and vibrations into several normal modes provides a powerful analysis method for understanding complex systems.

Appendix

3.A Solutions in Terms of Initial Conditions

The formulas (3.24), (3.25), and (3.28) give the form of the solutions in the overdamped, critically damped, and underdamped cases respectively. To obtain explicit expressions for the coefficients in terms of the initial conditions, we simply substitute the initial conditions $x(0)$ and $\dot{x}(0)$ in each of the formulas, and solve for the coefficients in terms of $x(0)$ and $\dot{x}(0)$.

[7] In the case of a MSD system of the form (3.42), the eigenvectors can be found analytically. However, for more general systems, eigenvectors can only be found with numerical methods. How exactly the eigenvectors are found is irrelevant to the present discussion.

Overdamped Case ($\zeta > 1$)

To find the coefficients a_1 and a_2 in terms of the initial conditions $x(0)$ and $\dot{x}(0)$, we have to solve a linear system of equations starting from

$$\begin{aligned} x(0) &= \mathsf{a}_1 e^{\bar{s}_1 t} + \mathsf{a}_2 e^{\bar{s}_2 t}\big|_{t=0} \;=\; \mathsf{a}_1 + \mathsf{a}_2 \\ \dot{x}(0) &= \frac{d}{dt}\left(\mathsf{a}_1 e^{\bar{s}_1 t} + \mathsf{a}_2 e^{\bar{s}_2 t}\right)\Big|_{t=0} \;=\; \left(\mathsf{a}_1\ \bar{s}_1 e^{\bar{s}_1 t} + \mathsf{a}_2\ \bar{s}_2 e^{\bar{s}_2 t}\right)\Big|_{t=0} \;=\; \mathsf{a}_1\ \bar{s}_1 + \mathsf{a}_2\ \bar{s}_2 \end{aligned}$$

This system of equations can be written in matrix-vector form and solved as

$$\begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \bar{s}_1 & \bar{s}_2 \end{bmatrix} \begin{bmatrix} \mathsf{a}_1 \\ \mathsf{a}_2 \end{bmatrix} \qquad \Rightarrow \quad \begin{bmatrix} \mathsf{a}_1 \\ \mathsf{a}_2 \end{bmatrix} = \frac{1}{\bar{s}_2 - \bar{s}_1} \begin{bmatrix} \bar{s}_2 & -1 \\ -\bar{s}_1 & 1 \end{bmatrix} \begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix}$$

We can now write the general solution in either of the two following forms

$$\begin{aligned} x(t) \;&=\; \frac{1}{\bar{s}_2 - \bar{s}_1}\left(\left(\bar{s}_2 e^{\bar{s}_1 t} - \bar{s}_1 e^{\bar{s}_2 t}\right)\ x(0) + \left(e^{\bar{s}_2 t} - e^{\bar{s}_1 t}\right)\ \dot{x}(0)\right) \\ &=\; \frac{1}{\bar{s}_2 - \bar{s}_1}\left(\left(\bar{s}_2 x(0) - \dot{x}(0)\right)\ e^{\bar{s}_1 t} + \left(-\bar{s}_1 x(0) + \dot{x}(0)\right)\ e^{\bar{s}_2 t}\right) \end{aligned}$$

Critically Damped Case

Here we see immediately that $\mathsf{a}_1 = x(0)$. To obtain a_2, we differentiate

$$\begin{aligned} \dot{x}(0) \;&=\; \frac{d}{dt}\left(\mathsf{a}_1 e^{\bar{s}t} + \mathsf{a}_2 t e^{\bar{s}t}\right)\Big|_{t=0} = \left(\mathsf{a}_1 \bar{s} e^{\bar{s}t} + \mathsf{a}_2 e^{\bar{s}t} + \mathsf{a}_2 t \bar{s} e^{\bar{s}t}\right)\big|_{t=0} = \mathsf{a}_1 \bar{s} + \mathsf{a}_2 \;=\; x(0)\ \bar{s} + \mathsf{a}_2 \\ \Rightarrow \qquad \mathsf{a}_2 \;&=\; \dot{x}(0) - x(0)\bar{s} \end{aligned}$$

In summary

$$x(t) \;=\; x(0)\ e^{\bar{s}t} \;+\; \left(\dot{x}(0) - x(0)\bar{s}\right)\ t e^{\bar{s}t}$$

Underdamped Case

To find the amplitude and phase of the response (3.28) in terms of initial conditions it is easiest to begin with the complex expression (3.27)

$$\begin{aligned} x(0) &= \mathsf{a}\ e^{(-\zeta\omega_{\rm n} + j\omega_{\rm d})t} \;+\; \mathsf{a}^*\ e^{(-\zeta\omega_{\rm n} - j\omega_{\rm d})t}\Big|_{t=0} \;=\; \mathsf{a} + \mathsf{a}^*, \\ \dot{x}(0) &= \frac{d}{dt}\left(\mathsf{a}\ e^{(-\zeta\omega_{\rm n} + j\omega_{\rm d})t} \;+\; \mathsf{a}^*\ e^{(-\zeta\omega_{\rm n} - j\omega_{\rm d})t}\right)\Big|_{t=0} \\ &= \left(\mathsf{a}\ (-\zeta\omega_{\rm n} + j\omega_{\rm d}) e^{(-\zeta\omega_{\rm n} + j\omega_{\rm d})t} \;+\; \mathsf{a}^*\ (-\zeta\omega_{\rm n} - j\omega_{\rm d}) e^{(-\zeta\omega_{\rm n} - j\omega_{\rm d})t}\right)\Big|_{t=0} \\ &= \mathsf{a}\ (-\zeta\omega_{\rm n} + j\omega_{\rm d}) \;+\; \mathsf{a}^*\ (-\zeta\omega_{\rm n} - j\omega_{\rm d}) \end{aligned}$$

This system of equations can be written in matrix-vector form and solved as

$$\begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -\zeta\omega_{\rm n} + j\omega_{\rm d} & -\zeta\omega_{\rm n} - j\omega_{\rm d} \end{bmatrix} \begin{bmatrix} \mathsf{a} \\ \mathsf{a}^* \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \mathsf{a} \\ \mathsf{a}^* \end{bmatrix} = \frac{j}{2\omega_{\rm d}} \begin{bmatrix} -\zeta\omega_{\rm n} - j\omega_{\rm d} & -1 \\ \zeta\omega_{\rm n} - j\omega_{\rm d} & 1 \end{bmatrix} \begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix}$$

$$\begin{aligned} \Rightarrow \quad 2\mathsf{a} \;&=\; \left(-j\zeta\tfrac{\omega_{\rm n}}{\omega_{\rm d}} + 1\right) x(0) \;-\; j\tfrac{1}{\omega_{\rm d}}\ \dot{x}(0) \\ &=\; x(0) \;-\; j\ \tfrac{\zeta\omega_{\rm n}\ x(0)\ +\ \dot{x}(0)}{\omega_{\rm d}}, \end{aligned} \tag{3.44}$$

where the last expression shows the real and imaginary parts of $2\mathsf{a}$.

Now recall that we can write a complex-conjugate exponential function like (3.27) in two equivalent real forms

$$\begin{aligned} x(t) &= \mathsf{a}\, e^{(-\zeta\omega_n + j\omega_d)t} + \mathsf{a}^*\, e^{(-\zeta\omega_n - j\omega_d)t} \\ &= e^{-\zeta\omega_n t}\, \mathsf{x}\, \cos\big(\omega_d t + \phi\big), && \mathsf{x} = 2|\mathsf{a}|, \quad \phi = \angle\mathsf{a} \\ &= e^{-\zeta\omega_n t}\Big(\alpha\, \cos(\omega_d t) - \beta\, \sin(\omega_d t)\Big), && \mathsf{a} = \alpha + j\beta. \end{aligned}$$

Comparing this last form to (3.44), we immediately read off the real and imaginary parts of a and write

$$x(t) = e^{-\zeta\omega_n t}\left(x(0)\, \cos(\omega_d t) + \frac{\zeta\omega_n\, x(0) + \dot{x}(0)}{\omega_d}\, \sin(\omega_d t)\right). \tag{3.45}$$

To obtain the other real form, we need to compute $|\mathsf{a}|$ and $\angle\mathsf{a}$. The expressions are a little too messy to be useful, but we include them here for completeness

$$\begin{aligned} \mathsf{x} = 2|\mathsf{a}| &= \frac{1}{\sqrt{1-\zeta^2}}\sqrt{x^2(0) + \frac{\dot{x}^2(0)}{\omega_n^2} + 2\frac{\zeta}{\omega_n}x(0)\dot{x}(0)} \\ \phi = \angle\mathsf{a} &= -\tan^{-1}\left(\frac{\zeta\omega_n\, x(0)+\dot{x}(0)}{x(0)\, \omega_n\sqrt{1-\zeta^2}}\right) \\ &= -\tan^{-1}\left(\frac{\zeta}{\sqrt{1-\zeta^2}} + \frac{\zeta}{\omega_n\sqrt{1-\zeta^2}}\frac{\dot{x}(0)}{x(0)}\right) \end{aligned}$$

Note that in the no damping case ($\zeta = 0$), these expression reduce to the previously calculated ones in (3.14).

Chapter 4

Harmonically Forced Vibrations: Frequency Response and Resonance

Harmonic forcing of a mechanical system is when it is subject to additive, sinusoidal forcing of some frequency. The response of the system to such a forcing will typically also be sinusoidal of the same frequency, but with a possibly different amplitude and phase relative to the forcing. Phasor analysis provides a very compact method for quantifying the relations between the amplitudes and phases of the forcing and the response. These relations are termed the "frequency response" of the system. Resonances occur when the forcing is in a range of frequencies close to the unforced system's natural frequencies. Around resonance, response amplitudes can be much larger than at other frequencies. Resonance phenomena are easily understood in terms of graphs of the frequency response. This analysis motivates an input-output point of view of mechanical vibrations where the "forcing" may also be sinusoidal displacements rather than actual forces. This input-output analysis is a powerful tool for systems described by constant-coefficient ODEs using representations of their "transfer functions".

Introduction and Motivation

In this chapter we study the behavior of systems with *external forcing* such as the forced Mass-Spring-Damper system in Figure 1.14. The differential equation for this system is

$$m\,\ddot{x}(t) \;+\; c\,\dot{x}(t) \;+\; k\,x(t) \;=\; f(t).$$

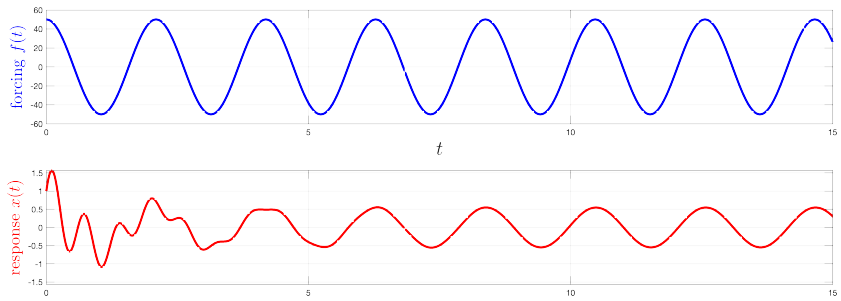
The dynamic (time-varying) force $f(t)$ is termed "external" to distinguish it from the "internal" spring and damper forces $-kx(t)$ and $-c\dot{x}(t)$. We restrict our attention in this chapter to "harmonic forcing", i.e. when the forcing is a pure sinusoid of some frequency ω, amplitude f and phase ϕ

$$m\,\ddot{x}(t) \;+\; c\,\dot{x}(t) \;+\; k\,x(t) \;=\; f(t) \;=\; \mathsf{f}\,\cos\big(\omega t+\phi\big). \tag{4.1}$$

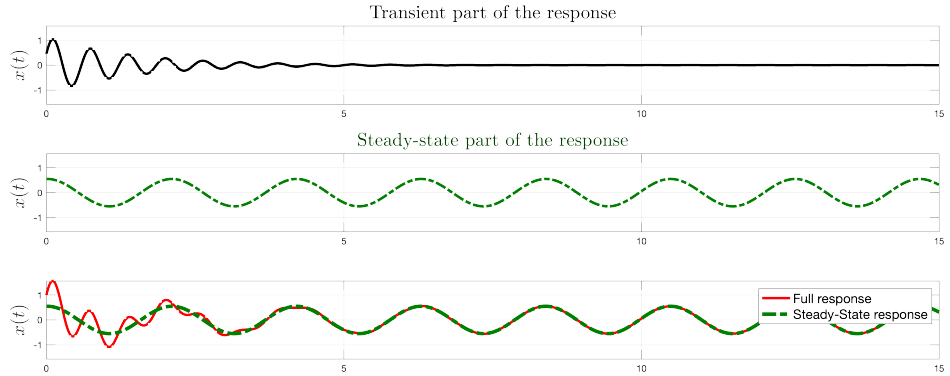
Figure 4.1 illustrates a typical displacement response $x(t)$ for a sinusoidal forcing $f(t)$. Figure 4.1a in particular shows that the response looks a little complicated initially, but after a "transient period" the response appears to be a pure sinusoid of the same frequency as the forcing $f(t)$, but with possibly a different amplitude and phase. Mathematically, it can be shown that if the parameters $m, c, k > 0$ are positive (which is the case for all mechanical systems), then (4.1) is a "stable" system, which means that effects of transients and initial conditions decay exponentially (recall Remark 3.4). The steady-state part of the response is of the form (Appendix 4.A describes this in detail)

$$\lim_{t\to\infty} x(t) \;=:\; x_{\mathrm{steady-state}}(t) \;=\; \mathsf{x}\,\cos\big(\omega t+\theta\big), \tag{4.2}$$

for some amplitude x and phase θ which will depend on the system parameters m, c, k as well as the parameters ω, f, ϕ of the forcing. The term "steady state" refers to the *long term response*, and



(a) A typical example of a forcing input (*top*) and the corresponding response (*bottom*). After an initial "transient time", the response limits to a pure sinusoid of the same frequency as the forcing input, but with a different amplitude and phase, and is referred to as the "steady-state response".



(b) The full response can be decomposed as the sum of two responses, the "transient response" (*top*) and the "steady-state" response (*middle*). The transient response is of the same frequency $\omega_{\rm d}$ as the unforced system, and decays to zero asymptotically. The steady-state response is a persistent pure sinusoid of the same frequency as the input. The bottom panel compares the full response (which is the sum of the transient and the steady-state) to the steady-state response.

Figure 4.1: Comparisons of the full and steady-state responses of the forced Mass-Spring-Damper system (4.1). If only long-time behavior is important, then only the steady-state response needs to be analyzed, and the transients can be ignored.

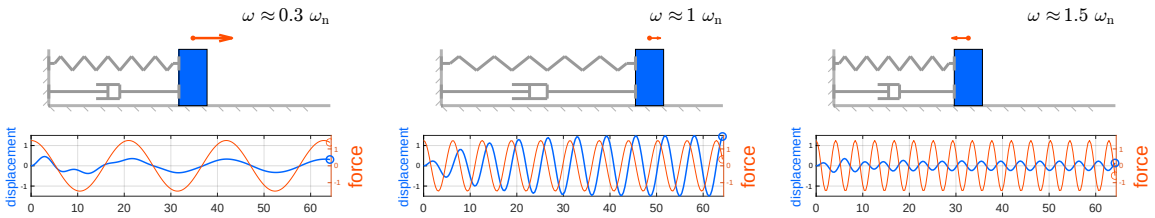
is defined mathematically in terms of the limit in (4.2). Figure 4.1b shows that the full response is decomposed as the sum of a transient (a decaying function) and a steady-state responses

$$x(t) \;=\; x_{\rm transient}(t) \;+\; x_{\rm steady-state}(t), \qquad\qquad \lim_{t\to\infty} x_{\rm transient}(t) \;=\; 0.$$

It is not difficult to show[1] that the decay rate of the transient response is the same as that of an unforced system, i.e. it is bounded by an exponentially decaying envelope proportional to $e^{-\omega_{\rm n}\zeta\ t}$. Thus transients decay faster in a highly damped system compared to a lightly damped one.

In harmonically forced systems that operate "persistently" (i.e. over long periods of time), the transient responses are much less important than steady-state responses. In this chapter we will ignore the transient response and develop tools to characterize the steady-state response which from now on will be denoted simply by $x(t)$ (rather than $x_{\rm steady-state}(t)$). Phasors provide a powerful method for calculation of those steady-state responses. This phasor analysis will also motivate an "input-ouput" point of view of mechanical systems. Equation (4.1) can be regarded as a *dynamical system* in the following sense; for each forcing function $f(t)$ and each set of initial conditions, the ODE gives a unique solution $x(t)$. In steady state, the effects of initial conditions disappear, and we can therefore regard the differential equation (4.1) as an "algorithm" that produces an "output" $x(t)$ given an "input" $f(t)$.

[1]The transient response can be expressed as a homogenous solution of (4.1), and therefore it is equivalent to the response of an unforced Mass-Spring-Damper system for some choice of initial conditions. Again, those details are in Appendix 4.A.



(a) At slow forcing frequencies (relative to the system's natural frequency), the response is similar to the static loading case where the mass' displacement follows the force's direction closely, i.e. they are "in phase". Click for animation.

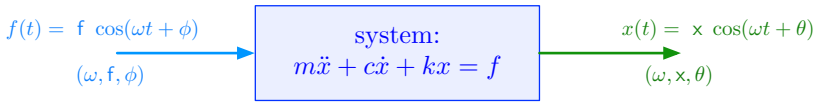
(b) When forcing at the natural frequency, oscillations build up until large amplitudes (relative to forcing at frequencies away from ω_n) are reached in steady state. Displacement appears to lag about 90° behind forcing. Click for animation.

(c) At high forcing frequencies, displacement amplitude is relatively small, and the mass barely moves due to its inertia. Displacement appears to lag by almost 180° behind forcing, i.e. they are completely "out of phase". Click for animation.

Figure 4.2: Illustrations and animations of the motion of a harmonically forced MSD system with forcing frequencies below ($\omega \approx 0.3\,\omega_n$), at ($\omega \approx \omega_n$), and above ($\omega \approx 1.5\,\omega_n$) the system's natural frequency ω_n (here $\omega_n = 1$ and $\zeta = 0.1$). The same forcing amplitudes is used in all three cases, and all plots have the same axis limits for comparison. In all cases, displacement oscillates harmonically at the same frequency as the forcing after an initial transient period has passed (the transients decay time constant here is $1/\zeta\omega_n \approx 10\,\text{s}$). A "resonance" phenomenon is clearly seen when the forcing frequency ω is near the system's natural frequency ω_n.

Before describing the analysis technique, we describe the physical phenomena that occur when the system (4.1) is forced with a variety of frequencies ω. The natural frequency ω_n of *free vibrations* does play an important role in characterizing the response in harmonically forced vibrations. In particular, forcing below ($\omega < \omega_n$), at ($\omega \approx \omega_n$), or above ($\omega > \omega_n$) the natural frequency will exhibit different response phenomena. Figure 4.14 illustrates the response of the system for various frequencies. Slow forcing frequencies are similar to the static loading case, and the mass' displacement follows the forcing phase closely (with amplitudes that are close to the static sag). When forcing with $\omega \approx \omega_n$, we see a clear "resonance phenomenon" where the mass' oscillations build up until large amplitudes are reached in steady state. We also observe that the mass' oscillations lag about 90° behind those of the forcing. At high forcing frequencies, the mass barely moves due to its inertia, which tends to "average" out the forcing oscillations. We also observe that forcing and displacement are completely "out of phase" (i.e. 180° phase difference).

We will develop a simple but powerful mathematical analysis tool, namely the *frequency response*, which quantifies the phenomena observed in the previous example, and is applicable to more general systems. The frequency response is an "input-output" point of view as depicted in the following block diagram



(4.3)

Unlike the case of free vibrations studied in Chapter 3, the vibrations we now study are the response (output) to harmonic (sinusoidal) forcings (inputs). Note that in the diagram (4.3) the forcing $f(t)$ and the response $x(t)$ are completely characterized the frequency ω (which is the same for both f and x), as well as (f, ϕ) and (x, θ), their respective amplitudes and phases. What we need to do is quantify the systems' response $(\mathsf{f}, \phi) \mapsto (\mathsf{x}, \theta)$, i.e. how to obtain the amplitude and phase (x, θ) of the response given (f, ϕ), the amplitude and phase of the forcing. This relation will vary with frequency ω, and this is the "frequency response" of the system, i.e. the relation $(\mathsf{f}, \phi) \mapsto (\mathsf{x}, \theta)$ as a function of frequency ω.

Given that the steady-state response is of the form (4.2), its amplitude x and phase θ can in principle be found by substituting the form (4.2) of the solution into the differential equation (4.1) as follows

$$\begin{aligned} & m\,\frac{d^2}{dt^2}\left(\mathsf{x}\cos\left(\omega t+\theta\right)\right) \;+\; c\,\frac{d}{dt}\left(\mathsf{x}\cos\left(\omega t+\theta\right)\right) \;+\; k\left(\mathsf{x}\cos\left(\omega t+\theta\right)\right) \;=\; \mathsf{f}\,\cos\left(\omega t+\phi\right) \\ \Rightarrow \quad & -m\,\omega^2\,\mathsf{x}\cos\left(\omega t+\theta\right) \;-\; c\,\omega\,\mathsf{x}\sin\left(\omega t+\theta\right) \;+\; k\,\mathsf{x}\cos\left(\omega t+\theta\right) \;=\; \mathsf{f}\cos\left(\omega t+\phi\right) \end{aligned} \tag{4.4}$$

If the system parameters m, c, k and the forcing parameters ω, f, ϕ are all given, then the only unknowns in this equation are x and θ. It is possible to use trigonometric identities to manipulate this equation and solve for them. This is the hard way to go about this solution, and if you want to see how messy this calculation is, see Appendix 4.B where it is illustrated in all of its ugliness.

We will instead employ a much simpler method using phasors, by combining the amplitudes and phases of the forcing and the response as complex numbers, i.e. $\hat{\mathsf{f}} := \mathsf{f}e^{j\phi}$ and $\hat{\mathsf{x}} := \mathsf{x}e^{j\theta}$, and then deriving the relation $\hat{\mathsf{f}} \mapsto \hat{\mathsf{x}}$ between these two complex numbers. This relation will be very simple. It will be a multiplication by another complex number $\mathsf{H}(\omega)$, i.e.

$$\hat{\mathsf{x}} \;=\; \mathsf{H}(\omega)\,\hat{\mathsf{f}},$$

where $\mathsf{H}(\omega)$ is a complex number for each ω that can be easily obtained from the system parameters. This quantity $\mathsf{H}(\omega)$ as a function of ω is the *frequency response* of the system.

We will first derive the method for the forced single Mass-Spring-Damper (4.1), and use it to do a thorough analysis of this system. Section 4.3 will then generalize this method to a much larger class of system that are described by higher-order, constant-coefficient forced ODEs.

4.1 Phasor Analysis of the Mass-Spring-Damper System

Recall the phasor representation of sinusoids of Section 2.3. Since both the forcing $f(t)$ in (4.1) and the steady-state solution (4.2) are cosines, and any cosine function can be written as the real part of a complex function, then we can write both $f(t)$ and $x(t)$ as

$$\begin{aligned} f(t) \;&=\; \mathsf{f}\cos\big(\omega t+\phi\big) \;=\; \mathbb{R}\Big(\hat{\mathsf{f}}\; e^{j\omega t}\Big), && \hat{\mathsf{f}} \;:=\; \mathsf{f}\; e^{j\phi} \\ x(t) \;&=\; \mathsf{x}\cos\big(\omega t+\theta\big) \;=\; \mathbb{R}\big(\hat{\mathsf{x}}\; e^{j\omega t}\big), && \hat{\mathsf{x}} \;:=\; \mathsf{x}\; e^{j\theta} \end{aligned}$$

Note that the complex numbers $\hat{\mathsf{f}} := \mathsf{f}\; e^{j\phi}$ and $\hat{\mathsf{x}} := \mathsf{x}\; e^{j\theta}$ "encode" the amplitudes and phases of $f(t)$ and $x(t)$ respectively, i.e.

$$\hat{\mathsf{f}} \;=\; \mathsf{f}\; e^{j\phi} \quad\Leftrightarrow\quad \begin{cases} \mathsf{f} = |\hat{\mathsf{f}}| \\ \phi = \angle(\hat{\mathsf{f}}) \end{cases}, \qquad\qquad \hat{\mathsf{x}} \;=\; \mathsf{x}\; e^{j\theta} \quad\Leftrightarrow\quad \begin{cases} \mathsf{x} = |\hat{\mathsf{x}}| \\ \theta = \angle(\hat{\mathsf{x}}) \end{cases}. \tag{4.5}$$

The procedure for phasor analysis is to now use the complex-valued functions of time $\hat{\mathsf{f}}\; e^{j\omega t}$ as the forcing function and $\hat{\mathsf{x}}\; e^{j\omega\theta}$ as the response, substitute those forms into the differential equation, and then solve for the complex number $\hat{\mathsf{x}}$ (which completely describes both the amplitude and phase of the response). First recall the formula for the derivative of a complex-valued exponential function from Section 2.3.1 which we repeat here for convenience

$$\begin{aligned} \tfrac{d}{dt}\; e^{j\omega t} \;=\; \tfrac{d}{dt}\;\big(\cos(\omega t)+j\sin(\omega t)\big) \;&=\; -\,\omega\;\sin(\omega t) \;+\; j\omega\;\cos(\omega t) \\ &=\; j\omega\;\big(\cos(\omega t) \;+\; j\sin(\omega t)\big) \;=\; j\omega\; e^{j\omega t}, \end{aligned} \tag{4.6}$$

where Euler's formula is used in the first and last equality. Thus differentiation of $e^{j\omega t}$ is equivalent to multiplying it by $j\omega$. The second time derivative can now obtained by applying the above formula twice

$$\tfrac{d^2}{dt^2}\; e^{j\omega t} \;=\; \tfrac{d}{dt}\;\big(j\omega\; e^{j\omega t}\big) \;=\; j\omega\;\frac{d}{dt}\; e^{j\omega t} \;=\; (j\omega)^2\;\; e^{j\omega t} \;=\; -\,\omega^2\; e^{j\omega t}.$$

Now we are ready to substitute $f(t) = \hat{\mathsf{f}}\; e^{j\omega t}$ and $x(t) = \hat{\mathsf{x}}\; e^{j\omega t}$ in the differential equation (4.1)

$$m\;\tfrac{d^2}{dt^2}\Big(\hat{\mathsf{x}}\; e^{j\omega t}\Big) \;+\; c\;\tfrac{d}{dt}\Big(\hat{\mathsf{x}}\; e^{j\omega t}\Big) \;+\; k\;\Big(\hat{\mathsf{x}}\; e^{j\omega t}\Big) \;=\; \hat{\mathsf{f}}\; e^{j\omega t} \tag{4.7}$$

$$\Rightarrow \qquad m\;\hat{\mathsf{x}}\;(j\omega)^2\; e^{j\omega t} \;+\; c\;\hat{\mathsf{x}}\;(j\omega)\; e^{j\omega t} \;+\; k\;\hat{\mathsf{x}}\; e^{j\omega t} \;=\; \hat{\mathsf{f}}\; e^{j\omega t} \tag{4.8}$$

$$\Rightarrow \qquad \hat{\mathsf{x}}\Big(-m\;\omega^2 \;+\; j\;c\;\omega \;+\; k\Big)\; \cancel{e^{j\omega t}} \;=\; \hat{\mathsf{f}}\; \cancel{e^{j\omega t}} \tag{4.9}$$

$$\Rightarrow \qquad \boxed{\hat{\mathsf{x}} \;=\; \frac{1}{(k-m\omega^2)+j\;(c\omega)}\;\hat{\mathsf{f}}.} \tag{4.10}$$

This is an expression for the response's phasor $\hat{\mathsf{x}}$ in terms of the forcing's phasor $\hat{\mathsf{f}}$, the system parameters m, c, k, and the forcing frequency ω. Note that the denominator has been rearranged to explicitly show its real and imaginary parts. It is important to note that the fact that both $f(t)$ and $x(t)$ have the same frequency ω allowed for the "removal" of the time dependence in the above relations by cancelling the common function of time $e^{j\omega t}$ from both sides.

Equation (4.10) is truly remarkable! The original differential equation between $f(t)$ and $x(t)$ is now replaced by an *algebraic equation* between the phasors $\hat{\mathsf{f}}$ and $\hat{\mathsf{x}}$. It is a simple algebraic relation involving only the product of the (frequency-dependent) complex number $1/\big((k-m\omega^2)+j(c\omega)\big)$ with $\hat{\mathsf{f}}$. This relation is important enough that it deserves its own name, which is referred to as the *complex frequency response* from the forcing f to the response x. This "complex" frequency response gives the amplitude and phase responses directly from the magnitude and phase of the complex number as we now describe in more detail.

To obtain the amplitude x and phase θ of the steady-state displacement oscillation $x(t) = \mathsf{x}\cos\big(\omega t+\theta\big)$, we simply take the magnitude and phase of $\hat{\mathsf{x}}$ in (4.10)

$$
\begin{aligned}
\mathsf{x} \;=\; |\hat{\mathsf{x}}| &= \left|\frac{1}{(k-m\omega^2)+j\,(c\omega)}\,\hat{\mathsf{f}}\right| \;=\; \left|\frac{1}{(k-m\omega^2)+j\,(c\omega)}\right|\left|\hat{\mathsf{f}}\right| && \big(\text{recall } |z_1 z_2| = |z_1||z_2|\big)\\
&= \left|\frac{1}{(k-m\omega^2)+j\,(c\omega)}\right|\mathsf{f} && \big(\text{since } |\hat{\mathsf{f}}| = |\mathsf{f}\,e^{j\phi}| = \mathsf{f}\,\cancel{|e^{j\phi}|}^{\,1}\big)\\
&= \frac{1}{\big|(k-m\omega^2)+j\,(c\omega)\big|}\,\mathsf{f} \;=\; \frac{1}{\sqrt{(k-m\omega^2)^2+(c\omega)^2}}\,\mathsf{f} && \big(\text{recall } |1/z| = 1/|z|\big)\\
\theta \;=\; \angle\hat{\mathsf{x}} &= \angle\left(\frac{1}{(k-m\omega^2)+j\,(c\omega)}\,\mathsf{f}e^{j\phi}\right) = \angle\left(\frac{1}{(k-m\omega^2)+j\,(c\omega)}\right) \;+\; \angle\big(\mathsf{f}e^{j\phi}\big) \\
& && \big(\text{recall } \angle(z_1 z_2) = \angle z_1 + \angle z_2\big)\\
&= -\angle\Big((k-m\omega^2)+j\,(c\omega)\Big) \;+\; \phi \;=\; -\tan^{-1}\left(\frac{c\omega}{k-m\omega^2}\right) \;+\; \phi && \big(\text{recall } \angle(1/z) = -\angle z\big)
\end{aligned}
$$

Note the use of magnitude and angle relations from (2.10). The reader should contrast the relative simplicity and compactness of these calculations with those based on trigonometric identities in Appendix 4.B.

We now summarize the above calculations and their implications

$$
\boxed{
\begin{aligned}
\frac{\mathsf{x}}{\mathsf{f}} = \frac{|\hat{\mathsf{x}}|}{|\hat{\mathsf{f}}|} &= \frac{1}{\sqrt{(k-m\omega^2)^2+(c\omega)^2}} && \textit{(the ratio of response amplitude to forcing amplitude:} \text{ Dynamic Compliance)}\\
\angle(\hat{\mathsf{x}}) - \angle(\hat{\mathsf{f}}) &= -\tan^{-1}\left(\frac{c\omega}{k-m\omega^2}\right) && \textit{(the "phase shift" from forcing to response. Negative phase shift is a time delay (lag).})
\end{aligned}
}
\tag{4.11}
$$

The ratio x/f has units of distance/force. To give intuition for what this ratio means, consider first the case when a a static force f is applied to a spring k. The resulting static deflection is $\mathsf{x} = \mathsf{f}/k$ (by definition of the spring constant), and the ratio $\mathsf{x}/\mathsf{f} = 1/k$ is called the *compliance* of the spring (i.e. the "compliance" $1/k$ is the reciprocal of the "stiffness" k). Under dynamic conditions such as sinusoidal forcing, the ratio x/f in (4.11) depends not just on k, but also on m, c as well as the frequency ω of forcing (dynamic loading). We can therefore think of the ratio x/f as a Dynamic Compliance. It is no longer a single number, but changes with the frequency of forcing.

One example of the ratio x/f and phase shift relations are illustrated in Figure 4.3, which the reader should now examine carefully. The numbers for this example are as follows

$$
\left.\begin{aligned} m &= 1\ \text{Kg},\\ k &= 10\ \tfrac{\text{N}}{\text{m}},\\ c &= .25\ \tfrac{\text{N}}{(\text{m/s})},\end{aligned}\right\} \Rightarrow \left\{\begin{aligned}\omega_{\text{n}} &= 3.16\ \tfrac{\text{rad}}{\text{s}},\\ \zeta &= 0.04,\end{aligned}\right. \quad \begin{matrix}\text{forcing frequency}\\ \omega = 0.9\ \omega_{\text{n}}\end{matrix} \quad \Rightarrow \left\{\begin{aligned}\frac{1}{\sqrt{(k-m\omega^2)^2+(c\omega)^2}} &\approx 0.49\ \tfrac{\text{m}}{\text{N}},\\ \tan^{-1}\left(\tfrac{c\omega}{k-m\omega^2}\right) &\approx 20.8^\circ.\end{aligned}\right.
$$

Examining the figure, we see that the amplitude ratio x/f is about $1/2$ as predicted by the formula. Note the transient time needed to reach steady state in first panel, which is roughly about 20 s. The

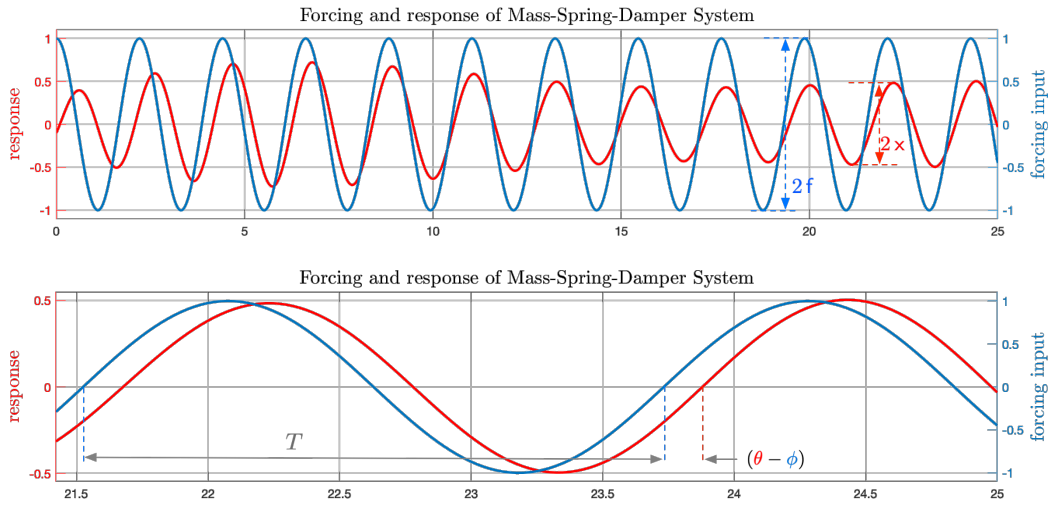


Figure 4.3: Depiction of the amplitude amplification ratio x/f and phase shift $\theta - \phi$ of the steady-state response given by Equations (4.11). (*Top*) The amplitude amplification ratio x/f is the ratio of the forcing function amplitude to the resulting response amplitude in steady state. It can be larger or smaller than one (it is about 1/2 here), and it has units of position/force. (*Bottom*) A zoomed-in view after transients have decayed (after about 20 s). The response sinusoid typically "lags" in time behind the forcing input's sinusoid. This lag is represented in terms of degrees, which is the difference between the forcing input's phase ϕ and the response's phase θ. Here the phase shift is approximately 20°. Since a cycle is 360°, and in terms of time units, a cycle is one period T, then the lag in terms of time units is $\frac{\theta-\phi}{360^\circ}T$, which in this case is approximately $0.056T$, i.e. about 5% of one period. Note that both the forcing input and the steady-state response have the same frequency (i.e. the same period T).

second panel shows a "zoomed in" view after steady state has been reached. The time and phase delays are clearly visible in the second panel. The time delay between forcing and displacement can be estimated from the figure to be approximately[2] 0.124 s. To compare that to the predicted phase delay of 20.8°, we need to convert the phase delay to a time delay. For this, we need the period of oscillation (this is the period of the forcing, which is $\omega = 0.9\ \omega_n$ for the simulation in Figure 4.3)

$$T = \frac{2\pi}{\omega} = \frac{2\pi \text{ rad}}{(0.9 \times 3.16\ \frac{\text{rad}}{\text{s}})} \approx 2.2 \text{ s} \quad \Rightarrow \quad \text{time shift} = \frac{\text{phase shift}}{360^\circ} T = \frac{20.8^\circ}{360^\circ} 2.2 \text{ s} \approx 0.127 \text{ s}.$$

This is close enough to the graphically-estimated delay of 0.124 s.

Dynamic, Frequency-dependent Stiffness and Compliance

As already mentioned, the frequency relations (4.11) have nice interpretations in terms of dynamic stiffness and compliance. Recall that in the static case, the stiffness of a mechanical system is the ratio of applied force to resulting displacement. Its compliance is the reciprocal, namely the ratio of displacement to applied force. The quantity x/f in (4.11) is the ratio of displacement to forcing amplitudes under periodically time-varying forcing, and thus can be interpreted as the dynamic compliance of the system. In the dynamic case however, this compliance changes with forcing frequency unlike the static case which is just a single number.

To understand dynamic compliance in the Mass-Spring-Damper system, it is useful to graph the quantities (4.11) as a function of applied force frequency ω. Figure 4.4 shows such a graph for a typical underdamped system. The behavior of the dynamic compliance relation

$$\text{dynamic compliance} := \frac{\mathsf{x}}{\mathsf{f}} = \frac{1}{\sqrt{(k - m\omega^2)^2 + c^2\omega^2}} \tag{4.12}$$

can be roughly roughly divided into three main frequency regions.

[2] This can be determined for example by using the "data cursor" in Matlab.

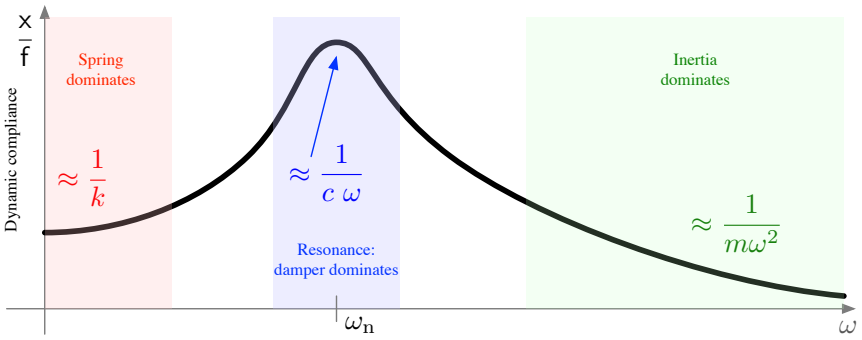


Figure 4.4: A typical graph of the frequency-dependent, dynamic compliance x/f for an underdamped system. Three different frequency regimes can be identified. The low frequency (almost static) regime is dominated by spring effects where the compliance is approximately $1/k$, similar to static compliance. Large compliance ratios $\mathsf{x}/\mathsf{f} \approx 1/c\omega$ are possible when the forcing frequency ω is near the system's free-vibrations natural frequency ω_n. This is the region of *resonance*, where the compliance ratio is largely dominated by damping effects. Lower damping implies larger resonances. In the high frequency regime, inertial effects dominate, and the mass' inertia "averages out" the effects of rapid forcing variations to near zero.

1. Low frequency (spring effects): In this region the dynamic compliance simplifies to

$$\omega \approx 0 \qquad \Rightarrow \qquad \frac{\mathsf{x}}{\mathsf{f}} \approx \frac{1}{k}.$$

The spring stiffness (or equivalently its compliance $1/k$) dominates the behavior. Intuitively, the low frequency regime is almost like the static regime. Since damping effects are velocity dependent, and the mass' inertia is related to acceleration, both of these effects are small in the low frequency regime. Thus the spring force is dominant at low frequencies.

2. Near resonance (damping effects): When the term $k - m\omega^2$ in the denominator of (4.12) is near zero, the compliance is near its maximum. This happens when the applied forcing frequency $\omega \approx \sqrt{k/m} =: \omega_n$ is close to the systems natural frequency ω_n of free vibrations. This is the phenomenon of resonance; when the compliance is high, small forces produce large displacements. The compliance near resonance is approximately

$$\omega \approx \sqrt{k/m} \qquad \Rightarrow \qquad k - m\omega^2 \approx 0 \qquad \Rightarrow \qquad \frac{\mathsf{x}}{\mathsf{f}} \approx \frac{1}{c\,\omega}.$$

This number is approximately the height of the resonance peak in the graph of dynamic compliance. It is inversely proportional to the damping coefficient c, and can be very large for systems with very low damping. Near resonance, the applied force pumps large amounts of power into the system producing increasingly larger oscillations until velocities are sufficiently high that the average power lost in the damper equalizes the average power delivered by the external force. In the absence of any damping, these oscillations can grow without bound. This will be analyzed quantitatively in the next section that studies power flows in this system.

3. High frequency (inertia effects): At very high frequencies, the compliance is dominated by the inertia term $(m\omega^2)^2$ under the square root

$$\omega \to \infty \qquad \Rightarrow \qquad \frac{\mathsf{x}}{\mathsf{f}} \approx \frac{1}{\sqrt{(m\omega^2)^2}} = \frac{1}{m\,\omega^2} \overset{\omega\to\infty}{\longrightarrow} 0.$$

This is physically intuitive. At high forcing frequencies, the inertia of the mass implies that it is effected by roughly the *average* of the applied force over the very short periods of oscillation. In other words, as the force oscillates between positive and negative, the mass' inertia prevents it from reacting instantly to those rapidly varying forces. The net result is very little compliance in the high frequency range.

A simple spring (with no mass or damper) reacts "instantly" to a force applied to it. The spring deflection x given a certain static force f is simply its compliance $\mathsf{x}/\mathsf{f} = 1/k$. If the force is time-varying, the ratio of deflection amplitude x to forcing amplitude f is still $1/k$. Thus for a simple spring by itself (without a mass or a damper), its compliance does not change with frequency. This idealization is never true in reality since any spring has some mass (and even some internal molecular damping). Thus for any mechanical system, its compliance is a function of the frequency of the applied force. In fact, plotting this frequency-dependent compliance (which in many cases can also be obtained experimentally) is one way to characterize the dynamic behavior of complex mechanical systems.

Non-dimensionalized Form of 2nd Order Frequency Response

It is very useful to rewrite the relations (4.11) in "non-dimensional" form so as to make comparisons across Mass-Spring-Damper systems with widely varying parameters using a single expression. Recall from the study of free vibrations that the natural frequency $\omega_{\mathrm{n}} := \sqrt{k/m}$ and the damping ratio $\zeta = \frac{c}{2m\omega_{\mathrm{n}}} = \frac{c}{2\sqrt{km}}$ played a significant role in characterizing free vibrations. Similarly, they will play a significant role in studying the frequency response and its resonances.

Rewriting[3] the parameters m, c, k in Equation (4.10) in terms of ω_{n} and ζ

$$\begin{aligned}\frac{\hat{\mathsf{x}}}{\hat{\mathsf{f}}} &= \frac{1}{(k - m\omega^2) + j(c\omega)} = \frac{1}{k\big(\,(1 - \frac{m}{k}\omega^2) + j\,(\frac{c}{k}\omega)\,\big)} = \frac{1}{k\big(\,(1 - (\omega/\omega_{\mathrm{n}})^2) + j\;2\zeta\,(\omega/\omega_{\mathrm{n}})\,\big)} \\ &= \frac{1}{k}\;\frac{1}{(1 - \Omega^2) + j\;2\zeta\Omega}, \qquad \Omega := \omega/\omega_{\mathrm{n}}, \end{aligned} \tag{4.13}$$

where we have defined a non-dimensional frequency Ω, the ratio of the *forcing frequency* ω to the *natural frequency* ω_{n} of the system without forcing. The expression above involves also the damping ratio $\zeta = c/c_c$ (c_c is the "critical damping" value), which is another non-dimensional number. The amplitude ratio $\mathsf{x}/\mathsf{f} = |\hat{\mathsf{x}}/\hat{\mathsf{f}}|$ however has dimensions of distance/force. It is sometimes useful to also rewrite the amplitude ratio in a non-dimensional form as follows

$$\frac{\hat{\mathsf{x}}}{\hat{\mathsf{f}}} = \frac{1}{k}\;\frac{1}{(1 - \Omega^2) + j\;2\zeta\Omega} \qquad \Leftrightarrow \qquad \boxed{\frac{\hat{\mathsf{x}}}{\hat{\mathsf{f}}/k} = \frac{1}{(1 - \Omega^2) + j\;2\zeta\Omega}}. \tag{4.14}$$

The magnitude $|\hat{\mathsf{f}}/k| = \mathsf{f}/k$ of the denominator has units of distance. It is sometimes referred to as the static sag, which is the magnitude of spring compression/extension (it is defined to always be a positive number) when a *constant force* of magnitude f is applied. A constant (static) applied force corresponds to *zero forcing frequency*[4] $\Omega = \omega/\omega_{\mathrm{n}} = 0$, and indeed if $\Omega = 0$ in the above expression, we get $\hat{\mathsf{x}}/(\hat{\mathsf{f}}/k) = 1$, i.e. the amplitude is precisely the static sag in the zero forcing frequency (i.e. static applied force) case.

The magnitude $\left|\hat{\mathsf{x}}/(\hat{\mathsf{f}}/k)\right| = \mathsf{x}/(\mathsf{f}/k)$ of the quantity $\hat{\mathsf{x}}/(\hat{\mathsf{f}}/k)$ in (4.14) is non-dimensional since it is a ratio of distances. It is best understood as $(\mathsf{x}/\mathsf{f})\,/(1/k)$, the ratio of dynamic-to-static compliance. At any given frequency, it is the amplitude of a mass' displacement relative to the displacement that an equal, but static force causes.

A summary of the above non-dimensional quantities, as well as plots of the frequency responses for various values of the damping ratio ζ are given in Figure 4.5. Examine the plots carefully and note the following observations.

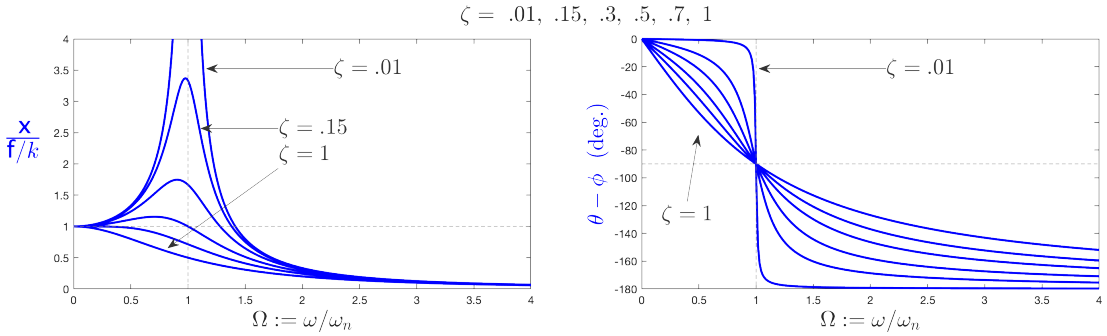
1. (Behavior at low frequencies). Beginning with the magnitude plot, we see that at low frequencies $\Omega \approx 0$, the non-dimensional amplitude ratio $\mathsf{x}/(\mathsf{f}/k) \approx 1$. That is expected from the definition since it is the ratio of dynamic to static displacement amplitudes, and $\Omega = 0$ corresponds to static loading.

[3] The relation that's needed here is $\frac{c}{k} = \frac{2\zeta\sqrt{km}}{k} = 2\zeta\sqrt{\frac{m}{k}} = 2\zeta\frac{1}{\omega_{\mathrm{n}}}$.

[4] Note that $\omega = 0$ corresponds to $f(t) = \mathsf{f}e^{j0t} = \mathsf{f}$), i.e. a signal with zero frequency is just a constant signal.

$$
\begin{aligned}
\textit{forcing frequency relative to natural frequency} &= \Omega := \frac{\omega}{\omega_{\mathrm{n}}} = \frac{\omega}{\sqrt{k/m}} \\
\textit{damping relative to critical damping: damping ratio} &= \zeta := \frac{c}{c_{\mathrm{c}}} = \frac{c}{2\sqrt{km}} \\
\frac{\textit{amplitude of dynamic (at frequency } \Omega\textit{) displacement}}{\textit{constant displacement due to static force (static sag)}} &= \frac{|\hat{\mathsf{x}}|}{|\hat{\mathsf{f}}|/k} = \frac{1}{\sqrt{(1-\Omega^2)^2 + (2\zeta\Omega)^2}} \\
&= \frac{\mathsf{x}/\mathsf{f}}{1/k} = \frac{\textit{dynamic compliance}}{\textit{static compliance}} \\
\textit{phase shift from forcing to response} &= \theta - \phi = -\tan^{-1}\left(2\zeta\,\frac{\Omega}{1-\Omega^2}\right)
\end{aligned}
$$

(a) A summary of non-dimensional quantities used in the dynamic analysis of an MSD system.



(b) The frequency response from forcing input to displacement output of the Mass-Spring-Damper system for various values of the damping ratio ζ. (*Left*) The non-dimensional ratio $\mathsf{x}/(\mathsf{f}/k) = (\mathsf{x}/\mathsf{f})/(1/k)$ of *dynamic-to-static compliance* as a function of the *normalized forcing frequency* $\Omega := \omega/\omega_{\mathrm{n}}$. The case $\omega \to 0$ corresponds to static loading, where by definition $\mathsf{x}/(\mathsf{f}/k) = 1$. Resonance occurs near $\Omega \approx 1$ ($\omega \approx \omega_{\mathrm{n}}$), where the response amplitude becomes infinite in the limit of no damping ($\zeta \to 0$). For moderate values of ζ, resonance manifests as a "peak" in the response ratio around the natural frequency. Note that for $\zeta \geq 0.5$, the resonance peak is barely perceptible. For very rapidly oscillating forcing ($\omega \to \infty$), the dynamic response amplitude tends to zero. (*Right*) The phase response represents the "phase shift" from forcing to displacement phases respectively. Here, it is a negative phase shift, implying that displacement "lags in time" after the forcing. The phase shift increases with frequency ω, but it is always -90° at resonance ($\omega = \omega_{\mathrm{n}}$). For very lightly damped systems ($\zeta \approx 0$), there is a sharp transition around resonance from a near zero phase shift to almost -180°.

Figure 4.5: A summary of the definitions and behavior of the frequency response of the single Mass-Spring-Damper system.

2. (Resonance). For low damping ratios $\zeta \ll 1$, there is a very large peak in the magnitude plot around $\Omega \approx 1$ ($\omega \approx \omega_{\mathrm{n}}$). This is the phenomenon of *resonance*: when the forcing frequency closely matches the system's natural frequency, large amplitude gains are exhibited. To see that mathematically, consider the magnitude response in the extreme case of zero damping $\zeta = 0$

$$\frac{|\hat{\mathsf{x}}|}{|\hat{\mathsf{f}}|/k} = \frac{1}{|1-\Omega^2|}.$$

This gain clearly goes to infinity when $\Omega \to 1$ ($\omega \to \omega_{\mathrm{n}}$). This means that without any damping, if the system is forced at its natural frequency, the displacement oscillations build up and increase without bound as $t \to \infty$. In any real system, there is always some damping $\zeta > 0$, no matter how small. Resonance in this case means that the displacement oscillations will build up to some large, but finite, steady-state amplitude. This is demonstrated in the responses shown in Figure 4.6.

Note that for large damping $\zeta \geq 0.5$, resonance is barely visible in magnitude frequency response plot, and indeed large damping inhibits resonance phenomena.

3. (Phase delay). The phase plot shows a negative phase shift from forcing to displacement at all frequencies. This means that the displacement response always "lags behind" the forcing. The

phase shift ranges from just below zero for low frequencies, which means that displacement is almost "in sync" with forcing if the forcing is sufficiently slow. As the forcing frequency becomes very high ($\Omega \to \infty$), there is an almost 180° phase delay from forcing to displacement.

Note that the phase shift is exactly $-90°$ at resonance $\Omega = 1$, and this is the case for any value of damping ζ. The meaning of this will become clearer when we study velocity and energy relations at resonance. This exact $-90°$ phase shift will be shown to mean that the "timing" of the forcing in this case produces the largest energy transfer from the applied force to the mass' motion.

Along with the frequency response plots of Figure 4.5, it is useful to examine the time-response plots shown in Figure 4.6. The system is forced at three different frequencies, at half resonance $\omega = \omega_n/2$, full resonance $\omega = \omega_n$, and at twice the resonance frequency $\omega = 2\omega_n$. The reader should examine those figures and their captions carefully to see the relations between the frequency-response plots and the time-response plots at various forcing frequencies. The basic idea is that all the features of the time-response plots (such as amplitude gain and phase/time shifts) are predictable from the frequency-response plots if interpreted correctly.

4.2 Energy and Power in Periodic Motion

To understand the phenomenon of resonance more thoroughly, it is useful to examine the energy flow in this system. There are 3 different forces acting on the mass

$$\begin{array}{lcll} f_e(t) & & : & \text{external force applied to mass} \\ f_s(t) & = & -kx(t) & : \text{ spring force} \\ f_d(t) & = & -c\dot{x}(t) & : \text{ damper's force} \end{array}$$

Note that we now use the term "external" and the notation f_e for the input force f in (4.1) to distinguish it from the "internal" forces f_s and f_d.

Now let $F(t)$ designate any one of those three forces. They each perform work on the mass. At any instant of time, the instantaneous power (time-rate of work) delivered *to the mass* by the force $F(t)$ is the product of the mass' velocity and the force applied to it. When both $v(t)$ and $F(t)$ are periodic with the same period T, the average power is defined as the normalized integral of the instantaneous power over any one cycle[5]

$$\underbrace{P(t) := v(t)F(t)}_{\text{instantaneous power}}, \qquad P_{av} := \frac{1}{T} \underbrace{\int_t^{t+T} v(\tau)F(\tau)\, d\tau}_{\substack{\text{energy delivered to mass} \\ \text{in one period}}}, \qquad v(t) := \dot{x}(t). \tag{4.15}$$

Over any one cycle, instantaneous power $P(t)$ can be positive at certain times and negative at others. The average power can also have either sign. When P_{av} is positive, the force is delivering net energy to the mass, and when it is negative, the force is extracting net energy from the mass. We will use the notation P^e_{av}, P^s_{av} and P^d_{av} for the average power delivered by the external f_e, spring f_s and damper's f_d forces respectively.

When a single mass is undergoing *periodic* motion (e.g. such as the forced Mass-Spring-Damper system in steady state), there is a conservation law of net energy or equivalently average power. For the mass subject to the three forces f_e, f_s and f_d it can be shown (Theorem 4.1 below) that the sum of average powers is zero

$$P^e_{av} + P^s_{av} + P^d_{av} = 0. \tag{4.16}$$

Since the spring force is a conservative force, the net power delivered by it to the mass $P^s_{av} = 0$ is zero. This then implies that the relation (4.16) reduces to

$$P^e_{av} = -P^d_{av},$$

[5] Note that since both $v(t)$ and $f(t)$ are T-periodic, then so is their product $v(t)f(t)$, and the integral in (4.15) is independent of the starting point t, and thus average power P_{av} does not depend on t.

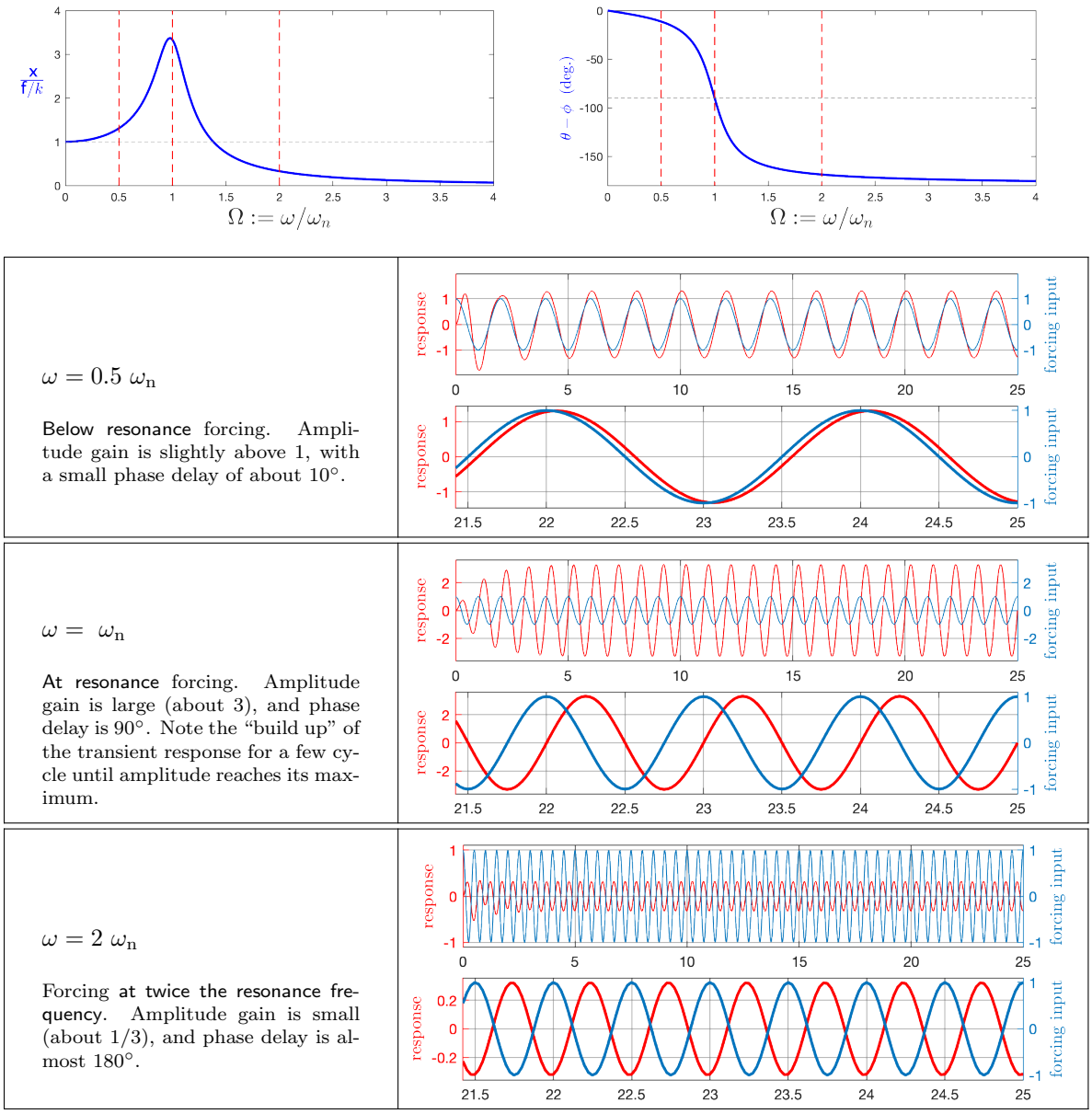


Figure 4.6: The time responses of a Mass-Spring-Damper system with $\zeta = 0.15$. Three forcing frequency scenarios of below, at, and above resonance are shown. The top panel shows the frequency-response plots of this system, with the red dashed lines marking the three frequencies of forcing used for the time-response plots in the panel below. In each of the three cases, the forcing has the same amplitude $\mathsf{f} = 1$, but the response steady-state amplitude varies depending on the frequency of the forcing.

i.e. the net power delivered by the external force to the mass is equal to the net power delivered by the mass to the damper. Thus in steady state, the externally supplied power is all used to counteract the power dissipated in the damper. On the other hand, the power exchange between the mass and the spring has a net value of zero over any one cycle.

The conservation law (4.16) follows from applying Newton's 2nd law to a single mass undergoing periodic motion. This is a special case of the following more general statement.

Theorem 4.1. *Let m be a single mass moving in one coordinate labeled x, and subject to n time-varying forces $f_1(t), \ldots, f_n(t)$ where the mass' position $x(t)$ and all forces are periodic with a common period. If $P^{\mathrm{k}}_{\mathrm{av}}$ denotes the average power supplied by each force f_k, then the sum total of average*

power provided by all forces is zero $\sum_{k=1}^{n} P_{\text{av}}^{\text{k}} = 0$.

The theorem can be understood to mean that under periodic motion, if some forces deliver net power to the mass, that same amount of power must be extracted from the mass by other forces. In the case of the MSD system, the external force delivers net power, while all of that power is extracted from the mass by the damper.

We next specialize to the case when the motion is not just periodic, but also *a sinusoid of a single frequency.* Much more can then be said about power flows, and we will see that relative phases play a central role in them.

4.2.1 Energy, Power and Velocity in Sinusoidal Steady State

In "steady state" (i.e. as $t \to \infty$) all signals in the Spring-Mass-Damper system are sinusoids of the same frequency. In particular $x(t)$ is a sinusoid of the same frequency as $f_{\text{e}}(t)$, and therefore also the derivative $\dot{x}(t)$ is a sinusoid of the same frequency as $f_{\text{e}}(t)$ and $x(t)$. Thus in steady state, all quantities listed above are sinusoids of the same frequency as the forcing $f_{\text{e}}(t)$, but they each have their own amplitude and phase. The average power can be quantified in terms of the amplitudes and relative phases of $v(t)$ and any force $F(t)$ as follows

$$
\begin{aligned}
& F(t) = \mathsf{F}\cos(\omega t+\phi), \qquad v(t) = \mathsf{v}\cos(\omega t+\theta) \\
\Rightarrow \quad & P_{\text{av}} = \frac{1}{T}\int_t^{t+T} \mathsf{F}\cos(\omega\tau+\phi)\ \mathsf{v}\cos(\omega\tau+\theta)\ d\tau \\
& = \frac{\mathsf{vF}}{2T}\int_t^{t+T} \big(\cos(2\omega\tau+\theta+\phi)+\cos(\theta\text{-}\phi)\big)d\tau \qquad \big(\cos(\phi_1)\cos(\phi_2) = \big(\cos(\phi_1+\phi_2)+\cos(\phi_1\text{-}\phi_2)\big)/2\big) \\
& = \frac{\mathsf{vF}}{2T}\int_t^{t+T} \cos(\theta-\phi)d\tau \qquad \left(\begin{array}{c}\text{the integral of a sinusoid over any multiple of its}\\ \text{period is always zero: } \int_t^{t+T}\cos\left(2\frac{2\pi}{T}\tau+\beta\right)\ d\tau = 0\end{array}\right) \\
& = \frac{\mathsf{vF}}{2T}\cos(\theta-\phi)\int_t^{t+T} d\tau \\
\Rightarrow \quad & \boxed{P_{\text{av}} = \tfrac{1}{2}\mathsf{vF}\cos(\theta-\phi).} \qquad (4.17)
\end{aligned}
$$

Thus power is a function of not only the amplitudes v and F (as would be expected), but also of the relative phasing between velocity and force. It is maximum when $\theta = \phi$, and equals its minimum, zero net power, when force and velocity are 90° out of phase. Figure 4.7 illustrates graphically the relationships between velocity, force and power for various phasings.

The formula (4.17) is expressed in terms of the amplitudes v, F and phases θ,ϕ of the signals $v(t)$ and $F(t)$ respectively. It can also be equivalently written in terms of their phasors as follows

$$
\begin{aligned}
P_{\text{av}} &= \tfrac{1}{2}\ \mathsf{vF}\ \cos(\theta-\phi) = \tfrac{1}{2}\ \mathbb{R}\Big(\mathsf{vF}\ e^{j(\theta-\phi)}\Big) = \tfrac{1}{2}\ \mathbb{R}\big(\mathsf{v}e^{j\theta}\ \mathsf{F}e^{-j\phi}\big) \\
&= \tfrac{1}{2}\ \mathbb{R}\Big(\mathsf{v}e^{j\theta}\ \big(\mathsf{F}e^{j\phi}\big)^*\Big) \qquad \Big(\text{using } e^{-j\phi} = \big(e^{j\phi}\big)^*\Big) \\
\Rightarrow \quad & \boxed{P_{\text{av}} = \tfrac{1}{2}\ \mathbb{R}\Big(\hat{\mathsf{v}}\ \hat{\mathsf{F}}^*\Big).} \qquad (4.18)
\end{aligned}
$$

Thus average power can also be obtained by taking the real part of the product of the two complex numbers $\hat{\mathsf{v}}$, the phasor of $v(t)$, and $\hat{\mathsf{F}}^*$, the *complex conjugate* of the phasor $\hat{\mathsf{F}}$ of $F(t)$. Note that $\Re(\hat{\mathsf{v}}\hat{\mathsf{F}}^*) = \mathsf{vF}\cos(\theta-\phi)$ is also equal to the *dot product* of the two vectors $\hat{\mathsf{v}}$ and $\hat{\mathsf{F}}$ when viewed as vectors in the plane. When the two vectors are aligned (equivalently, the phasors are "in phase"), this quantity is maximal, and it is zero when the two vectors are orthogonal (equivalently, when the phasors are 90° out of phase).

Now let's examine the average power delivered to the mass by each of the three forces f_{s}, f_{d} and the external force f_{e}. Begin with the spring force. It is proportional to displacement, which is always 90° out of phase with velocity, i.e. f_{s} and v are always 90° out of phase, resulting in zero net

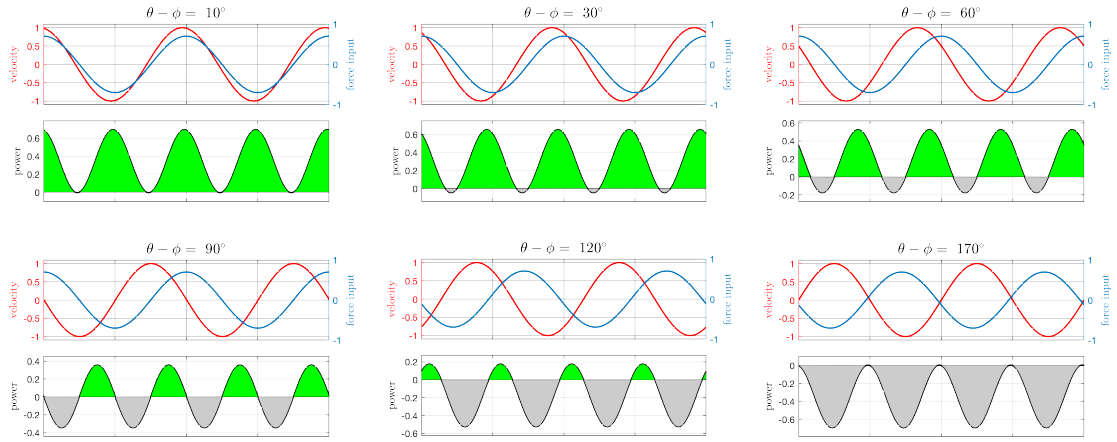


Figure 4.7: Various cases where velocity and force have relative phasings $\theta - \phi$ in the range 0° to 180°. The shaded areas represent the time-integral of power, i.e. energy, where green represents positive energy delivered to the mass by the force, and grey represents positive energy extracted from the mass to the applied force. The average power is the total area under the power curve in one cycle divided by the period. When velocity and force are in phase $\theta - \phi = 0$, maximum power is delivered to the mass. As the phase difference $\theta - \phi$ increases, the net power delivered is reduced. At $\theta - \phi = 90°$, the force does work on the mass in one half of the cycle, while the mass does an equal amount of work on the force in the other half of the cycle, resulting in a net power transfer of zero. For $90° < \theta - \phi \leq 180°$, the force extracts a positive net work from the mass rather than the other way around.

power flow

$$\begin{array}{lll} f_s(t) = -kx(t), & & \\ x(t),\ v(t) = \dot{x}(t) \text{ are } 90° \text{ out-of-phase} & \Rightarrow & f_s(t),\ v(t) = \dot{x}(t) \text{ are } 90° \text{ out-of-phase} \\ & \Rightarrow & P^s_{av} = 0. \end{array}$$

This is not surprising since the spring force is a conservative force. Note that over half the cycle, the spring does positive work on the mass, but over the other half of the cycle, the mass does an equal amount of work on the spring resulting in zero net work. This corresponds to the panel $\theta - \phi = 90°$ in Figure 4.7.

The damper's force on the other hand extracts power from the mass

$$\begin{array}{lll} f_d(t) = -c\dot{x}(t) = -cv(t), & \Rightarrow & f_d(t),\ v(t) \text{ are completely out-of-phase } (\theta - \phi = 180°) \\ & \Rightarrow & P^d_{av} = -\frac{1}{2}\ \mathsf{v}\ \mathsf{f}_d = -\frac{1}{2}\ \mathsf{v}\ (c\mathsf{v}) = -\frac{1}{2}\ c\ \mathsf{v}^2. \end{array}$$

This is again not too surprising since the damping force is a dissipative force. The average power extracted from the mass by the damper is proportional to the damping coefficient c and the square v^2 of the velocity amplitude. Note also that because $v(t)$ and $f_d(t)$ are 180° out of phase, the damper extracts power from the mass at all times, and at no time puts power back into the mass, unlike the spring force.

The conservation law (4.16) and the fact that average spring force power $P^s_{av} = 0$ is zero imply that the average power supplied by the external force must be exactly the power dissipated in the damper

$$P^e_{av} = -P^d_{av} = -\tfrac{1}{2}\ c\ \mathsf{v}^2. \tag{4.19}$$

This expression should be interpreted carefully. At first glance it seems that external power is proportional to the damping c. However, keep in mind that the velocity amplitude v also depends on c, so the dependence on the damping coefficient is not clear from this equation. In order to investigate this further, we should determine how v depends on both c and on the frequency of the input.

Root Mean Squared (RMS) "amplitude"

The formula (4.17) holds for sinusoidal functions with non-zero frequency. For *constant* (i.e. zero frequency) velocity and force the formula for average power is slightly different

$$\begin{array}{rl} \text{constant velocity} \\ \text{and force} \end{array} : \begin{array}{l} v(t) = \mathsf{v} \\ F(t) = \mathsf{F} \end{array} \quad \Rightarrow \quad P_{\text{av}} = \frac{1}{T}\int_{\bar{t}}^{\bar{t}+T} \mathsf{v}\ \mathsf{F}\ dt = \mathsf{vF}$$

$$\begin{array}{rl} \text{sinusoidal velocity} \\ \text{and force} \end{array} : \begin{array}{l} v(t) = \mathsf{v}\ \cos(\omega t + \phi) \\ F(t) = \mathsf{F}\ \cos(\omega t + \theta) \end{array} \quad \Rightarrow \quad P_{\text{av}} = \tfrac{1}{2}\mathsf{vF}\ \cos(\theta - \phi).$$

Note the factor of 2 difference between the two formulas. This is similar to the difference between AC and DC power formulas in electrical circuits. To avoid potential confusion (or maybe create more confusion), and have one single formula that applies to both oscillating and constant forces, it is common to define the Root Mean Square (RMS) *amplitude* of a signal $v(t)$ as follows

$$\mathsf{v}_{\text{rms}} := \sqrt{\frac{1}{T}\int_{\bar{t}}^{\bar{t}+T} v^2(t)\ dt}.$$

If $v(t)$ is a sinusoid of amplitude v, then its RMS amplitude is $\mathsf{v}/\sqrt{2}$ because

$$\begin{aligned} v(t) = \mathsf{v}\ \cos(\omega t) \quad \Rightarrow \quad \mathsf{v}_{\text{rms}}^2 &= \frac{1}{T}\int_0^T \mathsf{v}^2\left(\cos(\omega t)\right)^2\ dt = \mathsf{v}^2\ \frac{1}{T}\int_0^T \tfrac{1}{2}\left(1+\cos(2\omega t)\right)\ dt \\ &= \mathsf{v}^2\ \tfrac{1}{2}\ \frac{1}{T}\int_0^T dt = \frac{\mathsf{v}^2}{2} \quad \Rightarrow \quad \boxed{\mathsf{v}_{\text{rms}} = \mathsf{v}/\sqrt{2}}. \end{aligned}$$

In terms of the RMS amplitudes of sinusoidal forcing and velocity, the power formula (4.17) now becomes

$$\boxed{P_{\text{av}} = \mathsf{v}_{\text{rms}}\ \mathsf{F}_{\text{rms}}\ \cos(\theta - \phi) = \tfrac{1}{2}\ \mathsf{v}\ \mathsf{F}\ \cos(\theta - \phi)}.$$

Note that for constant signals, the RMS amplitude is just the regular amplitude, and therefore this single formula in terms of RMS amplitudes applies to *both sinusoidal or constant signals*. For this reason, RMS amplitudes of signals are used for power calculations. We should note however that this use more compelling for electrical circuits because they can operate in either DC or AC mode. In mechanical vibrations however, it is rare to consider constant forcing and velocity (no vibrations), thus the case for the use of RMS amplitudes for velocity and forcing is less compelling.

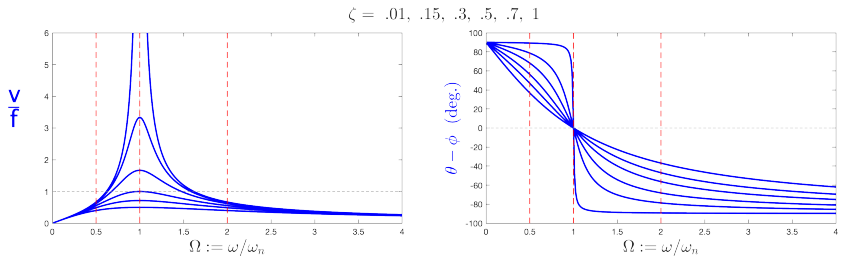
4.2.2 Velocity and Power Frequency Responses

To examine the dependence of $P_{\text{av}}^{\text{e}} = -P_{\text{av}}^{\text{d}}$ on forcing frequency, we need the relation between f_{e} and the mass' velocity $v = \dot{x}$. Again, phasor analysis makes this very easy. We already know the frequency response (4.13) from force to displacement. Since velocity is the time derivative of displacement, the velocity phasor is simply the displacement phasor multiplied by $j\omega$ (recall the the derivation (4.6) or the material in Section 2.3.1, which we now restate for the present case below)

$$\begin{aligned} & \dot{x}(t) =: v(t) && && \text{(velocity is the derivative of position)} \\ \text{if} \quad & x(t) = \mathsf{x}\ \cos(\omega t + \theta) && \longleftrightarrow \quad && \hat{\mathsf{x}} = \mathsf{x}\ e^{j\theta} \\ \text{then} \quad & v(t) = -\omega \mathsf{x}\ \sin(\omega t + \theta) \\ & \quad = \omega \mathsf{x}\ \cos(\omega t + \theta + 90^\circ) && \longleftrightarrow \quad && \hat{\mathsf{v}} = \omega \mathsf{x}\, e^{j(\theta+90^\circ)} = e^{j90^\circ}\omega\ \left(\mathsf{x}e^{j\theta}\right) = j\omega\ \hat{\mathsf{x}} \\ & && \Longrightarrow \quad && \boxed{\hat{\mathsf{v}} = j\omega\ \hat{\mathsf{x}}} \end{aligned}$$

We can now combine this with (4.13) to obtain the (complex) frequency response from f_{e} to v as

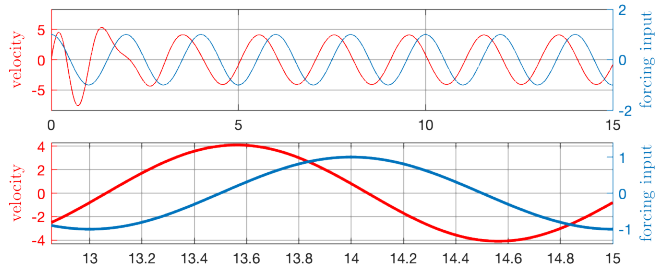
$$\begin{aligned} \frac{\hat{\mathsf{v}}}{\hat{\mathsf{f}}_{\text{e}}} &= j\omega\ \frac{\hat{\mathsf{x}}}{\hat{\mathsf{f}}_{\text{e}}} = j\omega\ \frac{1}{k}\ \frac{1}{\left((1-\Omega^2) + j\ 2\zeta\Omega\right)}, && \Omega := \omega/\omega_{\text{n}} \\ &= \frac{\omega_{\text{n}}}{k}\ \frac{j\ \omega/\omega_{\text{n}}}{\left((1-\Omega^2) + j\ 2\zeta\Omega\right)} = \frac{1}{\sqrt{km}}\ \frac{j\ \Omega}{\left((1-\Omega^2) + j\ 2\zeta\Omega\right)} && \omega_{\text{n}} = \sqrt{k/m} \end{aligned} \tag{4.20}$$



(a) The *velocity* frequency response of the Mass-Spring-Damper system for various values of the damping ratio ζ. The plots have many features that are similar to the displacement frequency response shown in Figure 4.5, namely resonance peaks at ω_n as well as the 180° change in phase around ω_n. However, the amplitude response goes to zero for $\omega \to 0$, and the phase plot goes from $+90°$ at low frequencies to $-90°$ at high frequencies. The three frequencies marked by dashed red lines are the frequencies used for the simulations shown in the panels below. The fact that the phase difference is always in the range $-90° \leq \theta - \phi \leq 90°$ means that net power delivered to the system is always positive. This is a consequence of the Mass-Spring-Damper system being a "passive" mechanical element, i.e. it can only absorb energy, it cannot create it.

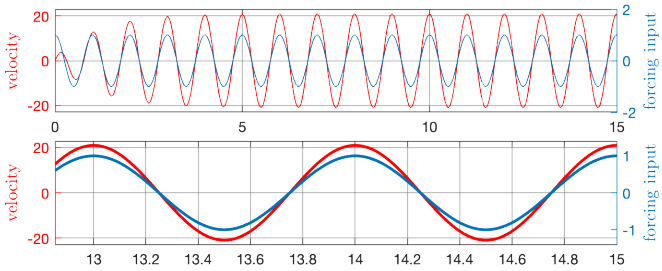
$\omega = 0.5\ \omega_n$

Below resonance forcing. Amplitude gain is about 1/4 of what it is at resonance (see below). Forcing and velocity are a little less than 90° out of phase, meaning that little net power is delivered to the mass.



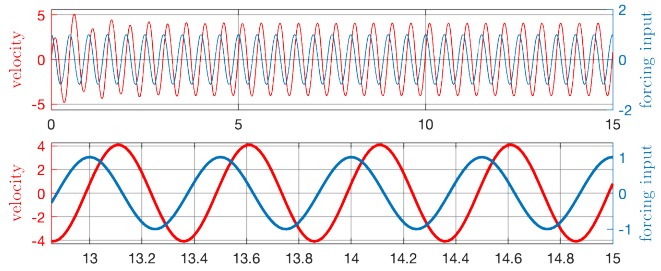
$\omega = \omega_n$

At resonance forcing. Amplitude gain is the largest (here approximately 20), and phase delay is 0°. Forcing and velocity are essentially completely in phase as predicted by the phase frequency response above. This is when maximum power is delivered to the mass.



$\omega = 2\ \omega_n$

Forcing at twice the resonance frequency. Amplitude gain is small (less than 1/4 of the maximum gain at resonance above). Phase delay is almost 90°, i.e. very little power is delivered to the mass.



(b) Forcing input and velocity output for three different forcing frequencies. The panels show the signals over a longer period of time as well as zoomed in views to highlight the phase shifts more easily. These simulations are done for the case $\zeta = 0.15$.

Figure 4.8: Frequency responses from-forcing-to-velocity of the Mass-Spring-Damper system, as well as three time response plots for below, at, and above resonance cases.

The magnitude and phase of this frequency response are shown in Figure 4.8 for various values of ζ. The magnitude frequency response is qualitatively similar to that for displacement with resonance peaks appearing around $\Omega = 1$. The main difference is that the response goes to zero for low forcing frequency $\Omega \to 0$. This is intuitive since for static loading, the steady state velocity is zero, and for very slowly varying forcing, the mass motion will also have very low velocity.

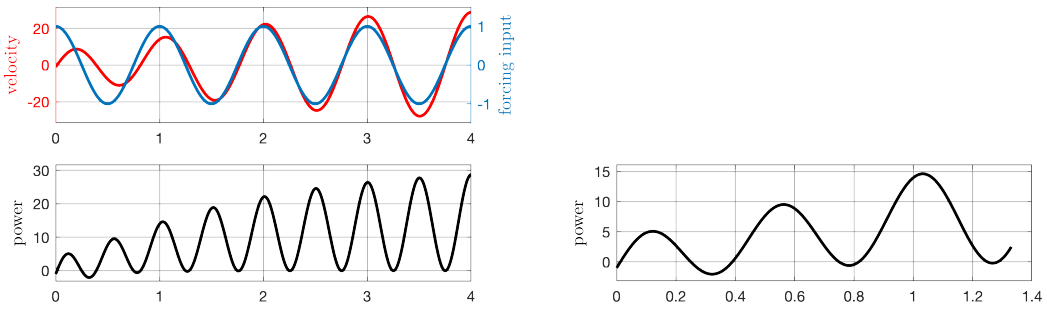


Figure 4.9: When a Mass-Spring-Damper system is forced at resonance, oscillations build up until they reach steady state. Forcing input and velocity response are completely in-phase at resonance, but that only occurs in steady state. Here the transient behavior is shown more clearly. It takes velocity a few cycles to get in-phase with forcing (*top*) during which average power increases with every cycle (*bottom left*). Even though in steady state instantaneous power is always non-negative (at resonance), it could be negative during certain parts of the cycle during the transients (*bottom right*) until the velocity locks in-phase with the forcing input. This simulation was started form non-zero initial conditions, and therefore there is a small mount of initially stored kinetic and potential energy in the Mass-Spring-Damper system. This results in a clearer transient exchange of energies between it and the applied force.

One notable difference between the displacement and velocity frequency responses is the phase plots. They are identical except for a 90° difference at all frequencies (again, this follows from $\hat{\mathsf{v}} = j\omega\hat{\mathsf{x}} \;\Rightarrow\; \angle\hat{\mathsf{v}} = \angle\hat{\mathsf{x}}+90°$, i.e. multiplying by j rotates a complex number by 90° counterclockwise). The velocity frequency response starts at 90° phase advance over forcing for low $\Omega \approx 0$ frequencies. It then decreases to exactly 0° at resonance $\Omega = 1$. This means that at resonance, forcing and velocity are perfectly in phase as shown in the time response plots of Figure 4.8. Recalling the power relation (4.17), we conclude that maximum power delivered to the mass by f_e occurs when forcing at resonance. At either low ($\Omega \to 0$) or high ($\Omega \to \infty$) forcing frequency, the phase difference approaches 90°, and the power relation (4.17) implies that very little average power is delivered to the mass. This is consistent with velocity oscillation amplitudes being close to zero in those frequency ranges. Note that the phase difference is always (at any frequency) in the interval $[-90°, 90°]$. The Mass-Spring-Damper is not an "active" mechanical element, it only absorbs energy, it cannot generate it, which means that the phase difference can never be more than 90° or less than $-90°$.

The previous statements are true under steady state conditions. Any mechanical system subject to forced vibrations exhibits some transients before it reaches steady state. Figure 4.9 shows the behavior of instantaneous power at initial times before steady state is reached. The main observation is that until velocity "locks into phase" with the applied forcing (which is what it does in steady state when forced at resonance), instantaneous power can be both positive or negative until steady state is reached, where it will always be non-negative.

Now we calculate P^{e}_{av} quantitatively as a function of frequency Ω. Equation (4.20) gives the phasor $\hat{\mathsf{v}}$ in terms of the phasor $\hat{\mathsf{f}}_{\text{e}}$, and Equation (4.18) gives average power in terms phasors. Combining the two

$$
\begin{aligned}
P^{\text{e}}_{\text{av}} \;=\; \tfrac{1}{2}\mathbb{R}\left(\hat{\mathsf{v}}\,\hat{\mathsf{f}}^{*}_{\text{e}}\right) \;&=\; \tfrac{1}{2}\,\mathbb{R}\left(\left(\frac{1}{\sqrt{km}}\,\frac{j\;\Omega}{\left((1-\Omega^2)+j\;2\zeta\Omega\right)}\;\hat{\mathsf{f}}_{\text{e}}\right)\hat{\mathsf{f}}^{*}_{\text{e}}\right) \\
&=\; \frac{\mathsf{f}^2_{\text{e}}\;\Omega}{2\sqrt{km}}\;\mathbb{R}\left(j\;\frac{1}{\left((1-\Omega^2)+j\;2\zeta\Omega\right)}\right) && \left(\text{since } \hat{\mathsf{f}}_{\text{e}}\hat{\mathsf{f}}^{*}_{\text{e}} = |\hat{\mathsf{f}}_{\text{e}}|^2 = \mathsf{f}^2_{\text{e}}\right) \\
&=\; \frac{\mathsf{f}^2_{\text{e}}\;\Omega}{2\sqrt{km}}\;\frac{2\zeta\Omega}{\left(1-\Omega^2\right)^2+(2\zeta\Omega)^2} && \left(\mathbb{R}(j/z) = \mathbb{I}(z)\,/|z|^2\right) \\
\Rightarrow \qquad P^{\text{e}}_{\text{av}} \;&=\; \frac{\zeta}{\sqrt{km}}\;\frac{\Omega^2}{\left(1-\Omega^2\right)^2+4\zeta^2\Omega^2}\;\mathsf{f}^2_{\text{e}} && (4.21)
\end{aligned}
$$

Note the use of the following complex number identities which simplified the calculation above. For

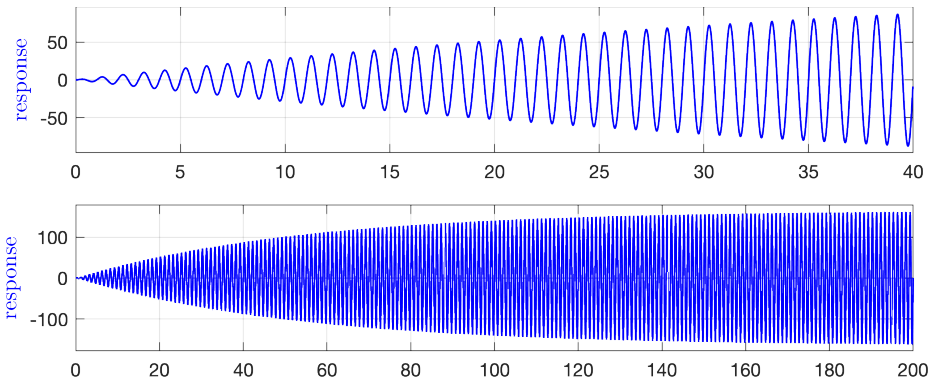


Figure 4.10: When a system with no damping $c = 0$ ($\zeta = 0$) is forced exactly at resonance $\omega = \omega_n$, the response (either displacement or velocity) is unbounded as $t \to \infty$. The amplitude grows linearly in time without bound. Therefore, there is *no steady state* in this case. In any real system however, there is always some damping $c > 0$ no matter how small. In the case of very small ζ, the response amplitude will grow linearly initially (*top*) , but then will eventually "saturate" to level out at some steady state amplitude (*bottom*). Note the different time spans between the two plots.

any complex number $z = \alpha + j\beta$

$$\mathbb{R}(jz) = \mathbb{R}(j(\alpha + j\beta)) = \mathbb{R}(j\alpha - \beta) = -\mathbb{I}(z)$$

$$zz^* = |z|^2 \qquad \Rightarrow \qquad \frac{1}{z} = \frac{z^*}{|z|^2} \qquad \Rightarrow \qquad \mathbb{R}\left(\frac{j}{z}\right) = -\mathbb{I}\left(\frac{1}{z}\right) = -\mathbb{I}\left(\frac{z^*}{|z|^2}\right) = \frac{-\mathbb{I}(z^*)}{|z|^2} = \frac{\mathbb{I}(z)}{|z|^2}.$$

The average externally supplied power (4.21) has a similar dependence on frequency as the velocity with some important differences. P^e_{av} goes to zero at low and high frequencies, and has a peak exactly at resonance. It is easy to verify by differentiation that the maximum power in (4.21) occurs at $\Omega = 1$, and therefore has value

$$\max P^e_{av} = \frac{1}{4\zeta\sqrt{km}}\,\mathsf{f}_e^2 = \frac{1}{4\frac{1}{2}\frac{c}{k}\sqrt{\frac{k}{m}}\sqrt{km}}\,\mathsf{f}_e^2 = \frac{1}{2c}\,\mathsf{f}_e^2. \qquad \text{(with forcing at resonance } \omega = \omega_n\text{)} \quad (4.22)$$

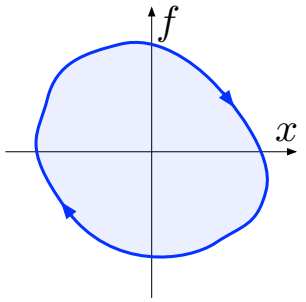
This expression implies that the lower the damping c, the higher the average external power transferred to the mass. This explains why the expression (4.19) is misleading since as c increases, the velocity amplitude v decreases[6].

The nonphysical case $c = 0$ deserves an explanation. The formula (4.22) seems to imply that maximum power flow into the system is infinite. The key to understanding this *apparent* paradox is that when $c = 0$, *there is no steady state*. In fact the response of the system grows unboundedly in this case. Recall the statement earlier that under the assumptions $k, m, c > 0$, the system has a steady state response. This assumption is violated when $c = 0$. Figure 4.10 shows the displacement response when a system with no damping $c = 0$ (equivalently $\zeta = 0$) is forced at exactly the resonance frequency. At every cycle, the external force transfers energy to the system to keep the oscillation magnitude growing unboundedly as $t \to \infty$. Any real system has some amount of damping $c > 0$, no matter how small, and thus this process of energy absorption will eventually reach an equilibrium where the average power delivered by the input is balanced by the power dissipated in the damper, and the amplitude of oscillations will reach some (possibly large) limit.

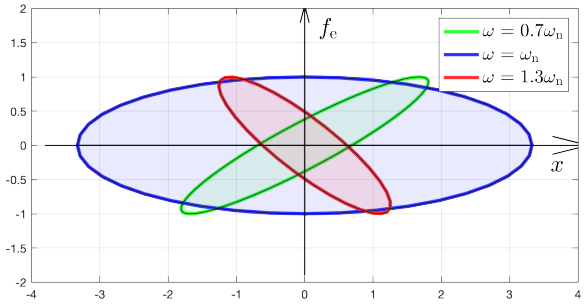
4.2.3 Energy from Force-Displacement Diagrams

Another way to visualize the work done in a system undergoing periodic motion is through force-versus-displacement diagrams. If an applied force $f(t)$ on a mass and the resulting displacement

[6]It can be seen from (4.20) that at resonance $\Omega = 1$, the velocity amplitude is proportional to $1/c$, i.e. v^2 is proportional to $1/c^2$, and therefore the expression $c\mathsf{v}^2$ in (4.19) is actually proportional to $1/c$, just like the expression above for maximum power at resonance.



(a) Under periodic motion, the force versus displacement trajectory $(f(t), x(t))$ forms a *closed curve* in a plane. The area enclosed clockwise by the curve is the total positive work done by the force over one cycle.



(b) Force-displacement curves for the harmonically forced Mass-Spring-Damper system forced at three different frequencies. Since $f(t)$ and $x(t)$ are sinusoids of the same frequency, the curves are all ellipses. A forcing amplitude of 1 is used in all three cases for comparison. The highest work done in one cycle is when forcing is at resonance $\omega = \omega_n$.

Figure 4.11: For a system undergoing periodic motion, the work done in one cycle can be visualized using a closed curve in a force-versus-displacement graph.

$x(t)$ are periodic with a common period, then plotting $f(t)$ versus $x(t)$ in the plane over any one period will always yield a *closed curve*. This is illustrated in Figure 4.11a. The work done by the force is the integral of force times incremental displacement

$$\text{work in one cycle} = \oint f\,dx \overset{1}{=} \int\int_{\mathcal{A}} dx\,df = \text{area of region } \mathcal{A} \text{ enclosed (clockwise) by } (x(t), f(t)) \text{ curve,}$$

where $\overset{1}{=}$ follows from Green's theorem. Thus we can visualize the work done as the enclosed area. Note that if the curve is oriented clockwise, then the work is positive, and negative if the curve is oriented counterclockwise.

For the externally forced Mass-Spring-Damper system in steady state, both $f(t)$ and $x(t)$ are sinusoids of the same frequency. It is not difficult to show that for such a case, the force-versus-displacement curve is the boundary of an ellipse. The orientation, eccentricity and enclosed area of the ellipse depend on the amplitudes and relative phases of $f(t)$ and $x(t)$. Figure 4.11b shows three such cases for forcing below, at and above resonance. The figure shows that the largest enclosed area, which represents the highest amount of work done over one cycle, corresponds to forcing at resonance.

4.3 Phasor Analysis of General Systems

The methods of the previous section are quite powerful and generalize to a much larger class of systems that the single Mass-Spring-Damper. In this section we develop the general method for any system that is described by a constant-coefficient ODEs with a harmonic input. These could be mechanical, electrical or other types of systems.

Consider the following p'th order differential equation for a function $y(t)$

$$\begin{aligned}\alpha_p\, y^{(p)}(t) \;+\; \alpha_{p\text{-}1}\, y^{(p\text{-}1)}(t) \;+\; \cdots \;+\; \alpha_1\, y^{(1)}(t) \;+\; \alpha_0\, y(t) \\ = \;\beta_q\, u^{(q)}(t) \;+\; \beta_{q\text{-}1}\, u^{(q\text{-}1)}(t) \;+\; \cdots \;+\; \beta_1\, u^{(1)}(t) \;+\; \beta_0\, u(t),\end{aligned} \tag{4.23}$$

where $u(t)$ is a harmonic forcing function (i.e. a sinusoid of some frequency, amplitude and phase). The numbers p and q are the highest order of differentiation that appears in the equation for x and f respectively. We typically assume $q \le p$, and call p the "order" of the system. The notation $y^{(k)}(t)$ stands for the k'th derivative of $y(t)$, and similarly for $u(t)$. Note also how the indices of the coefficients $\{\alpha_k\}$ and $\{\beta_k\}$ are chosen to match the order of differentiation for each term. For example, the equation (4.1)

$$m\,\ddot{x}(t) \;+\; c\,\dot{x}(t) \;+\; k\,x(t) \;=\; f(t)$$

is a special case of (4.23). We see that the order of (4.1) is $p = 2$, and $q = 0$ since there are no differentiations of the input f, and therefore this is 2nd order system. In this case $\alpha_2 = m$, $\alpha_1 = c$ and $\alpha_0 = k$. For the right hand side, $\beta_0 = 1$ and all other β coefficients are zero.

Equation (4.23) is general enough to model an incredibly large variety of physical dynamical systems. In AC electrical circuits, u could be a harmonic input voltage, and the equation can model a circuit with a large number of RLC components. In vibrations, u could be some applied harmonic force or the displacement of some mechanical element, and the equation models the dynamical response of a complex interconnection of masses, spring and damping elements. The larger the number of elements, the larger the order p of the equation. Typically, the order p equals twice the number of mass elements in the system. For circuits, p equals the total number of capacitors and inductors (the number of resistors does not contribute to the order p).

We now describe phasor analysis for the equation (4.23). Under the condition that its dynamics are stable, i.e. that transients decay in time, it is not difficult to show (as was seen in the Mass-Spring-Damper system) that if the forcing is a pure sinusoid of some frequency ω, then the *steady state response* x will also be a sinusoid of the same frequency ω, but with possibly a different amplitude and phase, i.e.

$$u(t) \;=\; \mathsf{u}\cos\big(\omega t+\phi\big) \qquad\qquad \Rightarrow \qquad\qquad \lim_{t\to\infty} y(t) \;=\; \mathsf{y}\cos\big(\omega t+\theta\big).$$

The task is to find an expression for the amplitude y and phase θ of the signal $y(t)$ as a function of the system (4.23) parameters and input frequency ω.

As in the Mass-Spring-Damper system case, we start with writing the forcing and steady-state response (we now refer to the latter as just $y(t)$) as the real parts of complex functions by

$$\begin{aligned} u(t) \;&=\; \mathsf{u}\cos\big(\omega t+\phi\big) \;=\; \mathbb{R}\big(\hat{\mathsf{u}}\; e^{j\omega t}\big)\,, \quad \text{where } \hat{\mathsf{u}} = \mathsf{u}\; e^{j\phi},\\ y(t) \;&=\; \mathsf{y}\cos\big(\omega t+\theta\big) \;=\; \mathbb{R}\big(\hat{\mathsf{y}}\; e^{j\omega t}\big)\,, \quad \text{where } \hat{\mathsf{y}} = \mathsf{y}\; e^{j\theta}. \end{aligned} \tag{4.24}$$

The idea of phasor analysis is to solve the differential equation for the complex-valued functions $\hat{\mathsf{u}}e^{j\omega t}$ and $\hat{\mathsf{y}}e^{j\omega t}$ instead of the real-valued ones. To justify this idea, we state it formally.

1. Substitute the complex-valued functions $u(t) = \hat{\mathsf{u}}\; e^{j\omega t}$ and $y(t) = \hat{\mathsf{y}}\; e^{j\omega t}$ in the differential equation (4.23).

2. From the resulting algebraic equation, find $\hat{\mathsf{y}}$ in terms of $\hat{\mathsf{u}}$.

3. It then follows that the real-valued functions $u(t) = \mathbb{R}\big(\hat{\mathsf{u}}\; e^{j\omega t}\big) = |\hat{\mathsf{u}}|\;\cos\big(\omega t + \angle\hat{\mathsf{u}}\big)$ and $y(t) = \mathbb{R}\big(\hat{\mathsf{y}}\; e^{j\omega t}\big) = |\hat{\mathsf{y}}|\;\cos\big(\omega t + \angle\hat{\mathsf{y}}\big)$ solve the same differential equation (4.23).

To justify that last statement, rewrite the differential equation in the following compact form and observe that if $u(t) = \hat{\mathsf{u}}\; e^{j\omega t}$ and $y(t) = \hat{\mathsf{y}}\; e^{j\omega t}$ satisfy the differential equation

$$\begin{aligned} &\sum_{n=0}^{p} \alpha_n\; \tfrac{d^n}{dt^n}\left(\hat{\mathsf{y}}\; e^{j\omega t}\right) = \sum_{m=0}^{q} \beta_m\; \tfrac{d^m}{dt^m}\left(\hat{\mathsf{u}}\; e^{j\omega t}\right)\\ \Rightarrow \qquad &\mathbb{R}\left(\sum_{n=0}^{p} \alpha_n\; \tfrac{d^n}{dt^n}\left(\hat{\mathsf{y}}\; e^{j\omega t}\right)\right) = \mathbb{R}\left(\sum_{m=0}^{q} \beta_m\; \tfrac{d^m}{dt^m}\left(\hat{\mathsf{u}}\; e^{j\omega t}\right)\right)\\ \Rightarrow \qquad &\sum_{n=0}^{p} \alpha_n\; \tfrac{d^n}{dt^n}\mathbb{R}\left(\hat{\mathsf{y}}\; e^{j\omega t}\right) = \sum_{m=0}^{q} \beta_m\; \tfrac{d^m}{dt^m}\mathbb{R}\left(\hat{\mathsf{u}}\; e^{j\omega t}\right), \end{aligned}$$

where the last implication follows from the facts that the real part of a sum is the sum of real parts, the coefficients α_n, β_m are real so e.g. $\mathbb{R}(\alpha_n\; f(t)) = \alpha_n\; \mathbb{R}(f(t))$, and finally that the real part of a derivative is the derivative of the real part.

We now follow this procedure to derive the algebraic equation relating $\hat{\mathsf{u}}$ and $\hat{\mathsf{y}}$. Substituting $u(t) = \hat{\mathsf{u}}\; e^{j\omega t}$ and $y(t) = \hat{\mathsf{y}}\; e^{j\omega t}$ in the differential equation (4.23) and using the fact $\frac{d^k}{dt^k} e^{j\omega t} =$

$(j\omega)^k e^{j\omega t}$ gives

$$\begin{aligned} \alpha_p \tfrac{d^p}{dt^p}\left(\hat{\mathsf{y}}\, e^{j\omega t}\right) + \cdots + \alpha_0\left(\hat{\mathsf{y}}\, e^{j\omega t}\right) &= \beta_q \tfrac{d^q}{dt^q}\left(\hat{\mathsf{u}}\, e^{j\omega t}\right) + \cdots + \beta_0\left(\hat{\mathsf{u}}\, e^{j\omega t}\right) \\ \Rightarrow \quad \alpha_p \hat{\mathsf{y}}\,(j\omega)^p\, e^{j\omega t} + \cdots + \alpha_0 \hat{\mathsf{y}}\, e^{j\omega t} &= \beta_q \hat{\mathsf{u}}\,(j\omega)^q\, e^{j\omega t} + \cdots + \beta_0 \hat{\mathsf{u}}\, e^{j\omega t} \\ \Rightarrow \quad \Big(\alpha_p\,(j\omega)^p + \cdots + \alpha_0\Big)\,\hat{\mathsf{y}}\, e^{j\omega t} &= \Big(\beta_q\,(j\omega)^q + \cdots + \beta_0\Big)\,\hat{\mathsf{u}}\, e^{j\omega t} \\ \Rightarrow \quad \hat{\mathsf{y}} &= \frac{\beta_q\,(j\omega)^q + \cdots + \beta_1\,(j\omega) + \beta_0}{\alpha_p\,(j\omega)^p + \cdots + \alpha_1\,(j\omega) + \alpha_0}\,\hat{\mathsf{u}} \end{aligned}$$

Note again that the appearance of $e^{j\omega t}$ on both sides of the equation (since u and y have the same frequency) allows for the removal of that factor, resulting finally in a *purely algebraic relation*. We summarize this formally in the next statement.

Theorem 4.2. *Consider the following input-output ODE with constant coefficients*

$$\begin{aligned} \alpha_p\, y^{(p)}(t) \;+\; \alpha_{p\text{-}1}\, y^{(p-1)}(t) \;+\; \cdots \;+\; \alpha_1\, y^{(1)}(t) \;+\; \alpha_0\, y(t) \\ = \;\beta_q\, u^{(q)}(t) \;+\; \beta_{q\text{-}1}\, u^{(q-1)}(t) \;+\; \cdots \;+\; \beta_1\, u^{(1)}(t) \;+\; \beta_0\, u(t), \end{aligned} \tag{4.25}$$

where $\{\alpha_k\}$ and $\{\beta_k\}$ are real coefficients. If the input is a pure harmonic of the form $u(t) = \mathsf{u}\cos\big(\omega t + \phi\big)$, then the steady-state output is also a pure harmonic with the same frequency

$$\lim_{t\to\infty} y(t) \;=\; \mathsf{y}\cos\big(\omega t + \theta\big),$$

where the output's amplitude y *and its phase θ are given by*

$$\left.\begin{aligned} \mathsf{y} &= |\mathsf{H}(\omega)|\;\mathsf{u}, \quad \text{(amplitude amplification)}, \\ \theta &= \angle\mathsf{H}(\omega) \;+\; \phi, \quad \text{(phase shift)}, \end{aligned}\right\} \quad \Leftrightarrow \quad \hat{\mathsf{y}} \;=\; \mathsf{H}(\omega)\;\hat{\mathsf{u}}, \tag{4.26}$$

where $\mathsf{H}(\omega)$ *is the* complex frequency response *of the system (4.25) defined by*

$$\mathsf{H}(\omega) \;:=\; \frac{\beta_q(j\omega)^q \;+\; \cdots \;+\; \beta_1(j\omega) \;+\; \beta_0}{\alpha_p(j\omega)^p \;+\; \cdots \;+\; \alpha_1(j\omega) \;+\; \alpha_0}. \tag{4.27}$$

The non-negative real-valued function $|\mathsf{H}(\omega)|$ is called the amplitude frequency response*, and the function $\angle\mathsf{H}(\omega)$ is called the* phase frequency response*.*

The frequency response $\mathsf{H}(\omega)$ as a function of ω is completely determined by the coefficients $\alpha_0, \ldots, \alpha_p$ and $\beta_0, \ldots, \beta_q$ of the differential equation. In other words, the frequency response function $\mathsf{H}(\omega)$ can be"read off" directly from the differential equation (4.25). The quantities $|\mathsf{H}(\omega)|$ and $\angle\mathsf{H}(\omega)$ are called the *magnitude frequency response* and the *phase frequency response* respectively. Figure 4.5 derived earlier is a specific example of such responses. Observe how elegant and simple the relations (4.26) are. They state that at each input frequency ω, the magnitude $|\mathsf{H}(\omega)|$ of the complex number $\mathsf{H}(\omega)$ is the "amplitude amplification" from input to output, while its phase $\angle\mathsf{H}(\omega)$ is the "phase shift" from input to output.

The significance of Theorem 4.2 cannot be overstated. The system (4.25) is originally described as a differential equation with a forcing "input". This is called the time domain description of the system. When the inputs are sinusoids, the relation between the phasors of the input and output is the very simple algebraic relation (product of complex numbers)

$$\underbrace{\hat{\mathsf{y}}}_{\substack{\text{phasor of}\\ \text{output}}} \quad = \quad \underbrace{\mathsf{H}(\omega)}_{\substack{\text{complex}\\ \text{frequency response}}} \qquad \underbrace{\hat{\mathsf{u}}}_{\substack{\text{phasor of}\\ \text{input}}}$$

This is called the frequency domain description of the system. Clearly a much simpler relation than solving differential equations in the time domain. The time domain versus frequency domain descriptions are illustrated in Figure 4.12.

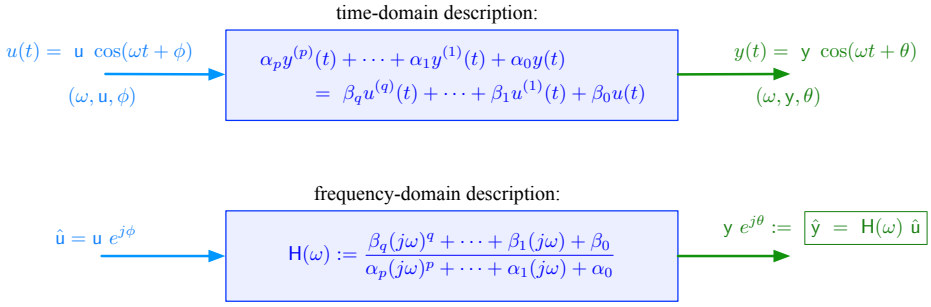
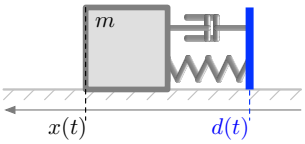


Figure 4.12: (*Top*) Any ODE with an input u describes an "input-output" system since for any function $u(t)$ there is a unique steady-state solution $y(t)$. Thus $u(t)$ can be considered an input, and for each such input there is a unique output, namely the solution $y(t)$ of the differential equation. The differential equation can thus be thought of as an "algorithm" for generating $y(t)$ given $u(t)$. (*Bottom*) When the equation has constant coefficients and we are concerned with only steady-state outputs (i.e. long-time behavior), effects of initial conditions disappear, and the relations between the input and output phasors ($\hat{\mathsf{u}}$ and $\hat{\mathsf{y}}$ respectively) is the simple complex-number product $\hat{\mathsf{y}} = \mathsf{H}(\omega)\,\hat{\mathsf{u}}$. This is called the frequency domain description of the system, and the quantity $\mathsf{H}(\omega)$ as a function of frequency ω is called the (complex) frequency response. The frequency response function $\mathsf{H}(\omega)$ can be "read off" directly from the coefficients of the differential equation without the need for any calculations.

4.3.1 Mass-Spring-Damper with Displacement as an Input

As already mentioned, inputs are more general than just forces. We now use Theorem 4.2 directly to find the frequency response of the system shown in the figure on the right. Here $d(t)$ is the position (not a force) of a "displacer" attached to one end of a spring and damper whose other ends are attached to a mass. The mass' position is measured by $x(t)$. We want to find the effect of the motion $d(t)$ of the displacer on the motion $x(t)$ of the mass. Clearly as $d(t)$ moves relative to $x(t)$, the spring and damper are stretched or compressed, exerting forces on the mass. Thus as $d(t)$ varies, it produces forces on the mass, but indirectly, and those forces depend on $x(t)$ as well. We first derive a differential equation for this system.



Let ℓ be the equilibrium length of the spring, then

$$m\,\ddot{x}(t) \;= -k\big(x(t) - d(t) - \ell\big) - c\big(\dot{x}(t) - \dot{d}(t)\big).$$

The constant term ℓ is eliminated by redefining the origin of the coordinate system for x by

$$\tilde{x}(t) := x(t) - \ell.$$

Substituting $x(t) = \tilde{x}(t) + \ell$ in the original differential equation yields a differential equation for $\tilde{x}$

$$\begin{aligned} m\,\ddot{\tilde{x}}(t) \;&=\; -k\,\big(\tilde{x}(t) - d(t)\big) \;-\; c\big(\dot{\tilde{x}}(t) - \dot{d}(t)\big) \\ \Rightarrow \qquad m\,\ddot{\tilde{x}}(t) \;+\; c\,\dot{\tilde{x}}(t) \;+\; k\,\tilde{x}(t) \;&=\; c\,\dot{d}(t) + k\,d(t), \end{aligned} \tag{4.28}$$

where we rearranged the equation so that all terms that involve the "input" $d(t)$ are on one side, and the terms involving the "output" $\tilde{x}(t)$ are on the other side. This is in the standard general form (4.23) introduced earlier. Comparing this system with the Mass-Spring-Damper system (4.1) with an actual force as an input, we see a distinction that here the input $d(t)$ appears with its derivative $\dot{d}(t)$, while the input $f(t)$ in (4.1) is not differentiated. This has some important consequences for the shape of the frequency response in the high frequency regime as we will see shortly.

To find the frequency response, there is no need to repeat the phasor analysis for this case from scratch. We can simply apply Theorem 4.2 to the second order system (4.28) (note that it is of a form for which the theorem is applicable). It simplifies things a bit to first divide by the mass, and the notation is simpler if we relabel $\tilde{x}$ to x

$$\begin{aligned} \ddot{x}(t) \;+\; \tfrac{c}{m}\,\dot{x}(t) \;+\; \tfrac{k}{m}\,x(t) \;&=\; \tfrac{c}{m}\,\dot{d}(t) \;+\; \tfrac{k}{m}\,d(t) \\ \Leftrightarrow \qquad \ddot{x}(t) \;+\; 2\zeta\omega_{\mathrm{n}}\,\dot{x}(t) \;+\; \omega_{\mathrm{n}}^2\,x(t) \;&=\; 2\zeta\omega_{\mathrm{n}}\,\dot{d}(t) \;+\; \omega_{\mathrm{n}}^2\,d(t), \end{aligned} \tag{4.29}$$

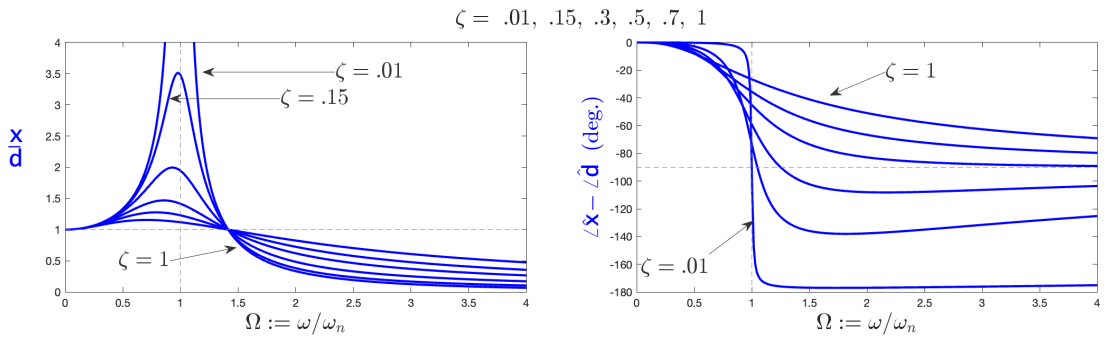
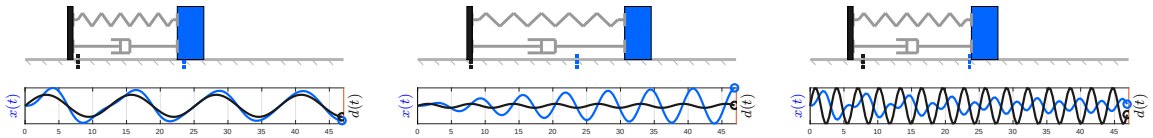


Figure 4.13: The frequency responses from displacer position $d(t)$ to mass position $x(t)$ of the system with displacement (rather than force) input. It shares many features of the force-input system's frequency response depicted in Figure 4.5 earlier. Namely that resonance occurs around $\Omega = 1$, with more pronounced resonance peaks for lower damping ratios ζ. The main difference is with the phase relations. In the high frequency region, the phase plots vary significantly depending on the damping ratio.



(a) At slow forcing frequencies the mass movement follows that of the displacer with approximately the same amplitude. The spring/damper link between the two appear to be almost like a rigid link.
Click for animation.

(b) Near resonance, the mass' oscillation amplitudes steadily grow until they are much larger than the displacer's oscillations.
Click for animation.

(c) At high forcing frequencies, the mass' oscillations are small relative to the displacer's. In this frequency regime, inertia dominates. The mass and displacer oscillations appear to be approximately "out of phase".
Click for animation.

Figure 4.14: Illustrations and animations of the motion of a displacement-input MSD system with input frequencies below ($\omega \approx 0.5\,\omega_n$), at ($\omega \approx \omega_n$), and above ($\omega \approx 1.5\,\omega_n$) the system's natural frequency ω_n (here $\omega_n = 1$ and $\zeta = 0.05$). All plots have the same axis limits for comparison, but the displacer's amplitudes are different for each plot. Note the relative amplitudes of displacer and mass in each case.

where ζ and ω_n are defined as before. Applying Theorem 4.2 to this last equation, the complex frequency response is given as

$$\frac{\hat{x}}{\hat{d}} = \frac{2\zeta\omega_n\,(j\omega) + \omega_n^2}{(j\omega)^2 + 2\zeta\omega_n\,(j\omega) + \omega_n^2} = \frac{j\;2\zeta\omega_n\;\omega + \omega_n^2}{(\omega_n^2 - \omega^2) + j\;2\zeta\omega_n\;\omega} = \frac{j\;2\zeta\;\omega/\omega_n + 1}{(1 - \omega^2/\omega_n^2) + j\;2\zeta\;\omega/\omega_n} \tag{4.30}$$

$$= \frac{j\;2\zeta\;\Omega + 1}{(1 - \Omega^2) + j\;2\zeta\;\Omega}, \qquad \text{(using normalized frequency } \Omega := \omega/\omega_n\text{)} \tag{4.31}$$

This is the frequency response from displacer to mass positions. A plot of its magnitude and phase is shown in Figure 4.13 for various values of ζ.

The behavior of this system can be understood by examining (4.31) in the low and high frequency regimes, as well as at resonance. For low Ω, the frequency response $\hat{x}/\hat{d} \approx 1$, which means that the mass' position follows that of the displacer. This is illustrated in Figure 4.14a. Thus in the low frequency regime, the spring/damper link between the displacer and mass acts almost like a "rigid link". Near resonance $\Omega \approx 1$, the expression (4.31) for the mass' phasor is

$$\hat{x} \;\approx\; \frac{j\;2\zeta + 1}{j\;2\zeta}\;\hat{u} \;\approx\; -\,j\;\frac{1}{2\zeta}\;\hat{u}, \qquad \text{when } \zeta \ll 1. \tag{4.32}$$

Thus the amplitude scales inversely with the damping ratio and can become arbitrarily large as $\zeta \to 0$. Because of the $-j$ factor, the mass' oscillations lag with 90° behind the displacer input. Figure 4.14b illustrates the resonant case and shows how the mass' oscillations build up to large amplitudes (relative to the displacer amplitude) until it reaches steady state.

When $d(t)$ oscillated rapidly (i.e. much faster than ω_n), then $\Omega \gg 1$, and the frequency response limits to

$$\frac{\hat{\mathsf{x}}}{\hat{\mathsf{d}}} = \frac{j\ 2\zeta\ \Omega + 1}{(1-\Omega^2) + j\ 2\zeta\ \Omega} \sim \frac{j\ 2\zeta\ \Omega}{-\Omega^2} = -j\,\frac{2\zeta}{\Omega}. \qquad \text{(when } \Omega \gg 1\text{)} \tag{4.33}$$

Thus for large Ω, the amplitude x is much smaller than the input amplitude d. This is again due to the mass' inertia which dominates the response in the high frequency regime. All of the phenomena mentioned so far are very similar to the case of force input analyzed earlier in Section 4.1. However, for this displacement input case we see somewhat different behavior in the high frequency region. The following two phenomena can be seen from (4.33) and the frequency response in Figure 4.13.

- *Effect of ζ on the high frequency behavior:* While a very lightly damped system ($\zeta \ll 1$) has higher resonance peaks near $\Omega \approx 1$, its high frequency response is actually *lower* with smaller ζ as evident from Equation (4.33) and Figure 4.13. This is in contrast to the high frequency behavior when the input is a force (see Figure 4.5b), which is largely independent of ζ. The effect of ζ on high frequency behavior will have important implications for the design of vibration isolation platforms as will be studied in Section 5.1.3.

- From (4.33) we see that for high $\Omega \gg 1$, the phase difference is about $\angle(-j) = -90°$ (unlike the force input case where the high frequency phase difference is 180°). Examining the phase frequency plot in Figure 4.13 confirms this, but we also see that the "rate" at which the phase approaches 90° for $\Omega \to \infty$ depends significantly on ζ. For lightly damped systems, the phase actually dips to almost 180° before it eventually limits to 90° as $\Omega \to \infty$.

Appendix

4.A Decomposition of Responses: Transients and Steady State

If you recall what you learned about Ordinary Differential Equations (ODEs) with forcing (the so-called non-homogenous ODEs)

$$m\,\ddot{x}(t) + c\,\dot{x}(t) + k\,x(t) = f(t), \qquad x(0) = a,\ \dot{x}(0) = b, \tag{4.34}$$

you may remember that the solution is obtained as the sum of two solutions, a homogenous solution and a particular solution. The terminology can be confusing because the role of initial conditions is not always explicitly stated, so let's state things precisely.

For each set of given initial conditions $x(0) = a,\ \dot{x}(0) = b$, Equation (4.34) has a unique solution. Thus if we vary the initial conditions over all possible numbers a and b, we get an infinite set of solutions parameterized by (a, b). Each such solution is called *a particular solution*. More formally

Definition 4.3. *A function $x_p(t)$ is called* a particular solution *of (4.34) if it is a solution for some pair of initial conditions a and b. Equivalently, if it satisfies (4.34) without regard to initial conditions, i.e. if*

$$m\,\ddot{x}_p(t) + c\,\dot{x}_p(t) + k\,x_p(t) = f(t).$$

Notice that we say *a particular solution* rather than *the particular solution* since there are many of them.

Now if we have two different particular solutions $x_{p,1}(t)$ and $x_{p,2}(t)$. They both satisfy the differential equation (4.34) each with different sets of initial conditions

$$\begin{aligned} m\,\ddot{x}_{p,1}(t) + c\,\dot{x}_{p,1}(t) + k\,x_{p,1}(t) &= f(t) \qquad & x_{p,1}(0) = a_1,\ \dot{x}_{p,1}(0) = b_1, \\ m\,\ddot{x}_{p,2}(t) + c\,\dot{x}_{p,2}(t) + k\,x_{p,2}(t) &= f(t) \qquad & x_{p,2}(0) = a_2,\ \dot{x}_{p,2}(0) = b_2. \end{aligned}$$

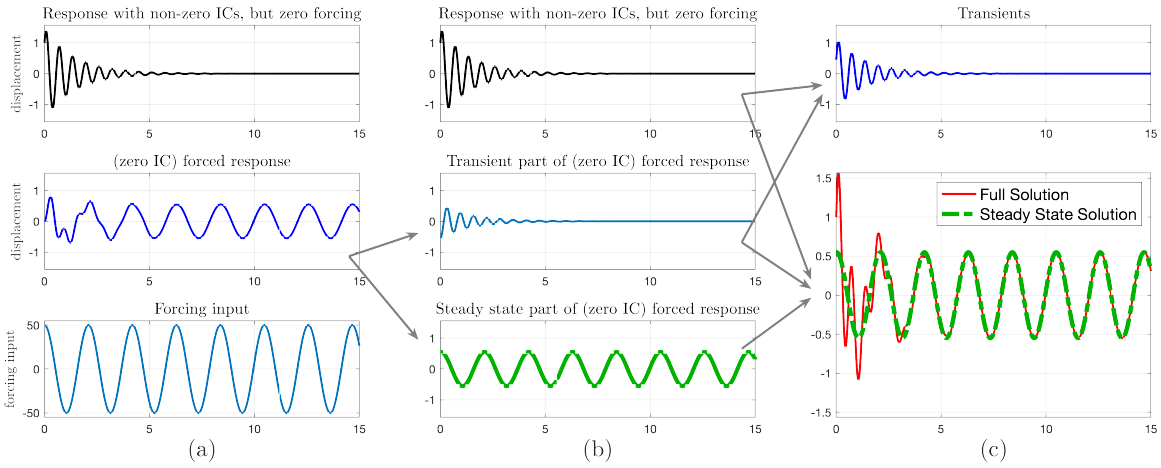


Figure 4.15: Decomposition of the full response (solution) of the Mass-Spring-Damper system into various pieces. (a) Shows the response due to Initial Conditions (ICs), the response due to forcing (with zero ICs), and the forcing input. (b) The response due to forcing can be further decomposed into a transient (middle) and a steady state (bottom) part. (c) The full solution and the steady-state solution are shown (bottom). The difference of those two functions would be the transient (top), which is the sum of the two transients in (b).

Because these equations are *linear*, we can subtract them and obtain a so-called homogenous equation (i.e. with no forcing)

$$m\Big(\ddot{x}_{\mathrm{p},1}(t)-\ddot{x}_{\mathrm{p},2}(t)\Big)+c\Big(\dot{x}_{\mathrm{p},1}(t)-\dot{x}_{\mathrm{p},2}(t)\Big)+k\Big(x_{\mathrm{p},1}(t)-x_{\mathrm{p},2}(t)\Big) \;=\; 0, \qquad \begin{aligned} x_{\mathrm{p},1}(0)-x_{\mathrm{p},1}(0) &= a_1-a_2,\\ \dot{x}_{\mathrm{p},1}(0)-\dot{x}_{\mathrm{p},2}(0) &= b_1-b_2, \end{aligned}$$

Thus any two particular solutions of (4.34) differ by a solution $x_{\mathrm{h}}(t)$ of the homogenous equation

$$m\;\ddot{x}_{\mathrm{h}}(t) \;+\; c\;\dot{x}_{\mathrm{h}}(t) \;+\; k\;x_{\mathrm{h}}(t) \;=\; 0, \qquad\qquad x_{\mathrm{h}}(0)=a_1-a_2,\;\; \dot{x}_{\mathrm{h}}(0)=b_1-b_2.$$

Another way to say this is that the solution can be written as the sum of a particular solution and a homogenous solution, provided the initial conditions satisfy a certain relationship. More precisely

Theorem 4.4. *Let $x(t)$ be the unique solution of the differential equation (4.34) for some given forcing function $f(t)$ and initial conditions a and b. $x(t)$ can be written as the sum of two functions $x(t)=x_{\mathrm{p}}(t)+x_{\mathrm{h}}(t)$, each of which solve the following differential equations*

$$m\;\ddot{x}_{\mathrm{h}}(t) \;+\; c\;\dot{x}_{\mathrm{h}}(t) \;+\; k\;x_{\mathrm{h}}(t) \;=\; 0, \qquad\qquad x_{\mathrm{h}}(0)=a_{\mathrm{h}},\;\; \dot{x}_{\mathrm{h}}(0)=b_{\mathrm{h}}, \tag{4.35}$$

$$m\;\ddot{x}_{\mathrm{p}}(t) \;+\; c\;\dot{x}_{\mathrm{p}}(t) \;+\; k\;x_{\mathrm{p}}(t) \;=\; f(t), \qquad\qquad x_{p}(0)=a_{\mathrm{p}},\;\; \dot{x}_{\mathrm{p}}(0)=b_{\mathrm{p}}, \tag{4.36}$$

provided that $a_{\mathrm{h}}+a_{\mathrm{p}}=a$ and $b_{\mathrm{h}}+b_{\mathrm{p}}=b$.

Therefore the unique solution can be decomposed into pairs of homogenous and particular solutions in many different ways. The utility of this theorem is that certain particular solutions are easy to obtain, as in the case of the "steady-state" solution derived in Section 4.1. Once an easy particular solution $x_{\mathrm{p}}(t)$ has been found, we can find the initial conditions of its corresponding homogenous problem by

$$a_{\mathrm{h}}=a-x_{\mathrm{p}}(0), \qquad b_{\mathrm{h}}=b-\dot{x}_{\mathrm{p}}(0),$$

where a and b are the given initial conditions of the problem. The homogenous problem (4.35) can then be solved.

In problems of linear dynamical systems such as those that occur in vibrations and AC circuit analysis, there are two different decompositions into particular and homogenous solutions that are used. The decomposition we already encountered in Figure 4.1 is a decomposition into a *steady-state*

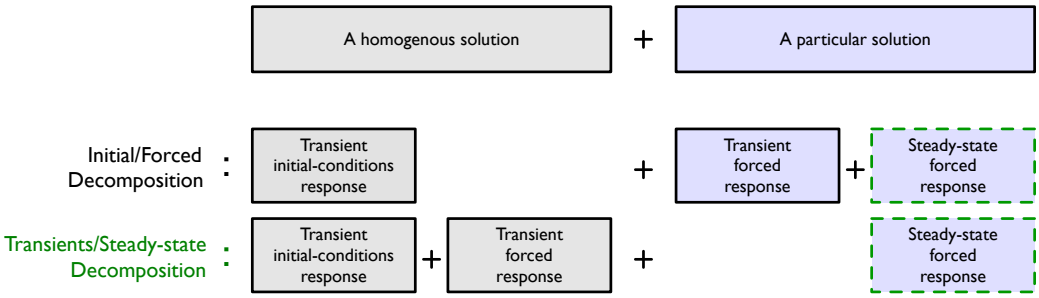


Figure 4.16: The solution of any ODE with given initial conditions can be decomposed into "a homogenous solution" and "a particular solution" in an infinite number of ways. Here we depict two of the most common decompositions. The decomposition used in this chapter is where the steady-state solution is the particular solution, while transients of two types are combined together as the homogenous solution. The utility of this decomposition is that initial conditions do not effect the steady-state solution, and can therefore be ignored to greatly simplify the calculations.

and *transient solution*. Since we were only interested in the steady-state solution, we didn't have to compute the corresponding initial conditions for either. Another commonly used decomposition (which we do not use in this chapter) is into the response due to initial conditions and no forcing, and the response due to forcing but with zero initial conditions. More precisely, given the problem (4.34), the homogenous and particular solutions in this case are

$$m\,\ddot{x}_{\rm h}(t) \;+\; c\,\dot{x}_{\rm h}(t) \;+\; k\,x_{\rm h}(t) \;=\; 0, \qquad x_{\rm h}(0)=a,\ \dot{x}_{\rm h}(0)=b, \qquad \textit{(response due to initial conditions)}$$

$$m\,\ddot{x}_{\rm p}(t) \;+\; c\,\dot{x}_{\rm p}(t) \;+\; k\,x_{\rm p}(t) \;=\; f(t), \qquad x_{\rm p}(0)=0,\ \dot{x}_{\rm p}(0)=0. \qquad \textit{(response due to forcing)}$$

Note that the initial conditions for the homogenous problem are the same as the original problem, while those for the particular problem are zero. This decomposition is illustrated in Figure 4.15.a.

An examination of Figure 4.15.a shows that the response due to forcing has some transients in it, even though the initial conditions for that response are zero. That response due to forcing can be further decomposed into another pair of particular and homogenous solutions where the steady state is the particular solution, and the transient is the homogenous solution. This decomposition is illustrated in Figure 4.15.b. In this case, the initial conditions for the homogenous problem (which is the "transient" part of the forced solution) are not the same as the original (true) initial conditions. They will be some "fictitious" initial conditions obtained just for that particular decomposition.

In Section 4.1 we derived the steady-state response by regarding it as a particular solution, and we used a method that did not depend on the initial conditions. It helped that we knew that the solution was of the form $A\cos\left(\omega t+\theta\right)$, and we solved for A and θ such that the differential equation was satisfied.

The two different types of decompositions are illustrated in figure 4.16.

4.B Solution of (4.4) Using Trigonometric Identities

Recall Equation (4.4)

$$-m\,\omega^2\,\mathsf{x}\cos\left(\omega t+\theta\right) \;-\; c\,\omega\,\mathsf{x}\sin\left(\omega t+\theta\right) \;+\; k\,\mathsf{x}\cos\left(\omega t+\theta\right) \;=\; \mathsf{f}\cos\left(\omega t\right), \tag{4.37}$$

and that we want to solve for x and θ in terms of the other parameters. We can disentangle the sine and cosine functions using the trigonometric identities

$$\begin{aligned}\cos\left(\omega t+\theta\right) &= \cos\left(\omega t\right)\cos\theta-\sin\left(\omega t\right)\sin\theta\\ \sin\left(\omega t+\theta\right) &= \sin\left(\omega t\right)\cos\theta+\cos\left(\omega t\right)\sin\theta.\end{aligned}$$

Substituting these expressions in (4.37) (and dividing through by x) gives

$$-m\omega^2\Big(\cos(\omega t)\cos\theta-\sin(\omega t)\sin\theta\Big)-c\omega\Big(\sin(\omega t)\cos\theta+\cos(\omega t)\sin\theta\Big)$$
$$+k\Big(\cos(\omega t)\cos\theta-\sin(\omega t)\sin\theta\Big) \;=\; \frac{\mathsf{f}}{\mathsf{x}}\;\cos(\omega t).$$

Now collect terms with $\cos(\omega t)$ and $\sin(\omega t)$ separately

$$\Big(-m\omega^2\cos\theta-c\omega\sin\theta+k\cos\theta\Big)\;\cos(\omega t)+\;\Big(m\omega^2\sin\theta-c\omega\cos\theta-k\sin\theta\Big)\;\sin(\omega t)\;=\;\frac{\mathsf{f}}{\mathsf{x}}\cos(\omega t). \tag{4.38}$$

This equation has to hold for all times t, and since the cosine and sine functions are "independent"[7], then we must have

$$-m\omega^2\cos\theta-c\omega\sin\theta+k\cos\theta \;=\; \frac{\mathsf{f}}{\mathsf{x}}, \tag{4.39}$$
$$m\omega^2\sin\theta-c\omega\cos\theta-k\sin\theta \;=\; 0. \tag{4.40}$$

These equations can be rearranged into a matrix-vector equation for $\cos\theta$ and $\sin\theta$

$$\begin{bmatrix}(k-m\omega^2) & -c\omega\\ -c\omega & -(k-m\omega^2)\end{bmatrix}\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \;=\; \begin{bmatrix}\mathsf{f}/\mathsf{x}\\ 0\end{bmatrix},$$

which can be rewritten using the inverse of the matrix as

$$\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \;=\; \frac{-1}{(k-m\omega^2)^2+(c\omega)^2}\begin{bmatrix}-(k-m\omega^2) & c\omega\\ c\omega & (k-m\omega^2)\end{bmatrix}\begin{bmatrix}\mathsf{f}/\mathsf{x}\\ 0\end{bmatrix}=\begin{bmatrix}\frac{-(k-m\omega^2)}{(k-m\omega^2)^2+(c\omega)^2}\\ \frac{c\omega}{(k-m\omega^2)^2+(c\omega)^2}\end{bmatrix}\frac{\mathsf{f}}{\mathsf{x}}. \tag{4.41}$$

We can eliminate θ using the identity $\cos^2\theta+\sin^2\theta=1$ which implies

$$1 \;=\; \cos^2\theta+\sin^2\theta \;=\; \left(\frac{(k-m\omega^2)^2}{\big((k-m\omega^2)^2+(c\omega)^2\big)^2}+\frac{(c\omega)^2}{\big((k-m\omega^2)^2+(c\omega)^2\big)^2}\right)\frac{\mathsf{f}^2}{\mathsf{x}^2}$$
$$=\;\left(\frac{(k-m\omega^2)^2+(c\omega)^2}{\big((k-m\omega^2)^2+(c\omega)^2\big)^2}\right)\frac{\mathsf{f}^2}{\mathsf{x}^2} \;=\; \left(\frac{1}{(k-m\omega^2)^2+(c\omega)^2}\right)\frac{\mathsf{f}^2}{\mathsf{x}^2},$$

and we finally obtain

$$\mathsf{x} \;=\; \frac{1}{\sqrt{(k-m\omega^2)^2+(c\omega)^2}}\;\mathsf{f}.$$

To obtain θ, we again use (4.41) to find $\tan\theta$ by

$$\tan\theta \;=\; \frac{\sin\theta}{\cos\theta} \;=\; -\frac{c\omega}{(k-m\omega^2)}.$$

Note that the above two expressions are identical to (4.11) derived using phasor analysis.

4.C Proof of Theorem 4.1

Theorem 4.5. *Let m be a single mass moving in one coordinate labeled x, and subject to n time-varying forces $f_1(t),\ldots,f_n(t)$ where the mass' position $x(t)$ and all forces are periodic with a common period. If $P^{\mathrm{k}}_{\mathrm{av}}$ denotes the average power supplied by each force f_k, then the total average power is zero*

$$\sum_{k=1}^{n} P^{\mathrm{k}}_{\mathrm{av}} \;=\; 0.$$

[7] For example, evaluating (4.38) at $t=0$ gives (4.39), and evaluating it at $t=\frac{2}{\pi\omega}$ gives (4.40).

Proof. Newton's 2nd law states

$$m\,\ddot{x}(t) \;=\; \sum_{k=1}^{n} f_k(t). \tag{4.42}$$

Applying formula (4.15) to the sum of the powers, we obtain for the integral over any one period

$$\begin{aligned}
\sum_{k=1}^{n} P_{\text{av}}^{\text{k}} \;&=\; \sum_{k=1}^{n}\left(\frac{1}{T}\int_t^{t+T} \dot{x}(\tau)\; f_k(\tau)\; d\tau\right) \;=\; \frac{1}{T}\int_t^{t+T} \dot{x}(\tau)\;\left(\sum_{k=1}^{n} f_k(\tau)\right)\; d\tau \\
&=\; \frac{m}{T}\int_t^{t+T} \dot{x}(\tau)\;\ddot{x}(\tau)\; d\tau \qquad\qquad \text{(using Newton's 2nd law (4.42))} \\
&=\; \frac{m}{T}\,\dot{x}(\tau)\;\dot{x}(\tau)\Big|_t^{t+T} \;-\; \frac{m}{T}\int_t^{t+T} \ddot{x}(\tau)\;\dot{x}(\tau)\; d\tau \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(integration by parts)} \\
\Rightarrow \qquad 2\int_t^{t+T} \ddot{x}(\tau)\;\dot{x}(\tau)\; d\tau \;&\overset{1}{=}\; \dot{x}(\tau)\;\dot{x}(\tau)\Big|_t^{t+T} \;=\; \dot{x}^2(\tau)\Big|_t^{t+T} \;\overset{2}{=}\; 0,
\end{aligned}$$

where $\overset{1}{=}$ follows from the previous equation by regrouping terms (and cancelling the common factor m/T), and $\overset{2}{=}$ follows from the fact that motion is T-periodic, which implies that

$$\dot{x}^2(\tau)\Big|_t^{t+T} \;=\; \dot{x}^2(t+T)-\dot{x}^2(t) \;=\; 0. \qquad \text{(since } T\text{-periodic implies } \dot{x}(t)=\dot{x}(t+T) \text{ for any } t)\square$$

Note the critical role that periodicity of motion played in the above proof. The theorem does not necessarily hold if the motion is not periodic.

Chapter 5

Harmonically Forced Vibrations: Applications

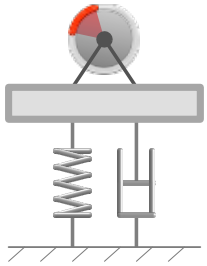
The frequency response enables analysis and design in a variety of applications of vibrations theory. The vibrations in mechanical machinery caused by rotation of unbalanced masses can be effectively modeled as periodic forcing with frequency related to the rotation rate. Design guidelines for vibration attenuation and suppression platforms, or seismic vibrations in general can be informed by studying the frequency response of systems subject to displacement inputs. Vehicles traveling on undulating roads experience resonant vibrations in the suspension system depending on the vehicles velocity and the road's undulation wavelength. Accelerometers use the vibrations of an internal proof mass to estimate acceleration in an inertial frame. In all of these vibrations problems, system design is informed by analysis of the respective frequency response. Some systems need to operate below resonance (e.g. rotating imbalance, vehicles, accelerometers), while others (e.g. vibration attenuation) need to operate above resonance.

Introduction

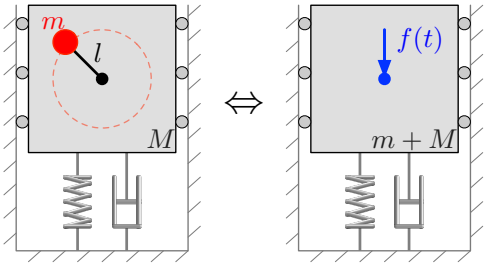
In the previous chapter, we introduced frequency response analysis as a powerful tool for studying forced vibrations in single degree of freedom (1-DOF) systems such as the Mass-Spring-Damper. Despite its simplicity, this method is widely used in engineering analysis and design. In this chapter, we explore some applications of frequency response analysis to gain valuable insights and design principles. Although the models used in this chapter are highly simplified, they still provide useful guidelines.

To illustrate the effectiveness of frequency response analysis, we will consider four different physical systems in this chapter: (Section 5.1.1) rotating machinery imbalance and its effect on supporting platforms, (Section 5.1) vibration control and attenuation such as in vibration-isolation optical tables, (Section 5.2) suspension system vibrations when traveling on undulating roads, and (Section 5.3) accelerometer design. By examining the frequency responses in these systems, we can derive useful design guidelines. However, it is important to note that the interpretation of the frequency response in each problem will be different. For example, in rotating imbalance, the natural frequency of the support platform needs to be higher than any potential imbalance frequency, which means the system needs to operate below resonance. On the other hand, in floor vibration isolation, the natural frequency of the support platform needs to be lower than any possible floor vibrations, i.e., the system needs to operate above resonance.

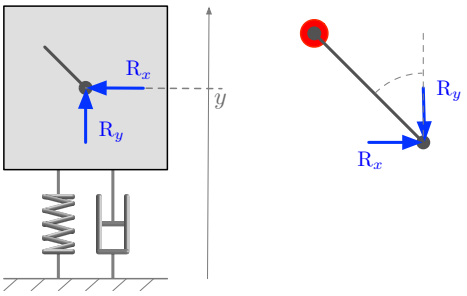
While the examples we consider in this chapter are based on 1-DOF systems, it is important to note that more complex systems may require additional tools presented in later chapters for n-DOF systems. Nonetheless, the concepts and principles discussed in this chapter provide a foundation for understanding frequency response analysis and its application to engineering design.



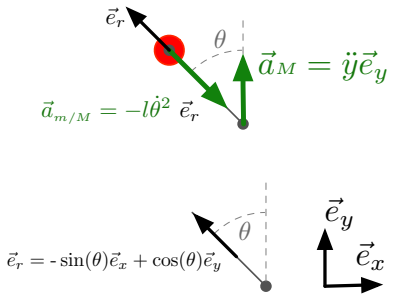
(a) Rotating imbalance is typically caused by unbalanced rotating machinery anchored to a platform or other piece of machinery that is itself connected to a rigid platform (here shown as ground) by flexible attachments (here depicted as a spring and damper). Unbalanced rotation is caused by the rotating element being slightly off-center, depicted here as an unbalanced mass in red.



(b) A simplified model of rotating imbalance is shown here where only vertical motion is possible. A mass M (which is the sum of the platform's mass and the "balanced" portion of the rotating mass) is constrained (shown here as constrained by vertical "guides") to move only vertically. (*Left*) The imbalance is modeled as a mass m constrained to move in a circular path of radius l around the point connecting M and m. l is the radius of gyration of the unbalanced mass. (*Right*) With the reaction force analysis below, it is possible to replace the rotating mass m with an effective time-varying force $f(t)$ on M.



(c) Free Body Diagrams of M and m showing only the reaction forces R_x and R_y exchanged between them. The connecting rod is assumed rigid, and therefore it just transmits those reaction forces from M to m. Since M is constrained to move vertically, only R_y is needed to model the effects of rotating imbalance.



(d) R_y can be found by modeling the centripetal acceleration of m relative to M. (*Top*) Since the motion of m relative to M is constrained to be circular with constant $\dot{\theta}$, then the acceleration $\vec{a}_{m/M}$ of m relative to M is purely centripetal and is equal to $-l\dot{\theta}^2\ \vec{e}_r$. The acceleration $\vec{a}_M$ of M is constrained to be purely vertical. (*Bottom*) The moving radial coordinate unit vector $\vec{e}_r$ can be expressed in terms of the stationary cartesian coordinate unit vectors $\vec{e}_x$, $\vec{e}_y$, and the angle coordinate θ.

Figure 5.1: Rotating imbalance and a simplified version of its analysis.

5.1 Vibration Control

In this section we look at simple examples of vibration control in the form of suppression or attenuation. This is a topic of wide applicability in industrial and laboratory settings, as well as in the analysis of seismic effects on structures and buildings.

5.1.1 Rotating Imbalance

Rotating imbalance is a type of mechanical vibration caused by uneven distribution of mass around an axis of rotation. This results in a time-varying force acting on the rotating body. A familiar example to most readers would be unbalanced rotation of a washing machine drum, which may cause excessive vibrations (and even a "walking behavior") if the resonance conditions are just right. Rotating imbalance in industrial settings can lead to excessive wear, decreased efficiency, and reduced lifespan of components. It is typically corrected by re-balancing the rotating assembly to distribute mass more evenly, but that may not always be possible, and a full understanding of the phenomenon requires a vibration analysis using the tools developed so far.

We consider the simplified set up shown in Figure 5.1. The first Figure 5.2 shows a platform whose support structure is modeled as a spring/damper combination. An unbalanced rotating element is attached to the platform. In the following analysis we replace the rotating unbalanced mass with

an effective time-varying reaction force (see Figure 5.1b). The setting then becomes a Mass-Spring-Damper (MSD) system with an externally applied harmonic force whose frequency can be related to the angular velocity of the rotating element. This is the setting of harmonically forced MSD systems analyzed in the previous chapter, and thus the tools we developed are applicable to this problem. The following assumptions are made to simplify the analysis.

1. The assembly (platform and rotating mass) can only move vertically in the y direction. This means the lateral balance of forces need not be analyzed.
2. The rotating mass m rotates with uniform angular velocity $\dot{\theta} = \omega = \text{constant}$. This means that the motion of m relative to M is purely circular. Those two assumptions together imply that the acceleration $\vec{a}_{m/M}$ of m relative to M is purely centripetal, and therefore (see Figure 5.1d)

$$\vec{a}_{m/M} \;=\; \vec{a}_m - \vec{a}_M \;=\; -\,l\dot{\theta}^2\;\vec{e}_r. \tag{5.1}$$

Now for the modeling

- The effect of the rotating mass m on the mass M is completely described by the vertical reaction force R_y since lateral motion is zero. We need to find R_y in terms of the acceleration of the rotating mass.
- Refer to Figure 5.1d to see that the moving radial unit vector $\vec{e}_r$ can be written in terms of the stationary cartesian unit vectors $\vec{e}_x$, $\vec{e}_y$ and the angular coordinate θ as

$$\vec{e}_r(t) \;=\; -\sin\big(\theta(t)\big)\;\vec{e}_x \;+\; \cos\big(\theta(t)\big)\;\vec{e}_y \tag{5.2}$$

The vertical acceleration of m is is obtained from (5.1) and (5.2) as follows

$$\begin{aligned}\vec{a}_m \;=\; \vec{a}_M \;-\; l\dot{\theta}^2\;\vec{e}_r \;&=\; \ddot{y}(t)\;\vec{e}_y \;-\; l\dot{\theta}^2\left(-\sin\big(\theta(t)\big)\;\vec{e}_x \;+\; \cos\big(\theta(t)\big)\;\vec{e}_y\right)\\ &=\; \Big(l\dot{\theta}^2\sin\big(\theta(t)\big)\Big)\;\vec{e}_x + \Big(\ddot{y}(t) - l\dot{\theta}^2\cos\big(\theta(t)\big)\Big)\;\vec{e}_y\end{aligned}$$

- The vertical component of the acceleration $\vec{a}_m$ must be provided by the vertical reaction force R_y, and the Newton's 2nd law for the small mass m states

$$m\left(\ddot{y}(t) - l\dot{\theta}^2\cos\big(\theta(t)\big)\right) \;=\; -\,\mathrm{R}_y$$

- The same vertical reaction force R_y acts on the mass M (along with the spring and damper), and Newton's 2nd law for M is

$$\begin{aligned}M\,\ddot{y}(t) \;&=\; -ky(t) \;-\; c\dot{y}(t) \;+\; \mathrm{R}_y\\ &=\; -ky(t) \;-\; c\dot{y}(t) \;-\; m\left(\ddot{y}(t) - l\dot{\theta}^2\cos\big(\theta(t)\big)\right)\end{aligned}$$

Now substituting $\dot{\theta} = \omega$ and $\theta(t) = \omega t$ and gathering terms

$$\begin{aligned}&(M+m)\,\ddot{y}(t) \;+\; c\,\dot{y}(t) \;+\; k\,y(t) \;=\; ml\omega^2\;\cos(\omega t)\\ \Rightarrow\quad &\ddot{y}(t) \;+\; 2\zeta\omega_{\mathrm{n}}\,\dot{y}(t) \;+\; \omega_{\mathrm{n}}^2\,y(t) \;=\; \tfrac{m}{M+m}l\omega^2\;\cos(\omega t), \qquad \omega_{\mathrm{n}} := \sqrt{\tfrac{k}{M+m}},\; \zeta := \tfrac{c}{2\sqrt{k(M+m)}}.\end{aligned} \tag{5.3}$$

Equation (5.3) is that of a harmonically forced MSD system with mass $M+m$. Note that the forcing term has factor of ω^2 in it, i.e. the forcing amplitude increases with increasing frequency. This has a physical explanation. The term $l\cos(\omega t)$ is the vertical position of m relative to M. Therefore $-l\omega^2\cos(\omega t)$ is the vertical *acceleration* of m relative to M, and consequently $ml\omega^2\cos(\omega t)$ is the reaction force on M producing that relative acceleration. We can therefore say that the system (5.3) is driven by the acceleration of m.

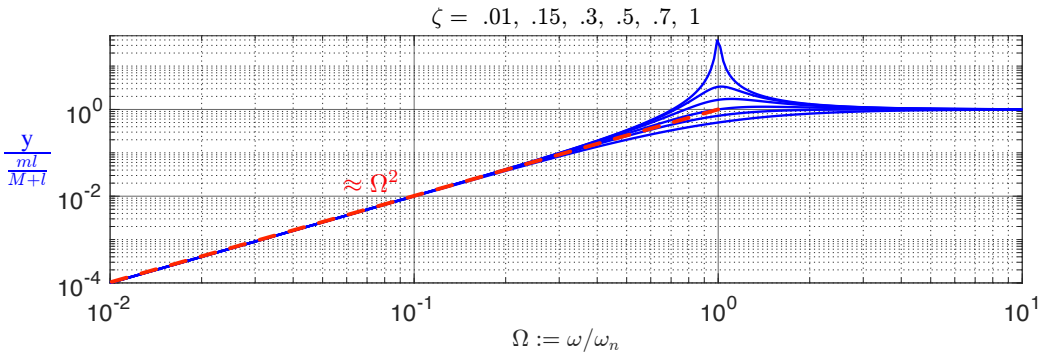


Figure 5.2: The frequency response (5.4) of the rotating imbalance problem, where y is the platform's oscillation amplitude. To as low y as possible, operation in the regime $\Omega \ll 1$ is needed. This implies that the resonant frequency ω_n of the platform's supports needs to be much higher than the operating imbalance frequency ω_n. In this low frequency range, the response is very well approximated by Ω^2 as shown.

Analysis of the Frequency Response

Since (5.3) has a pure sinusoid as an input, the platform's steady-state oscillation is also a sinusoid of the same frequency

$$y(t) \;=\; \mathsf{y}\,\cos(\omega t + \theta).$$

The amplitude of oscillations y is obtained from the frequency response (4.14) (note that $ml\omega^2$ is the amplitude of the forcing) as

$$\begin{aligned}
\mathsf{y} \;&=\; \frac{1}{k}\frac{1}{\sqrt{(1-\Omega^2)^2+4\zeta^2\,\Omega^2}}\; ml\omega^2 \;=\; \frac{1}{k}\frac{\omega^2/\omega_n^2}{\sqrt{(1-\Omega^2)^2+4\zeta^2\,\Omega^2}}\; ml\;\omega_n^2 \qquad (\Omega := \omega/\omega_n)\\
&=\; \frac{1}{k}\frac{\Omega^2}{\sqrt{(1-\Omega^2)^2+4\zeta^2\,\Omega^2}}\; ml\;\tfrac{k}{M+m} \;=\; \frac{\Omega^2}{\sqrt{(1-\Omega^2)^2+4\zeta^2\,\Omega^2}}\;\tfrac{ml}{M+m} \qquad (5.4)
\end{aligned}$$

This frequency response is shown in Figure 5.2. The main features are that the frequency response goes to zero for low Ω (it behaves like Ω^2 for $\Omega \ll 1$), and in that range, it is relatively insensitive to the choice of ζ. Thus a design guideline emerges. The operating condition should be *below resonance*, i.e. in the regime

$$\Omega \;=\; \frac{\omega}{\omega_n} \;\ll\; 1 \qquad \Leftrightarrow \qquad \text{imbalance frequency} : \omega \;\ll\; \omega_n : \text{platform's natural frequency}$$

If the angular velocity of the rotating imbalanced element is fixed. This fixes ω, and we therefore need to make the resonant frequency ω_n high so that the system is excited much below resonance. Since $\omega_n = \sqrt{k/(M+m)}$ this means either making M small (perhaps not a feasible option) or "stiffen" the support structure thus making k large.

As Figure 5.2 indicates, for low Ω the magnitude frequency response is well approximated by

$$\mathsf{y} \;=\; \Omega^2\;\tfrac{ml}{M+m}, \qquad\qquad \Omega < 0.5.$$

This is the regime of typical design. As the log-log plot shows, to achieve a factor of 100 reduction in amplitude y, the natural frequency ω_n needs to increase by a factor of 10. If the masses are fixed, then the only design freedom is k ($\omega_n^2 = k/(m+M)$). Thus for a 100-fold reduction in amplitude, the support's k would have to be stiffened by a factor of 100 also.

5.1.2 Transferred Force

An important consideration in the design of heavy machinery is quantifying the various forces that support structures are subjected to during operation. Consider the simple system shown in Figure 5.3a where some unbalanced rotating machinery is mounted on a platform that is in turn

anchored by a support structure to the ground. As in Section 5.1.1, the rotating imbalance can be modeled as a periodically time-varying force on an equivalent mass. This force will cause the platform to vibrate, and in turn the spring and damper modeling the support structure will exert forces on both the platform as well as the ground. It is important to quantify those forces since in some situations, they could be sufficiently strong to damage the ground anchors.

The analysis of this system will demonstrate another use for phasor analysis, namely when the output to be analyzed is not a displacement or velocity, but rather a force, which is a linear combination of both. Referring to Figure 5.3a again, note that the ground force f_g is the sum of the spring and damper forces. Combine this with the differential equation for the system developed earlier, and rewrite them in terms of ζ and ω_n

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f_b(t) \qquad \ddot{x}(t) + 2\zeta\omega_n\, \dot{x}(t) + \omega_n^2\, x(t) = \tfrac{1}{m}\, f_b(t), \tag{5.5}$$

$$f_g(t) = c\dot{x}(t) + kx(t) \qquad \tfrac{1}{m}\, f_g(t) = 2\zeta\omega_n\, \dot{x}(t) + \omega_n^2\, x(t). \tag{5.6}$$

We are interested in the frequency response from f_b to f_g rather than to the displacement x. However, since f_g is a linear combination of x and its derivatives, the phasor analysis done earlier for displacement is useful. Let the forcing $f_b(t)$ be a sinusoid of frequency ω, then we know that the steady state response $x(t)$ will be a sinusoid of the same frequency, and consequently both $\dot{x}(t)$ and $f_g(t)$ will also be sinusoids of that same frequency. Specifically

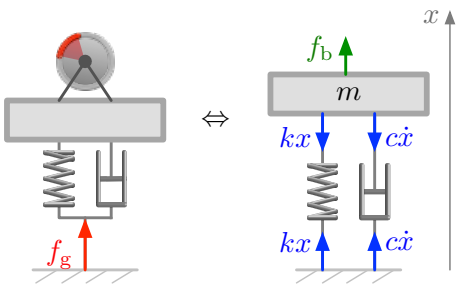
$$\begin{aligned}
& f_b(t) = \mathsf{f}_b \cos(\omega t) \\
\Rightarrow\ & x(t) = \mathsf{x}\cos(\omega t + \theta), && \mathsf{x} = |\hat{\mathsf{x}}| = \left|\frac{1/m}{(j\omega)^2 + 2\zeta\omega_n\,(j\omega) + \omega_n^2}\right| \mathsf{f}_b && \text{(frequency response } f_b \mapsto x\text{)} \\
& f_g(t) = \mathsf{f}_g \cos(\omega t + \phi), && \tfrac{1}{m}\mathsf{f}_g = \tfrac{1}{m}\left|\hat{\mathsf{f}}_g\right| = \left|2\zeta\omega_n\,(j\omega)\,\hat{\mathsf{x}} + \omega_n^2\hat{\mathsf{x}}\right| && \text{(from (5.6), phasor of } \dot{x} \text{ is } (j\omega)\hat{\mathsf{x}}\text{)} \\
& && = \left|2\zeta\omega_n\,(j\omega) + \omega_n^2\right| |\hat{\mathsf{x}}|
\end{aligned}$$

The above two equations together give the amplitude frequency response from $\mathsf{f}_b \mapsto \mathsf{f}_g$ as the product of the response $\mathsf{f}_b \mapsto \mathsf{x}$ with the response $\mathsf{x} \mapsto \mathsf{f}_g$ (note the cancellation of the common factor m)

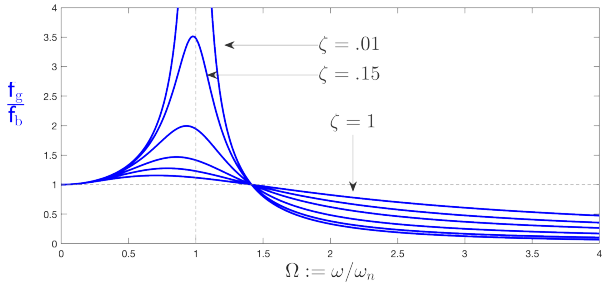
$$\begin{aligned}
\frac{\mathsf{f}_g}{\mathsf{f}_b} &= \left|\frac{2\zeta\omega_n\,(j\omega) + \omega_n^2}{(j\omega)^2 + 2\zeta\omega_n\,(j\omega) + \omega_n^2}\right| = \frac{\sqrt{\omega_n^4 + 4\zeta^2\omega_n^2\,\omega^2}}{\sqrt{(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\,\omega^2}} \\
&= \sqrt{\frac{1 + 4\zeta^2\,\Omega^2}{(1-\Omega^2)^2 + 4\zeta^2\,\Omega^2}}, \qquad \Omega := \omega/\omega_n.
\end{aligned} \tag{5.7}$$

A plot of this frequency response is shown in Figure 5.3b for various values of $0 \le \zeta \le 1$. It is interesting to examine how this response serves as a design guide. If large support forces are to be avoided at all frequencies, then we should have a lot of damping with $\zeta \to 1$. In this case however, the transfer ratio is close to 1 over a large range of frequencies (including those above resonance). If it is desired to keep the transfer ratio significantly less than 1, then ω_n should be lower than the operating machinery frequency ω (i.e $\Omega = \omega/\omega_n \gg 1$, and ζ should be less than 1. This would give a significant reduction in the transfer ratio as seen in the frequency response plot. However, this comes with a risk. If for some reason the operating frequency ω is reduced to near resonance, the support anchors will experience much larger forces than if $\zeta \approx 1$. This is a design tradeoff that needs to take into account realistic operating conditions as well as contingencies. Having frequency response graphs like that of Figure 5.3b enables a systematic approach to design tradeoffs.

We now compare the above design guidelines with those in the previous subsection for platform vibrations due to rotating imbalance. For platform vibrations reduction, the guidelines are the opposite of what we have here, i.e. that we should make ω_n higher than the operating frequency (by stiffening the supports). However, this leads to transferring larger forces to the ground. Thus the two guidelines are *conflicting*! If we want to minimize transferred forces, we will have to live with larger platform vibrations. If we want to minimize platform vibrations, we need to insure that ground supports are sufficiently strong to withstand the large transferred forces.



(a) Any structure anchored to ground and subject to forces that cause vibrations will "transfer" some of those forces to its ground support. Here a rotating imbalance subjects the platform to a force f_b. The ground attachment of the platform is then subjected to the transferred force f_g.



(b) Magnitude frequency response of the "transfer ratio" f_g/f_b between the two forces. At very low frequencies $\Omega \approx 0$ the ratio is one since the spring simply transfers the static loading on m to an equal static loading on the ground support. At resonance, the ratio can be much larger than one. At very high frequencies $\Omega \to \infty$, the ratio goes to zero, but much more "slowly" as $\zeta \to 1$. There is a sort of "waterbed effect"; if ζ is used as a design parameter, then increasing ζ reduces the ratio at resonance, but it *increases it* at frequencies higher than resonance. Note that the ratio $f_g/f_b = 1$ at exactly two frequencies $\Omega = 0$ and $\Omega = \sqrt{2}$.

Figure 5.3: Analysis of the transferred force from a vibrating machine to the ground support of a platform.

5.1.3 Vibration Isolation/Attenuation

In an earthquake, the ground moves beneath your feet. The ground is no longer an inertial frame, but rather a moving platform. We have already encountered this setting with "displacement forcing" in Section 4.3.1, and we recap it here in the context of vibration attenuation and isolation. For seismic applications, it is important to quantify the effect of ground movement on a structure. On a smaller scale in laboratory settings, certain sensitive equipment and experiments need to be mounted on a vibration isolation platform, i.e. a specially designed tabletop that reduces the effect of floor vibrations on the equipment mounted on the tabletop. Optical experiments for example are particularly susceptible to such floor vibrations, and vibration control platforms for such experiments are called "optical tables".

A simple 1-DOF vibration isolation set up is shown in Figure 5.4a, where a spring and damper element model the support structure of a platform. The vertical ground motion $u(t)$ is the input to this system, and the analysis goal is to characterize the dynamic effects of ground vibrations on the platform vibrations $x(t)$. We will see that by appropriately designing m, k, c, we can achieve a certain level of ground vibration attenuation but only in certain frequency ranges. Figure 5.4b show a more sophisticated system where an electronically controlled actuator produces a force in addition to the spring and damper. Such a system performs what is generally termed "active" vibration attenuation. Other vibration attenuation and isolation schemes involve multiple masses, and those will be considered in later chapter when analyzing systems with multiple degrees of freedom.

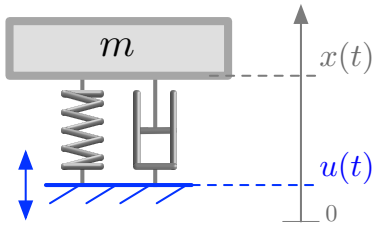
The mathematical model for the system in Figure 5.4a is written down as follows. The vertical axis shows the coordinates in an inertial frame. The ground's position is $u(t)$ (in this case, the ground is no longer an inertial frame), and the platform's position is $x(t)$. If the spring is assumed to have an equilibrium length of L, then Newton's 2nd law for the platform reads

$$m\,\ddot{x}(t) \;=\; -k\big(x(t) - u(t) - L\big) - c\big(\dot{x}(t) - \dot{u}(t)\big).$$

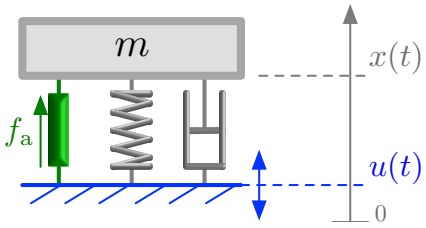
where $x(t) - u(t) - L$ is the extension of the spring length above equilibrium. The constant factor of L can be eliminated by defining the new position variable

$$\tilde{x}(t) \;:=\; x(t) - L,$$

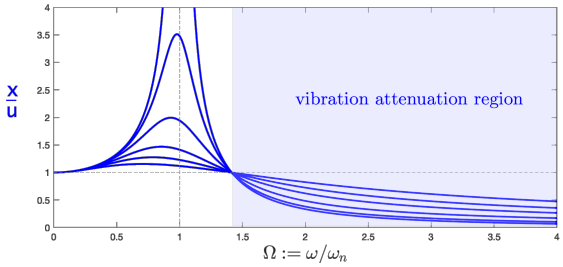
which is a simple offset from the platform's position. The differential equation above can now be



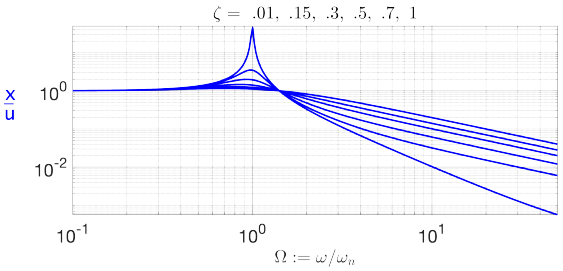
(a) When a platform is mounted on a floor or ground that is vibrating, both coordinates of the platform as well as ground have to be measured in an inertial frame as shown here. This diagram depicts a *passive* vibration attenuation system which in a 1-DOF model consists of a spring and a damper.



(b) Active vibration attenuation systems have an "actuator" (depicted here in green) in addition to the passive components modeled as a spring and damper. The actuator exerts a force f_a on the platform that is typically electronically controlled through some feedback mechanism that measures the platform's vibrations and reacts using an active vibration feedback control algorithm.



(c) The frequency response from ground vibrations amplitude u to platform motion amplitude x. The vibration attenuation region where the amplitude ratio is < 1 is for frequencies $\Omega > \sqrt{2}$. Note that higher damping ratios ζ gives *lower* x/u ratios in the vibration attenuation region at the expense of higher resonance peaks.



(d) For quantitative analysis and design, the frequency response is best viewed on log-log plots (also known as Bode plots) as shown here. For example, if $\zeta = .15$ and vibration attenuation by at least a factor of 100 is desired, then $\Omega > 30$. Since the floor vibration is not a design parameter, the systems resonance frequency ω_n must be 30 times lower than any expected floor vibration frequency.

Figure 5.4: Analysis of the effects of floor vibrations on a platform. Typically vibration attenuation ratios of much less than 1 are needed. This means that the platform's m, k, c should be designed so that its resonance ω_n is lower than the expected ground vibration frequency. This makes for heavy platforms (large m) and lower spring constants k as is the case for optical tables.

rewritten in terms of $\tilde{x}$ and u as

$$\begin{aligned} m\,\ddot{\tilde{x}}(t) &= -k\,\big(\tilde{x}(t)-u(t)\big) - c\,\big(\dot{\tilde{x}}(t)-\dot{u}(t)\big) \\ \Rightarrow \quad m\,\ddot{\tilde{x}}(t)+c\,\dot{\tilde{x}}(t)+k\,\tilde{x}(t) &= c\,\dot{u}(t)+k\,u(t) \\ \Rightarrow \quad \ddot{\tilde{x}}(t)+2\zeta\omega_n\,\dot{\tilde{x}}(t)+\omega_n^2\,\tilde{x}(t) &= 2\zeta\omega_n\,\dot{u}(t)+\omega_n^2\,u(t). \end{aligned} \tag{5.8}$$

The system (5.8) has the ground displacement $u(t)$ as an input, and platform displacement $x(t)$ as an output. To find the frequency response, we note that it is in a form to which Theorem 4.2 directly applies. The "transmission ratio" of ground-to-platform vibrations is given by the Theorem as

$$\frac{\mathsf{x}}{\mathsf{u}} = \left|\frac{\hat{\mathsf{x}}}{\hat{\mathsf{u}}}\right| = \left|\frac{2\zeta\omega_n\,(j\omega)+\omega_n^2}{(j\omega)^2+2\zeta\omega_n\,(j\omega)+\omega_n^2}\right| = \sqrt{\frac{1+4\zeta^2\,\Omega^2}{(1-\Omega^2)^2+4\zeta^2\,\Omega^2}}, \qquad \Omega := \omega/\omega_n, \tag{5.9}$$

where the last equality follows from observing that this fraction is mathematically exactly the same as that in Equation (5.7). Only in this case, it is a ratio of displacements rather than a ratio of forces as in (5.7).

The frequency response of x/u is shown in Figure 5.4c (note that it is exactly the same as that of Figure 5.3b, but interpreted now as ratio of displacements rather than forces). The purpose of this system is to insure that platform vibrations are smaller in amplitudes than ground vibrations. If that were not a requirement, we can simply connect the platform with a rigid structure to ground, and then $\mathsf{x} = \mathsf{u}$. The design requirement is therefore that $\mathsf{x}/\mathsf{u} \ll 1$, i.e. "vibration attenuation". By

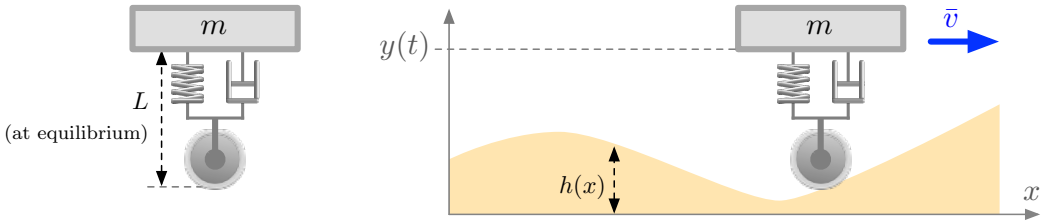


Figure 5.5: (*Left*) A suspension system where the length L is the distance between the road surface and some designated point on the vehicle's body at equilibirum. (*Right*) The vertical elevation of a terrain is described by a function $\mathrm{h}(x)$ where x is the horizontal coordinate. A vehicle moving at constant velocity $\bar{v}$ is at location $x = \bar{v}t$ at time t. To the vehicle, the road surface appears to be a time-varying function $u(t) = \mathrm{h}(\bar{v}t)$, and the suspension spring extension is then $y(t) - u(t) - L$, where $y(t)$ is that designated point on the vehicle's body.

looking at Figure 5.4c we see that this is only possible in the region $\Omega > \sqrt{2}$ which we label as the "vibration-attenuation region". Since the frequency range of vibration attenuation is *higher* than the natural frequency of the platform, we say that the system must operate *above resonance*. This is in contrast to the rotating imbalance problem, where the system must be designed to operate below resonance.

It is important to note that the the frequency of ground vibrations $\omega = \Omega\omega_{\mathrm{n}}$ is not a design parameter but rather a design specification. To suppress ground vibrations in a particular frequency range, the platform's parameters m, c, k must be designed so that the natural frequency is significantly *below* that frequency range. For instance, in large buildings where a laboratory optical table is to be housed, vertical floor vibrations are in the range of 5-50 Hz. Therefore, the natural frequency of the platform should be below 2 Hz, which requires platforms with large masses m and rather soft springs k. It is also worth noting that lower damping ratios ζ provide better attenuation in the vibration-attenuation frequency region shown in Figure 5.4. However, if a very low damping ratio is used, one must ensure that there is no possibility of floor vibrations at the resonance frequency, as lower damping ratios produce higher resonance peaks.

5.2 Vehicles over Undulating Terrains

Certain roads are designed with regularly spaced features, such as expansion joints between segments on concrete highways, typically spaced 10-30 meters apart. After some time, the road may experience buckling which causes those segments to deform slightly in the vertical direction. This produces a road terrain with a slight vertical oscillation with a well-defined spacing or "wavelength" λ. A vehicle traveling at constant velocity $\bar{v}$ on such a road will experience these undulations as a vertical time-varying force of temporal frequency $\bar{v}/\lambda$. Indeed, to travel a distance λ (e.g. in meters) at velocity $\bar{v}$ (e.g. in meters/second) takes a time of $\lambda/\bar{v}$ seconds. This is the period of the apparent force, so its frequency would be $\bar{v}/\lambda$ Hz. If this frequency is close to the natural frequency of the suspension system, then resonance phenomena occur where the vehicle will bounce up and down uncomfortably. The reader may have experienced this phenomenon first hand.

We consider a simple analysis of this setting as illustrated in Figure 5.5. The following quantities are used in the modeling.

- $y(t)$: Height of vehicle relative to absolute ground (i.e. in an inertial frame).
- $h(x)$: Height of terrain (relative to absolute ground) as a function of the horizontal coordinate x.
- $\bar{v}$: Constant forward velocity of the vehicle $\Rightarrow$ the vehicle's horizontal position at time t is $x = \bar{v}t$.
- L: Equilibrium length of the suspension's spring.

As the vehicle moves with position $x = \bar{v}t$, the height of the terrain as s function of time is $h(\bar{v}t)$. Assume a sinusoidal terrain

$$h(x) \;=\; \mathsf{h}\,\cos\big((2\pi/\lambda)\,x\big)$$

The spatial wavelength of the terrain is then λ. A vehicle moving with constant velocity $\bar{v}$ over such a terrain will experience this as a *time-varying height*

$$u(t) \;:=\; h(\bar{v}t) \;=\; \mathsf{h}\,\cos\big((2\pi/\lambda)\bar{v}\;t\big) \;=\; \mathsf{u}\,\cos\big(\omega t\big), \qquad\qquad \omega := 2\pi\;\bar{v}/\lambda, \tag{5.10}$$

with frequency $\omega := 2\pi\;\bar{v}/\lambda$.

Given the definitions above, the suspension's extension (positive extension means upward movement) is given by $(y(t) - u(t) - L)$. Newton's 2nd law for the vehicle is then (ignoring gravity)

$$\begin{aligned} m\;\ddot{y}(t) \;&=\; -\,k\,\Big(y(t) - u(t) - L\Big) - c\frac{d}{dt}\Big(y(t) - u(t) - L\Big) \\ &=\; -\,k\,\big(y(t) - u(t) - L\big) - c\,\big(\dot{y}(t) - \dot{u}(t)\big)\,. \end{aligned} \tag{5.11}$$

To analyze the vehicle's vibrations and in particular to asses their effects on passenger comfort, we need the vehicle's position in an inertial frame. If we define a new vertical position variable

$$\tilde{y}(t) \;:=\; y(t) - L,$$

then $\tilde{y}(t)$ is $y(t)$ in a new inertial frame with a shifted origin. However, Equation (5.11) now simplifies to

$$\begin{aligned} m\;\ddot{\tilde{y}}(t) \;&=\; -\,k\,\big(\tilde{y}(t) - u(t)\big) - c\,\big(\dot{\tilde{y}}(t) - \dot{u}(t)\big) \\ \Rightarrow \qquad m\ddot{\tilde{y}}(t) + c\dot{\tilde{y}}(t) + k\tilde{y}(t) \;&=\; c\dot{u}(t) + ku(t) \\ &\qquad\qquad \big(\text{now divide by } m \text{ and rewrite in terms of } \zeta \text{ and } \omega_{\rm n}\big) \\ \Rightarrow \qquad \ddot{\tilde{y}}(t) + 2\zeta\omega_{\rm n}\;\dot{\tilde{y}}(t) + \omega_{\rm n}^2\;\tilde{y}(t) \;&=\; 2\zeta\omega_{\rm n}\;\dot{u}(t) + \omega_{\rm n}^2\;u(t) \end{aligned} \tag{5.12}$$

This equation represents a dynamical system whose input is the road undulation $u(t)$, and the output is the vehicle's vertical displacement in an inertial frame $\tilde{y}(t)$. Observe that Equation (5.12) has exactly the same form as Equation (5.8) encountered earlier in modeling floor vibration attenuation. This is not surprising since both equations are modeling very similar physical settings, namely the effect of a vibrating ground on the vibrations of a supported mass.

If $u(t)$ is sinusoidal, then $\tilde{y}(t)$ is sinusoidal with the same frequency, and the amplitude ratio $\tilde{\mathsf{y}}/\mathsf{u}$ is given by the amplitude frequency response, which we have already calculated (Equation (5.9)) in the previous section as

$$\frac{\tilde{\mathsf{y}}}{\mathsf{u}} \;=\; \frac{\sqrt{1 \;+\; 4\zeta^2\;\Omega^2}}{\sqrt{(1-\Omega^2)^2 \;+\; 4\zeta^2\;\Omega^2}}, \qquad\qquad \Omega := \omega/\omega_{\rm n}. \tag{5.13}$$

The frequency response is the same as that shown in Figure 5.4. If the road undulations and the vehicle's speed conspire so that the forcing frequency ω is at resonance $\omega_{\rm n}$, then we expect large amplitude of oscillation for $\tilde{y}$. Let's do a quick calculation. Assuming a suspension system with damping ratio of $\zeta \approx 0.25$, the amplification near resonance ($\omega = \omega_{\rm n} \;\Leftrightarrow\; \Omega = 1$) is

$$\left.\frac{\tilde{\mathsf{y}}}{\mathsf{u}}\right|_{\Omega=1} \;=\; \frac{\sqrt{1+4\zeta^2}}{\sqrt{4\zeta^2}} \;=\; \frac{\sqrt{1+4(0.25)^2}}{\sqrt{4(0.25)^2}} \;\approx\; 2.24. \qquad\qquad (\text{if } \zeta \approx 0.25)$$

To see what that means for a typical vehicle, we use the following parameters

$$\begin{aligned} \bar{v} \;&=\; 65\;\frac{\text{miles}}{\text{hr}} \;\times\; \frac{1\;\text{hr}}{3600\;\text{s}} \;\times\; \frac{1609\;\text{m}}{1\;\text{mile}} \;\approx\; 30\;\frac{\text{m}}{\text{s}} \\ k \;&=\; 80,000\;\text{N/m} \qquad\qquad \big(\text{assuming } k = 20,000\;\text{N/m at each of the 4 tires}\big) \\ m \;&=\; 2,000\;\text{Kg} \qquad\qquad (\Rightarrow \quad \omega_{\rm n} = \sqrt{k/m} = 6.32\;\text{rad/s} \;\approx\; 1\;\text{Hz}) \end{aligned}$$

Now using the relation (5.10) $\omega = 2\pi\;\bar{v}/\lambda$ we see that at a velocity of 65 miles/hr resonance will occur when λ is approximately

$$\lambda_{\rm res} \;\approx\; 2\pi\;\frac{\bar{v}}{\omega_{\rm n}} \;=\; 2\pi\;\frac{30\;\text{m/s}}{6.32\;\text{rad/s}} \;\approx\; 30\;\text{m}.$$

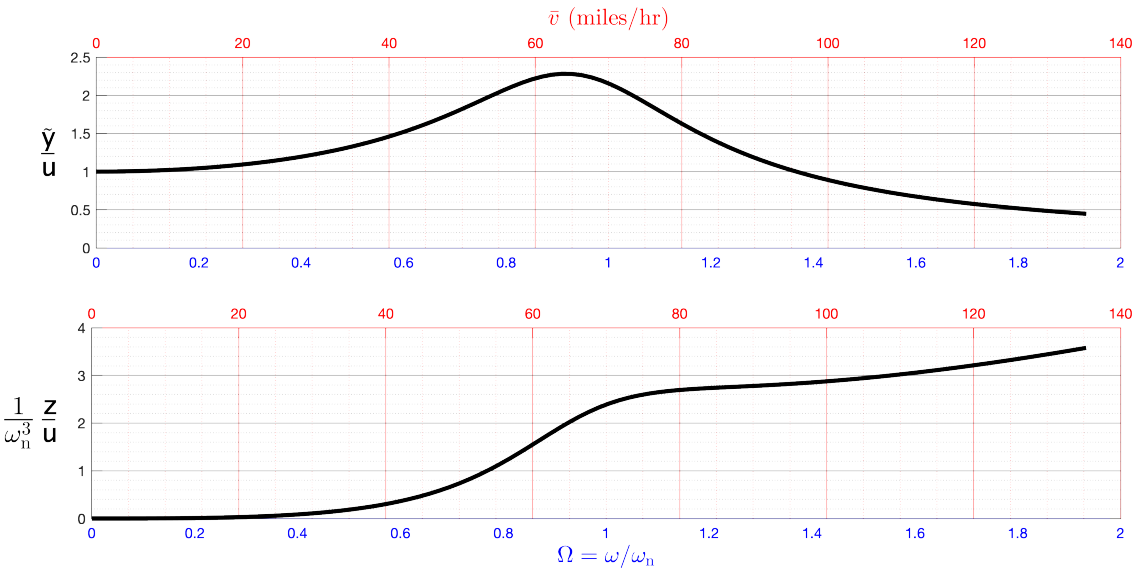


Figure 5.6: The amplitude frequency responses from road undulations u to (*top*) vertical vibrations $\tilde{y}$ and (*bottom*) vertical jerk z. In each figure, the bottom axis (in blue) shows the normalized frequency $\Omega = \omega/\omega_n$, while the top axis (in red) shows the equivalent vehicle speed $\bar{v} = \frac{\lambda}{2\pi}\,\omega \;=\; \frac{\lambda}{2\pi}\omega_n\,\Omega$ in miles/hour assuming road undulations wavelength of $\lambda = 30$ m, which results in a "resonance velocity" of approximately 65 miles/hr. It is the jerk amplitude z, rather than the vibrations amplitude $\tilde{\mathsf{y}}$ that quantifies passenger comfort. The jerk plot demonstrates that the only way to reduce passenger discomfort is to lower the vehicle's velocity. The damping ratio $\zeta = 0.25$ is used for these plots.

Thus if traveling at 65 miles/hr on a road whose undulations have wavelength of 30 m, the road undulations will be amplified by a factor of two in the vibrations of the suspension system. A road height variation of say 10 cm (which is relatively small) will cause the vehicle to oscillate $10\text{cm} \times 2.24 \approx 22\text{cm}$ vertically with a frequency of 1 Hz (i.e. up, down and up again in 1 second). This can make for a rather uncomfortable ride.

When faced with such a road, the only way to reduce the vibrations is to move away from the resonance condition. The only control the driver has over the parameters is the vehicle's velocity $\bar{v}$. Should the vehicle go faster or slower to avoid resonance? Recall that the frequency of $u(t)$ is given by

$$\omega \;=\; 2\pi\;\bar{v}/\lambda.$$

Thus the faster the vehicle the higher the frequency and vice versa. Let's set the road undulations wavelength to 30 m, and vary the vehicle's velocity instead. Figure 5.6 (*top*) shows the frequency response with axes labeled by velocity $\bar{v}$ as well as normalized frequency Ω. It shows that the higher the frequency above resonance, the lower the vibration amplitude of $\tilde{y}$ would be, implying that perhaps the vehicle should go faster. However, Figure 5.6 also shows that at low frequencies, the amplitude frequency response $\tilde{\mathsf{y}}/\mathsf{u}$ is close to 1. This is similar to the static case; if the road height moves up by a certain amount, then so does the vehicle's vertical position in an inertial frame. However, this would not be considered uncomfortable by any passenger. The reason for this "apparent paradox" is that we are using the wrong criterion for passenger comfort as explained next.

In suspension system design, passenger comfort is most directly related not to the amplitude of displacement, but rather to the *jerk*, which is the derivative of acceleration

$$\text{vertical jerk:} \quad z(t) \;:=\; \frac{d^3}{dt^3}\,\tilde{y}(t).$$

We therefore need to find the frequency response from road undulations u to the resulting vertical jerk z. A differential equation for the jerk can be derived directly from the equation (5.12) for $\tilde{y}$ by

differentiating both sides three times

$$\begin{aligned} & \ddot{\tilde{y}}(t) + 2\zeta\omega_{\rm n}\,\dot{\tilde{y}}(t) + \omega_{\rm n}^2\,\tilde{y}(t) \;=\; 2\zeta\omega_{\rm n}\,\dot{u}(t) + \omega_{\rm n}^2\,u(t) \\ \Rightarrow \qquad & \tilde{y}^{(5)}(t) + 2\zeta\omega_{\rm n}\,\tilde{y}^{(4)}(t) + \omega_{\rm n}^2\,\tilde{y}^{(3)}(t) \;=\; 2\zeta\omega_{\rm n}\,u^{(4)}(t) + \omega_{\rm n}^2\,u^{(3)}(t) \\ & \qquad\qquad \text{(differentiating both sides 3 times)} \\ \Rightarrow \qquad & \dddot{z}(t) + 2\zeta\omega_{\rm n}\,\ddot{z}(t) + \omega_{\rm n}^2\,\dot{z}(t) \;=\; 2\zeta\omega_{\rm n}\,u^{(4)}(t) + \omega_{\rm n}^2\,u^{(3)}(t) \end{aligned} \tag{5.14}$$

This is an input-output differential equation of the form to which Theorem 4.2 is applicable, and we can therefore write down the complex frequency response directly as

$$\begin{aligned} \frac{\hat{\mathsf{z}}}{\hat{\mathsf{u}}} \;&=\; \frac{2\zeta\omega_{\rm n}\,(j\omega)^4 \;+\; \omega_{\rm n}^2\,(j\omega)^3}{(j\omega)^2 \;+\; 2\zeta\omega_{\rm n}\,(j\omega) \;+\; \omega_{\rm n}^2} \;=\; (j\omega)^3\,\frac{2\zeta\omega_{\rm n}\,(j\omega) \;+\; \omega_{\rm n}^2}{(j\omega)^2 \;+\; 2\zeta\omega_{\rm n}\,(j\omega) \;+\; \omega_{\rm n}^2} \\ &=\; -\,j\omega^3\,\frac{1 \;+\; j\;2\zeta\;\omega/\omega_{\rm n}}{(\omega_{\rm n}^2 - \omega^2) \;+\; j\;2\zeta\omega_{\rm n}\;\omega} \;=\; -\,j\omega^3\,\frac{1 \;+\; j\;2\zeta\;\omega/\omega_{\rm n}}{(1 - \omega^2/\omega_{\rm n}^2) \;+\; j\;2\zeta\;\omega/\omega_{\rm n}} \qquad \text{(dividing by } \omega_{\rm n}^2\text{)} \\ &=\; -\,j\omega_{\rm n}^3\,\Omega^3\,\frac{1 \;+\; j\;2\zeta\;\Omega}{(1 - \Omega^2) \;+\; j\;2\zeta\;\Omega} \qquad\qquad (\Omega := \omega/\omega_{\rm n} \;\Rightarrow\; \omega = \omega_{\rm n}\Omega) \end{aligned}$$

Taking the magnitude of this complex ratio gives the amplitude frequency response (the phase response is not relevant here since we are only interested in the amplitude of the jerk)

$$\frac{\mathsf{z}}{\mathsf{u}} \;=\; \frac{|\hat{\mathsf{z}}|}{|\hat{\mathsf{u}}|} \;=\; \omega_{\rm n}^3\,\Omega^3\,\frac{\sqrt{1 \;+\; 4\zeta^2\,\Omega^2}}{\sqrt{(1 - \Omega^2)^2 \;+\; 4\zeta^2\,\Omega^2}} \tag{5.15}$$

Observe that the fractional part of this frequency response is exactly the same as the response to $\tilde{y}$ calculated earlier in (5.13). The response to jerk calculated above has an extra factor of Ω^3, which is not surprising since the jerk is the 3rd derivative of position. Note that the ratio z/u above has dimensions of acceleration over distance $1/s^2$, as are the units of $\omega_{\rm n}^2$.

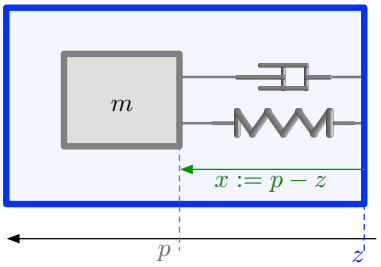
The frequency response (5.15) is shown in Figure 5.6 (*bottom*). It is different from the frequency response to position $\tilde{y}$ shown in the same figure. The extra factor of Ω^3 keeps the response very low at low frequencies as might be expected from physical intuition. Low vehicle velocities produce lower jerks. Another notable difference is that the jerk response keeps increasing beyond the resonance frequency $\Omega = 1$, unlike the position response which decreases beyond resonance. This also makes physical sense. Although the position response y/u decreases above resonance, the frequency is also higher above resonance, and since the jerk is the 3rd derivative, its amplitude (relative to displacement) is much higher due to differentiating a high frequency signal three times. The jerk response shown in Figure 5.6 implies that the only way to reduce uncomfortable vibrations is to lower the vehicle's speed to below resonance.

5.3 Accelerometers and other Vibration Sensors

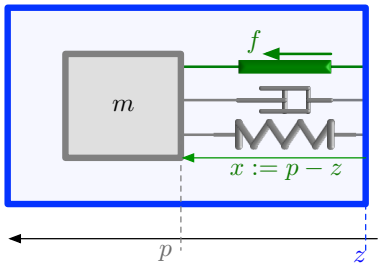
An accelerometer is a device that measures acceleration using the motion of a suspended proof mass inside an enclosure. Imagine you are in a car or a plane that is accelerating at a constant rate (this is of course only true for short times such as when the plane is taking off). Without looking outside, you can estimate the acceleration by holding up a pendulum and observing the angle the pendulum makes with the vertical axis. More sophisticated arrangements are required when measuring rapidly time varying accelerations as is possible with modern MEMS accelerometers. We begin by using basic vibration theory to examine the simplest possible accelerometer model, and then give a brief introduction to high-performance accelerometer design using the the principle of "force balance" feedback.

Consider the arrangement shown in Figure 5.7a. An enclosure (shown in blue) contains a "proof mass" m connected to the enclosure through a spring and damper[1]. In our simple model we assume

[1] Often the connection is through some elastic element like a flexure, which has some small inherent damping, but actual damping elements are typically not used.



(a) A mass m is suspended inside an enclosure and the position $x(t) := p(t) - z(t)$ of the mass relative to an enclosure can be measured by some type of sensor. This relative distance obeys the dynamics (5.17) of an MSD system "forced" by the inertial acceleration $\ddot{z}(t)$. Over a certain frequency range (called the bandwidth of the sensor), the signal $x(t)$ is proportional to the signal $\ddot{z}(t)$.



(b) Much better performance is possible when the accelerometer is operated using the "force balance" principle. An actuator (shown in green) exerts a force $f(t)$ between the enclosure and the proof mass. The force is generated by some feedback algorithm on the sensed distance $x(t) = p(t) - z(t)$ such that $x(t)$ is kept as close to equilibrium as possible. The inertial acceleration can then be "read off" from the force $f(t)$ required to keep the mass motionless relative to the enclosure. This scheme increases both the sensitivity and the bandwidth of the accelerometer.

Figure 5.7: Simplified schematics of accelerometers which use the motion of a "proof mass" inside an enclosure to estimate the acceleration $\ddot{z}$ of the enclosure in an inertial frame. The coordinate axis outside the enclosure represents an inertial frame (inaccessible to direct measurements), and the coordinate inside the accelerometer is the position of the proof mass relative to the enclosure, which is something that can be readily measured.

motion in only one dimension. The coordinate of the enclosure in an inertial frame is denoted by z. The purpose of this device is to measure the inertial acceleration $\ddot{z}(t)$ of the enclosure, even though we don't have access to an inertial position measurement like $z(t)$. Inside the enclosure however, we can measure the *relative distance* $p(t) - z(t)$ between the two inertial coordinates $p(t)$ of the proof mass and $z(t)$ of the enclosure. This can be implemented by a variety of schemes depending on the type of device[2].

Now since $p(t)$ is the position of the mass in an inertial frame, Newton's 2nd law says

$$m\,\ddot{p}(t) \;=\; -\,k\,(p(t)-z(t)) \;-\; c\,(\dot{p}(t)-\dot{z}(t))\,, \tag{5.16}$$

where we have assumed the origin of the coordinate system for measuring $p - z$ inside the enclosure is such that $p - z = 0$ corresponds to the spring's equilibrium. The only thing we can directly measure is the relative position $p(t) - z(t)$, so let's rewrite the equations in terms of this quantity

$$\begin{aligned} x(t) := p(t) - z(t) \qquad &\Leftrightarrow \qquad p(t) = x(t) + z(t) \\ &\Rightarrow \qquad \text{Equation (5.16) becomes} \qquad m\,(\ddot{x}(t) + \ddot{z}(t)) \;=\; -\,kx(t) - c\dot{x}(t). \end{aligned}$$

Rearranging this equation so that $\ddot{z}(t)$ is the input and dividing my m so that the parameters are expressed in terms of ζ and ω_n

$$\begin{aligned} m\ddot{x}(t) + c\dot{x}(t) + kx(t) \;&=\; -\,m\ddot{z}(t) \\ \Rightarrow \qquad \ddot{x}(t) + 2\zeta\omega_n\,\dot{x}(t) + \omega_n^2\,x(t) \;&=\; -\,a(t), \qquad\qquad a(t) := \ddot{z}(t). \end{aligned} \tag{5.17}$$

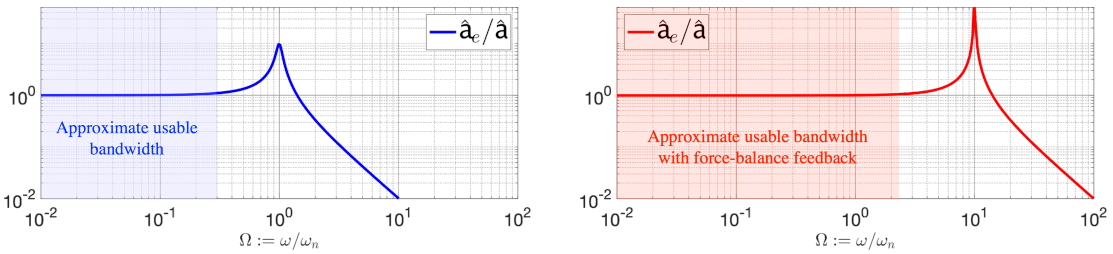
The inertial acceleration $\ddot{z}(t)$ has been relabeled as $a(t)$ to emphasize that it is the quantity to be estimated rather than $z(t)$.

Equation (5.17) represents an MSD system where the inertial acceleration $a(t)$ is the input and the (measurable) relative distance is the output. The complex frequency response of this system is

$$\frac{\hat{\mathsf{x}}}{\hat{\mathsf{a}}} \;=\; \frac{--1}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_n^2} \;=\; \frac{-1}{(\omega_n^2 - \omega^2) + j\;2\zeta\omega_n\;\omega} \;=\; \frac{-1/\omega_n^2}{(1-\omega^2/\omega_n^2) + j\;2\zeta\;\omega/\omega_n}$$

The input to this system is the inertial acceleration $a(t)$ that we want to estimate, but what is directly measured is the *output* x of the system. In the low frequency region $\Omega \ll 1$, this relation is

[2]For example, some MEMS accelerometers use "capacitive sensing" where the distance between two capacitor plates is measured by measuring the capacitance electrically.



(a) The frequency response from acceleration $\hat{a}$ to its estimate $\hat{a}_e = \omega_n^2 \hat{x}$ based on the proof mass displacement $\hat{x}$ (relative to enclosure). In the frequency range where the response is "flat", $a_e(t)$ is an "undistorted" estimate of $a(t)$. In this scheme, the flat region is primarily determined by the system's natural frequency ω_n.

(b) With proportional force-balance feedback, the estimate of acceleration $a_e(t) = f(t)/m$ is obtained from the force $f(t)$ required to keep the proof mass motionless. This scheme amounts to adding "virtual stiffness" to the system and therefore increases the bandwidth. It does however reduce the damping ratio, as can be seen from the larger resonance peak. This may cause stability problems inside the sensor.

Figure 5.8: Frequency responses from actual inertial acceleration a to its estimate a_e for (a) an accelerometer without force balance, and (b) an accelerometer with proportional force balance. The shaded areas represent the frequency regions over which the response is "flat", which is a requirement for the estimate of $a_e(t)$ to be an undistorted version of $a(t)$. The range of such frequencies is called the bandwidth of the sensor.

approximately a static gain, which can be inverted

$$\text{for } \Omega \ll 1, \qquad x(t) \approx \frac{-1}{\omega_n^2}\, a(t) \qquad \Rightarrow \qquad a_e(t) \approx -\omega_n^2\, x(t), \tag{5.18}$$

$$a_e(t) := \text{estimate of true acceleration } a(t)$$

Now the frequency response from actual acceleration $\hat{a}$ to its estimate $\hat{a}_e$ is

$$\frac{\hat{a}_e}{\hat{a}} \;=\; -\,\omega_n^2\,\frac{\hat{x}}{\hat{a}} \;=\; -\,\omega_n^2\,\frac{-1/\omega_n^2}{(1-\omega^2/\omega_n^2)+j\;2\zeta\;\omega/\omega_n} \;=\; \frac{1}{(1-\omega^2/\omega_n^2)+j\;2\zeta\;\omega/\omega_n}$$

This frequency response is shown in Figure 5.8a. In the low frequency regime, this frequency response is approximately 1, which means that the signal $a_e(t)$ will be similar to the signal $a(t)$ with little distortion. The frequency range over which this relationship is accurate is called the bandwidth of this sensor. This is the region where the frequency response is almost "flat", i.e. the gain is the same for all frequencies in this region. To fully understand this point requires understanding of system responses for more complex signals than just pure sinusoids. We will study this in Chapter 6 where we will demonstrate a beautiful relationship between the frequency response of a system and the Fourier transform of its input and output signals. For the time being, the point to keep in mind is that for high fidelity sensing in a certain frequency range, the frequency response of the sensor needs to be flat. For signals with frequency content beyond the flat region, the reconstruction of $a_e(t)$ using (5.18) will result in a "distorted" version of $a(t)$.

Figure 5.8a shows the flat region of the frequency response to be approximately $\Omega \leq 0.4$. Modern day MEMS accelerometers have a nominal natural frequency in the low KHz, about 1-5 KHz, which then restricts the bandwidth to approximately 0.4-2 KHz. One way to increase this bandwidth is of course is to increase the system's natural frequency ω_n by stiffening the spring or decreasing the mass. There is always a limit to how high ω_n can be (even if its possible to fabricate such a device) due to noise considerations which we now briefly discuss.

We can demonstrate the effect of noise on the reconstruction formula (5.18) as follows. Recall that the relative distance $x(t)$ is measured by some sort of sensing device which converts that quantity to a voltage, e.g. capacitive sensing in MEMS accelerometers. There is always a certain amount of electronic noise in any measurement, and we can model that as an additive noise signal

$$x_m(t) \;=\; x(t) \;+\; n(t),$$

where $x(t)$ is the *true* relative position, $n(t)$ is unknown measurement noise (i.e it cannot be measured directly), and $x_m(t)$ is the measured relative position. Since $x_m(t)$ is the only measurement we have,

we can try to reconstruct $a(t)$ using the formula (5.18) applied to it (instead of $x(t)$, which we don't have). Thus the estimate $a_z^{\rm e}(t)$ of the true acceleration $a(t)$ is

$$a_z^{\rm e}(t) \;=\; \omega_{\rm n}\, x_{\rm m}(t) \;=\; \omega_{\rm n}\big(x(t)+n(t)\big) \;=\; \omega_{\rm n}\, x(t) \;+\; \omega_{\rm n}\, n(t) \;=\; a(t) \;+\; \omega_{\rm n}\, n(t)$$

The term $\omega_{\rm n} n(t)$ is therefore the *error* in the estimate of $a(t)$, which is the quantity of interest. If $\omega_{\rm n}$ is chosen too high, then this reconstruction formula will significantly amplify the noise to produce large errors in the estimate.

Remark 5.1. There are further issues with the reconstruction of acceleration just described which are beyond the scope of the current discussion. We describe them briefly for a larger context.

1. The *phase frequency response* of the sensor's frequency response has not been examined. The phase response also plays a role in whether the estimate of acceleration is distorted or not. For the flat regions shown in Figures 5.8, the phase response is rather minimal.

2. The reconstruction of acceleration can be further enhanced over a somewhat larger bandwidth than described above by "deconvolution". This is a signal processing technique which "inverts" the frequency response over a slightly larger frequency bandwidth than those shown in Figures 5.8. It uses a frequency dependent version of the inversion formula (5.18), but implemented in the time domain. However, this comes at the expense of a time-delay in the estimate. Whether this delay is acceptable or not depends on the application. For data gathering such as characterizing a structure or other system identification applications, delays are tolerable. If the accelerometer is used as a sensor in a feedback control system, delays may be very detrimental to the stability of the system being controlled.

5.3.1 Accelerometers with Force Balance

The performance of the accelerometer can be greatly enhanced if operated with *force-balance feedback*. The *force-balance principle* is an ingenious idea that has been rediscovered many times over. In the case of the accelerometer, it operates as illustrated in Figure 5.7b. An actuator applies an additional force $f(t)$ to the proof mass. This force is determined by a *feedback algorithm* that senses the relative motion $x(t)$ and applies the force necessary to keep it at zero. In other words, it tries to keep the mass motionless relative to the enclosure, and the force required for this must be proportional to the acceleration of the enclosure. In this scheme, the estimate of $a(t)$ is actually obtained from the force $f(t)$ required to keep $x(t)$ near zero. At first, this scheme might seem to be unnecessarily circuitous, as well as needing the expense of an additional component (the actuator), and the design of the feedback algorithm. It does however have two very significant advantages; it improves the sensitivity and accuracy of the sensor (see Appendix 5.A), and it can increase its bandwidth by sometimes an order of magnitude. As already mentioned, many MEMS accelerometers have a natural frequency $\omega_{\rm n} \approx 1-5$ KHz, which results in a usable bandwidth of $\approx 0.4-2$ KHz when operated without feedback. High performance MEMS accelerometer that use force-balance feedback can easily achieve bandwidths in the 20-30 KHz range.

In this subsection, we will carry out the analysis to show how the accelerometer bandwidth can be increased using two different feedback schemes. This will also serve as an introduction to the analysis of composite dynamical systems that have more than one component. We will show how various frequency responses of "subsystems" can be combined using "block diagrams" to obtain descriptions of an overall composite system.

The proof mass subject to the additional force $f(t)$ (Figure 5.7b) has the following dynamics

$$\begin{aligned} m\ddot{x}(t)+c\dot{x}(t)+kx(t) \;&=\; -\,m\ddot{z}(t) \;+\; f(t) \\ \Rightarrow \qquad \ddot{x}(t)+2\zeta\omega_{\rm n}\dot{x}(t)+\omega_{\rm n}^2 x(t) \;&=\; -\,a(t) \;+\; \tfrac{1}{m}f(t), \qquad\qquad a(t):=\ddot{z}(t). \end{aligned} \tag{5.19}$$

The task of this force is to "cancel" the effects of inertial acceleration, so if $f(t)/m$ is equal to $a(t)$, then the mass will be motionless. Thus a read out of $f(t)/m$ would give the inertial acceleration. As already mentioned, the actuator force is determined by some feedback scheme on the displacement

$x(t)$. The simplest such scheme is to have it proportional to the mass' motion as a "restoring force"

$$f(t) \;=\; -\,\mathrm{g}\;x(t), \tag{5.20}$$

where g is some constant called the feedback gain. This kind of scheme is called proportional feedback. Note that $x(t)$ is a position measurement typically obtained electronically, and the actuator force $f(t)$ is determined by an electronic signal (e.g. a voltage command to an actuator), and therefore the constant g in (5.20) is typically the "gain" of some electronic amplifier (e.g. built with an Op-amp). Note the similarity of the force law (5.20) to that of a spring (Hook's law) where g is analogous to a spring constant The feedback scheme (5.20) can therefore be thought of as a "virtual spring", i.e. an electronically simulated spring force.

Now we combine the proof mass dynamics (5.19) with the feedback scheme (5.20) to analyze the overall dynamics

$$\ddot{x}(t)+2\zeta\omega_{\mathrm{n}}\dot{x}(t)+\omega_{\mathrm{n}}^2x(t) \;=\; -\,a(t)+\tfrac{1}{m}f(t) \;=\; -\,a(t)+\tfrac{1}{m}\;\mathrm{g}\;x(t) \tag{5.21}$$

$$\Rightarrow\qquad \ddot{x}(t)+2\zeta\omega_{\mathrm{n}}\;\dot{x}(t)+\left(\omega_{\mathrm{n}}^2+\tfrac{\mathrm{g}}{m}\right)x(t) \;=\; -\,a(t)$$

$$\Rightarrow\qquad \ddot{x}(t)+2\left(\zeta\tfrac{\omega_{\mathrm{n}}}{\omega_{\mathrm{g}}}\right)\omega_{\mathrm{g}}\;\dot{x}(t)+\omega_{\mathrm{g}}^2\;x(t) \;=\; -\,a(t), \qquad \omega_{\mathrm{g}}^2:=\omega_{\mathrm{n}}^2+\tfrac{\mathrm{g}}{m}=\tfrac{k}{m}+\tfrac{\mathrm{g}}{m}=\tfrac{k+\mathrm{g}}{m}. \tag{5.22}$$

Note how the constant g enters the coefficients of this equation as "spring constant" equivalent to an additional spring in parallel to k. Equation (5.22) is that of an MSD system forced by $a(t)$ and having a new natural frequency of ω_{g} and a new damping ratio of $\zeta(\omega_{\mathrm{n}}/\omega_{\mathrm{g}})$. Either g or ω_{g} can be considered a "design parameter" since specifying one determines the other

$$\omega_{\mathrm{g}}^2:=\omega_{\mathrm{n}}^2+\tfrac{\mathrm{g}}{m} \qquad\Leftrightarrow\qquad \mathrm{g}=m\left(\omega_{\mathrm{g}}^2-\omega_{\mathrm{n}}^2\right).$$

The estimate of $a(t)$ is obtained from the feedback force $f(t)$. If the actuator force insures the mass is nearly motionless, then the terms on the right hand side of equation (5.22) must add up to nearly zero, i.e. $f(t)/m$ must be almost equal to $a(t)$ to cancel it. Thus we define $f(t)/m$ as the estimate $a_e(t)$ of the actual acceleration $a(t)$

$$a_e(t) \;:=\; \tfrac{1}{m}f(t) \;\text{ (estimate of inertial acceleration } a(t) \text{ when using force-balance feedback } f(t))$$

The next question is how should ω_{g} (or equivalently the feedback gain g) be designed to improve the performance of the accelerometer? To answer the question, we need to find the frequency response from $a(t)$ to $a_e(t)$ and see what role the constant g plays. First note that $a_e(t)$ is proportional to the relative displacement $x(t)$

$$a_e(t) \;=\; \tfrac{1}{m}\;f(t) \;=\; -\,\tfrac{1}{m}\;\mathrm{g}\;x(t).$$

The frequency response from $a(t)$ to $x(t)$ can be read off from (5.22) and combined with this proportionality relation to give

$$\frac{\hat{\mathsf{a}}_e}{\hat{\mathsf{a}}} \;=\; \frac{\hat{\mathsf{a}}_e}{\hat{\mathsf{x}}}\;\frac{\hat{\mathsf{x}}}{\hat{\mathsf{a}}} \;=\; \frac{-\mathrm{g}}{m}\;\frac{-1}{(j\omega)^2+2\left(\zeta(\frac{\omega_{\mathrm{n}}}{\omega_{\mathrm{g}}}\right)\omega_{\mathrm{g}}\;(j\omega)+\omega_{\mathrm{g}}^2} \;=\; \frac{\omega_{\mathrm{g}}^2-\omega_{\mathrm{n}}^2}{(j\omega)^2+2\left(\zeta(\frac{\omega_{\mathrm{n}}}{\omega_{\mathrm{g}}}\right)\omega_{\mathrm{g}}\;(j\omega)+\omega_{\mathrm{g}}^2} \tag{5.23}$$

where the last equality $\mathrm{g}/m=\omega_{\mathrm{g}}^2-\omega_{\mathrm{n}}^2$ follows from the definition (5.22) of ω_{g}. This is a frequency response of MSD system with natural frequency ω_{g}. Recall that the bandwidth of such a system is largely determined by its natural frequency (refer to Figure 5.8a). This gives a design guideline; to increase the sensor's bandwidth by a factor of α, choose g such that $\omega_{\mathrm{g}}=\alpha\omega_{\mathrm{n}}$. Note however that if the natural frequency is increased by a factor of α, then the damping ratio $\zeta(\omega_{\mathrm{n}}/\omega_{\mathrm{g}})=\zeta/\alpha$ is actually decreased by the same factor of α.

Figure 5.8b shows this frequency response when ω_{g} is chosen to be a factor of 10 times ω_{n} for an attempt to increase the bandwidth by a factor of 10. Note that the bandwidth in indeed increased,

but since the damping ratio has also been decreased by also a factor of 10, a higher resonance peak is seen, which tends to reduce the range where the frequency response is flat. In this case the bandwidth appears to have increased by about a factor of 7, which is still significant. Note also that the "low frequency gain"

$$\text{at } \omega \approx 0, \qquad \frac{(\hat{\mathsf{f}}_{\mathrm{f}}/m)}{\hat{\mathsf{a}}_z} \;\approx\; \frac{\omega_g^2 - \omega_{\mathrm{n}}^2}{\omega_{\mathrm{g}}^2} \;=\; \frac{10^2\omega_{\mathrm{n}}^2 - \omega_{\mathrm{n}}^2}{10^2\omega_{\mathrm{n}}^2} \;=\; \frac{100-1}{100}\,\omega_{\mathrm{n}} \;\approx\; 1,$$

which also holds approximately in the flat frequency range. Thus over the usable bandwidth, the estimate $\hat{\mathsf{a}}_e$ is almost identical to the actual acceleration $\hat{\mathsf{a}}$.

Analysis of Force-Balance Feedback Schemes Using Block Diagrams

It is now useful to "zoom out" to get a better idea of the system "architecture" to really understand what we are doing. The "block diagram" illustrated in Figure 5.9 will be a useful guide. We will repeat the derivation of the overall frequency response (5.23) using this block diagram, and this analysis will be useful for any other more elaborate feedback scheme such as the Proportional-Integral (PI) feedback presented later. Block diagram analysis is a very useful tool, so it is worth spending some time on it using the accelerometer example as a motivation.

The proof mass dynamics (5.19) and its frequency response can be rewritten as follows (recall that $a_e(t) = \frac{1}{m} f(t)$)

$$\begin{aligned} u(t) &:= a_e(t) - a(t) \\ \ddot{x}(t) + 2\zeta\omega_{\mathrm{n}}\dot{x}(t) + \omega_{\mathrm{n}}^2 x(t) &= u(t) \end{aligned} \qquad \Leftrightarrow \qquad \begin{aligned} \hat{\mathsf{u}} &:= \hat{\mathsf{a}}_e - \hat{\mathsf{a}} \\ \frac{\hat{\mathsf{x}}}{\hat{\mathsf{u}}} &= \frac{1}{(j\omega)^2 + 2\zeta\omega_{\mathrm{n}}(j\omega) + \omega_{\mathrm{n}}^2} =: \mathsf{M}(\omega) \end{aligned}$$

The frequency response $\mathsf{M}(\omega)$ is to be regarded as a system that takes the input $\hat{\mathsf{u}} = \hat{\mathsf{a}}_e - \hat{\mathsf{a}}$ to produce the output $\hat{\mathsf{x}}$. The system $\mathsf{M}(\omega)$ is represented in Figure 5.9 as the grey block. Its output is labeled $\hat{\mathsf{x}}$, and note how the summing junction forces its input to be $\hat{\mathsf{u}} = \hat{\mathsf{a}} - \hat{\mathsf{a}}_e$. The signal $\hat{\mathsf{a}}$ (the inertial acceleration) is provided externally, while the signal $\hat{\mathsf{a}}_e$ (the acceleration estimate) is *fed back* from the output of the overall system. Each block in a block diagram represents an equation. For example

$$\boxed{\hat{\mathsf{u}} := \hat{\mathsf{a}}_e - \hat{\mathsf{a}}} \qquad \text{(represented by the summing junction)}$$

$$\frac{\hat{\mathsf{x}}}{\hat{\mathsf{u}}} = \frac{1}{(j\omega)^2 + 2\zeta\omega_{\mathrm{n}}(j\omega) + \omega_{\mathrm{n}}^2} =: \mathsf{M}(\omega) \qquad \Leftrightarrow \qquad \boxed{\hat{\mathsf{x}} \;=\; \mathsf{M}(\omega)\;\hat{\mathsf{u}}} \qquad \text{(represented by the grey block)}$$

The remaining blocks in Figure 5.9 are the force equation (5.20) and the estimate equation

$$f(t) = \mathrm{g}\; x(t) \qquad \Leftrightarrow \qquad \boxed{\hat{\mathsf{f}} = \mathsf{F}(\omega)\;\hat{\mathsf{x}}} \qquad \mathsf{F}(\omega) := \mathrm{g} \qquad \text{(green block } \mathsf{F}(\omega)\text{)} \tag{5.24}$$

$$a_e(t) = \tfrac{1}{m}\; f(t) \qquad \Leftrightarrow \qquad \boxed{\hat{\mathsf{a}}_e = \tfrac{1}{m}\;\hat{\mathsf{f}}} \qquad \text{(green block } \tfrac{1}{m}\text{)}$$

The phasor relations in this case are simple since force equation $\mathsf{F}(\omega) := \mathrm{g}$ is just a multiplication by the the constant g (proportional feedback). However, we defined it here more generally as a possibly ω-dependent multiplication to analyze more elaborate force feedbacks in the next scheme.

The feedback diagram 5.9 gives some intuition. The quantity $\hat{\mathsf{a}}_e - \hat{\mathsf{a}}$ is the *estimation error* since it is the difference between the acceleration estimate $\hat{\mathsf{a}}_e$ and the true acceleration $\hat{\mathsf{a}}$. If $\hat{\mathsf{a}}_e$ is higher than $\hat{\mathsf{a}}$, then the error $\hat{\mathsf{a}}_e - \hat{\mathsf{a}}$ positive. This positive error is then amplified in the "forward loop" first by going through the mass dynamics $M(\omega)$, and then going through $F(\omega) = -\mathrm{g}$. Thus it becomes a "negative feedback" that reduces $\hat{\mathsf{a}}_e$ as a corrective action. Let's test this intuition mathematically by calculating the frequency response from $\hat{\mathsf{a}}$ to $\hat{\mathsf{a}}_e$, which is the frequency response of the overall system depicted as the dashed block in the figure. The following calculation will be done for a general force law $\mathsf{F}(\omega)$ which will be useful in analyzing more elaborate feedback schemes.

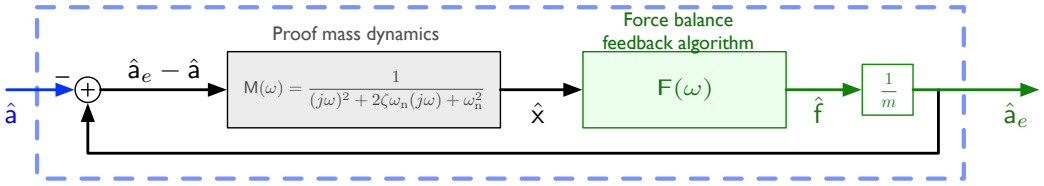


Figure 5.9: Any force-balance feedback scheme for the accelerometer can be represented using a block diagram. Each block in this diagram represents the frequency response of an input-output system. Each of the labels $\hat{a}$, $\hat{x}$, $\hat{f}_f$, $\hat{a}_e$ represent the phasors of the signals $a(t)$, $x(t)$, $f(t)$, and $a_e(t)$ respectively. The summation junction (depicted as the circle with the + sign) models the input to the proof mass dynamics which is "estimation error" $\hat{a} - \hat{a}_e$ between the actual acceleration $\hat{a}$ and its estimate $\hat{a}_e$. The description of the overall system as a feedback between subsystem blocks $\mathsf{M}(\omega)$ and $\mathsf{F}(\omega)$ guides the calculation of the overall "closed-loop" frequency response from a to $\hat{a}_e$ depicted as the blue dashed block.

To obtain the frequency response from $\hat{a}$ to $\hat{a}_e$, begin with the output signal $\hat{a}_e$ and "work backwards" through each of the blocks as follows

$$\begin{aligned} \hat{a}_e \;=\; \tfrac{1}{m}\,\hat{f} \;=\; \tfrac{1}{m}\,\mathsf{F}(\omega)\,\hat{x} \;=\; \tfrac{1}{m}\,\mathsf{F}(\omega)\,\mathsf{M}(\omega)\,\big(\hat{a}_e - \hat{a}\big) \;&=\; \tfrac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)\,\hat{a}_e \;-\; \tfrac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)\,\hat{a} \\ \Rightarrow \qquad \Big(1 - \tfrac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)\Big)\,\hat{a}_e \;&=\; -\,\tfrac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)\,\hat{a} \\ \Rightarrow \qquad \boxed{\frac{\hat{a}_e}{\hat{a}} \;=\; \frac{-\frac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)}{1 - \frac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)}} \end{aligned} \tag{5.25}$$

This elegant and simple formula gives the frequency response of the overall feedback system from the frequency responses of the subsystems' blocks $\mathsf{M}(\omega)$, $\mathsf{F}(\omega)$ and $\frac{1}{m}$. Note that the analysis leading to this formula is quite general since it didn't depend on the details of the systems $\mathsf{M}(\omega)$ or $\mathsf{F}(\omega)$.

Applying formula (5.25) to the proportional feedback scheme

$$\begin{aligned} \frac{\hat{a}_e}{\hat{a}} &= \frac{-\frac{1}{m}\,(\text{-g})\,\frac{1}{(j\omega)^2+2\zeta\omega_n(j\omega)+\omega_n^2}}{1 - \frac{1}{m}\,(\text{-g})\,\frac{1}{(j\omega)^2+2\zeta\omega_n(j\omega)+\omega_n^2}} = \frac{\frac{\text{g}/m}{(j\omega)^2+2\zeta\omega_n(j\omega)+\omega_n^2}}{1 + \frac{\text{g}/m}{(j\omega)^2+2\zeta\omega_n(j\omega)+\omega_n^2}} = \frac{\text{g}/m}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_n^2 \;+\; \text{g}/m} \\ &= \frac{\omega_g^2 - \omega_n^2}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_g^2}. \qquad \Big(\omega_g^2 := \omega_n^2 + \tfrac{g}{m} = \tfrac{k}{m} + \tfrac{g}{m} = \tfrac{k+g}{m}\Big) \end{aligned}$$

This is precisely the frequency response (5.23) obtained earlier. However the calculations used earlier to arrive at (5.23) were only possible because of the simple nature of the proportional feedback scheme. For more elaborate feedback laws (such as the PI feedback discussed next), it is necessary to use the formula (5.25).

Force Balance with Proportional-Integral (PI) Feedback

The force-balance proportional feedback scheme presented earlier is not the best one. More elaborate schemes can actually do much better, and those are indeed employed in high-performance accelerometers. One such scheme is described in this subsection.

A common feedback control scheme is the so-called Proportional-Integral (PI) control, which for the force-balance feedback would be of the form

$$f(t) \;=\; -\left(\text{g}_\text{p}\; x(t) \;+\; \text{g}_\text{i}\int_0^t x(\tau)\; d\tau\right), \tag{5.26}$$

where the "proportional gain" g_p and the "integral gain" g_i are design parameters to be determined. This feedback scheme has the following intuitive interpretation. The first term $-\text{g}_\text{p}x(t)$ is a restoring force proportional to the instantaneous position deviation $x(t)$. The second term represents a restoring force that is proportional to the integral of position deviation, i.e. the "past history" of this deviation. This gives better behavior of the feedback at slow frequencies. If the deviation $x(t)$

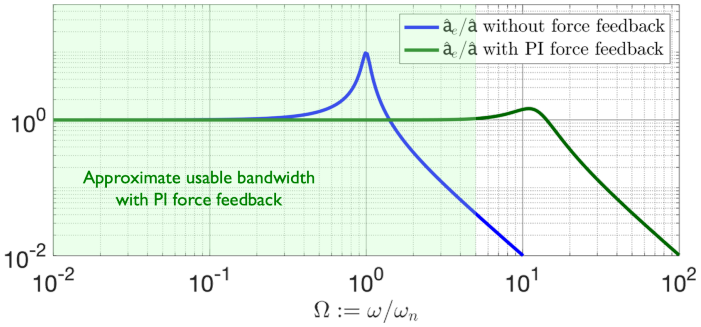


Figure 5.10: A comparison of the frequency response of the accelerometer without force-balance feedback (blue curve) and with Proportional-Integral (PI) force-balance feedback (green curve). The bandwidth is clearly increased by more than an order of magnitude without increasing the resonance peak that the proportional-feedback scheme of Figure 5.8b suffers from. The resonance peak of the PI scheme is in fact smaller than that without any feedback. The useful bandwidth of the PI feedback accelerometer is about 1.5 orders of magnitude than without feedback.

is small, but lasts a long time without switching sign, then the integral term will be large (despite $x(t)$ being small) and therefore result in a larger overall restoring force.

To understand the performance of this scheme, we need to look at the overall frequency response. For this we need the frequency response from measurement x to actuator forcing f. We know how to compute frequency responses for input-output systems described by ODEs. The relation (5.26) however involves an integral. This can be easily converted to a differential equation by differentiating both sides of (5.26) once to arrive at

$$\dot{f}(t) \;=\; -\,\mathrm{g_p}\,\dot{x}(t) \;-\; \mathrm{g_i}\,x(t).$$

This is an input-output ODE to which Theorem 4.2 is applicable, and the (complex) frequency response from x to f is readily obtained as

$$\mathsf{F}(\omega) \;:=\; \frac{\hat{\mathsf{f}}}{\hat{\mathsf{x}}} \;=\; -\,\frac{\mathrm{g_p}\,(j\omega) \;+\; \mathrm{g_i}}{(j\omega)}. \tag{5.27}$$

Note that in contrast to the case of proportional feedback (5.24), $\mathsf{F}(\omega)$ is no longer constant in ω, but it is frequency dependent. It is this additional design freedom that gives it superior properties.

To obtain the frequency response of this system, we simply use formula (5.25) using $\mathsf{F}(\omega)$ from (5.27) and $\mathsf{M}(\omega)$ from the mass dynamics

$$\begin{aligned}
\frac{\hat{\mathsf{a}}_e}{\hat{\mathsf{a}}} \;&=\; \frac{-\frac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)}{1-\frac{1}{m}\mathsf{F}(\omega)\mathsf{M}(\omega)} \;=\; \frac{-\frac{1}{m}\left(-\frac{\mathrm{g_p}(j\omega)+\mathrm{g_i}}{(j\omega)}\right)\frac{1}{(j\omega)^2+2\zeta\omega_\mathrm{n}(j\omega)+\omega_\mathrm{n}^2}}{1-\frac{1}{m}\left(-\frac{\mathrm{g_p}(j\omega)+\mathrm{g_i}}{(j\omega)}\right)\frac{1}{(j\omega)^2+2\zeta\omega_\mathrm{n}(j\omega)+\omega_\mathrm{n}^2}} \\
&=\; \frac{\mathrm{g_p}\;(j\omega)+\mathrm{g_i}}{m(j\omega)\left((j\omega)^{+}2\zeta\omega_\mathrm{n}(j\omega)+\omega_\mathrm{n}^2\right) \;+\; \mathrm{g_p}(j\omega)+\mathrm{g_i}} \\
&=\; \frac{\mathrm{g_p}\;(j\omega)+\mathrm{g_i}}{m\;(j\omega)^3+2m\zeta\omega_\mathrm{n}\;(j\omega)^2+(m\omega_\mathrm{n}^2+\mathrm{g_p})\,(j\omega)+\mathrm{g_i}}
\end{aligned}$$

Note that the denominator of this response is 3rd order in ω. This is because the differential equation for the mass is 2nd order, while the differential equation for the force feedback is 1st order. It can be shown that in a composite system, the order of the overall differential equation is the sum of the orders of the differential equations of all subsystems, which in this case is $2+1=3$. This results in a 3rd order frequency response.

The design parameters in this scheme are the proportional and integral gains $\mathrm{g_p}$ and $\mathrm{g_i}$ respectively. This is a standard exercise in *feedback control system design* and is beyond the scope of the current discussion. By way of illustration, Figure 5.10 shows an example design with $\mathrm{g_p} = 100$

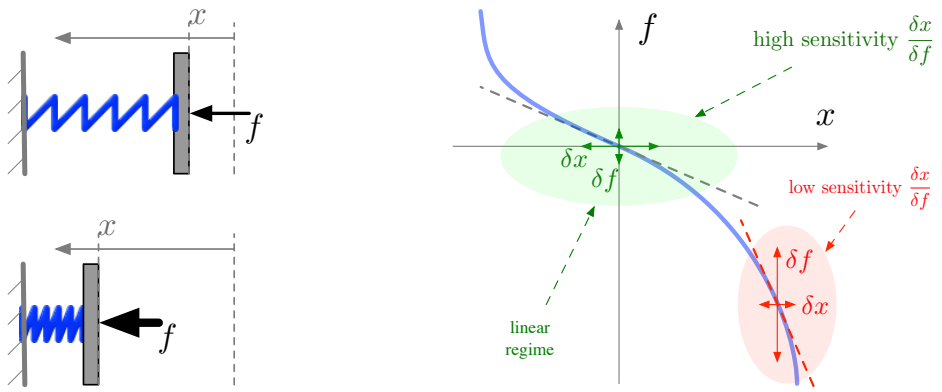


Figure 5.11: A force or pressure measurement device (e.g. load cell) operates on the principle of measuring the deflection of an elastic element in response to an applied force. Deflection can be measured with a variety of sensors such as strain gauges, interferometers, optical encoders, etc. (*Left*) A model of such a device is a (nonlinear spring) whose deflection x can be related to the applied force using a calibrated (x, f) curve (*right*) describing the deflection as a function of the applied force. The (x, f) curve has the typical characteristics of a "stiffening spring". The sensitivity of this force sensor is given by $\delta x/\delta f$, where it is highest in the "linear regime" where small deviations in f give the largest possible deviation in x (since the accuracy of sensing is determined by the displacement sensors). For such a sensor, good accuracy is in general not guaranteed when large forces are applied. The sensor thus has a limited "dynamic range".

and $g_i = 1000$. The figure clearly shows a significant bandwidth increase of about 1.5 orders of magnitude over the case of no feedback. Unlike the case of pure proportional feedback, the increase in the resonance peak that occurs with that scheme is avoided here. In fact, the PI scheme reduces the resonance peak as well as increases the bandwidth simultaneously.

Appendix

5.A The Force-Balance Principle

The force-balance principle is a method for enhancing the performance of sensors. It is an ingenious technique that has been rediscovered repeatedly in many different types of sensors. The basic idea is summarized here for force/acceleration sensors, but other instances of its use are briefly mentioned towards the end to emphasize that this is a widely applicable method.

A common force sensor is the "load cell". It is basically a relatively stiff elastic element which deforms under the application of a force, its deformation is measured by a displacement sensors (e.g. a strain gauge), which then gives an indirect measurement of the force. A schematic of such a sensor is shown in Figure 5.11. The elastic element is modeled as a (nonlinear) spring. When a force f is applied, it deforms by x amount. The applied force can then be obtained from the deformation x using a calibrated (x, f) curve (shown on the right in the figure) for that particular elastic element.

The (x, f) curve shown in Figure 5.11 has the typical behavior of a "stiffening" spring, the larger the displacement, the larger the incremental force δf required for an incremental displacement δx. If this sensor is to be operated over a large range of possible forces, the entire (x, f) curve needs to be known by some calibration procedure.

A key consideration in the design of any sensor is its sensitivity. Referring to the (x, f) curve, if a force of value $\bar{f}$ is applied, the spring will deform with displacement $\bar{x}$ given by the (x, f) curve. If the displacement sensor has an error of δx, then the error in determining the force is given by

$$\delta f \approx f'(\bar{x})\, \delta x,$$

where $f'(\bar{x})$ is the derivative of that curve at the displacement $\bar{x}$. In regions (of displacement) where $f'(\bar{x})$ is large, a larger error in determining the force is obtained compared to where $f'(\bar{x})$ is small.

Figure 5.12: The force sensor operated using the force-balance principle. The deflection x is sensed, and is *fed back* to produce the actuator force f_a required to counteract the externally applied force f so displacement is kept at zero $x \approx 0$. This required actuator force is then the measurement of the externally applied force. In this manner, the sensor is always operating in the linear regime shown in Figure 5.11, which is the regime of highest sensitivity. When operated in this manner, the force-balance sensor has much higher sensitivity over a large dynamic range.

The sensitivity of the sensor at displacement $\bar{x}$ is defined as the inverse of the error ratio

$$\text{sensor's sensitivity at } \bar{x} \;:=\; \left.\frac{\delta x}{\delta f}\right|_{\text{at } \bar{x}} \;=\; \frac{1}{f'(\bar{x})}.$$

If a small change in the force produces a large change in displacement, then the accuracy in determining the force is higher for a fixed displacement sensor's accuracy. This is depicted in Figure 5.11 for two cases where the $f'(\bar{x})$ is low (high sensitivity) and $f'(\bar{x})$ is high (low sensitivity). For a typical stiffening spring type sensor, the area of highest sensitivity is around zero displacement, sometimes referred to as the "linear regime" of sensing. There two major difficulties with such a sensor.

1. The sensor requires extensive calibration. The (x, f) curve needs to be obtained accurately through a collection of experiments where a calibrated force is applied and the resulting displacement recorded. Each elastic element is likely to have an (x, f) curve that is slightly different from another.

2. The behavior of many elastic elements is that of a "stiffening" spring, so that when larger forces are applied, the derivative f' becomes larger (stiffer), and the sensitivity $1/f'$ becomes lower. This limits the "dynamic range" of the sensor to a maximum force for which acceptable sensitivity is required.

A force sensor that uses the force-balance principle resolves both of the two difficulties listed above. Such a sensor is depicted in Figure 5.12. It requires an actuator capable of providing an opposing force f_a, as well as a *feedback* mechanism that works to keep the elastic element at close to zero deformation. The externally applied force f then is equal to the required actuator force f_a, and therefore the force measurement is read off from f_a. This means that the sensor is always operating in the region around $\bar{x} \approx 0$, which is the region of highest sensitivity. The dynamic range of force measurements is thus determined by the largest force f_a the actuator can provide rather than the stiffening of the elastic element. Thus the sensor is always operating in the "linear regime" regardless of the applied force. Furthermore, the only calibration required is the single number of the stiffness $f'(0)$ around zero deflection rather than the entire (x, f) curve.

The scenarios described above are static ones, and they show that a force-balance sensor gives the highest sensitivity over as large an applied force magnitudes as the actuator is able to provide. There is also the very important dynamic property of significantly increasing the bandwidth of the sensor as demonstrated for accelerometers in Section 5.3.1.

As mentioned, the force-balance principle is used in many other types of sensors. Below is another notable example.

Hot-wire Anemometers in Constant-temperature Mode: The Constant Temperature Anemometer (CTA)

A hot-wire anemometer is a local gas velocity sensor. A very thin wire placed in a gas flow is resistively heated. Due to convective heat transfer to the flow, the faster the flow, the lower the temperature of the wire when heated with constant power. The wire's temperature can be read off

from its resistance, which can be determined from its voltage and current. The bandwidth of this sensor is related to its heat capacity, which determines how fast its temperature can react to changes in gas velocity. Thin wires of diameters in the few microns have a bandwidth in the 100's of Hz when operated in this mode.

An alternative mode of operation is in "constant temperature mode", and is more commonly referred to as the Constant Temperature Anemometer (CTA). As mentioned, the wire's temperature can be sensed electronically by monitoring its voltage and current. A feedback circuit then alters the power delivered to the wire so that its temperature stays constant regardless of flow velocity fluctuations (and consequently convective heat transfer fluctuations to the gas). The power needed to keep the temperature constant is then "read off" to compute the gas' velocity. The force-balance principle here is that the wire's temperature is akin to the position of the spring in a load cell. The variable power applied to keep that temperature constant is akin to the actuator force applied to keep the load cell deflection constant. When hot wires are operated in this mode, their bandwidth can be dramatically increased by 2 or 3 orders of magnitude to the range of 10's of KHz. The CTA was invented in the 1960's, and remains as the state-of-the-art measurement technique for high temporal resolution turbulence studies. The dynamic response of the anemometer can be analyzed with similar frequency response techniques used earlier for force-balanced accelerometers.

Chapter 6

General Forced Vibrations

The superposition property of linear systems enables response analysis for general and complicated-looking input signals when they are written as a sum of simpler signals. A periodic forcing signal can be expressed as a Fourier series sum, and the superposition property implies that the response to the periodic input is the total sum of the responses to each of the individual harmonics in its Fourier series. More generally, any not-necessarily periodic forcing signal can be written as an integral with respect to its Fourier transform, also known as its "spectrum". The systems we deal with have the remarkable property that the spectrum of the output is simply the product of the input's spectrum with the system's frequency response. This "frequency domain" analysis gives much insight into the resulting vibrations with the "shape" of the system's frequency response encoding all of its dynamical properties. Transfer functions evaluated on the imaginary axis give the system's frequency response, and the poles and zeros of the transfer function give insight, as well as design guidelines, for the system dynamics.

6.1 Introduction: Linearity and the Superposition Principle

In previous chapters we analyzed the steady-state response of constant-coefficient ODEs to a single sinusoidal input. The main tools used were phasor analysis and the frequency response. These same mathematical tools can be used in much greater generality to analyze the response to more complex signals. The key property that enables this generalization is that of *linearity*, otherwise known as the *principle of superposition*. This property implies that once we know the response to some simple inputs like pure sinusoids, then we can obtain the response to inputs that are made up of *all possible linear combinations* of such sinusoids. Most signals of interest can be represented as linear combinations of a finite or infinite number of sinusoids, and when we combine the frequency response with this linearity property, we obtain a powerful method for describing the response of a system to very general types of inputs.

The differential equations we have encountered so far have this linearity property. This follows from the following two observations.

1. The derivative operation (of any order) is a linear operation. Indeed, take any two signals $u_1(t)$ and $u_2(t)$, and any linear combination of them

   $$\frac{d}{dt}\Big(a\ u_1(t)\ +\ b\ u_2(t)\Big)\ =\ a\ \frac{d}{dt}u_1(t)\ +\ b\ \frac{d}{dt}u_2(t),$$

   where a and b are arbitrary (scalar) coefficients. Applying this repeatedly, we see that differentiation of any order k is also a linear operation

   $$\frac{d^k}{dt^k}\Big(a\ u_1(t)\ +\ b\ u_2(t)\Big)\ =\ a\ \frac{d^k}{dt^k}u_1(t)\ +\ b\ \frac{d^k(t)}{dt^k}u_2(t).$$

2. Any linear combination of derivative operations (of any order) is also a linear operation. For

example

$$\begin{aligned}
&\left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \beta_0\right)\Big(a\,u_1(t) + b\,u_2(t)\Big) \\
&= \beta_m \tfrac{d^m}{dt^m}\Big(a\,u_1(t) + b\,u_2(t)\Big) + \cdots + \beta_1 \tfrac{d}{dt}\Big(a\,u_1(t) + b\,u_2(t)\Big) + \beta_0\Big(a\,u_1(t) + b\,u_2(t)\Big) \\
&= a\left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \beta_0\right)u_1(t) + b\left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \beta_0\right)u_2(t)
\end{aligned}$$

There are other operations on signals that also have the linearity property (integration for example), but the above two facts are the ones we need for analysis system responses.

Consider now an input-output system described by and ODE of the form used earlier in Theorem 4.2. The differential equation can be rewritten slightly to emphasize the linearity property as follows

$$\begin{aligned}
&\alpha_n y^{(n)}(t) + \cdots + \alpha_1 y^{(1)}(t) + \alpha_0 y(t) = \beta_m u^{(m)}(t) + \cdots + \beta_1 u^{(1)}(t) + \beta_0\, u(t) \\
\Leftrightarrow\quad &\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) y(t) = \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) u(t) \qquad (6.1)
\end{aligned}$$

Thus the left hand side is a linear operation on the output y and similarly the right hand side is a linear operation on the input u. This has a remarkable consequence: suppose we know the outputs y_1 and y_2 to the inputs u_1 and u_2 respectively. If a new input to the system is some linear combination of u_1 and u_2, there is no need to re-solve the system to find the output to this new input. The new output is simply the same linear combination of the respective outputs y_1 and y_2. This is a fundamental principle that we state precisely next.

The Superposition (Linearity) Principle

Let u_1 and u_2 be two inputs to a system producing outputs y_1 and y_2 respectively, i.e.

$$\begin{aligned}
&\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) y_1(t) = \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) u_1(t), \\
\text{and}\quad &\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) y_2(t) = \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) u_2(t).
\end{aligned}$$

Let a, b be scalars, then the output corresponding to the linear combination input $a\,u_1 + b\,u_2$ is the *same linear combination* of the respective outputs, i.e.

$$\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right)\Big(a\,y_1(t) + b\,y_2(t)\Big) = \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right)\Big(a\,u_1(t) + b\,u_2(t)\Big).$$

The proof of this principle is done by simply adding the top two equations to obtain the third one. This also implies that the linearity principle applies not only to linear combinations of two input, but any number of inputs. More precisely, if $a_1, \ldots, a_n$ are any scalars, and $u_1, \ldots, u_n$ are input signals with corresponding outputs $y_1, \ldots, y_n$, then the input $a_1\,u_1 + \cdots + a_n\,u_n$ will produce the output $a_1\,y_1 + \cdots + a_n\,y_n$. We will see that this argument applies also to combinations of an *infinite number* of signals as will be the case of Fourier series studied later in this chapter.

Example 6.1. Given a Mass-Spring-Damper system with a forcing that is made up of sinusoids of two different frequencies

$$m\,\ddot{x}(t) + c\,\dot{x}(t) + k\,x(t) = \mathsf{f}_1 \cos(\omega_1 t + \theta_1) + \mathsf{f}_2 \cos(\omega_2 t + \theta_2), \qquad (6.2)$$

we can use the linearity property to find the steady-state response by finding the response to each of the two forcing terms separately. Let x_1 and x_2 be the steady-state response to the following two inputs

$$\begin{aligned}
m\,\ddot{x}_1(t) + c\,\dot{x}_1(t) + k\,x_1(t) &= \mathsf{f}_1 \cos(\omega_1 t + \theta_1) \quad \Rightarrow \quad x_1(t) = |\hat{\mathsf{x}}_1|\, \cos(\omega_1 t + \angle \hat{\mathsf{x}}_1), \qquad (6.3) \\
m\,\ddot{x}_2(t) + c\,\dot{x}_2(t) + k\,x_2(t) &= \mathsf{f}_2 \cos(\omega_2 t + \theta_2) \quad \Rightarrow \quad x_2(t) = |\hat{\mathsf{x}}_2|\, \cos(\omega_2 t + \angle \hat{\mathsf{x}}_2), \qquad (6.4)
\end{aligned}$$

where the phasors $\hat{\mathsf{x}}_1$ and $\hat{\mathsf{x}}_2$ of $x_1(t)$ and $x_2(t)$ must be obtained from the frequency response at the *two different frequencies* ω_1 and ω_2 as follows

$$\begin{array}{l} \hat{\mathsf{x}}_1 \;=\; \mathsf{H}(\omega_1)\; \mathsf{f}_1 e^{j\theta_1} \\ \hat{\mathsf{x}}_2 \;=\; \mathsf{H}(\omega_2)\; \mathsf{f}_2 e^{j\theta_2} \end{array}, \qquad \text{where} \qquad \mathsf{H}(\omega) \;=\; \frac{1}{m(j\omega)^2 + c(j\omega) + k}.$$

The linearity property then implies that the solution to (6.2) is simply the sum of the solutions to (6.3) and (6.4)

$$x(t) \;=\; x_1(t) + x_2(t) \;=\; |\hat{\mathsf{x}}_1|\; \cos(\omega_1 t + \angle \hat{\mathsf{x}}_1) \;+\; |\hat{\mathsf{x}}_2|\; \cos(\omega_2 t + \angle \hat{\mathsf{x}}_2).$$

The example above is not restricted to inputs with two different frequencies. The same linearity property is applied to an input with any number of sinusoids with different frequencies (we now use complex exponentials rather than cosines for simpler expressions of amplitude and phase relations)

$$m\,\ddot{x}(t) + c\,\dot{x}(t) + k\,x(t) \;=\; \sum_{k=1}^{N} \hat{\mathsf{f}}_k\; e^{j\omega_k t} \tag{6.5}$$

$$\Rightarrow \qquad x(t) \;=\; \sum_{k=1}^{N} \hat{\mathsf{x}}_k\; e^{j\omega_k t}, \qquad \text{where} \qquad \hat{\mathsf{x}}_k = \mathsf{H}(\omega_k)\; \hat{\mathsf{f}}_k. \tag{6.6}$$

Therefore, provided we can evaluate the frequency response $\mathsf{H}(\omega)$ at any desired frequency, the formula (6.6) gives a simple method for determining the steady-state response. The linearity property can be used in this manner for any system described by an input-output ODE of the form (6.1), and the corresponding frequency response is readily obtained from Theorem 4.2.

The linearity property can be applied to even more general inputs than finite linear combinations like (6.5). We will see in the next section that it can be applied to infinite series such as the Fourier series decomposition of arbitrary periodic signals. We can go even further and think of a *continuum of linear combinations*. We will also see in the next sections that if the input is a "linear combination" of a "continuum" of components with different frequencies

$$u(t) \;=\; \int_{-\infty}^{\infty} \hat{\mathsf{u}}(\omega)\; e^{j\omega t} d\omega, \qquad \Big(\hat{\mathsf{u}}(\omega) \text{ is the "component" of } u(t) \text{ at frequency } \omega\Big), \tag{6.7}$$

then the corresponding output can be expressed as a similar continuum linear combination in terms of the frequency response

$$y(t) \;=\; \int_{-\infty}^{\infty} \hat{\mathsf{y}}(\omega)\; e^{j\omega t} d\omega, \qquad \text{where} \qquad \hat{\mathsf{y}}(\omega) = \mathsf{H}(\omega)\; \hat{\mathsf{u}}(\omega). \tag{6.8}$$

Note the similarity between the expression (6.6) and (6.8). The first is a sum over a finite set of frequencies $\{\omega_k\}_{k=1}^N$, while the second is an integral over all possible frequencies. The expression (6.8) is the most general one since it allows for signals that contain *all* possible frequencies $-\infty < \omega < \infty$. The reasoning behind (6.8) is again the superposition property. If we plug in the Fourier representation (6.7) of the input in the system's differential equation

$$\left(\alpha_n \frac{d^n}{dt^n} + \cdots + \alpha_1 \frac{d}{dt} + \alpha_0\right) y(t) \;=\; \left(\beta_m \frac{d^m}{dt^m} + \cdots + \beta_1 \frac{d}{dt} + \alpha_0\right) \left(\int_{-\infty}^{\infty} \hat{\mathsf{u}}(\omega)\; e^{j\omega t} d\omega\right),$$

and observe that $\hat{\mathsf{y}}(\omega)\, e^{j\omega t} = H(\omega)\; \hat{\mathsf{u}}(\omega)\, e^{j\omega t}$ is the response to each "component" $\hat{\mathsf{u}}(\omega)\, e^{j\omega t}$ of the input, then the total response $y(t)$ is given by the integral (6.8).

We will see that the *Fourier Transform* provides the right mathematical tool to justify the above arguments, to describe signals with rich "content", and to characterize the response of dynamical systems to such complex-looking signals.

6.2 Response in Frequency Domain: Fourier Series & Transforms

We first treat *periodic signals* using their Fourier *series* and express a system's response using Fourier series and the frequency response. We then consider a system's response to *arbitrary signals* using the Fourier *transform*.

We will use the "complex form" of Fourier series where any T-periodic signal can be written as the infinite sum

$$u(t) \;=\; \sum_{k=-\infty}^{\infty} \hat{\mathsf{u}}_k \; e^{jk\bar{\omega}t}, \qquad\qquad \bar{\omega} = \tfrac{2\pi}{T}, \tag{6.9}$$

where $\bar{\omega}$ is the fundamental frequency and the term $e^{jk\bar{\omega}t}$ of frequency $k\bar{\omega}$ is called the k'th harmonic component of $u(t)$. The complex coefficients $\{\hat{\mathsf{u}}_k\}$ have the conjugacy property $\hat{\mathsf{u}}_{-k} = \hat{\mathsf{u}}_k^*$ when $u(t)$ is a real function (which is always the case here). Appendix 6.A contains a more extensive background on Fourier series and justifications of the previous statements.

Now consider the response of a general ODE input-output system to the periodic input (6.9). We know from the frequency response Theorem 4.2 that the response to each term $\hat{\mathsf{u}}_k\, e^{jk\bar{\omega}t}$ is $\hat{\mathsf{y}}_k\, e^{jk\bar{\omega}t}$ where the relation between the phasors $\hat{\mathsf{u}}_k$ and $\hat{\mathsf{y}}_k$ is given by the frequency response, i.e. for any k

$$\begin{aligned} \left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) \hat{\mathsf{y}}_k\, e^{jk\bar{\omega}t} \;&=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) \hat{\mathsf{u}}_k\, e^{jk\bar{\omega}t}, \\ \text{where} \quad \hat{\mathsf{y}}_k \;&=\; \mathsf{H}(k\bar{\omega})\, \hat{\mathsf{u}}_k. \end{aligned}$$

Combining this with the linearity property, we conclude that the response to the infinite sum (6.9) is also an infinite sum of responses at the same respective frequencies $k\bar{\omega}$, i.e.

$$\begin{aligned} \left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) \left(\sum_{k=-\infty}^{\infty} \hat{\mathsf{y}}_k\, e^{jk\bar{\omega}t} \right) \;&=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) \left(\sum_{k=-\infty}^{\infty} \hat{\mathsf{u}}_k\, e^{jk\bar{\omega}t} \right), \\ \text{where} \quad \hat{\mathsf{y}}_k \;&=\; \mathsf{H}(k\bar{\omega})\, \hat{\mathsf{u}}_k, \qquad\qquad k = \ldots, -2, -1, 0, 1, 2, \ldots. \end{aligned}$$

We summarize this conclusion next.

Theorem 6.2. *Consider the input-output system*

$$\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) y(t) \;=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) u(t),$$

where the input $u(t)$ is a T-periodic signal with Fourier series

$$u(t) = \textstyle\sum_{k=-\infty}^{\infty} \hat{\mathsf{u}}_k\; e^{jk\bar{\omega}t}, \qquad\qquad \bar{\omega} = 2\pi/T. \tag{6.10}$$

The steady-state output $y(t)$ is also T-periodic with Fourier series

$$y(t) = \textstyle\sum_{k=-\infty}^{\infty} \hat{\mathsf{y}}_k\; e^{jk\bar{\omega}t},$$

and its Fourier series coefficients $\{\hat{\mathsf{y}}_k\}_{k=-\infty}^{\infty}$ are given by

$$\hat{\mathsf{y}}_k \;=\; \mathsf{H}(k\bar{\omega})\; \hat{\mathsf{u}}_k, \qquad\qquad \mathsf{H}(\omega) \;:=\; \tfrac{\beta_m (j\omega)^m \,+\, \cdots \,+\, \beta_1 (j\omega) \,+\, \beta_0}{\alpha_n (j\omega)^n \,+\, \cdots \,+\, \alpha_1 (j\omega) \,+\, \alpha_0}, \tag{6.11}$$

where $\mathsf{H}(\omega)$ is the complex frequency response of the system.

Note that the theorem statement refers to the coefficients $\{\hat{\mathsf{y}}_k\}_{k=-\infty}^{\infty}$ for both positive and negative k. However, the conjugacy relation $\hat{\mathsf{y}}_{-k} = \hat{\mathsf{y}}_k^*$ holds, and therefore the products $\hat{\mathsf{y}}_k = \mathsf{H}(k\bar{\omega})\, \hat{\mathsf{u}}_k$ in (6.11) need only be computed for $k \geq 0$. Then the coefficients $\hat{\mathsf{y}}_k$ for negative k are obtained from the conjugacy relation.

Theorem 6.2 has a nice graphical interpretation illustrated in Figure 6.1. For each signal, say $u(t)$, the set of Fourier series coefficients $\{\hat{\mathsf{u}}_k\}_{k=0}^{\infty}$ are depicted as a "bar graph" over the frequency axis, with bars of size $\hat{\mathsf{u}}_k$ at the harmonic frequencies $k\bar{\omega}$ respectively. This bar graph is referred to

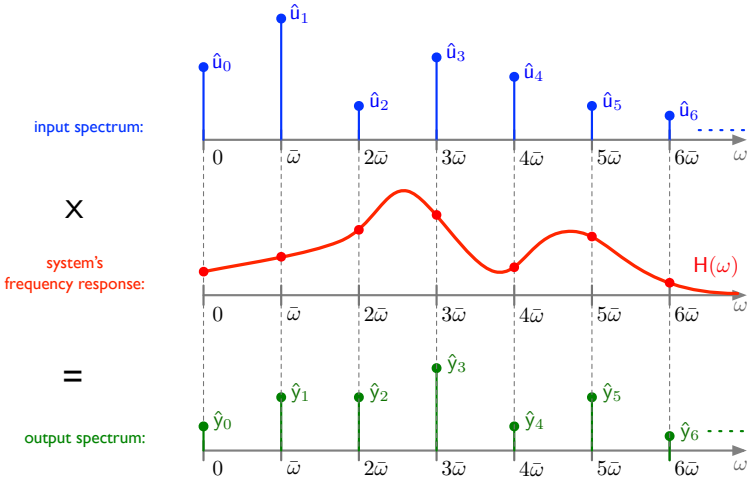


Figure 6.1: A graphical depiction of Theorem 6.2 expressing the relation between the Fourier series coefficients of the input and output, and the frequency response $\mathsf{H}(\omega)$ of the system. The Fourier coefficients of a signal are represented as a "bar graph" over the harmonic frequencies $\{k\bar{\omega}\}_{k=0}^{\infty}$. This bar graph is referred to as the spectrum of the signal. The frequency response relation $\hat{y}_k = \mathsf{H}(k\bar{\omega})\ \hat{u}_k$ is represented here as the k'th Fourier coefficient $\hat{u}_k$ of the input multiplied by the frequency response $\mathsf{H}(k\bar{\omega})$ *evaluated at the k'th harmonic frequency* $k\bar{\omega}$ to yield the k'th Fourier coefficient $\hat{y}_k$ of the output. Note that all quantities depicted above are complex numbers. Theorem 6.2 can therefore be summarized by the simple statement: output spectrum = input spectrum × system's frequency response.

as the spectrum of the signal. Note that since the coefficients are complex numbers, this bar graph is somewhat of a cartoon since it does not show both magnitude and phase information. The point of this diagram is to show the relation $\hat{y}_k = \mathsf{H}(k\bar{\omega})\ \hat{u}_k$, which states that the spectrum of the output is obtained by multiplying the spectrum of the input by the values of the system's frequency response $\mathsf{H}(\omega)$ *evaluated at the harmonic frequencies* $\{k\bar{\omega}\}_{k=0}^{\infty}$.

Example 6.3. The Fourier series of a square wave u is worked out in Example 6.20 as

$$u(t) \ = \sum_k \hat{u}_k\ e^{jk\bar{\omega}t} \ = \ \sum_{k \text{ odd}} \tfrac{-2j}{\pi k}\, e^{jk\frac{2\pi}{T}t}, \qquad \bar{\omega} = \tfrac{2\pi}{T}.$$

Thus the spectrum of u has non-zero components at frequencies $\bar{\omega}k$ for $k = \ldots, -3, -1, 1, 3, \ldots$. The amplitude of the k'th component is $\frac{-2j}{\pi k}$. Note that the amplitude decreases with increasing harmonic index k.

Let this signal be the input to a MSD system with frequency response (4.14)

$$\mathsf{H}(\omega) \ = \ 1 \Big/ \left(\left(1 - \left(\frac{\omega}{\omega_{\mathrm{n}}}\right)^2 \right) + j\ 2\zeta \left(\frac{\omega}{\omega_{\mathrm{n}}}\right) \right). \tag{6.12}$$

If y is the steady-state output of this system to the square wave input, then y is a T-periodic signal with Fourier series

$$y(t) \ = \ \sum_{k \text{ odd}} \mathsf{H}(k\bar{\omega})\ \tfrac{-2j}{\pi k}\ e^{jk\bar{\omega}t},$$

where $\mathsf{H}(k\bar{\omega})$ is the frequency response function (6.12) evaluated at the input's harmonics frequencies $\{\bar{\omega}k\}$. Thus the Fourier series coefficients of y are the *product* of the numbers $\mathsf{H}(k\bar{\omega})$ and the Fourier series coefficients of u. Depending on where the resonances of H are, different harmonics of the square wave will be prominent in the output. Figure 6.2 shows two different possibilities depending on how the natural frequency of the system ω_{n} is related to the fundamental frequency $\bar{\omega}$ of the square wave.

6.2.1 The Fourier Transform

The Fourier series is only applicable to signals that are periodic. The *Fourier transform* is a generalization of Fourier series that gives a "frequency decomposition" of non-periodic signals that have

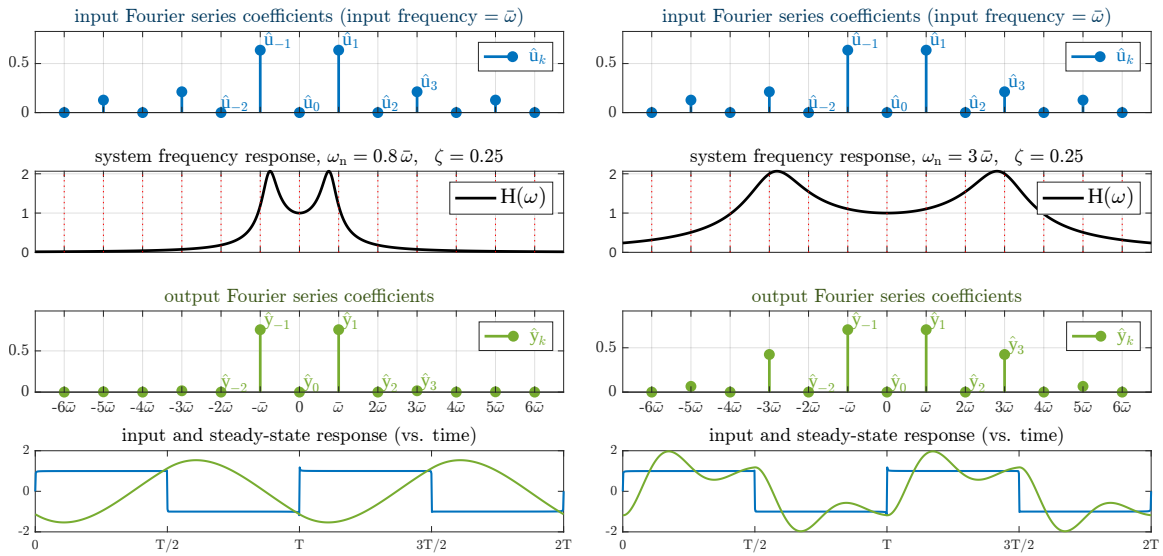


Figure 6.2: Fourier series analysis of a square wave and corresponding system responses. A square wave is the input to two different MSD systems with the natural frequencies and damping ratios shown (ω_n shown in terms of the square wave fundamental frequency $\bar{\omega}$). The Fourier spectrum of the square wave input (*top*) is shown along with the two different systems frequency responses (*second from top*), and the resulting Fourier spectra of the two corresponding outputs (*third from top*). Only the magnitudes (and not phases) are shown. The input and steady-state outputs are shown (*bottom*) versus time. Note the significant 3rd harmonic content in the second case since the system's resonance amplifies this harmonic.

"frequency content" which is not necessarily harmonic. To motivate the definitions, recall that a Fourier series of the form

$$u(t) \;=\; \sum_{k=-\infty}^{\infty} \hat{u}_k \; e^{jk\bar{\omega}t},$$

is necessarily a periodic signal of period $T = 2\pi/\bar{\omega}$. Only harmonics $k\bar{\omega}$ of the fundamental frequency $\bar{\omega}$ are present in this sum. A more general expression that gives a much larger variety of signals is to take linear combinations over "all frequencies" possible. One way to do this is to replace the summation with an integral and range over all possible frequencies

$$u(t) \;=\; \int_{-\infty}^{\infty} \hat{u}(\omega) \; e^{j\omega t} \; d\omega. \tag{6.13}$$

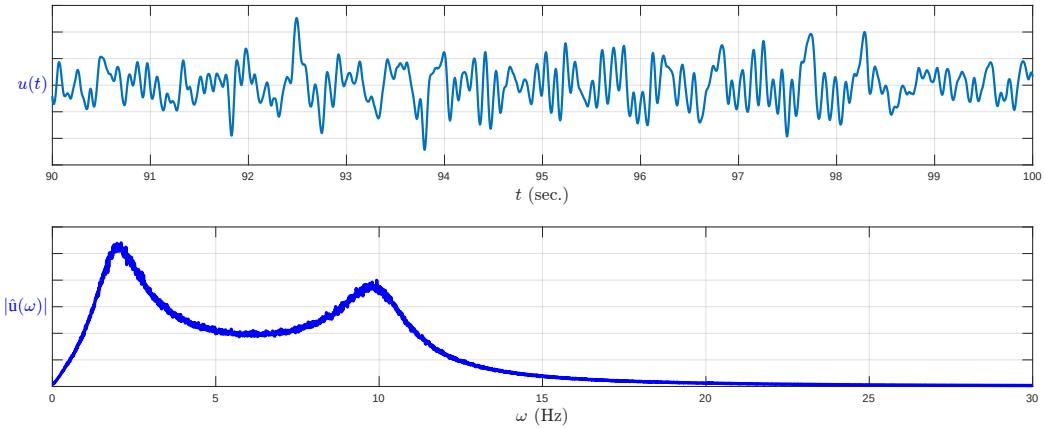
Note that now instead of Fourier coefficients $\hat{u}_k$, we now have a whole *function* of "coefficients" $\hat{u}(\omega)$ defined over all frequencies $-\infty < \omega < \infty$. There is no such thing as a fundamental frequency or harmonics in this expression any more since all possible frequencies are allowed. The function $\hat{u}(\omega)$ can be given the interpretation of the "frequency content" of the signal $u(t)$.

If a signal can be represented in terms of the integral (6.13), how should the "coefficient function" $\hat{u}(\omega)$ be found from the signal $u(t)$? The answer is given by the *Fourier transform* which we define together with its relation to the representation (6.13) in the following lemma.

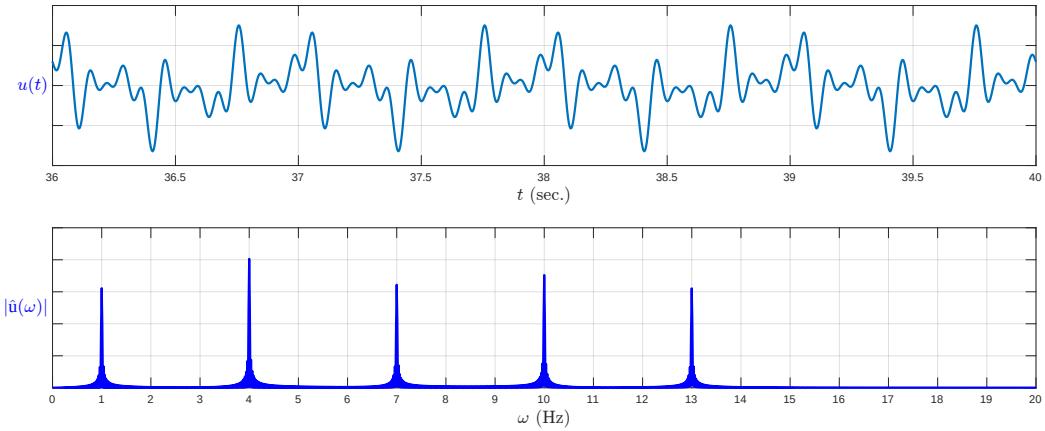
Lemma 6.4. *Given a signal $u(t)$ defined for all $-\infty < t < \infty$, its* Fourier transform *is a (complex-valued) function defined for all frequencies $-\infty < \omega < \infty$ by*

$$\hat{u}(\omega) \;:=\; \tfrac{1}{2\pi}\int_{-\infty}^{\infty} u(t) \; e^{-j\omega t} \; dt. \tag{6.14}$$

The signal $u(t)$ can be reconstructed from its Fourier transform $\hat{u}(\omega)$ using the inverse Fourier transform *(6.13).*



(a) A non-periodic signal $u(t)$ and the magnitude $|\hat{u}(\omega)|$ of its Fourier transform (its spectrum). The spectrum shows that the signal has frequency content mostly between 0 and 15-20 Hz. It also has two prominent peaks around 2 and 10 Hz. None of these features are obvious from observing the time signal $u(t)$, but they are clearly seen in its spectrum $|\hat{u}(\omega)|$.



(b) A periodic signal $u(t)$ and the magnitude $|\hat{u}(\omega)|$ of its Fourier transform (its spectrum). By visual inspection of $u(t)$, it may be possible to ascertain that this signal is periodic with period 1. However, the spectrum $|\hat{u}(\omega)|$ has very "sharp" lines, indicating that this signal is a sum of five pure sinusoids. In addition, those sinusoids are all at frequencies which are integer multiples of 1 Hz, thus the fundamental frequency is 1 Hz. Here we see the fundamental frequency, as well as the 4'th, 7'th, 10'th and 13'th harmonic.

Figure 6.3: Examples of, and contrasts between the Fourier transform of a non-periodic (a) and a periodic (b) signal.

The proof of this lemma is not difficult, but is given in Appendix **??** so as not to distract from the main point here which is the interpretation of the Fourier transform and its uses.

The Fourier transform is an incredibly useful tool in Engineering and Science. It is the basis of what is termed "frequency domain" analysis, which is the suite of methods for analysis of a wide variety of oscillatory phenomena including mechanical vibrations, electrical and electromagnetic oscillations, and acoustics amongst many others. The Fourier transform has many properties whose study would take much more space than what is presented here. For the current topic, we are mainly interested in the relationship between Fourier transforms of signals and the frequency response of systems described by ODEs, which is the subject of the next section.

Example 6.5. An example of non-periodic signal and its numerically computed Fourier transform is shown in Figure 6.3a. Note that the mathematical definition (6.14) of the Fourier transform requires knowing the signal over all time $-\infty < t < \infty$ (the range of integration in that formula). For any real signal, only a finite duration of a signal can be recorded. To compute a reasonable approximation of

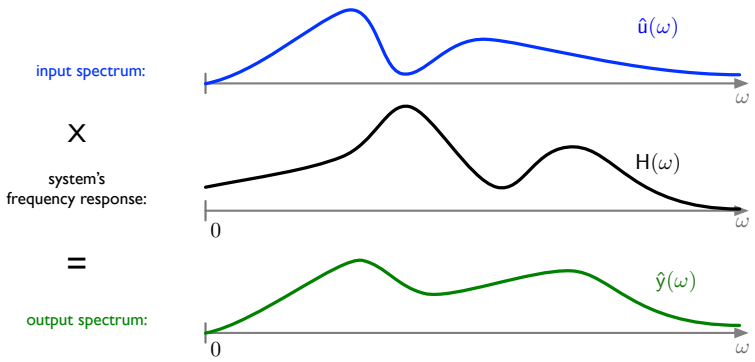


Figure 6.4: A graphical depiction of Theorem 6.7. The spectrum (the Fourier transform) $\hat{u}(\omega)$ of the input signal gets multiplied (complex number multiplication) by the frequency response $H(\omega)$ of the system to produce the spectrum $\hat{y}(\omega)$ of the output signal. The diagram above is a cartoon since each of the three functions $\hat{u}(\omega)$, $H(\omega)$ and $\hat{y}(\omega)$ are complex-valued, and full graphical depiction would require separating into magnitude and phase graphs.

the integral (6.14), a signal needs to be recorded over a long time interval[1]. The signal in Figure 6.3 is collected over a time span of length 1000 seconds, and sampled at the rate of 1 millisecond. The integral (6.14) is then approximated by a truncated Riemann sum. The Riemann sum can be calculated in the usual way, but is more commonly done with an algorithm called the Fast Fourier Transform (FFT), which exploits certain symmetries in the sum to substantially speed up the computation.

Figure 6.3 shows a small section of the signal for clarity, as well as its Fourier transform. Unlike a periodic signal, this signal has no obvious repeating pattern. This is because its spectrum (its Fourier transform) is "broad", i.e. it has no clear sharp peaks at evenly spaced harmonics as a periodic signal would. The spectrum shows that it has significant frequency content roughly between 0 and 15 Hz. Note that while it is difficult to say something quantitative about the signal by looking at its time traces, more insight into its frequency content is obtained from looking at the spectrum.

Example 6.6. An example of *periodic* signal and its numerically computed Fourier transform is shown in Figure 6.3b. The spectrum of periodic signals always has sharp well-defined lines at (some of the) harmonics of a fundamental frequency. It is almost zero at all other frequencies.

6.2.2 Steady-State Response to Arbitrary Inputs

Theorem 6.2 describes the response of an input-output system to periodic signals in terms of their spectrums and the frequency response of the system. The next theorem gives a similar statement for the response to arbitrary inputs (not just periodic ones) in terms of the Fourier transforms of the input and output signals, and the frequency response of the system. The main difference is that now we have responses at *all* frequencies.

Theorem 6.7. *Consider the input-output system*

$$\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) y(t) \;=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) u(t). \tag{6.15}$$

Denote the Fourier transforms of the input and output respectively by

$$\hat{u}(\omega) \;=\; \tfrac{1}{2\pi}\int_{-\infty}^{\infty} u(t)\; e^{-j\omega t}\; dt, \qquad \hat{y}(\omega) \;=\; \tfrac{1}{2\pi}\int_{-\infty}^{\infty} y(t)\; e^{-j\omega t}\; dt. \tag{6.16}$$

Those Fourier transforms are related by the (complex) multiplicative relation

$$\text{for all } \omega, \qquad \boxed{\hat{y}(\omega) \;=\; H(\omega)\; \hat{u}(\omega),} \qquad H(\omega) \;:=\; \tfrac{\beta_m (j\omega)^m \;+\; \cdots \;+\; \beta_1 (j\omega) \;+\; \beta_0}{\alpha_n (j\omega)^n \;+\; \cdots \;+\; \alpha_1 (j\omega) \;+\; \alpha_0}, \tag{6.17}$$

where $H(\omega)$ is the complex frequency response of the system (6.15).

[1] To make this more precise one needs to consider the "frequency content" of the signal. We do not address these approximation issues here.

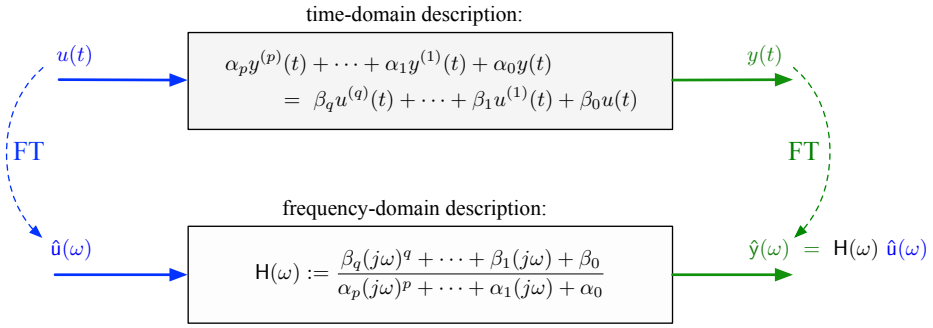


Figure 6.5: The contrast between the "time-domain" description of a system and its "frequency domain" description. In the time domain (*top*), the input $u(t)$ is a non-homogenous term in an ODE, while the output $y(t)$ is the steady-state solution to this ODE. Thus the ODE is a system (can also be thought of as an algorithm) which takes a signal $u(t)$ to a signal $y(t)$ by solving a differential equation. The frequency domain description is much simpler. The relation between the Fourier transforms $\hat{u}(\omega)$ and $\hat{y}(\omega)$ of the input and output respectively is simply given by the complex multiplication $\hat{y}(\omega) = \mathsf{H}(\omega)\hat{u}(\omega)$ with the system's frequency response $\mathsf{H}(\omega)$. The system's frequency response is a complex, rational function of ω whose coefficients can be immediately read-off from the ODE.

The boxed formula in (6.17) has a simple interpretation. It should be read as follows

$$\text{spectrum of output signal} \;=\; \text{frequency response of system} \;\times\; \text{spectrum of input signal.}$$

This statement is illustrated graphically in Figure 6.4. Note the contrast between this and Figure 6.1, where only a discrete set of frequencies are involved. In the general case here, the spectra of the input and output could be non-zero at all frequencies.

Proof. Starting from the differential equation (6.15) and substituting Fourier representations (inverse F.T.) (6.13) for both $u(t)$ and $y(t)$

$$\begin{aligned}
\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_0\right) \int_{-\infty}^{\infty} \hat{y}(\omega)\, e^{j\omega t}\, d\omega \;&=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_0\right) \int_{-\infty}^{\infty} \hat{u}(\omega)\, e^{j\omega t}\, d\omega \\
\Rightarrow \quad \int_{-\infty}^{\infty} \hat{y}(\omega) \left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_0\right) e^{j\omega t}\, d\omega \;&=\; \int_{-\infty}^{\infty} \hat{u}(\omega) \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_0\right) e^{j\omega t}\, d\omega \\
\Rightarrow \quad \int_{-\infty}^{\infty} \underbrace{\hat{y}(\omega) \left(\alpha_n (j\omega)^n + \cdots + \alpha_0\right)}_{=:\ \hat{a}(\omega)} e^{j\omega t}\, d\omega \;&=\; \int_{-\infty}^{\infty} \underbrace{\hat{u}(\omega) \left(\beta_m (j\omega)^m + \cdots + \beta_0\right)}_{=:\ \hat{b}(\omega)} e^{j\omega t}\, d\omega \\
\Rightarrow \text{ for all } t, \qquad a(t) \;&=\; b(t),
\end{aligned}$$

where the time functions $a(t), b(t)$ are the inverse Fourier transforms of $\hat{a}(\omega), \hat{b}(\omega)$ respectively. Since the two functions are equal $a(t) = b(t)$ for all time, this means that their Fourier transforms are equal for all frequencies ω, and therefore

$$\begin{aligned}
\text{for all } \omega, \qquad \hat{a}(\omega) = \hat{b}(\omega) \qquad \Rightarrow \qquad \hat{y}(\omega) \left(\alpha_n (j\omega)^n + \cdots + \alpha_0\right) \;&=\; \hat{u}(\omega) \left(\beta_m (j\omega)^m + \cdots + \beta_0\right) \\
\Rightarrow \qquad \hat{y}(\omega) \;&=\; \tfrac{\beta_m (j\omega)^m + \cdots + \beta_0}{\alpha_n (j\omega)^n + \cdots + \alpha_0}\ \hat{u}(\omega),
\end{aligned}$$

which is the statement (6.17). □

To appreciate the utility of this theorem, and the contrast between time-domain analysis and frequency-domain analysis, the reader should examine Figure 6.5. The time-domain description (6.15) is an input-output relation that involves solving an ODE. Given any input $u(t)$, the right hand side of the ODE is specified, and by solving the ODE and waiting until steady state is reached, the output is obtained. The frequency domain description however is much simpler. The Fourier transforms of the input and output are related by a simple multiplication of complex functions, without any differentiation operations. This is another manifestation of frequency response and phasor analysis,

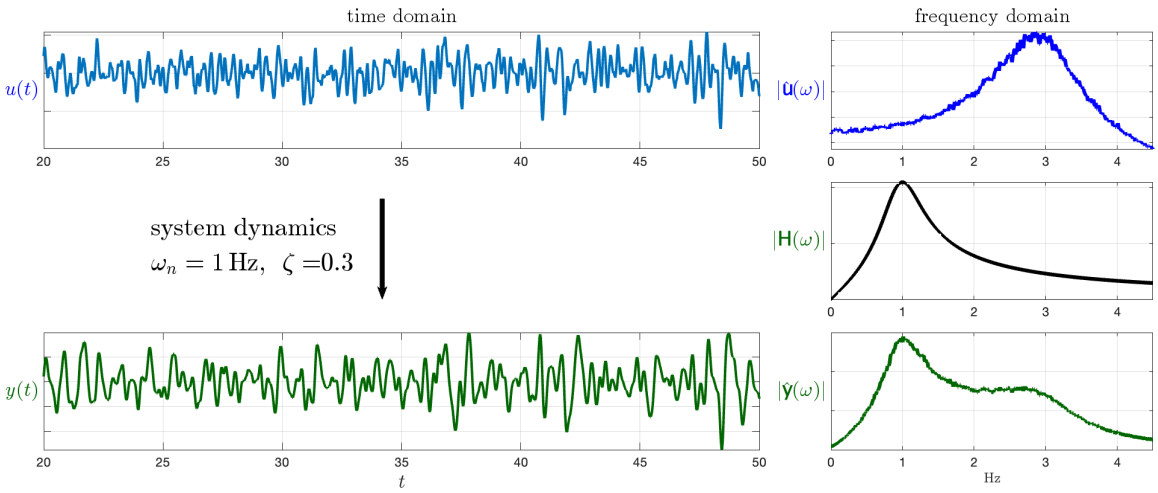


Figure 6.6: (Example 6.8): Time-domain (*left*) and frequency-domain (*right*) descriptions of non-periodic input and output signals. It's difficult to get much insight into the signals and the system dynamics in the time domain. On the other hand, the frequency-domain picture is much clearer. The spectrum of the input gets multiplied by the system's frequency response to produce the spectrum of the output. This gives insight into how the frequency content of the input, and the system's resonances contribute to the frequency content of the output.

where differential equations in the time domain are converted to algebraic equations in the frequency domain.

Example 6.8. Figure 6.6 illustrates an example of the utility of frequency domain analysis. A particular system (whose ODE is not shown, but it is a 2nd order input-output ODE) is fed the input $u(t)$ shown. This input is not periodic, but a more complex-looking signal. The ODE is solved (numerically), and the resulting output is shown as $y(t)$. This output signal is also not periodic and complex-looking. From looking at the time traces of the input and the output signals, it is difficult to get either a qualitative or a quantitative understanding of the signals and the system's action on them.

The figure also shows the Fourier transforms of the input and output signals, as well as the frequency response of the system. The frequency content of both signals, as well as the simple multiplicative relation between the signal's spectrum and the system's frequency response is now clearly visible. In particular it is clear how the output spectrum magnitude $|\hat{y}(\omega)|$ is the product of the magnitude $|\hat{u}(\omega)|$ of the input's spectrum and $|H(\omega)|$, the magnitude of the frequency response of the system. Note also how the input spectrum has only small magnitudes around 1 Hz, while the system has a strong resonance at 1 Hz, resulting in the output spectrum having high frequency content near both the system's resonance as well as where the input has the bulk of its frequency content.

6.3 Impulse (Shock) Response and Convolutions

As seen in the previous section, the frequency response provides a much simpler analysis tool (compared to the original ODE) of a system whose inputs are signals that persist over "long times". Long times are mathematically idealized as $-\infty < t < \infty$, and the signals were either periodic as represented with Fourier series, or more general as represented using the Fourier transform. In this section, we want to study a system's response to signals that "start" at a given time rather than signals that have been entering the system for long times in the past. In other words, we want to now study the *transient response*. The mathematical idealization is therefore to study signals and system responses over the time interval $0 \leq t < \infty$. It is particularly useful for studying stability of systems where responses may grow unboundedly with time and therefore the setting $-\infty < t < \infty$ would not be appropriate.

One particularly important type of input is the so-called shock or impulsive input. A typical "shock response" experiment is depicted in Figure 6.7. An impulsive force is applied to the system,

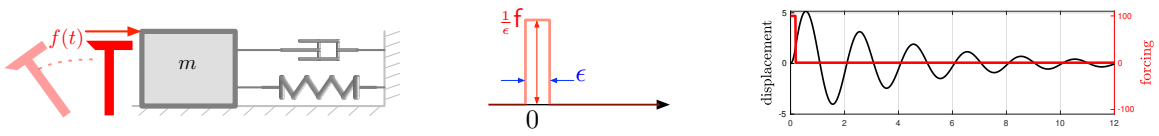


Figure 6.7: (*Left*) A typical shock response experiment depicted as "hammering" on the system. The hammering force is modeled as a pulse of duration ϵ and height f/ϵ so that the *total impulse* (integral of force in time) is f. The duration ϵ should be as short as possible. (*Right*) The response is then recorded and the system's parameters can be identified from this response.

and its responses (e.g. displacements or velocities or any other quantity of interest) are recorded as signals over time. It turns out that if the applied impulse is of *very short* duration, then this "impulse response" will actually allow us to characterize the system response to any other input u starting at a particular time. We will show that the response of the system to any other input is a "convolution" operation of its impulse response with that input. First let's define impulsive inputs more precisely.

A very short duration impulse is depicted in Figure 6.7 (middle) with duration ϵ and magnitude f/ϵ. The unit approximate impulse δ_ϵ is such an impulse with $\mathsf{f} = 1$

$$\delta_\epsilon(t) \;:=\; \begin{cases} 1/\epsilon, & 0 \le t \le \epsilon, \\ 0, & t > \epsilon. \end{cases}$$

Note that no matter what ϵ is, the total impulse of δ_ϵ (the integral of force over time) is 1, so we refer to δ_ϵ has an impulse of "strength" 1 (note that the total integral is independent of ϵ). We are interested in very short impulses, and ideally we want $\epsilon \to 0$. We can now define the impulse response of the system.

Definition 6.9. *Let h_ϵ be the response of the following system when the input is the approximate unit impulse δ_ϵ, i.e.*

$$\left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) h_\epsilon(t) \;=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \alpha_0\right) \delta_\epsilon(t), \tag{6.18}$$

with zero initial conditions $h_\epsilon(0) = 0$, …, $h_\epsilon^{(n\text{-}1)}(0) = 0$.

The (unit) impulse response *of the system is defined as the limit $h(t) := \lim_{\epsilon \to 0} h_\epsilon(t)$.*

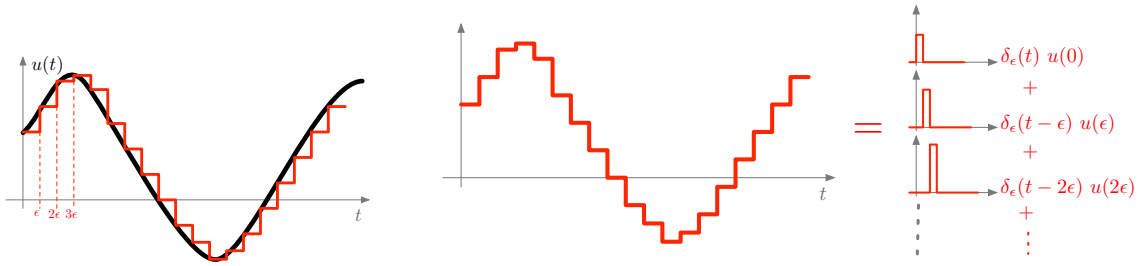
The unit impulse response $h(t)$ can be thought of as the response to an impulse of infinitesimally short duration, but of strength 1. Since such an infinitesimally short duration pulse is non-physical (it is an idealization), we have to obtain its response as the limit of responses to shorter and shorter duration impulses. Mathematically we should point out that the limit of the input $\lim_{\epsilon \to 0} \delta_\epsilon(t)$ is not well defined, but the limit of the responses $\lim_{\epsilon \to 0} h_\epsilon(t)$ is usually well defined.

Note that the above is just a definition, and we have not yet described how this differential equation can be solved. This will be done in the next subsection after introducing partial fraction expansions. For now, we assume that the unit impulse response can be found (perhaps even by an experiment), and show how we can use it to obtain the (zero initial conditions) response to an arbitrary input.

Figure 6.8 illustrates how an arbitrary function $u(t)$ can be approximated using a "modulated train" of approximate impulses $\delta_\epsilon(t)$ as

$$u(t) \;\approx\; \sum_{k=0}^{\infty} u(k\epsilon) \;\left(\epsilon\, \delta_\epsilon(t - k\epsilon)\right), \qquad t \ge 0. \tag{6.19}$$

Here $u(k\epsilon)$ are the values of the signal u at grid points spaced ϵ apart, and $\epsilon\, \delta(t - k\epsilon)$ are pulses of width ϵ, total area ϵ, and starting at the grid point $k\epsilon$. As the figure indicates, the approximation gets better as $\epsilon \to 0$. This should remind the reader of approximations to integrals using Riemann sums.



(a) A signal $u(t)$ is approximated by a piecewise-constant signal shown here in red.

(b) The piecewise-constant signal can be written as a sum of delayed impulses $\delta_\epsilon(t-k\epsilon)$ each with height given by the samples $u(k\epsilon)$ of the original signal $u(t)$.

Figure 6.8: Illustration of the approximation formula (6.19) where a signal is approximated by a piece-wise constant signal, which in turn can be written as a sum of delayed pulses with varying heights.

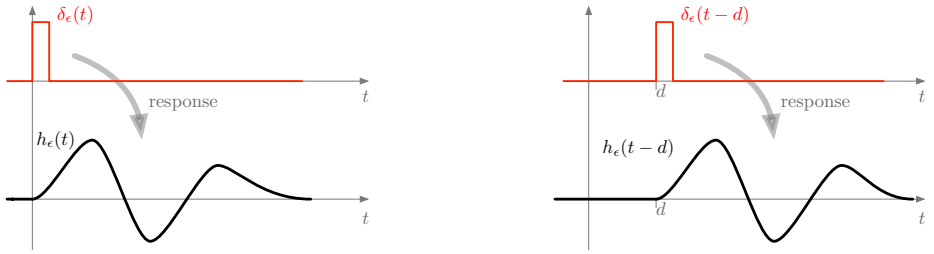


Figure 6.9: Illustration of the *time invariance* property shown here when the input is the approximate impulse δ_ϵ. (*Left*) $h_\epsilon(t)$ is the response to the impulse $\delta_\epsilon(t)$. (*Right*) If the input is a now delayed by d time units, then the output is exactly the same as h_ϵ but also delayed by exactly d time units, i.e. the response to the input $\delta_\epsilon(t-d)$ is $h_\epsilon(t-d)$.

Now to compute the response to an input like (6.19), we can use linearity and decompose it as the sum of responses to each *delayed* impulse $\delta_\epsilon(t-k\epsilon)$. Note that because of linearity, the response to the scaled impulse $\epsilon\,\delta_\epsilon(t)$ is simply $\epsilon\,h_\epsilon(t)$, i.e. the response $h_\epsilon(t)$ scaled by ϵ. It remains to characterize the response to a *delayed* impulse. This is done using an additional property that ODEs with constant coefficients poses, which is *time invariance*.

Definition 6.10. *Let u and y be the input and output signal respsectively of a system. The system is called* time invariant *if the response to a delayed version of the input u is the same output delayed by exactly the same amount. More precisely*

$$\begin{aligned} &\textit{if the input } u \textit{ produces the output } y, \textit{ i.e.} && u \longrightarrow y, && \\ &\textit{then for any } d>0, \textit{ the delayed input } u_{\mathrm{d}} \textit{ produces } y_{\mathrm{d}} && u_{\mathrm{d}} \longrightarrow y_{\mathrm{d}}, && \begin{aligned} u_{\mathrm{d}}(t) &:= u(t-d), \\ y_{\mathrm{d}}(t) &:= y(t-d). \end{aligned} \end{aligned}$$

This property is illustrated in Figure 6.9. In particular, when the input is the delayed impulse $\delta_\epsilon(t-k\epsilon)$, then the response is simply the delayed approximate impulse response $h_\epsilon(t-k\epsilon)$. Now we can combine the linearity and the time-invariance properties to find the general form of a system's response using convolutions. There are two steps to this.

1. Write an arbitrary input u in terms of its approximation (6.19)

$$u(t) \;\approx\; \sum_{k=0}^{\infty} \epsilon\, u(k\epsilon)\; \delta_\epsilon(t-k\epsilon), \qquad t \geq 0. \tag{6.20}$$

2. We know that the response to each $\delta_\epsilon(t-k\epsilon)$ is $h_\epsilon(t-k\epsilon)$, the system's impulse response *delayed by* $k\epsilon$. This together with the linearity property give the total output to be

$$y(t) \;\approx\; \sum_{k=0}^{\infty} \epsilon\, u(k\epsilon)\; h_\epsilon(t-k\epsilon). \tag{6.21}$$

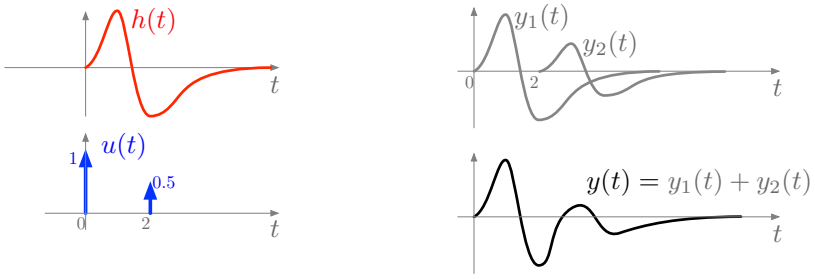


Figure 6.10: When the input $u(t)$ is a collection of impulses of different strengths arriving at different times, the systems response is the sum of all the corresponding scaled and delayed copies of the impulse response $h(t)$. Here y_1 is the response to the first impulse in u, and since that impulse has strength 1, y_1 is just a copy of the systems impulse response $h(t)$. y_2 is the response to the second impulse in u, and it is just $h(t)$ scaled by 0.5 and delayed by 2 time units. The total response $y(t)$ is the sum of y_1 and y_2.

This expression has the following interpretation. At time $k\epsilon$, the system receives an impulse of strength $\epsilon\, u(k\epsilon)$. Its response to this input is $\epsilon\, u(k\epsilon)\ h_\epsilon(t-k\epsilon)$, i.e. the impulse response $h_\epsilon(t)$ scaled by $\epsilon\, u(k\epsilon)$ and delayed according to the time $k\epsilon$ when that impulse arrives. At the next time $(k+1)\epsilon$, another impulse arrives, and the corresponding scaled and delayed impulse response is added to the previous response, and so on. Thus the expression (6.21) represents the accumulated response to all the delayed and modulated impulses arriving at the input at different times and with different strengths. This idea is illustrated in Figure 6.10.

Now observe that (6.21) is the Riemann sum approximation to an integral. Taking the limit $\epsilon \to 0$ gives the integral

$$\lim_{\epsilon\to 0} \epsilon \sum_{k=0}^{\infty} u(k\epsilon)\ h_\epsilon(t-k\epsilon) \ = \ \int_0^\infty u(\tau)\ h(t-\tau)\ d\tau$$

$$\Rightarrow \qquad y(t) \ = \ \int_0^\infty u(\tau)\ h(t-\tau)\ d\tau \ = \ \int_0^t u(\tau)\ h(t-\tau)\ d\tau, \tag{6.22}$$

where the second equality follows from the fact that $h(t-\tau)=0$ for $t-\tau<0$, i.e. for $\tau>t$. This is because the impulse response is zero for $t<0$, i.e. there is no response until the time the impulse is applied.

Note that the impulse response $h(.)$ used in (6.22) is the unit impulse response of Definition 6.9, i.e. it is the limit of the approximate impulse responses $h_\epsilon(.)$ as $\epsilon \to 0$. This is the response to an infinitesimally short duration impulse.

We now summarize all the above formally.

Theorem 6.11. *Let $h(t)$ be the unit impulse response of a linear time-invariant system. Then the zero-initial conditions response $y(t)$ to an arbitrary input $u(t)$, $t \geq 0$, is given by the convolution*

$$y(t) \ = \int_0^t h(t-\tau)\ u(\tau)\ d\tau. \tag{6.23}$$

Figure 6.11 gives a graphical illustration of the convolution formula (6.23). However, the best visualization of convolutions is by observing animation of the process. A good tool to experiment with such animations is found here: https://lpsa.swarthmore.edu/Convolution/CI.html.

6.4 The Transfer Function

In this section, we introduce anther analysis tool called the *transfer function* which is based on the *unilateral Laplace transform*. Like the impulse response analysis of the previous section, It is particularly suited for analysis of system response to signals that "start" at a given time rather than

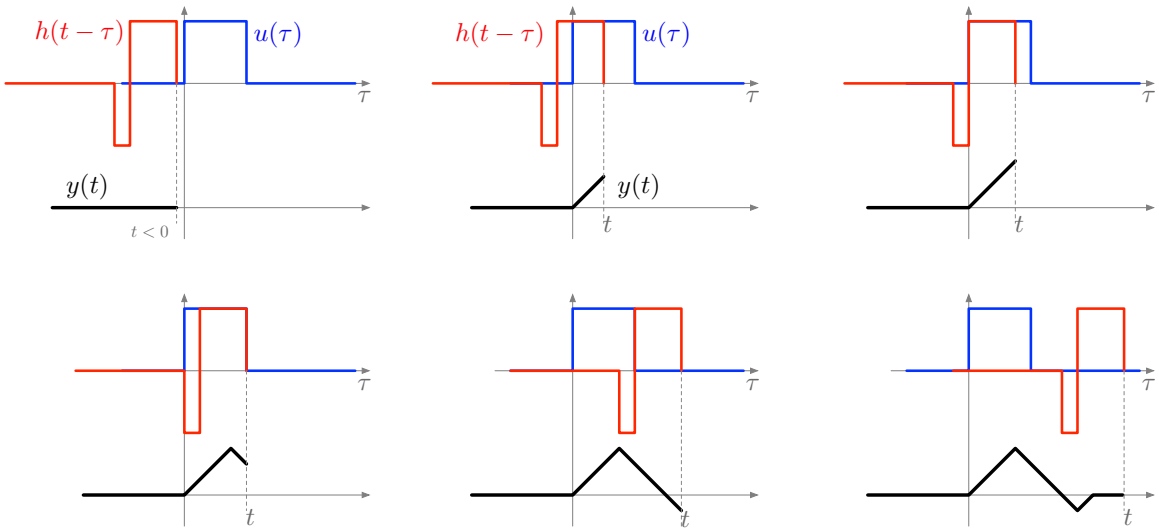


Figure 6.11: A graphical illustration of convolution where one of the signals is "flipped" in time and slides rightwards as t increases. The value of y at time t is given by the integral of the product of the flipped $h(t-\tau)$ with $u(\tau)$. Animations of this process: https://lpsa.swarthmore.edu/Convolution/CI.html help clarify it.

signals that have been entering the system for long times in the past. The mathematical idealization is therefore to study signals and system response over the time interval $0 \leq t < \infty$. It also turns out that the frequency response of a system forms part of the transfer function (i.e. it can be obtained from the transfer function), and therefore the transfer function is a more general analysis tool than the frequency response. We will also see that the Laplace transform of the impulse response of a system (studied in the previous section) is the same as the transfer function.

Given the importance of the concept of transfer functions, we will introduce them using two equivalent, but different ways. The first is a simple technique sometimes known[2] as the "operational calculus", and avoids the direct use of the Laplace transform. The second approach uses the Laplace transform and requires setting up a little bit of mathematical machinery.

6.4.1 The Transfer Function via Operational Calculus

The basic idea of the operational calculus is to replace the derivative operator $\frac{d}{dt}$ in any ODE with the "symbol" $\boldsymbol{s}$, use the linearity property, and algebraically manipulate expressions with the operation $\boldsymbol{s}$ in them. For example, the simplest differential equation stating that the input is the derivative of the output implies

$$\frac{d}{dt}y(t) \;=\; u(t) \qquad \Rightarrow \qquad y(t) \;=\; \int_0^t u(\tau)\, d\tau,$$

i.e. that the output is the integral of the input (assuming zero initial conditions on y). In the operational calculus, the equivalent statement would be

$$\boldsymbol{s}\, y \;=\; u \qquad \Rightarrow \qquad y \;=\; \frac{1}{\boldsymbol{s}}\, u.$$

In other words, if $\boldsymbol{s}$ denotes the operation of differentiation, then $1/\boldsymbol{s}$ is its inverse, which is integration. Note that we write $\boldsymbol{s}\, y$ rather than $\boldsymbol{s}\, y(t)$ to indicate that $\boldsymbol{s}$ operates on the whole function y rather than on its value $y(t)$ at time t.

[2] In the mathematics literature, this is also known as a "functional calculus". The operational calculus and transform analysis was introduced into Engineering by Oliver Heaviside (1850-1925), a brilliant Engineer/Mathematician who invented much of the analysis methods used in this chapter. Unfortunately for him, he is much more appreciated today than he was during his lifetime.

Similarly, the operation $\frac{d^k}{dt^k}$ of repeated differentiation is expressed as $\boldsymbol{s}^k$, and integration k times as $1/\boldsymbol{s}^k$. We can go further and say that an operation like $\left(\frac{d}{dt}+1\right)$, which is represented as $(\boldsymbol{s}+1)$, has as its inverse the operation $1/(\boldsymbol{s}+1)$. But how should we interpret $1/(\boldsymbol{s}+1)$? What kind of an operation on signals is it? The clarification is obtained using the Laplace transform, but for now we will proceed formally with algebraic manipulations.

Consider the following input-output ODE, and rearrange each side as a "polynomial" in the operation $\frac{d}{dt}$

$$\alpha_n \tfrac{d^n}{dt^n} y(t) + \cdots + \alpha_1 \tfrac{d}{dt} y(t) + \alpha_0 y(t) \;=\; \beta_m \tfrac{d^m}{dt^m} u(t) + \cdots + \beta_1 \tfrac{d}{dt} u(t) + \beta_0\, u(t) \tag{6.24}$$

$$\Leftrightarrow \qquad \left(\alpha_n \tfrac{d^n}{dt^n} + \cdots + \alpha_1 \tfrac{d}{dt} + \alpha_0\right) y(t) \;=\; \left(\beta_m \tfrac{d^m}{dt^m} + \cdots + \beta_1 \tfrac{d}{dt} + \beta_0\right) u(t). \tag{6.25}$$

Now replace each $\frac{d^k}{dt^k}$ with its representation $\boldsymbol{s}^k$ to obtain polynomials in $\boldsymbol{s}$ acting on y and u respectively

$$(\alpha_n \boldsymbol{s}^n + \cdots + \alpha_1 \boldsymbol{s} + \alpha_0)\, y \;=\; (\beta_m \boldsymbol{s}^m + \cdots + \beta_1 \boldsymbol{s} + \beta_0)\, u. \tag{6.26}$$

So far this equation is simply another notation for the original ODE (6.24). However, now we proceed (bravely) with algebraic manipulations to find y in terms of some operation on y

$$y \;=\; \frac{\beta_m \boldsymbol{s}^m + \cdots + \beta_1 \boldsymbol{s} + \beta_0}{\alpha_n \boldsymbol{s}^n + \cdots + \alpha_1 \boldsymbol{s} + \alpha_0}\; u \;=:\; \mathcal{H}(\boldsymbol{s})\; u, \tag{6.27}$$

where $\mathcal{H}(\boldsymbol{s})$ is the rational function[3] of $\boldsymbol{s}$ defined above. If we can make sense of this operation $\mathcal{H}(\boldsymbol{s})$, then we have in effect "solved" the ODE, i.e. obtained the output y explicitly as the operation $\mathcal{H}(\boldsymbol{s})$ on the input u.

The expression $\mathcal{H}(\boldsymbol{s})$ in (6.27) is called the transfer function of the system (6.25). How should we interpret this rational function of $\boldsymbol{s}$ as an operation on signals? There are several interpretations possible, which we examine next.

Solutions through Decompositions

Any rational function of a complex number s has a *Partial Fraction Expansion* (PFE, see Appendix **??**) of the following form

$$\frac{\beta_m s^m + \cdots + \beta_1 s + \beta_0}{\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0} \;=\; \frac{b_1}{s-p_1} + \cdots + \frac{b_n}{s-p_n}, \tag{6.28}$$

where $p_1, \ldots, p_n$ are the roots of the denominator polynomial $\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0$ (assumed here to have distinct roots), and $b_1, \ldots, b_n$ are coefficients that can be calculated from the rational function, i.e. they can be calculated from the coefficients $\alpha_n, \ldots, \alpha_0, \beta_m, \ldots, \beta_0$ (See Appendix **??** for the calculation procedure).

Since this is purely algebraic expression, we can apply it to the operation $\mathcal{H}(\boldsymbol{s})$ to decompose it into simpler operations like

$$y \;=\; \frac{\beta_m \boldsymbol{s}^m + \cdots + \beta_1 \boldsymbol{s} + \beta_0}{\alpha_n \boldsymbol{s}^n + \cdots + \alpha_1 \boldsymbol{s} + \alpha_0}\; u \;=\; \left(\frac{b_1}{\boldsymbol{s}-p_1} + \cdots + \frac{b_n}{\boldsymbol{s}-p_n}\right) u \;=\; \frac{b_1}{\boldsymbol{s}-p_1}\, u + \cdots + \frac{b_n}{\boldsymbol{s}-p_n}\, u$$
$$=:\; y_1 + \cdots + y_n. \tag{6.29}$$

Thus if we can explicitly compute what the function $\frac{b_k}{\boldsymbol{s}-p_k}\, u$ is in terms of u, then we can obtain the full solution $y = y_1 + \cdots + y_n$ in terms of a decomposition into simpler problems for each y_k, $k = 1, \ldots, n$. Now to find out what $y_k := \frac{b_k}{\boldsymbol{s}-p_k}\, u$ is, convert this expression back to a simple 1st order differential equation

$$y_k \;=\; \frac{b_k}{\boldsymbol{s}-p_k}\, u \qquad \Leftrightarrow \qquad (\boldsymbol{s}-p_k)\, y_k \;=\; b_k\, u \qquad \Leftrightarrow \qquad \dot{y}_k(t) - p_k\, y_k(t) \;=\; b_k\, u(t).$$

[3] A function is called "rational" if it is the ratio of two polynomials.

Recall from Section 6.3 that the impulse response of this system was found to be $b_k\, e^{p_k t}$, and therefore the general response to any input u is given by the convolution

$$y_k(t) \;=\; b_k \int_0^t e^{p_k(t-\tau)}\, u(\tau)\, d\tau. \tag{6.30}$$

This is the answer to the question asked earlier about what kind of operation is $1/(\boldsymbol{s}+1)$ on signals. From (6.30) we see that it is the following convolution operation (again assuming zero initial conditions)

$$y \;=\; \frac{1}{\boldsymbol{s}+1}\, u \qquad \Leftrightarrow \qquad y(t) \;=\; \int_0^t e^{-(t-\tau)}\, u(\tau)\, d\tau.$$

For the more general case of (6.29), the solution is the sum of n terms like (6.30) and is therefore

$$\begin{aligned} y(t) \;=\; y_1(t) + \cdots + y_n(t) \;&=\; b_1 \int_0^t e^{p_1(t-\tau)} u(\tau)\, d\tau + \cdots + b_n \int_0^t e^{p_n(t-\tau)} u(\tau)\, d\tau \\ &=\; \int_0^t \left(b_1 e^{p_1(t-\tau)} + \cdots + b_n e^{p_n(t-\tau)} \right) u(\tau)\, d\tau. \end{aligned}$$

Note that this last expression is the convolution of the input u with the function

$$h(t) \;:=\; b_1 e^{p_1 t} + \cdots + b_n e^{p_n t},$$

which is the impulse response of the system.

We summarize the above as a the following procedure.

Impulse Response from Partial Fraction Expansion

1. Given the input-output ODE

   $$\alpha_n \tfrac{d^n}{dt^n} y(t) + \cdots \;+\; \alpha_1 \tfrac{d}{dt} y(t) + \alpha_0 y(t) \;=\; \beta_m \tfrac{d^m}{dt^m} u(t) + \cdots + \beta_1 \tfrac{d}{dt} u(t) + \beta_0\, u(t),$$

   form the *transfer function* $\mathcal{H}(s) \;=\; \dfrac{\beta_m s^m + \cdots + \beta_1 s + \beta_0}{\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0}$.

2. Find the coefficients $p_1, \ldots, p_n$ (the roots of the denominator polynomial) and $b_1, \ldots, b_n$ of the partial fraction expansion $\mathcal{H}(s) \;=\; \dfrac{b_1}{s-p_1} + \;\cdots\; + \dfrac{b_n}{s-p_n}$.

3. The general solution y for any input u (and zero initial conditions on y) is given by the convolution of u with the system's *impulse response* h

   $$y(t) = \int_0^t h(t-\tau)\; u(\tau)\; d\tau, \qquad\qquad h(t) := b_1\, e^{p_1 t} + \cdots + b_n\, e^{p_n t}. \tag{6.31}$$

Remark 6.12. Observe the following about the procedure outlined above.

1. The partial fraction expansion in 2. shows that the transfer function $\mathcal{H}(s)$ is the Laplace transform of the impulse response $h(t)$ in (6.31). The Laplace transform pair $e^{pt} \longleftrightarrow \frac{1}{s-p}$ is a standard one from the tables.

2. The fact that $h(t)$ is the *response to an impulse* is clear from the convolution expression in (6.31). If the input is an impulse $u(t) = \delta(t)$, then the response to this input is

   $$y(t) = \int_0^t h(t-\tau)\, \delta(\tau)\, d\tau \;=\; h(t-\tau)\Big|_{\tau=0} \;=\; h(t).$$

The partial fraction expansion shown above is valid for denominator polynomials with no repeated roots. The case of repeated roots requires a slightly different treatment and more bookkeeping, and is therefore relegated to Appendix **??**.

The Frequency Response from the Transfer Function

The observant reader will have noticed the similarity between the expression for $\mathcal{H}(\boldsymbol{s})$ above and the frequency response expression (6.17). In fact, the frequency response is exactly $\mathcal{H}(\boldsymbol{s})$ with the term $\boldsymbol{s}$ replaced by the complex number $j\omega$. Indeed

$$\mathcal{H}(\boldsymbol{s})|_{\boldsymbol{s}=j\omega} = \left.\frac{\beta_m \boldsymbol{s}^m + \cdots + \beta_1 \boldsymbol{s} + \beta_0}{\alpha_n \boldsymbol{s}^n + \cdots + \alpha_1 \boldsymbol{s} + \alpha_0}\right|_{\boldsymbol{s}=j\omega} = \mathcal{H}(j\omega) = \frac{\beta_m (j\omega)^m + \cdots + \beta_1 (j\omega) + \beta_0}{\alpha_n (j\omega)^n + \cdots + \alpha_1 (j\omega) + \alpha_0} = \mathsf{H}(\omega).$$

6.4.2 The Transfer Function via the Laplace Transform

To being with we recall the definition and some important properties of the Laplace transform.

Definition 6.13. *Let $u(t)$ be a* one-sided *signal (i.e. defined over the time interval $0 \leq t < \infty$). The* unilateral Laplace transform[4] *U of u is a complex-valued function of a complex variable s defined by*

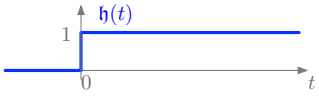
$$U(s) := \int_0^\infty u(t)\, e^{-st} dt. \tag{6.32}$$

We will adopt the notational convention of lower-case letters for signals and upper-case letters for their Laplace transforms, e.g. the Laplace transform of a signal y will be denoted by Y. Notice the similarity of the Laplace transform definition (6.32) to that of the Fourier transform (6.14), the difference being that instead of integrating the signal against the function $e^{j\omega t}$, we integrate it against e^{-st} where s is now allowed to be any complex number rather than only imaginary numbers $j\omega$. We will return to this important observation shortly.

The Laplace transform (6.32) takes a function of a real variable (namely time t) and converts it to a function of a complex variable s. Just like the Fourier transform, its utility for analyzing ODEs is that it converts differentiations of time signals to algebraic operations on their Laplace transforms, but first let's see some examples.

The canonical one-sided signal is the constant, or "unit-step" Heaviside[5] function $\mathfrak{h}(t)$

$$\mathfrak{h}(t) := \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}$$



Its Laplace transform is easily computed to be

$$\mathfrak{H}(s) = \int_0^\infty \mathfrak{h}(t)\, e^{-st} dt = \int_0^\infty e^{-st} dt \qquad \text{(since } \mathfrak{h}(t) = 1 \text{ for } t \geq 0\text{)}$$

$$= \left.\frac{-1}{s} e^{-st}\right|_0^\infty = \frac{-1}{s}(0 - 1) = \frac{1}{s}. \tag{6.33}$$

Note that the integral is finite only if s is such that $\mathbb{R}(s) > 0$, and in that case $\lim_{t\to\infty} e^{-st} = 0$ as shown above. Thus the Laplace transform $\mathfrak{H}(s)$ is only defined over the subset of the complex plane where $\mathbb{R}(s) > 0$. This is called the *region of convergence* of the Laplace transform of the signal $\mathfrak{h}(t)$. We will largely ignore issues of regions of convergence here since we will be interested in the Laplace transform primarily to encode differentiation operations in ODEs. Another easily computable Laplace transform is for the exponential function that "starts" at time $t = 0$

$$u(t) := \mathfrak{h}(t)\, e^{at} \qquad \Rightarrow \qquad U(s) = \int_0^\infty e^{at}\, e^{-st} dt = \int_0^\infty e^{(a-s)t} dt = \tfrac{1}{a-s}\, e^{(a-s)t}\Big|_0^\infty = \tfrac{1}{a-s}(0-1)$$

$$= \frac{1}{s-a}.$$

[4] Also called the "one-sided" Laplace transform to distinguish it from the "two-sided" Laplace transform used for signals defined over $-\infty < t < \infty$.

[5] named after Oliver Heaviside (1850-1925), a brilliant Engineer/Mathematician who invented much of the transform analysis methods used in this chapter. He is much more appreciated today than he was during his lifetime.

Note that for $a = 0$, the exponential function e^{at} is just he unit step function, and in that case the Laplace transform $1/(s-a)$ is simply $1/s$ as derived earlier in (6.33). Further important examples are shown in the table of Laplace transforms is included in Appendix **??**

The most important property of Laplace transforms is the differentiation property. Let u be any one-sided signal and denote its Laplace transform as usual by U. What is the Laplace transform of its derivative $y(t) := \frac{du}{dt}(t)$? This important relation can be calculated as follows

$$\begin{aligned} & U(s) = \int_0^\infty u(t)\, e^{-st} dt, \\ y(t) = \tfrac{du}{dt}(t) \quad \Rightarrow \quad & Y(s) = \int_0^\infty \tfrac{du}{dt}(t)\, e^{-st} dt = u(t)e^{-st}\Big|_0^\infty + s\int_0^\infty u(t)\, e^{-st} dt \quad \text{(integration by parts)} \\ & \qquad\qquad = -u(0) \;+\; s\, Y(s) \end{aligned}$$

If you're familiar with the Laplace transform, then you see that if we take the Laplace transform of both sides of the ODE (6.24) (and assuming zero initial conditions) we get the following relations between the Laplace transforms Y and U of the functions y and u respectively

$$\begin{aligned} & \alpha_n \tfrac{d^n}{dt^n} y(t) + \cdots + \alpha_1 \tfrac{d}{dt} y(t) + \alpha_0\, y(t) \;=\; \beta_m \tfrac{d^m}{dt^m} u(t) + \cdots + \beta_1 \tfrac{d}{dt} u(t) + \beta_0\, u(t) \qquad (6.34) \\ \Rightarrow \quad & \alpha_n\, s^n Y(s) + \cdots + \alpha_1\, sY(s) + \alpha_0\, Y(s) \;=\; \beta_m\, s^m U(s) + \cdots + \beta_1\, sU(s) + \beta_0\, U(s) \\ \Rightarrow \quad & \big(\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0\big)\, Y(s) \;=\; \big(\beta_m s^m + \cdots + \beta_1 s + \beta_0\big)\, U(s) \\ \Rightarrow \quad & \frac{Y(s)}{U(s)} \;=\; \frac{\beta_m s^m + \cdots + \beta_1 s + \beta_0}{\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0} \;=:\; \mathcal{H}(s). \qquad (6.35) \end{aligned}$$

The complex function $\mathcal{H}(s)$ is called the transfer function of the system (6.34), and it is defined as the ratio of the Laplace transform Y of the output to the Laplace transform U of the input. By the simple calculation above, we see that it is completely determined by the coefficients of the differential equation. A comparison of (6.35) and (4.27) shows that the *frequency response is the transfer function evaluated on the imaginary axis*, i.e. evaluated at $s = j\omega$.

Theorem 6.14. *Consider the system described by the following constant-coefficient ODE*

$$\begin{aligned} & \alpha_p\, y^{(p)}(t) \;+\; \alpha_{p\text{-}1}\, y^{(p\text{-}1)}(t) \;+\; \cdots \;+\; \alpha_1\, y^{(1)}(t) \;+\; \alpha_0\, y(t) \\ & \qquad = \; \beta_q\, u^{(q)}(t) \;+\; \beta_{q\text{-}1}\, u^{(q\text{-}1)}(t) \;+\; \cdots \;+\; \beta_1\, u^{(1)}(t) \;+\; \beta_0\, u(t), \qquad (6.36) \end{aligned}$$

where the signal u is an input and y is an output. With zero initial conditions on y, the relation between the Laplace transform U of the input and the Laplace transform Y of the output is given by

$$Y(s) \;=\; \mathcal{H}(s)\, U(s), \qquad \boxed{\mathcal{H}(s) \;:=\; \frac{\beta_m s^m + \cdots + \beta_1 s + \beta_0}{\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0},} \qquad (6.37)$$

where $\mathcal{H}(s)$ is called the transfer function *of the system. The frequency response $\mathsf{H}(\omega)$ of this system is the transfer function evaluated on the imaginary axis $s = j\omega$*

$$\textit{frequency response} \;=\; \mathsf{H}(\omega) \;=\; \mathcal{H}(j\omega) \;=\; \textit{transfer function evaluated on the imaginary axis.}$$

6.4.3 Poles and Zeros: Visualizing the Transfer Function

The transfer function $\mathcal{H}(s)$ of a system is a complex-valued function of a complex variable s, and therefore it is a map $\mathcal{H} : \mathbb{C} \to \mathbb{C}$ from a complex plane $\mathbb{C}$ to a complex plane $\mathbb{C}$. Since a complex plane has two real dimensions, a full visualization of the "graph" of the function $\mathcal{H}(s)$ requires visualizing a graph in four real dimensions, which is very difficult. However, the *magnitude* $|\mathcal{H}| : \mathbb{C} \to \mathbb{R}$ is a real-valued function of a complex variable s, and this can be visualized as a graph (a surface) over a two-dimensional plane. The so-called "rubber sheet" analogy provides an intuitive geometric description of this graph in terms of the transfer function's *poles* and *zeros*. This in turn helps explain resonances that appear in the system's frequency response.

A complex function of the form (6.37) is a ratio of two polynomials (in the complex variable s), i.e. it is a *rational* function of the variable s. Recall that any polynomial can be factored using its roots. For example, the denominator polynomial in (6.37) has n roots (since it's an n'th order polynomial)

$$\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0 \;=\; \alpha_n \left(s^n + \cdots + \tfrac{\alpha_1}{\alpha_n} s + \tfrac{\alpha_0}{\alpha_n} \right) \;=\; \alpha_n \, (s - p_1) \cdots (s - p_n) \, ,$$

where $p_1, \ldots, p_n$ are the roots of the polynomial. Similarly, the numerator polynomial can also be factored in terms of its roots. The roots of both the numerator and the denominator polynomials have special significance, so they deserve their own names as described next.

Definition 6.15. *Consider a rational transfer function $\mathcal{H}(s)$ where its numerator and denominator polynomials are factored as follows*

$$\mathcal{H}(s) \;=\; \frac{\beta_m s^m + \cdots + \beta_1 s + \beta_0}{\alpha_n s^n + \cdots + \alpha_1 s + \alpha_0} \;=\; \frac{\beta_m}{\alpha_n} \, \frac{(s - z_1) \;\; \cdots \;\; (s - z_m)}{(s - p_1) \;\; \cdots \;\; (s - p_n)} . \tag{6.38}$$

The (possibly complex) roots $z_1, \ldots, z_m$ of the numerator are the zeros *of the transfer function since*

$$\mathcal{H}(z_i) \;=\; 0, \qquad\qquad i = 1, \ldots, m.$$

The (possibly complex) roots $p_1, \ldots, p_n$ of the denominator are the poles *of the transfer function*

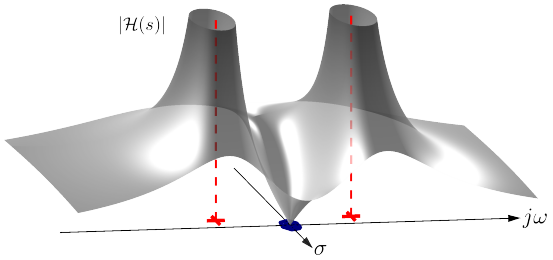
$$\lim_{s \to p_i} \mathcal{H}(s) \;=\; \infty, \qquad\qquad i = 1, \ldots, n.$$

In other words, the transfer function magnitude is zero when evaluated at the zeros $z_1, \ldots, z_m$ (thus the name "zeros"), and it is infinite when evaluated at the poles $p_1, \ldots, p_n$.

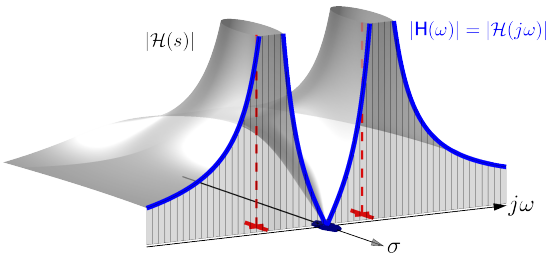
We have already encountered the roots of the denominator (the poles) when studying solutions of homogenous ODEs. They are exactly the *characteristic roots* of the ODE. Recall that those give the form of the general solution of the homogenous ODE. On the other hand, the roots of the numerator (the zeros of the transfer function) are something new. They are a property of input-output ODEs and do not occur if the ODE is homogenous (note that the denominator polynomial comes from the right hand side of the ODE (6.36), which would be zero if there were no inputs). We will see in Chapter 8 that zeros of a transfer function have an important dynamical interpretation in terms of a phenomenon called *vibration absorption*. Thus, both poles and the zeros of a transfer function have significance for the dynamic behavior of the system. In addition, they also provide a geometric visualization of the transfer function and the frequency response as discussed next.

A real-valued function of a real variable is easy to visualize in terms of its graph, which is normally plotted as a curve in a two-dimensional plane with one axis representing the independent variable, and the other axis representing the value of the function. Similarly, a real-valued function of *two* real variables can also be visualized in terms of its graph in three dimensions, which is a surface over the two-dimensional plane of the two variables. On the other hand, the graph of a *complex-valued* function $\mathcal{H}(s)$ of a *complex variable* s requires four real dimensions to visualize, which is beyond the imagination of most people. However, the *magnitude* $|\mathcal{H}(s)|$ of a transfer function is a real-valued function of a complex variable s, and can therefore be visualized as a *surface* drawn over the complex plane.

This visualization of the magnitude $|\mathcal{H}(s)|$ of a transfer function is depicted in Figure 6.12a. Note that the magnitude is always non-negative, so the surface of the graph is always above the complex plane. A very useful intuition is given by the so-called "rubber sheet" analogy. Imagine the surface first as a "flat" rubber sheet stretched over the complex plane. Mark the pole locations $p_1, \ldots, p_n$ as points in the complex plane where "poles" (like the poles of a tent) are placed vertically, and the rubber sheet is forced to go to infinity at those locations. Similarly, the zero locations $z_1, \ldots, z_m$ are points in the complex plane where the rubber sheet is "pinned to the ground" (the "ground" is the complex plane, and the zeros are like the ground pins of a tent). The rubber sheet then interpolates between poles and zeros throughout the rest of the complex plane. Finally, if the order of the



(a) The "rubber sheet" or "tent" analogy visualizing the magnitude $|\mathcal{H}(s)|$ as a surface over the the whole of the complex plane $s = \sigma + j\omega$. The poles of the transfer function (the two red crosses) act like "poles of a tent" (where $|\mathcal{H}(s)|$ goes to ∞), and the zeros (the blue disk at $s = 0$) act like "pins to the ground".



(b) The frequency response $\mathsf{H}(\omega)$ is the transfer function $\mathcal{H}(s)$ evaluated on the imaginary axis $s = j\omega$. This is visualized as taking a "slice" of the graph of $\mathcal{H}(s)$ over the imaginary axis. In particular, this gives a geometric visualization of how poles of $\mathcal{H}(s)$ near the imaginary axis (the so-called lightly-damped poles) produce resonances in the frequency response $\mathsf{H}(\omega)$.

Figure 6.12: The "rubber sheet" visualization of a rational transfer function $\mathcal{H}(s) = \frac{(s-z_1)\cdots(s-z_m)}{(s-p_1)\cdots(s-p_n)}$, where $z_1, \ldots, z_m$ and $p_1, \ldots, p_n$ are its zeros and poles respectively. This analogy demonstrates how poles and zeros near the imaginary axis influence the resonances of the system's frequency response $\mathsf{H}(\omega)$.

numerator $m < n$ is less than the order of the denominator (a typical case), then $|\mathcal{H}(s)| \overset{s\to\infty}{\longrightarrow} 0$, and therefore the surface stretches down to the ground for large values of s in all directions.

The justification for the rubber sheet analogy follows from the product and ratio rules of complex numbers. Recall that for any two complex numbers z_1, z_2, the magnitudes of their products and ratios are given by

$$|z_1 z_2| \;=\; |z_1|\,|z_2|\,, \qquad \left|\frac{z_1}{z_2}\right| \;=\; \frac{|z_1|}{|z_2|}.$$

Applying these two rules repeatedly to the products and ratios in the rational function (6.38)

$$|\mathcal{H}(s)| \;=\; \left|\frac{\beta_m}{\alpha_n}\,\frac{(s-z_1)\;\cdots\;(s-z_m)}{(s-p_1)\;\cdots\;(s-p_n)}\right| \;=\; \frac{|\beta_m|}{|\alpha_n|}\,\frac{|s-z_1|\;\cdots\;|s-z_m|}{|s-p_1|\;\cdots\;|s-p_n|}.$$

Recall also that for any two complex numbers s and z_k for example, the quantity $|s - z_k|$ is the *distance* in the complex plane between s and z_k. Therefore as the variable s varies in the complex plane, the ratio above can be interpreted as

$$|\mathcal{H}(s)| \;=\; \frac{|\beta_m|}{|\alpha_n|}\,\frac{|s-z_1|\;\cdots\;|s-z_m|}{|s-p_1|\;\cdots\;|s-p_n|} \;=\; \frac{|\beta_m|}{|\alpha_n|}\,\frac{\text{product of distances from } s \text{ to the zeros}}{\text{product of distances from } s \text{ to the poles}}.$$

As s gets close to a zero z_k of $\mathcal{H}(s)$, the numerator of $|\mathcal{H}(s)|$ goes to zero, i.e. $|\mathcal{H}(s)|$ gets "pinned" to the ground. Similarly as s gets close to a pole p_k of $\mathcal{H}(s)$, the denominator of $|\mathcal{H}(s)|$ goes to zero, which means that $|\mathcal{H}(s)|$ goes to infinity.

The rubber sheet analogy also gives a very useful geometric intuition for the frequency response $\mathsf{H}(\omega)$ of the system. Since $\mathsf{H}(\omega) = \mathcal{H}(j\omega)$, the *frequency response magnitude* $|\mathsf{H}(\omega)|$ *is just the magnitude* $|\mathcal{H}(j\omega)|$ *of the transfer function evaluated on the imaginary axis.* Thus if we take a "vertical slice" of the surface of $|\mathcal{H}(s)|$ over the imaginary axis $s = j\omega$ (see Figure 6.12b), we obtain the frequency response. Note how in Figure 6.12b the two poles close to the imaginary axis[6] "induce" peaks in the frequency response (i.e. resonances) at nearby frequencies. Similarly zeros near or on the imaginary axis will force the frequency response to be close to zero at nearby frequencies. For example, the transfer function in Figure 6.12 has a zero at $s = 0$, which forces the frequency response to be exactly zero at $\omega = 0$.

[6] Poles close to the imaginary axis are called underdamped poles since those arise for example in a Mass-Spring-Damper system with very low damping ratio ζ.

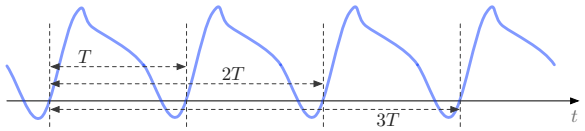


Figure 6.13: A signal that is T-periodic is also $2T$-periodic. It is also kT-periodic for any $k = 1, 2, 3, \ldots$. The fundamental period is the smallest number T such that the signal is T-periodic.

Appendix

6.A Background on Fourier Series

Let $u(t)$ be any T-periodic signal. This means that starting at any time t, the value of the signal at time $t + T$ is the same as its value at time t

$$\text{for all } t \qquad u(t) = u(t+T),$$

i.e. the signal "repeats" after T time units. There is some ambiguity in defining the period because it is also true that this signal is $2T$-periodic as well since

$$u(t) \;=\; u(t+T) \;=\; u((t+T)+T) \;=\; u(t+2T).$$

Repeating this argument, we see that a T-periodic signal is actually kT-periodic where k is any integer

$$\text{for all } t, \quad u(t) = u(t+T) \qquad \Rightarrow \qquad \text{for all } t, \quad u(t) = u(t \pm kT), \qquad k = 0, 1, 2, \ldots$$

This fact is illustrated in Figure 6.13. Thus a periodic signal has many periods. If it is periodic with period T, then it is periodic with period kT where k is any integer. The smallest such period is however special and has a definition.

Definition 6.16. *For a periodic signal $u(t)$, the* fundamental period *is the smallest number T such that*

$$\textit{for all } t, \qquad u(t) = u(t+T),$$

i.e. the smallest possible period. The fundamental frequency *is obtained from the fundamental period T as $2\pi/T$ in rad/s or $1/T$ in Hz.*

A pure sinusoid like

$$\cos(\bar{\omega} t), \qquad\qquad \bar{\omega} = 2\pi/T,$$

is periodic with fundamental period $T = 2\pi/\bar{\omega}$. A sinusoid of twice the frequency has fundamental period of $T/2$, but is also T-periodic as well

$$\cos(2\bar{\omega}t) = \cos\left(2\tfrac{2\pi}{T}t\right) \qquad \Rightarrow \qquad \cos\left(2\tfrac{2\pi}{T}(t+T)\right) = \cos\left(2\tfrac{2\pi}{T}t + 2\tfrac{2\pi}{T}T\right) = \cos\left(2\tfrac{2\pi}{T}t + 4\pi\right) = \cos\left(2\tfrac{2\pi}{T}t\right).$$

Similarly, the signal $\cos(k\bar{\omega}t)$ has fundamental period $2\pi/k\bar{\omega}$, but it is also periodic with period $2\pi/\bar{\omega}$, which is the fundamental period of $\cos(\bar{\omega}t)$, i.e. they are both T-periodic.

Definition 6.17. *Consider a sinusoidal signal $u(t)$ with fundamental frequency $\bar{\omega}$*

$$u(t) \;=\; \mathsf{u}\;\cos(\bar{\omega}t + \theta).$$

A sinusoid with k times (k an integer) that frequency, i.e.

$$y(t) \;=\; \mathsf{y}\;\cos(k\bar{\omega}t + \phi)$$

is called a k'th harmonic *of the signal $u(t)$. While $y(t)$ has fundamental period $2\pi/(k\bar{\omega})$, it is also periodic with period $2\pi/\bar{\omega}$, the fundamental period of $u(t)$. Therefore $u(t)$ and any of its harmonics share the common period $2\pi/\bar{\omega}$, the fundamental period of $u(t)$.*

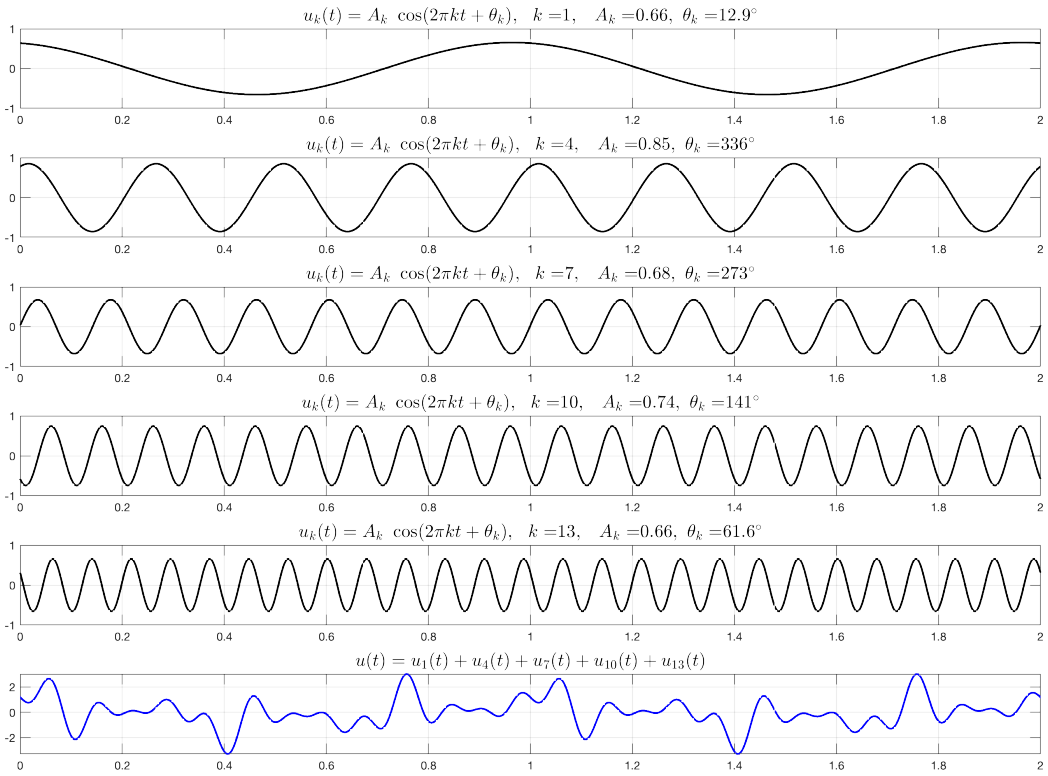


Figure 6.14: A "randomly constructed" periodic signal (bottom panel in blue) as the sum of of five signals all of which have a common period. The top panel shows $u_1(t)$ which has frequency 1 Hz and therefore a period of 1 second. $u_4(t)$ (second panel) has 4 times the frequency of $u_1(t)$ and therefore 1/4th of the period. Thus $u_2(t)$ has a period of 1/4 seconds, but it is also periodic with period 1 second. Similarly all other signals are harmonics of $u_1(t)$, and therefore are also periodic with period 1 second, but their amplitudes and phases are chosen randomly. The bottom panel shows the sum of all those signals, which must also be periodic with period of 1 second, but does not have any smaller period.

Figure 6.14 shows a signal and some of its harmonics. Note that while its harmonics have higher frequencies, and thus smaller periods, they all have the common period of the first signal. A key observation is that the *sum of any number of T-periodic signals is another T-periodic signal.* In particular, a signal of the form

$$\sum_{k=0}^{\infty} \mathsf{u}_k \cos(k\bar{\omega}t + \theta_k) \tag{6.39}$$

for any set of amplitudes u_k and phases θ_k is periodic with the fundamental period $T = 2\pi/\bar{\omega}$. Figure 6.14 shows an example of such a sum of a sinusoid and some of its harmonics where the amplitudes and phases are chosen randomly. The sum is clearly a periodic signal with the same period as the fundamental. We therefore arrive at the conclusion that any sum of sinusoids with a common period T is a T-periodic signal. The idea of *Fourier Series* is the "converse" of this; given any T-periodic signal, write it as a linear combination of harmonics like (6.39)!

There are several different forms of Fourier series. The most convenient form is in terms of complex exponentials. Fix a fundamental frequency $\bar{\omega}$, and consider an infinite sum of the form

$$u(t) \;=\; \sum_{k=-\infty}^{\infty} \hat{\mathsf{u}}_k \, e^{jk\bar{\omega}t}, \tag{6.40}$$

where the coefficients $\{\hat{\mathsf{u}}_k\}_{k=-\infty}^{\infty}$ are possibly complex numbers. The complex-valued signal $e^{jk\bar{\omega}t}$ has frequency $k\bar{\omega}$, and note that k can be either positive or negative. A "negative frequency" term

is the complex conjugate of the corresponding positive frequency term as can be seen by applying Euler's formula

$$\left(e^{jk\bar{\omega}t}\right)^* \;=\; \left(cos(k\bar{\omega}t) + j\sin(k\bar{\omega}t)\right)^* \;=\; \cos(k\bar{\omega}t) - j\sin(k\bar{\omega}t) \;=\; e^{-jk\bar{\omega}t} \;=\; e^{j(-k)\bar{\omega}t}.$$

We are only interested in real-valued signals $u(t)$ in (6.40). If we take the two k'th harmonic terms in (6.40) corresponding to k and $-k$, their sum must be real, and therefore must equal its own complex conjugate

$$\begin{aligned} \text{requirement:} \qquad \left(\hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} + \hat{\mathrm{u}}_{-k}\, e^{-jk\bar{\omega}t}\right)^* &= \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} + \hat{\mathrm{u}}_{-k}\, e^{-jk\bar{\omega}t} \\ \Rightarrow \qquad \hat{\mathrm{u}}_k^*\, e^{-jk\bar{\omega}t} + \hat{\mathrm{u}}_{-k}^*\, e^{jk\bar{\omega}t} &= \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} + \hat{\mathrm{u}}_{-k}\, e^{-jk\bar{\omega}t} \qquad \Rightarrow \qquad \hat{\mathrm{u}}_{-k} = \hat{\mathrm{u}}_k^*. \end{aligned} \tag{6.41}$$

Thus the coefficient $\hat{\mathrm{u}}_{-k}$ at the negative frequency $-k\bar{\omega}$ must be the complex conjugate of the coefficient $\hat{\mathrm{u}}_k$ at the positive frequency $k\bar{\omega}$. In other words, the set of coefficients $\{\hat{\mathrm{u}}_k\}_{k=-\infty}^{-1}$ are completely determined by the set of coefficients $\{\hat{\mathrm{u}}_k\}_{k=1}^{\infty}$. The former are the complex conjugates of the latter. Finally, since the "zeroth frequency" $\hat{\mathrm{u}}_0$ multiplies the constant function $e^{j0\bar{\omega}t} = 1$, it must be a real number. We summarize the above in the following definition.

Definition 6.18. *A* complex Fourier series *of a real-valued signal $u(t)$ is*

$$u(t) \;=\; \sum_{k=-\infty}^{\infty} \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t}, \tag{6.42}$$

where the complex coefficients $\{\hat{\mathrm{u}}_k\}_{k=-\infty}^{\infty}$ satisfy the conjugacy relations $\hat{\mathrm{u}}_{-k} = \hat{\mathrm{u}}_k^*$. The signal $u(t)$ is periodic with period $2\pi/\bar{\omega}$.

It is not difficult to find the other forms of the Fourier series from the complex form. The key is to use the conjugacy relations. We first find the relation between the complex form and what might be called the "phased-cosine" form as follows

$$\begin{aligned} \sum_{k=-\infty}^{\infty} \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} &= \hat{\mathrm{u}}_0 + \sum_{k=1}^{\infty} \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} + \sum_{k=-\infty}^{-1} \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} \;=\; \hat{\mathrm{u}}_0 + \sum_{k=1}^{\infty} \left(\hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} + \hat{\mathrm{u}}_{-k}\, e^{-jk\bar{\omega}t}\right) \\ &= \hat{\mathrm{u}}_0 + \sum_{k=1}^{\infty} \left(\hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} + \hat{\mathrm{u}}_k^*\, \left(e^{jk\bar{\omega}t}\right)^*\right) \qquad \left(\text{since } \hat{\mathrm{u}}_{-k} = \hat{\mathrm{u}}_k^* \text{ and } e^{-jk\bar{\omega}t} = \left(e^{jk\bar{\omega}t}\right)^*\right) \\ &= \hat{\mathrm{u}}_0 + \sum_{k=1}^{\infty} 2\,\mathbb{R}\left(\hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t}\right) \qquad \left(\text{recall } z + z^* = 2\,\mathbb{R}(z)\right) \\ \Rightarrow \quad \sum_{k=-\infty}^{\infty} \hat{\mathrm{u}}_k\, e^{jk\bar{\omega}t} &= \hat{\mathrm{u}}_0 + \sum_{k=1}^{\infty} 2\,|\hat{\mathrm{u}}_k|\;\cos\left(k\bar{\omega}t + \angle\hat{\mathrm{u}}_k\right) \end{aligned} \tag{6.43}$$

Observe that the magnitude and phase of the coefficient $\hat{\mathrm{u}}_k$ in the complex series gives the magnitude and phase of the k'th harmonic in the phased-cosine series (with the factor of 2 taken into account). The remaining Fourier series forms can now be easily derived from (6.43) as follows

$$\begin{aligned} c_0 + \sum_{k=1}^{\infty} c_k\;\cos\left(k\bar{\omega}t + \theta_k\right) &= c_0 + \sum_{k=1}^{\infty} c_k\;\sin(k\bar{\omega}t + \phi_k), \qquad \left(\phi_k = \theta_k + \pi/2 \text{ since } \cos(\psi) = \sin(\psi + \pi/2)\right) \\ &= c_0 + \sum_{k=1}^{\infty} \alpha_k\;\cos(k\bar{\omega}t) + \beta_k\;\sin(k\bar{\omega}t), \qquad \begin{aligned} \alpha_k &= c_k\cos(\theta_k), \\ \beta_k &= -c_k\sin(\theta_k). \end{aligned} \\ &\qquad \left(\text{using } \cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y) \text{ on } \cos(k\bar{\omega}t + \theta_k)\right) \end{aligned}$$

These relations are summarized and their simple geometrical relations are illustrated in Figure 6.15. We will be primarily using the complex exponential form as it is the most convenient for algebraic manipulations and differentiation.

$$\sum_{k=-\infty}^{\infty} \hat{u}_k\, e^{jk\bar{\omega}t} = c_0 + \sum_{k=1}^{\infty} c_k\, \cos\left(k\bar{\omega}t + \theta_k\right), \qquad c_0 = \hat{u}_0,\ c_k = 2\,|\hat{u}_k|,\ \theta_k = \angle \hat{u}_k$$
$$= c_0 + \sum_{k=1}^{\infty} c_k\, \sin(k\bar{\omega}t + \phi_k), \qquad \phi_k = \theta_k + \tfrac{\pi}{2}$$
$$= c_0 + \sum_{k=1}^{\infty} \alpha_k\, \cos(k\bar{\omega}t) + \beta_k\, \sin(k\bar{\omega}t)$$

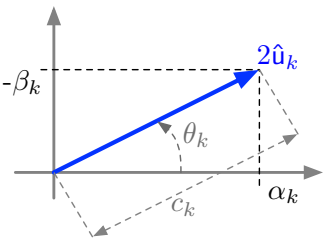


Figure 6.15: The four different forms of Fourier series: the complex-exponential, phased-cosine, phased-sine, and the cosine-sine forms respectively, and the relations between them. The diagram on the right illustrates the relations geometrically for $k \neq 0$. For the constant terms $k = 0$, we simply have $\hat{u}_0 = c_0$.

6.A.1 Obtaining Fourier Series Coefficients for any Periodic Signal

The previous discussion described how a series expression like (6.42) produces a periodic signal. The converse question is: given a periodic signal, how to obtain its Fourier series coefficients? This is given by the next statement.

Lemma 6.19. *Let $u(t)$ be a T-periodic signal. Then its Fourier series coefficients are given by the following integral*

$$\hat{u}_k \;=\; \tfrac{1}{T}\int_{\tau}^{\tau+T} u(t)\; e^{-jk\bar{\omega}t} dt \;=\; \tfrac{1}{T}\int_{\tau}^{\tau+T} u(t)\; e^{-jk\frac{2\pi}{T}t} dt, \tag{6.44}$$

where $\bar{\omega} = 2\pi/T$ is the fundamental frequency, and τ is any starting point of the integral. From these coefficients, the signal can be reconstructed by the (possibly infinite) sum

$$u(t) \;=\; \sum_{k=-\infty}^{\infty} \hat{u}_k\; e^{jk\bar{\omega}t}.$$

The proof of this statement is not difficult, but is relegated to Appendix **??** where a beautiful geometric interpretation of Fourier series as an "expansion in an orthogonal" basis of functions is presented. Finally, note that once the complex Fourier series coefficients $\hat{u}_k$ are found, the coefficients of any of the other three forms can be found using the relations summarized in Figure 6.15.

Example 6.20. A square wave of unit amplitude and period T is depicted in Figure 6.16. Its Fourier series can be calculated from the formula (6.44) as

$$\hat{u}_k = \tfrac{1}{T}\int_0^T u(t)\, e^{-jk\frac{2\pi}{T}t}\, dt = \tfrac{1}{T}\left(\int_0^{\frac{T}{2}} e^{-jk\frac{2\pi}{T}t}\, dt - \int_{\frac{T}{2}}^{T} e^{-jk\frac{2\pi}{T}t}\, dt\right)$$
$$= \tfrac{1}{T}\frac{1}{-jk\frac{2\pi}{T}}\left(e^{-jk\frac{2\pi}{T}t}\Big|_0^{\frac{T}{2}} - e^{-jk\frac{2\pi}{T}t}\Big|_{\frac{T}{2}}^{T}\right) = \tfrac{j}{2\pi\, k}\left(e^{-j\pi\, k} - e^{0} - e^{-j2\pi\, k} + e^{-j\pi\, k}\right)$$
$$= \tfrac{j}{\pi\, k}\left((-1)^k - 1\right) \;=\; \begin{cases} 0 & k \text{ even}, \\ -\frac{2}{\pi\, k}\, j & k \text{ odd}. \end{cases}$$

First note that the coefficients $\hat{u}_k$ are purely imaginary, thus by the diagram in Figure 6.15 the Fourier series of this square wave has only sine terms[7]. This is to be expected since this particular square wave is an odd function of t, and therefore can only have sine terms in its expansion. The Fourier series of this square wave is then written in both complex and cosine-sine form as

$$u(t) \;=\; \sum_{k \text{ odd}} \frac{-2j}{\pi\, k}\, e^{jk\frac{2\pi}{T}t} \;=\; \sum_{k\geq 1,\ \text{odd}} \frac{4}{\pi\, k}\, \sin\left(k\,\tfrac{2\pi}{T}\,t\right). \tag{6.45}$$

[7] A periodic signal which is an even function of time (i.e. $u(t) = u(-t)$ has only cosine terms in its Fourier series. If it is an odd function (i.e. $u(-t) = -u(t)$, then it only has sine terms in its Fourier series.

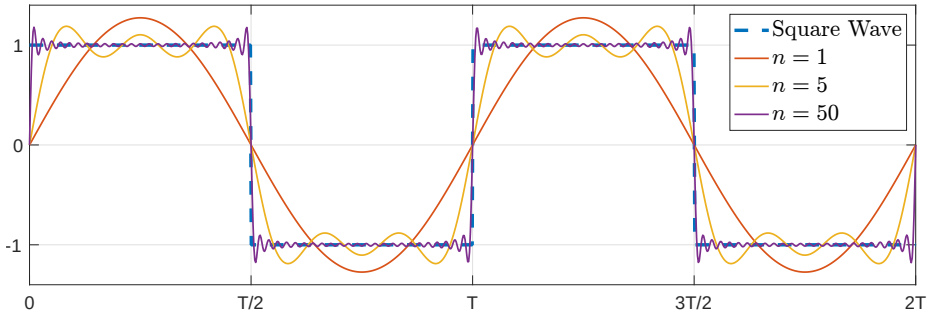


Figure 6.16: A square wave and its Fourier series $u(t) = \sum_{-n \leq k \leq n} \hat{u}_k \, e^{jk\frac{2\pi}{T}t}$ reconstructions with various series sizes n. The larger n is, the more accurate the reconstruction is. The only exception is the "Gibbs" phenomenon of peaked oscillations at the point of discontinuity of the square wave.

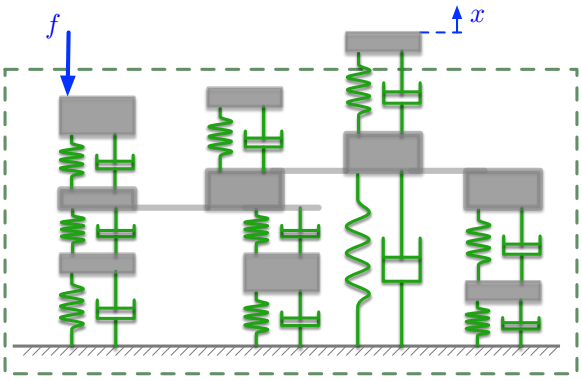
Although the Fourier series has an infinite number of terms, a "truncation" where the summation in (6.45) is taken for $|k| \leq n$ gives an approximation of the original square wave. Examples of such truncated series approximations are shown in Figure 6.16. As n increases, we see that the approximation becomes better and closer to the original square wave at every point in time, with the exception of the isolated times when the square wave has a discontinuity (the jump points) where the so-called "Gibbs phenomenon" appears.

Part III

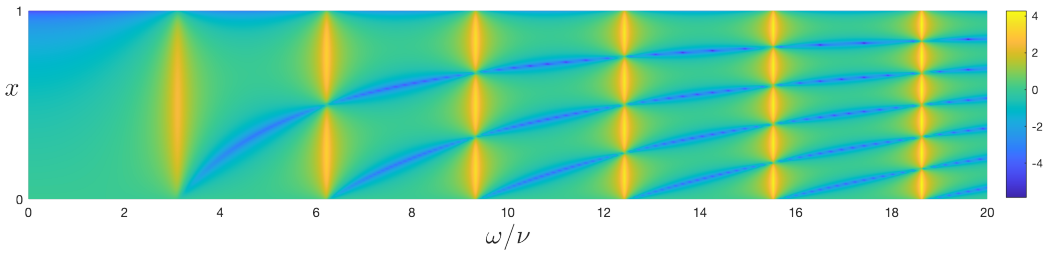
Higher-Order Systems

Higher-Order Systems

Vibrational analysis of complex systems with many degrees of freedom is best done using vector and matrix notation and methods. These "matrix methods" uses concepts from linear algebra, such as change of bases, eigenvalues and eigenvectors, and diagonalization to encode the dynamical properties of higher-order systems. In contrast to single degree of freedom systems, a higher-order system has more than one natural frequency, as well as specific vibration shapes associated with each natural frequency. These are called the normal modes of vibration. The frequencies are found as eigenvalues of certain matrices, while the "shape of vibrations" are obtained from the corresponding eigenvectors. Externally forced systems admit a frequency response analysis, in which the frequency response is a matrix of individual frequency responses from each forcing input to each mechanical degree of freedom. This frequency domain analysis complements the normal mode analysis, and direct correspondences between external resonances and internal normal modes are established.



The limit of an infinite number of degrees of freedom are continuum systems, such as with the vibrations of strings, beams, plates and acoustic waves. Those can be treated with similar linear algebra methods as finite number of degrees of freedom systems.



Chapter 7

Normal Modes of Free Vibrations: Matrix Methods

Systems with several mechanical degrees of freedom are best analyzed with matrix methods. The coupled differential equations describing the dynamics are reorganized using matrix-vector products into input-output systems where both the input and output signals are vector valued. Mass and stiffness and other parameters now become matrices rather than scalars. With these matrix methods, complicated-looking coupled differential equations can be reorganized in a unified form for which analysis techniques are developed. The main tool for analysis is based on matrix diagonalizations, which in turn is based on eigenvalues and eigenvectors of matrices. An N-DOF system has n possibly different natural frequencies, and each of those frequencies correspond to a particular mode of vibration which describes how the N coordinates move relative to each other. Eigenvalues give the natural frequencies and eigenvectors give the "shape" of the corresponding motion, which are also called the modal shapes.

Introduction with a 2-DOF Example: Sprung Beam

We begin with a brief introduction to the methods of this chapter using the 2-DOF example of the (rigid) Sprung Beam, or equivalently the "half car" model introduced in an earlier chapter in Figure 1.15c. This will serve as a preview of the matrix methods developed further in this chapter for vibrations analysis of N-Degree Of Freedom (N-DOF) systems.

Figure 7.1 shows a diagram of the sprung beam. Let x be the vertical displacement of the center of mass, and x_1 and x_2 be the vertical displacements of the support points. Assume the origin of the coordinate systems for x_1, x_2 and θ are chosen so that $x_1 = 0$, $x_2 = 0$ and $\theta = 0$ correspond to the two springs being at equilibrium. Let m be the mass of the rigid body, then Newton's second law for vertical and rotational motions are

$$\begin{aligned} m\,\ddot{x} &= -k_1 x_1 - k_2 x_2, \\ \mathrm{J}\,\ddot{\theta} &= l\,k_1 x_1 - l\,k_2 x_2, \end{aligned} \tag{7.1}$$

where J is the moment of inertia about the center of mass, and we have assumed $l_1 = l_2 = l$ for simplicity. Note the signs on the torques $l\,k_1 x_1$ and $l\,k_2 x_2$ which are due to the sign convention on θ being measured positively in the counter-clockwise direction. For example, when x_1 is positive, spring k_1 exerts a downwards force on its support point, which is a positive torque on the mass. On the other hand, when x_2 is positive, spring k_2 exerts a downwards force, which results in a clockwise (negative) torque on the mass.

We need to rewrite the equations above in terms of only one set of coordinates, either (x, θ) or (x_1, x_2), but not both. The first option is to use the kinematics to express (x_1, x_2) in terms of (x, θ). To do this, we make the simplifying assumption (reasonable for small oscillations in θ) that

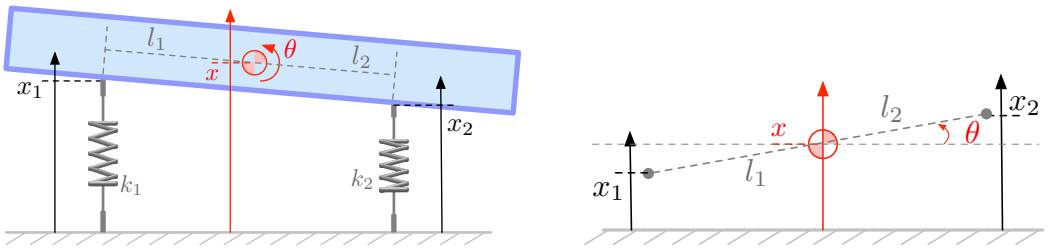


Figure 7.1: (*Left*) Schematic of the sprung-beam model, which has two degrees of freedom. The coordinates are either (x_1, x_2), the vertical displacements of the support points, or equivalently (x, θ) the vertical displacement of the center of mass and the rotation angle of the mass respectively. (*Right*) The kinematics used to relate the two coordinate systems (x, θ) and (x_1, x_2).

the springs and dampers move only vertically[1] (see Figure 7.1 (*Right*))

$$\left.\begin{aligned} x_1 &\approx x - l\sin\theta \approx x - l\theta \\ x_2 &\approx x + l\sin\theta \approx x + l\theta \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} m\,\ddot{x} &= -k_1x_1 - k_2x_2 \\ &= -(k_1+k_2)\,x + l(k_1-k_2)\,\theta \\ \mathrm{J}\,\ddot{\theta} &= l\,k_1x_1 - l\,k_2x_2 \\ &= l(k_1-k_2)\,x - l^2(k_1+k_2)\,\theta \end{aligned}\right. \tag{7.2}$$

Note an important feature of these two equations. The first one is for the derivative $\ddot{x}$, but it depends on both x and θ and their derivatives. Similarly, the second equation is for $\ddot{\theta}$, but it also depends on both x and θ and their derivatives. The equations are therefore *coupled*. Each equation cannot be solved separately, they have to be solved together. In the language of dynamics, the equations for $\ddot{x}$ and $\ddot{\theta}$ describe vertical and rotational dynamics respectively. The fact that their equations are coupled means that vertical vibrations effect rotational vibrations and vice versa. A very useful technique that helps reveal the underlying structure of complex equations like (7.2) is to rewrite them as a *single matrix differential equation* as follows.

$$\begin{bmatrix} m\,\ddot{x} + (k_1+k_2)\,x - l(k_1-k_2)\,\theta \\ \mathrm{J}\,\ddot{\theta} - l(k_1-k_2)\,x + l^2(k_1+k_2)\,\theta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{7.3}$$

$$\Leftrightarrow \quad \begin{bmatrix} m & 0 \\ 0 & \mathrm{J} \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} k_1+k_2 & l(k_2-k_1) \\ l(k_2-k_1) & l^2(k_1+k_2) \end{bmatrix}\begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{7.4}$$

The equation (7.3) is exactly the two equations (7.2) written as the two components of a vector. This vector equation can be further rearranged into the matrix-vector equation (7.4) in which each term represents all derivatives of a given order. Here the two terms collect derivatives of second and zeroth orders respectively.

An alternate model is to rewrite equations (7.1) using the coordinates (x_1, x_2) by substituting for (x, θ) in terms of (x_1, x_2). First note that the mapping $(x, \theta) \mapsto (x_1, x_2)$ in (7.2) can be written in matrix-vector form and inverted as follows

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -l \\ 1 & l \end{bmatrix}\begin{bmatrix} x \\ \theta \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} x \\ \theta \end{bmatrix} = \frac{1}{2l}\begin{bmatrix} l & l \\ -1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (x_1+x_2)/2 \\ (x_2-x_1)/2l \end{bmatrix}$$

Substituting these expressions for (x, θ) in (7.1) gives

$$\begin{aligned} m\,(\ddot{x}_1+\ddot{x}_2)/2 &= -k_1x_1 - k_2x_2 \\ \mathrm{J}\,(-\ddot{x}_1+\ddot{x}_2)/2l &= l\,k_1x_1 - l\,k_2x_2 \end{aligned}$$

These equations can be reorganized into a matrix-vector form similar to (7.4) as follows

$$\begin{bmatrix} m & m \\ -\mathrm{J} & \mathrm{J} \end{bmatrix}\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} + 2\begin{bmatrix} k_1 & k_2 \\ l^2k_1 & -l^2k_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0. \tag{7.5}$$

[1] This is clearly not the case if the beam is assumed to be a rigid body. If $x_1 \neq x_2$, then the support points will have to move laterally. This motion can be assumed negligible for small oscillations θ. An alternative way to view this is that if we write the full two dimensional model that allows for lateral motion, then linearizing around $\theta \approx 0$ will give the model above that ignores lateral motion.

N-DOF Mass-Spring-Damper System in Matrix Form

Note the similarity in the structures of the equations (7.4) and (7.5). They are both of the following form

$$\begin{bmatrix} \quad \mathsf{M} \quad \end{bmatrix} \begin{bmatrix} \ddot{\boldsymbol{x}}(t) \end{bmatrix} + \begin{bmatrix} \quad \mathsf{K} \quad \end{bmatrix} \begin{bmatrix} \boldsymbol{x}(t) \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \qquad \Leftrightarrow \qquad \mathsf{M}\,\ddot{\mathsf{x}}(\mathsf{t}) \;+\; \mathsf{K}\,\mathsf{x}(\mathsf{t}) \;=\; \mathsf{0}, \tag{7.6}$$

where $\boldsymbol{x}(t)$ is a *vector* of coordinates, and M, K are *matrices* of system's coefficients. Such equations are the N-DOF generalizations of the single DOF equation for a Mass-Spring matrix. M is referred to as the "mass matrix", and similarly K is called the stiffness matrix. The two equations in (7.6) are exactly the same. The first equation graphically emphasizes the dimensions of the matrices and vectors for illustration. The second equation is written in the much more compact matrix-vector form, but the reader should always keep in mind that despite the simplicity of writing it in that form, a lot of information is "coded into" the entries of the matrices M and K as can be seen by examining (7.4) and (7.2).

Decoupling and Normal Modes

When analyzing N-DOF systems written in a matrix form, certain coordinates are much better than others. Take for example the equations (7.4) written in (x, θ) coordinates, and assume a special case where $k_1 = k_2 = k$ and $l_1 = l_2 = l$. The equations then simplify to

$$\begin{bmatrix} m & 0 \\ 0 & \mathrm{J} \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} 2k & 0 \\ 0 & 2kl^2 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} = 0. \tag{7.7}$$

Note that the matrices M and K are now *diagonal* (i.e. all off-diagonal entries are zero). This matrix equation is therefore *decoupled* into two scalar equations

$$m\ \ddot{x}(t) \;+\; 2k\ x(t) \;=\; 0, \tag{7.8}$$

$$\mathrm{J}\ \ddot{\theta}(t) \;+\; 2kl^2\ \theta(t) \;=\; 0. \tag{7.9}$$

Note that the differential equation for $x(t)$ does not involve $\theta(t)$, and similarly, the differential equation for $\theta(t)$ does not involve $x(t)$. This means each differential equation can be solved independently of the other. In this case we say that the *dynamics of x and θ are decoupled* in the sense that they do not influence each other. We call the motion of $x(t)$ the *vertical (or heaving) mode*, and that of θ the *rotational mode* of vibration respectively. Those are the *normal modes* of vibration of this 2-DOF system. In general, an n-DOF system will have n different normal modes of vibration.

Let's make more simplifying assumptions to get some intuition for the distinction between vertical and rotational vibrations. Assume that the support points of the sprung beam in Figure 7.1 are at each end. Additionally, assume the beam is slender so that its moment of inertia (about the center of mass) is that of a bar of uniform length $2l$

$$\mathrm{J} \;=\; m(2l)^2/12 \;=\; ml^2/3$$

With these assumptions, the rotational motion equation (7.9) becomes

$$ml^2/3\ \ddot{\theta} \;+\; 2kl^2\ \theta \;=\; 0 \qquad \Rightarrow \qquad \ddot{\theta} \;+\; \frac{6k}{m}\ \theta \;=\; 0. \tag{7.10}$$

Now comparing the vertical motion equation (7.8) and the rotational motion equation (7.10), we see that they are both of the Mass-Spring type, but with two different natural frequencies

$$\begin{aligned} \text{vertical vibrations:} \quad & \omega_{\mathrm{n}} \;=\; \sqrt{2k/m} \\ \text{rotational vibrations:} \quad & \omega_{\mathrm{n}} \;=\; \sqrt{6k/m} \;=\; \sqrt{3}\ \sqrt{2k/m}. \end{aligned} \tag{7.11}$$

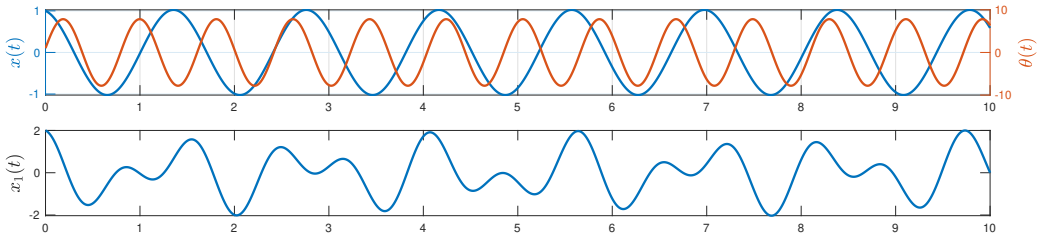


Figure 7.2: Vibrations of the sprung beam model in the special case where vertical and rotational dynamics (7.8)-(7.9) are decoupled. (*Top*) Both rotational $\theta(t)$ and vertical $x(t)$ vibrations are pure sinusoids, but with different frequencies. The frequency of rotational vibrations $\theta(t)$ is higher than that of vertical vibrations $x(t)$ as predicted by the formulas (7.11) for their respective natural frequencies. (*Bottom*) The vibration of the support point $x_1(t)$ is a superposition of the two modes (vertical and rotational) of vibrations.

Thus rotational vibrations have a natural frequency that is $\sqrt{3}$ times higher than that of vertical vibrations.

Any free vibrations of this 2-DOF system will be a *superposition* of those two modes of vibrations with those two different frequencies. For example, the motion of one of the support points, say $x_1(t)$ is linear combination of both $x(t)$ and $\theta(t)$ since (recall (7.2))

$$x_1(t) \;=\; x(t) \;-\; l\;\theta(t).$$

Figure 7.2 shows an example of the vibration of such a system. Note how vertical motion $x(t)$ and rotational motion $\theta(t)$ are both pure sinusoids, but with different natural frequencies. The vibration of $x_1(t)$ appears to be some superposition (a linear combination) of the two motions.

The fact that the matrix equation (7.7) involved only *diagonal* matrices is what led to the decoupling of the dynamics of x and θ as described by the two mutually-independent equations (7.8) and (7.9). Although this looks like a "lucky accident" due to the symmetry of the problem, there is a general method to take any coupled matrix problem representing an n-DOF system, and find the fundamental *normal modes* of vibration in a similar manner to what was done above. The mathematical technique is based on matrix diagonalization, which in turn is based on finding eigenvalues of eigenvectors of matrices. In Section 7.2 we will see how to do this in general without having a "lucky guess" of what the right coordinates should be for each problem.

7.1 Modeling with Vector Differential Equations

The dynamics of any N-degree of freedom mechanical systems with either linear or linearized dynamics can be expressed as a *vector* differential equation of the form

$$\mathsf{M}\,\ddot{\boldsymbol{x}}(t) \;+\; \mathsf{C}\,\dot{\boldsymbol{x}}(t) \;+\; \mathsf{K}\,\boldsymbol{x}(t) \;=\; \mathsf{B}_2\,\ddot{\boldsymbol{u}}(t) \;+\; \mathsf{B}_1\,\dot{\boldsymbol{u}}(t) \;+\; \mathsf{B}_0\,\boldsymbol{u}(t) \tag{7.12}$$

$$\Leftrightarrow \quad \begin{bmatrix} \\ \mathsf{M} \\ \\ \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \vdots \\ \ddot{x}_N(t) \end{bmatrix} + \begin{bmatrix} \\ \mathsf{C} \\ \\ \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix} + \begin{bmatrix} \\ \mathsf{K} \\ \\ \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix}$$

$$= \begin{bmatrix} \\ \mathsf{B}_2 \\ \\ \end{bmatrix} \begin{bmatrix} \ddot{u}_1(t) \\ \vdots \\ \ddot{u}_q(t) \end{bmatrix} + \begin{bmatrix} \\ \mathsf{B}_1 \\ \\ \end{bmatrix} \begin{bmatrix} \dot{u}_1(t) \\ \vdots \\ \dot{u}_q(t) \end{bmatrix} + \begin{bmatrix} \\ \mathsf{B}_0 \\ \\ \end{bmatrix} \begin{bmatrix} u_1(t) \\ \vdots \\ u_q(t) \end{bmatrix},$$

where $u_1(t), \ldots, u_q(t)$ are q "input signals", which might be externally applied forces, displacements of other externally applied variables. All the inputs are combined into the vector signal $\boldsymbol{u}(t)$.

Let's parse through the notation above carefully. Matrix-vector notation allows for very compact expressions, but it hides details! If you're not too familiar with it, you may not notice the details at first. The expressions above are written in compact matrix notation on the first line, and this notation is expressed more explicitly on the second line. Capital letters M, C, K and B_k are used

for matrices, bold small letters are used to express vectors (which may be functions of time as well) such as

$$\boldsymbol{x}(t) \;=\; \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad \dot{\boldsymbol{x}}(t) \;=\; \begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix}, \qquad \text{etc.}$$

Thus when we write a matrix-vector equation like (7.12), a lot of detail is contained in the matrix entries of M, C, K and $\{\mathsf{B}_k\}$. In fact, all the dynamical properties of this system are "encoded" in those matrices.

The equation (7.12) is the most general form for the dynamics of an n-degree of freedom mechanical system with linear dynamics. If a system has k masses where each can move in l coordinates, then $n = kl$. Those masses can be connected by any arrangement of springs and dampers and possibly kinematic constraints. It is also the form of the linearization of multi-link systems that arise from e.g. modeling robotic manipulators. Since equation (7.12) is a generalization of the single degree of freedom (Mass-Spring-Damper) system, we therefore refer to M, C and K as the "generalized" mass, damping and spring constant *matrices* respectively. The matrix K is also more commonly referred to as the "stiffness" matrix.

The advantage of expressing all mechanical systems in the matrix-vector form (7.12) cannot be overstated. It allows us to develop convenient formulas for solutions and frequency responses using compact matrix algebra, which can also be easily programmed in software like MATLAB. In addition, *normal mode* analysis which we will see in the next section is best expressed in linear algebra language in terms of eigenvalues and eigenvectors of matrices.

As already mentioned, the vector $\boldsymbol{u}(t)$ contains all the external "inputs" that may be acting on the system, which may be external forces, or kinematic quantities like displacements or velocities. Figure 7.3 shows three examples with the corresponding matrix-vector form of their dynamics. Note that in the first two examples, the external inputs are forces, while in the third example, the external input is actually a displacement. This variety of types of inputs is why we denote them generally by u rather than f in (7.12), the latter being notation for forces. The size q of the vector u in (7.12) is the *number of inputs*, which is not usually the same as the number n of mechanical degrees of freedom. When $q < n$, we call the system *underactuated*, and when $q = n$, it is *fully actuated*.

We now detail the derivations of the differential equations for each of the examples in that figure.

Example 7.1. (Example I, Figure 7.3) A convenient coordinate system to choose is where $x_1 = 0$ and $x_2 = 0$ when all three springs are in equilibrium. The equations of motion for the two masses are then

$$\begin{aligned} m_1\,\ddot{x}_1(t) \;&=\; -\,k_1 x_1(t) \;-\; k_c\big(x_1(t) - x_2(t)\big) \;+\; f_1(t), \\ m_2\,\ddot{x}_2(t) \;&=\; -\,k_2 x_2(t) \;+\; k_c\big(x_1(t) - x_2(t)\big) \;+\; f_2(t). \end{aligned}$$

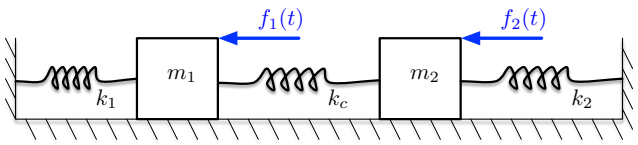
These two equations can be rewritten as a single matrix-vector equation as follows

$$\begin{aligned} \begin{bmatrix} m_1\ddot{x}_1(t) \\ m_2\ddot{x}_2(t) \end{bmatrix} &= -\begin{bmatrix} k_1 x_1(t) + k_c\big(x_1(t) - x_2(t)\big) \\ k_2 x_2(t) - k_c\big(x_1(t) - x_2(t)\big) \end{bmatrix} \;+\; \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} \\ \Rightarrow \qquad \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} &= -\begin{bmatrix} k_1 + k_c & -k_c \\ -k_c & k_2 + k_c \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \;+\; \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} \\ \Rightarrow \qquad \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} &+ \begin{bmatrix} k_1 + k_c & -k_c \\ -k_c & k_2 + k_c \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \;=\; \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}. \end{aligned}$$

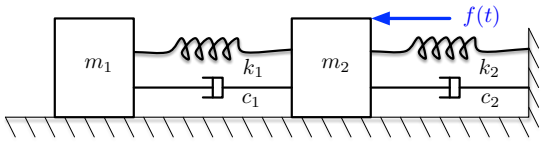
This last equation is of the general form (7.12)

$$\begin{aligned} &\mathsf{M}\,\ddot{\boldsymbol{x}}(t) \;+\; \mathsf{K}\,\boldsymbol{x}(t) \;=\; \boldsymbol{f}(t), \\ &\text{where} \qquad \mathsf{M} := \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad \mathsf{K} := \begin{bmatrix} k_1 + k_c & -k_c \\ -k_c & k_2 + k_c \end{bmatrix}, \quad \boldsymbol{x}(t) := \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad \boldsymbol{f}(t) := \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}. \end{aligned}$$

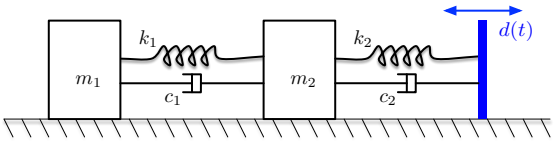
Comparing this equation with the general form (7.12), we see that in this case the damping matrix $C = 0$ (since there is no damping in this problem), the input $\boldsymbol{u}(t) = \boldsymbol{f}(t)$ vector of forces enters with no differentiation, thus $\mathsf{B}_2 = 0$, $\mathsf{B}_1 = 0$ and $\mathsf{B}_0 = \mathrm{I}$, the identity matrix.



(I):
$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1+k_c & -k_c \\ -k_c & k_2+k_c \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}$$



(II):
$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} c_1 & -c_1 \\ -c_1 & c_1+c_2 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1+k_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} f(t)$$



(III):
$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} c_1 & -c_1 \\ -c_1 & c_1+c_2 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1+k_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ c_2 \end{bmatrix} \dot{d}(t) + \begin{bmatrix} 0 \\ k_2 \end{bmatrix} d(t)$$

Figure 7.3: Three examples of mechanical systems with their dynamics written in the matrix-vector form (7.12). Note that in Systems (i) and (II), the "external inputs" are forces, while in System (III) the external input is a displacement.

Example 7.2. (Example II, Figure 7.3) The most convenient choice of coordinates here is again where $x_1 = 0$ and $x_2 = 0$ when both springs are in equilibrium. The equations of motion are

$$\begin{aligned} m_1\, \ddot{x}_1(t) &= -k_1\big(x_1(t) - x_2(t)\big) - c_1\big(\dot{x}_1(t) - \dot{x}_2(t)\big) \\ m_2\, \ddot{x}_2(t) &= k_1\big(x_1(t) - x_2(t)\big) + c_1\big(\dot{x}_1(t) - \dot{x}_2(t)\big) - k_2\, x_2(t) - c_2\, \dot{x}_2(t) + f(t), \end{aligned} \tag{7.13}$$

which can be rewritten in matrix-vector form as

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} c_1 & -c_1 \\ -c_1 & c_2 + c_1 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_2 + k_1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} f(t).$$

This equation is again of the form (7.12) with

$$\mathsf{M}\, \ddot{\boldsymbol{x}}(t) + \mathsf{C}\, \dot{\boldsymbol{x}}(t) + \mathsf{K}\, \boldsymbol{x}(t) = \mathsf{B}_0\, \boldsymbol{f}(t).$$

The input $f(t)$ is a scalar-valued function, but it enters the matrix equation through multiplication by the matrix

$$\mathsf{B}_0 := \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

since it only enters the second equation in (7.13). This is an example of when the B_i matrices are non-square because the number of inputs is not equal to the number of mechanical degrees of freedom.

Note that even though the external forcing $f(t)$ enters only in the dynamics of x_2, it does have (indirect) effects on the dynamics of x_1, since the latter are influenced by x_2 through the coupling in the equations. This point will be emphasized when we analyze this example later as a prototype of a certain vibration isolation technique termed "tuned mass dampers".

Example 7.3. (Example III, Figure 7.3) In this example, the input is a displacement rather than a force. In such problems, the choice of convenient coordinates requires a little care. Consider measuring x_1, x_2 and u with one coordinate system, then the equations of motion are

$$\begin{aligned} m_1\ddot{x}_1(t) &= -k_1\left(x_1(t) - x_2(t) - L_1\right) - c_1\left(\dot{x}_1(t) - \dot{x}_2(t)\right) \\ m_2\ddot{x}_2(t) &= \ k_1\left(x_1(t) - x_2(t) - L_1\right) + c_1\left(\dot{x}_1(t) - \dot{x}_2(t)\right) - k_2\left(x_2(t) - d(t) - L_2\right) - c_2\big(\dot{x}_2(t) - \dot{d}(t)\big), \end{aligned}$$

where L_1 and L_2 are constants such that when $x_1 - x_2 - L_1 = 0$, spring k_1 is in equilibrium, and when $x_2 - u = 0$, spring k_2 is in equilibrium. Now redefining x_1 as $x_1 - L_1$ and x_2 as $x_2 - L_2$, the new equations of motion are written more nicely (without the constant terms) as

$$\begin{aligned} m_1\ddot{x}_1(t) &= -k_1\Big(x_1(t) - x_2(t)\Big) - c_1\Big(\dot{x}_1(t) - \dot{x}_2(t)\Big) \\ m_2\ddot{x}_2(t) &= \ k_1\Big(x_1(t) - x_2(t)\Big) + c_1\Big(\dot{x}_1(t) - \dot{x}_2(t)\Big) - k_2\Big(x_2(t) - d(t)\Big) - c_2\Big(\dot{x}_2(t) - \dot{d}(t)\Big). \end{aligned}$$

Those two equations can now be rewritten as a single matrix equation

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}\begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_2 + k_1 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} c_1 & -c_1 \\ -c_1 & c_1 + c_2 \end{bmatrix}\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} \ = \ \begin{bmatrix} 0 \\ c_2 \end{bmatrix}\dot{d}(t) + \begin{bmatrix} 0 \\ k_2 \end{bmatrix} d(t)$$

This equation is again of the form (7.12) with

$$\mathsf{M}\,\ddot{\boldsymbol{x}}(t) \ + \ \mathsf{C}\,\dot{\boldsymbol{x}}(t) \ + \ \mathsf{K}\,\boldsymbol{x}(t) \ = \ \mathsf{B}_1\,\dot{d}(t) \ + \ \mathsf{B}_0\,d(t),$$

where the matrices B_0 and B_1 are

$$\mathsf{B}_0 \ := \ \begin{bmatrix} 0 \\ k_2 \end{bmatrix}, \qquad \mathsf{B}_1 \ := \ \begin{bmatrix} 0 \\ c_2 \end{bmatrix}.$$

Note again that the input in this case (which was denoted as a general signal u in (7.12)) is a displacement d rather than a force.

This system has a single-input $d(t)$ and 2 mechanical degrees of freedom, thus the matrices B_0 and B_1 are 2×1, while the matrices $\mathsf{M}, \mathsf{C}, \mathsf{K}$ are 2×2.

In many systems (though not all, see e.g. the next example), the "mass matrix" M will typically be diagonal as is the case in the three examples just described. In this case, every "row" in equation (7.12) represents a scalar equation of the form $m_i\ddot{x}_i = \sum_j f_{ij}$ for the i'th mass, with $\sum_j f_{ij}$ being the sum of forces on the i'th mass. Note that we have as many equations as masses, and therefore the reader should check that this implies that M, C and K must be square ($N \times N$) matrices. The "B matrices" however can be non-square, i.e. $N \times q$, where q is the number of external forces (or inputs) acting on the system, which may not be the same as the number of masses.

In general, the mass matrix need not be a diagonal matrix as the following example of the combined mass and pendulum of Figure 7.4 illustrates.

Example 7.4. (Example V: Linear Mass Motion with Pendulum) Consider the system of a block M with linear motion coupled to a pendulum as shown in Figure 7.4.

The equations of motion of this system are best derived using the Lagrangian formulation of mechanics. The total kinetic and potential energies of this system (see diagrams in Figure 7.4) and the Lagrangian are

$$\begin{aligned} \text{KE} &= \tfrac{1}{2}\left(M\,\dot{x}^2 + m\left(\left(\dot{x} + l\dot{\theta}\cos\theta\right)^2 + (l\dot{\theta}\sin\theta)^2\right)\right) = \tfrac{1}{2}\left((M+m)\,\dot{x}^2 + 2ml\,\dot{x}\dot{\theta}\cos\theta + ml^2\dot{\theta}^2\right) \\ \text{PE} &= \tfrac{1}{2}k\,x^2 + mgl\,(1 - \cos(\theta)) \\ \mathcal{L}(x,\theta,\dot{x},\dot{\theta}) &= \text{KE} - \text{PE} = \tfrac{1}{2}\left((M+m)\,\dot{x}^2 + 2ml\,\dot{x}\dot{\theta}\cos\theta + ml^2\dot{\theta}^2 - k\,x^2\right) - mgl\,(1 - \cos(\theta)) \end{aligned}$$

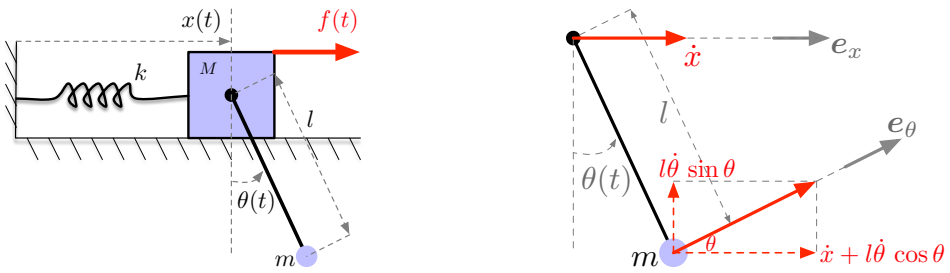


Figure 7.4: (*Left*) The system of Example 7.4 consisting of a mass constrained to move laterally, and an attached pendulum swinging freely from a frictionless pivot. (*Right*) The velocity of m relative to M is $V_{m/M} = l\dot{\theta}\,\boldsymbol{e}_\theta$. Thus the velocity of m in the inertial frame is $V_m = V_M + V_{m/M} = \dot{x}\,\boldsymbol{e}_x + l\dot{\theta}\,\boldsymbol{e}_\theta$ has the lateral and vertical components shown.

The first EL equation and its linearization is

$$\begin{aligned} -f \;=\; \frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}} &= -k\,x - \tfrac{d}{dt}\Big((M+m)\dot{x} + ml\,\dot{\theta}\cos\theta\Big) \\ &= -k\,x - (M+m)\ddot{x} - ml\,\Big(\ddot{\theta}\cos\theta - \dot{\theta}^2\sin\theta\Big) \\ &\approx -k\,x - (M+m)\ddot{x} - ml\,\ddot{\theta}. \end{aligned}$$

The second EL equation and its linearization is similarly derived

$$\begin{aligned} 0 \;=\; \frac{\partial \mathcal{L}}{\partial \theta} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\theta}} &= -ml\,\dot{x}\,\dot{\theta}\,\sin\theta - mgl\sin\theta - \tfrac{d}{dt}\Big(ml\,\dot{x}\,\cos\theta + ml^2\dot{\theta}\Big) \\ &= -ml\,\dot{x}\,\dot{\theta}\,\sin\theta - mgl\sin\theta - ml\,\ddot{x}\,\cos\theta + ml\,\dot{x}\,\dot{\theta}\,\sin\theta - ml^2\ddot{\theta} \\ &\approx -mgl\,\theta - ml\,\ddot{x} - ml^2\ddot{\theta} \end{aligned}$$

The linearized equations can be written in a matrix-vector form as

$$\begin{bmatrix} M+m & ml \\ 1 & l \end{bmatrix}\begin{bmatrix} \ddot{x}(t) \\ \ddot{\theta}(t) \end{bmatrix} + \begin{bmatrix} k & 0 \\ 0 & g \end{bmatrix}\begin{bmatrix} x(t) \\ \theta(t) \end{bmatrix} \;=\; \begin{bmatrix} 1 \\ 0 \end{bmatrix} f(t). \tag{7.14}$$

Here we see that while the stiffness matrix is diagonal, the mass matrix is not, thus providing the coupling between the two equations. Note that it is possible to rewrite those equations in an equivalent form with a diagonal mass matrix (in fact a mass matrix which is the identity) by multiplying both sides of the equation by the inverse of the mass matrix

$$\begin{bmatrix} M+m & ml \\ 1 & l \end{bmatrix}^{-1} \;=\; \frac{1}{Ml}\begin{bmatrix} l & -ml \\ -1 & M+m \end{bmatrix},$$

which then renders Equation (7.14) as

$$\begin{bmatrix} \ddot{x}(t) \\ \ddot{\theta}(t) \end{bmatrix} + \frac{1}{Ml}\begin{bmatrix} lk & -mlg \\ -k & (M+m)g \end{bmatrix}\begin{bmatrix} x(t) \\ \theta(t) \end{bmatrix} \;=\; \frac{1}{Ml}\begin{bmatrix} l \\ -1 \end{bmatrix} f(t). \tag{7.15}$$

The mass matrix is now the identity matrix, which is not explicitly written out.

We note that in the previous example, Equations (7.14) and (7.15) represent the same exact system. Which set of equations we use for analysis is a matter of convenience. In fact, since any N-DOF system involves N coupled differential equations, by taking linear combinations of these linear equations, we can generate a second set of *equivalent* differential equations. This is precisely what was done by the matrix multiplication that converted (7.14) to (7.15). There is an infinite number of ways one can take linear combinations of coupled equations and rewrite equivalent equations. There is however one special way of doing this that renders the equations very simple to analyze, and in fact renders them decoupled. This is the subject of normal mode analysis to which we now turn.

7.2 Free Vibrations via Normal Mode Analysis

Recall the motivating example (7.7) in the introduction where the matrices were diagonal due to some very special assumptions on the system. Any system with interesting dynamics will typically not be uncoupled like that in (7.7). However, it turns out that it is always possible to rewrite the equations in a different coordinate system such that the equations become decoupled in the new coordinates. There are two diagonalization procedures depending on whether $C = 0$ (no damping), or otherwise. In this section we present the simpler case of no damping. The most general case with damping requires the use of state-space methods described in Section **??**.

Consider now a system of the general form (7.12) but with no damping ($C = 0$), and no external inputs $\boldsymbol{u}(t) = 0$. Equation (7.12) can then be manipulated as follows

$$\mathsf{M}\,\ddot{\boldsymbol{x}}(t) + \mathsf{K}\,\boldsymbol{x}(t) = 0 \qquad \Leftrightarrow \qquad \mathsf{M}\,\ddot{\boldsymbol{x}}(t) = -\mathsf{K}\,\boldsymbol{x}(t) \qquad \Leftrightarrow \qquad \boxed{\ddot{\boldsymbol{x}}(t) = -\left(\mathsf{M}^{-1}\mathsf{K}\right)\ \boldsymbol{x}(t)}, \tag{7.16}$$

where the last equation is obtained by multiplying both sides of the previous one by M^{-1} from the left[2]. Note again that this last equation is a *vector* differential equation. The solutions and dynamical properties of this equation are completely determined by the matrix $\mathsf{M}^{-1}\mathsf{K}$. This matrix is the counterpart of the stiffness-to-mass ratio k/m we encountered when studying single DOF system. Thus in the vector case, the dynamical properties are determined by an entire $N \times N$ matrix rather than a single number. However, it turns out that the important dynamical properties are not revealed by looking at the individual N^2 entries of the matrix $\mathsf{M}^{-1}\mathsf{K}$, but rather by looking at its *eigenvalues and eigenvectors*. To illustrate this point, we first consider very special solutions related to eigenvectors and known as the normal modes of vibrations. Then in the remainder of the section, we show how general solutions can be regarded as a "superposition" of normal modes by using the technique of matrix diagonalization.

For the illustration, assume for simplicity that $\mathsf{M} = \mathrm{I}$ (the identity matrix) and therefore $\mathsf{M}^{-1}\mathsf{K} = \mathsf{K}$. Recall that an eigenvector of a matrix K is a vector $\boldsymbol{v}$ such that

$$\mathsf{K}\,\boldsymbol{v} \;=\; \lambda\,\boldsymbol{v},$$

for some (possibly complex) number λ. It can be shown (see Appendix 7.B) that stiffness matrices K of mechanical systems always have real and non-negative eigenvalues, i.e. $\lambda \geq 0$ always. For any given eigenvector $\boldsymbol{v}$ of K, we can show by direct verification that the following vector function of time

$$\begin{bmatrix} \boldsymbol{x}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} \end{bmatrix} \mathsf{a}\cos(\omega t + \theta), \qquad \omega = \sqrt{\lambda} \tag{7.17}$$

for any amplitude a and phase θ is a solution of the differential equation $\ddot{\boldsymbol{x}}(t) = -\mathsf{K}\,\boldsymbol{x}(t)$. Indeed, given the form above, we see that $\ddot{\boldsymbol{x}}(t)$ is a scalar multiple of $\boldsymbol{x}(t)$

$$\begin{aligned} \ddot{\boldsymbol{x}}(t) \;&=\; \tfrac{d^2}{dt^2}\left(\boldsymbol{v}\ \mathsf{a}\cos(\omega t + \theta)\right) \;&=\; -\,\omega^2\ \left(\boldsymbol{v}\ \mathsf{a}\cos(\omega t + \theta)\right) \\ -\mathsf{K}\ \boldsymbol{x}(t) \;&=\; -\mathsf{K}\left(\boldsymbol{v}\ \mathsf{a}\cos(\omega t + \theta)\right) \;&=\; -\,\mathsf{K}\boldsymbol{v}\ \mathsf{a}\cos(\omega t + \theta) \\ & & =\; -\,\lambda\ \ \left(\boldsymbol{v}\ \mathsf{a}\cos(\omega t + \theta)\right). \qquad & (\text{since } \mathsf{K}\boldsymbol{v} = \lambda\boldsymbol{v}) \end{aligned}$$

Thus if we choose $\omega = \sqrt{\lambda}$, then the function (7.17) indeed satisfies the differential equation $\ddot{\boldsymbol{x}}(t) = -\mathsf{K}\,\boldsymbol{x}(t)$.

Note the "separation structure" of the solution (7.17) which is the product of a constant (independent of t) vector $\boldsymbol{v}$ and a scalar, oscillatory function of time $\mathsf{a}\cos(\omega t + \theta)$. If we think of $\boldsymbol{x}(t)$

[2] In this setting, the mass matrix M is always assumed to be non-singular. When M is diagonal, this is equivalent to all the masses being non-zero.

as a time-varying vector in N-dimensional space, then the solution (7.17) oscillates along a single direction $\boldsymbol{v}$ which does not change for all $t \geq 0$. Note also that the initial conditions of this solution are

$$\boldsymbol{x}(0) \;=\; \boldsymbol{v}\; \mathsf{a}\cos(\theta) \qquad \dot{\boldsymbol{x}}(0) \;=\; \boldsymbol{v}\; (-\mathsf{a}\omega \sin(\theta)),$$

i.e. both $\boldsymbol{x}(0)$ and $\dot{\boldsymbol{x}}(0)$ are vectors in the same direction as $\boldsymbol{v}$. The converse is also true as shown in Section 7.3; if the initial positions and velocities are in the direction of the eigenvector $\boldsymbol{v}$, then the solution $\boldsymbol{x}(t)$ remains (for all t) in the direction of the eigenvector $\boldsymbol{v}$! It oscillates in time with frequency $\omega = \sqrt{\lambda}$ determined by the *eigenvalue* λ corresponding to the *eigenvector* $\boldsymbol{v}$. This is a very special set of initial conditions and solutions, and they are called the *normal modes* of vibration. There are N of them since an $N \times N$ matrix like K will have N eigenvalues and vectors. Thus an N-DOF system has N *natural frequencies* and N *normal modes* of vibration. For more general initial conditions (i.e. those not corresponding to eigenvectors), the typical response will be a superposition of all the normal modes of the system. All these phenomena are illustrated in Figure 7.5 and the included animations for the example of System I. The full analysis of this example will be given later in Example 7.7 once the concepts of normal mode analysis are fully explained.

The systematic analysis of general n-DOF systems is done by diagonalizing the matrix $\mathsf{M}^{-1}\mathsf{K}$, which is equivalent to changing coordinates to a more convenient set of coordinates given by the eigenvectors of $\mathsf{M}^{-1}\mathsf{K}$. The concept of diagonalization of matrices is explored in detail in Appendix 7.A, which the reader is now highly encouraged to review. Here we are interested in showing what diagonalization implies for the dynamical equation. First we explore how the equation (7.16) transforms when we use a different coordinate system.

7.2.1 Change of Coordinates

Let V be some $N \times N$, non-singular, constant (i.e. does not depend on t) matrix. Suppose we define new variables by $\boldsymbol{y}(t) := V^{-1}\boldsymbol{x}(t)$, or equivalently $\boldsymbol{x}(t) = V\boldsymbol{y}(t)$. This has two possible interpretations. The first comes from writing the matrix-vector product in detail as

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \begin{bmatrix} \\ V \\ \\ \end{bmatrix} \begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \begin{bmatrix} v_{11}y_1(t) + \cdots + v_{1N}y_N(t) \\ \vdots \\ v_{N1}y_1(t) + \cdots + v_{NN}y_N(t) \end{bmatrix}, \tag{7.18}$$

where v_{ij} is the ij'th entry of the matrix V. The above equation says that each function $x_i(t)$ is written as a *linear combination* of the n functions $\{y_1(t), \ldots, y_N(t)\}$.

The second interpretation comes from partitioning the matrix V columnwise

$$\begin{bmatrix} \\ \boldsymbol{x}(t) \\ \\ \end{bmatrix} = \begin{bmatrix} \vdots & & \vdots \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_N \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{bmatrix} \\ \boldsymbol{x}(t) \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{v}_1 \\ \\ \end{bmatrix} y_1(t) \; + \cdots + \; \begin{bmatrix} \\ \boldsymbol{v}_N \\ \\ \end{bmatrix} y_N(t). \tag{7.19}$$

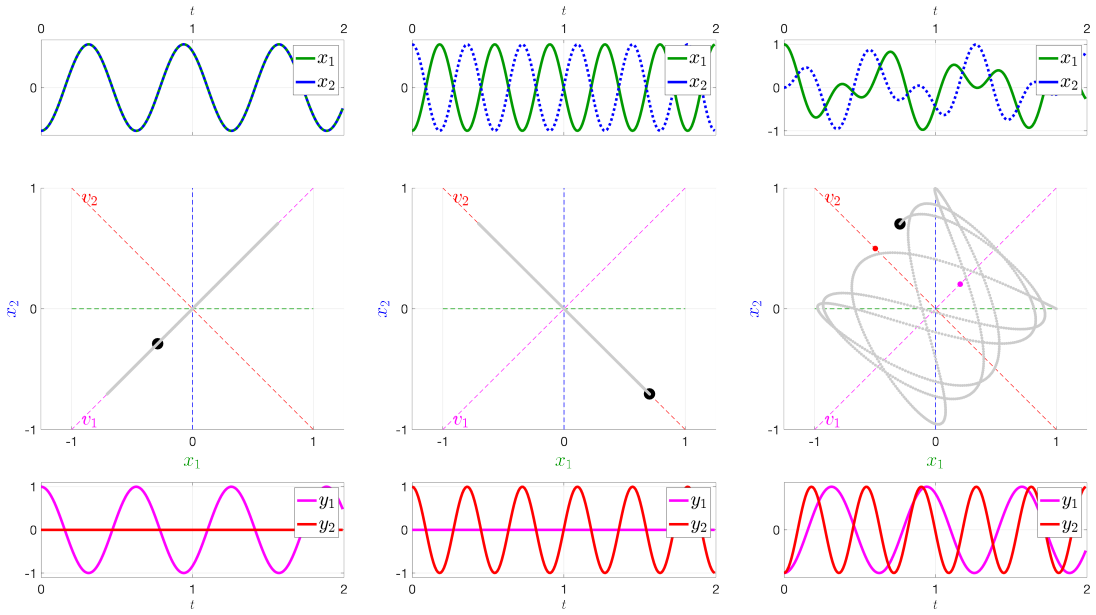
This means that at each time t, the vector $\boldsymbol{x}(t)$ is written in terms of the *basis* $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$, with the scalars $y_1(t), \ldots, y_N(t)$ as the coefficients in that basis. Since $\boldsymbol{x}(t)$ is time varying, so are the coefficients. The basis however is fixed in time. Viewing the motion of $\boldsymbol{x}(t)$ using either (7.18) or (7.19) gives two different insights. In the example of Figure 7.5b (the case of a generic initial condition), the plots of $x_1(t)$ and $x_2(t)$ show that each position is a linear combination of two pure sinusoidal modes. This corresponds to the form (7.19). On the other hand, the trajectories of $x_1(t)$ versus $x_2(t)$ plotted in (x_1, x_2) space show how the motion of the vector $\big(x_1(t), x_2(t)\big)$ can be expressed in the eigenvectors basis by $\boldsymbol{x}(t) = y_1(t)\boldsymbol{v}_1 + y_2(t)\boldsymbol{v}_2$. Note how the motions of the basis expansion coefficients $y_1(t)$ and $y_2(t)$ are pure sinusoids.

7.2.2 The Dynamics in New Coordinates

When we change coordinates as done above, we can derive the dynamics (i.e. the differential equation) that the vector $\boldsymbol{y}(t)$ obeys from the original differential equation for $\boldsymbol{x}(t)$. The reason for

click for animation click for animation click for animation

(a) Motion of System (I) starting from equally displaced positions (*left*), oppositely displaced positions (*middle*), and a generic initial displacement (*right*). These lead to unison motion, differential motion, and motion combining both unison and differential modes respectively. Note the differences in oscillation frequencies between unison and differential motion when viewing the animations.



click for animation click for animation click for animation

(b) Trajectories from the 3 different initial conditions shown in 3 different ways. Trajectories are shown as $\big(x_1(t), x_2(t)\big)$ versus time (*top*), $x_1(t)$ versus $x_2(t)$ in (x_1, x_2) space (black dot in *middle*), and its projections onto the eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ (magenta and red dots respectively). The coordinates $y_1(t), y_2(t)$ versus time of the point $\big(x_1(t), x_2(t)\big)$ in the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ (i.e. the modal coordinates) are shown at the (*bottom*). On the left and middle are "pure mode" initial conditions along $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ producing pure unison and differential motions respectively. On the right is a generic initial condition which produces a response with a *combination* of unison and differential modes. Note that the projections $y_1(t)$, $y_2(t)$ along the eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ are pure sinusoidal motions with frequencies $\omega_1 = \sqrt{\lambda_1}$, $\omega_2 = \sqrt{\lambda_2}$.

Figure 7.5: Trajectories of masses' positions for System (I) from three different sets of initial conditions. $m_1 = m_2 = 1$ and $k_1 = k_2 = k_c = 100$ in these simulations. Trajectories from equally displaced, oppositely displaced, and generic initial conditions are shown. These generate unison motion (in phase), differential motion (opposite phase), and a combination of the two modes respectively. The unison and differential modes have frequencies of $\omega_1 = \sqrt{\lambda_1}$ and $\omega_2 = \sqrt{\lambda_2}$ respectively, where λ_1, λ_2 are the eigenvalues of the matrix $M^{-1}K$, and $\boldsymbol{v}_1, \boldsymbol{v}_2$ are the corresponding eigenvectors.

changing coordinates is that with a judicious choice of the transformation V, the dynamics of $\boldsymbol{y}(t)$ will be simple and easily solvable. Once we find the solution for $\boldsymbol{y}(t)$, we can then *recover* the original variables from the relation $\boldsymbol{x}(t) = V\boldsymbol{y}(t)$.

In matrix-vector form, the equations (7.18)-(7.19) and their inverse is written as

$$\boldsymbol{x}(t) \;=\; V\,\boldsymbol{y}(t) \qquad \Leftrightarrow \qquad \boldsymbol{y}(t) \;=\; V^{-1}\,\boldsymbol{x}(t).$$

Since V and V^{-1} are constant matrices (do not depend on t), then the vector derivatives are simply related by

$$\begin{aligned} \dot{\boldsymbol{x}}(t) \;&=\; V\,\dot{\boldsymbol{y}}(t) \qquad \Leftrightarrow & \dot{\boldsymbol{y}}(t) \;&=\; V^{-1}\,\dot{\boldsymbol{x}}(t) \\ \ddot{\boldsymbol{x}}(t) \;&=\; V\,\ddot{\boldsymbol{y}}(t) \qquad \Leftrightarrow & \ddot{\boldsymbol{y}}(t) \;&=\; V^{-1}\,\ddot{\boldsymbol{x}}(t). \end{aligned}$$

Now derive the differential equation for $\boldsymbol{y}(t)$ using these relations and the differential equation (7.16) for $\boldsymbol{x}(t)$

$$\begin{aligned} & \ddot{\boldsymbol{x}}(t) = -\left(\mathsf{M}^{-1}\mathsf{K}\right)\ \boldsymbol{x}(t) && \text{(Starting from Eq. (7.16))} \\ \Rightarrow \quad & V^{-1}\ \ddot{\boldsymbol{x}}(t) = -V^{-1}\left(\mathsf{M}^{-1}\mathsf{K}\right)\ \boldsymbol{x}(t) && \text{(multiplying both sides from left by } V^{-1}\text{)} \\ \Rightarrow \quad & \ddot{\boldsymbol{y}}(t) = -\left(V^{-1}\left(\mathsf{M}^{-1}\mathsf{K}\right)V\right)\boldsymbol{y}(t) && \left(\begin{array}{r}\text{substituting } \ddot{\boldsymbol{y}}(t) = V^{-1}\ddot{\boldsymbol{x}}(t), \\ \text{and } \boldsymbol{x}(t) = V\boldsymbol{y}(t)\end{array}\right) \end{aligned} \tag{7.20}$$

At first it might seem that this last equation is more complicated than the one for $\boldsymbol{x}(t)$ since we replaced the matrix $\left(\mathsf{M}^{-1}\mathsf{K}\right)$ by the matrix $V^{-1}\left(\mathsf{M}^{-1}\mathsf{K}\right)V$. However, if we can choose V such $V^{-1}\left(\mathsf{M}^{-1}\mathsf{K}\right)V$ is simple, then we have a simpler differential equation. It is a fact that for the types of matrices that occur in mechanical vibrations problems, the matrix $\left(\mathsf{M}^{-1}\mathsf{K}\right)$ is always diagonalizable[3]. Diagonalization means (see Appendix 7.A) that we can find a non-singular matrix V such that

$$V^{-1}\left(\mathsf{M}^{-1}\mathsf{K}\right)V = \begin{bmatrix} & V^{-1} & \end{bmatrix}\begin{bmatrix} & \mathsf{M}^{-1}\mathsf{K} & \end{bmatrix}\begin{bmatrix} & V & \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} =: \Lambda, \tag{7.21}$$

where Λ is the diagonal matrix with the numbers $\{\lambda_1, \ldots, \lambda_N\}$ on the diagonal, and zero entries off the diagonal. The numbers $\{\lambda_1, \ldots, \lambda_N\}$ are the eigenvalues of $\left(\mathsf{M}^{-1}\mathsf{K}\right)$, and the columns of V are the corresponding eigenvectors.

Substituting the diagonalization (7.21) in the transformed equation (7.20), we finally arrive at the simple *decoupled* equation for $\boldsymbol{y}(t)$ which reads

$$\ddot{\boldsymbol{y}}(t) = -\Lambda\, \boldsymbol{y}(t) \quad \Leftrightarrow \quad \begin{bmatrix} \ddot{y}_1(t) \\ \vdots \\ \ddot{y}_N(t) \end{bmatrix} = \begin{bmatrix} -\lambda_1 & & \\ & \ddots & \\ & & -\lambda_N \end{bmatrix}\begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} \ddot{y}_1(t) = -\lambda_1 y(t) \\ \qquad\vdots \\ \ddot{y}_N(t) = -\lambda_N y(t) \end{cases}$$

The last set of equations are N, decoupled equations, each for a simple, undamped systems with natural frequency $\omega_i = \sqrt{\lambda_i}$. Therefore the solution of each is (written here in the "phased-cosine form")

$$y_i(t) = \mathsf{y}_i \cos\left(\sqrt{\lambda_i}\ t + \theta_i\right) \qquad i = 1, \ldots, N.$$

The exact values of the amplitudes $\{\mathsf{y}_i\}_{i=1}^N$ and phases $\{\theta_i\}_{i=1}^N$ depend on the initial conditions. However, the frequencies $\{\omega_i\}_{i=1}^N = \left\{\sqrt{\lambda_i}\right\}_{i=1}^N$ do not, and only depend on the eigenvalues of $\left(\mathsf{M}^{-1}\mathsf{K}\right)$, i.e. only on the system's parameters, and not on its initial conditions. We now summarize these conclusions.

Theorem 7.5. *Consider an N-degree of freedom, undamped, linear mechanical system of the form*

$$\mathsf{M}\ \ddot{\boldsymbol{x}}(t) + \mathsf{K}\ \boldsymbol{x}(t) = 0.$$

The solution is given as $\boldsymbol{x}(t) = V\,\boldsymbol{y}(t)$, where the components of the n-vector $\boldsymbol{y}(t)$ each satisfy the following scalar, uncoupled differential equations

$$\ddot{y}_i(t) = -\omega_i^2\ y(t), \qquad i = 1, \ldots, N, \tag{7.22}$$

where the oscillation frequencies $\omega_i := \sqrt{\lambda_i},\ i = 1, \ldots, N$ are called the normal modes *of the system, $\lambda_i,\ i = 1, \ldots, N$ are the (always non-negative) eigenvalues of the matrix $\left(\mathsf{M}^{-1}\mathsf{K}\right)$, and the columns of the matrix V are its eigenvectors.*

[3]The justification of this statement is beyond the scope of this chapter. The precise statement is that if M is symmetric positive definite, and K is symmetric, then $\mathsf{M}^{-1}\mathsf{K}$ is diagonalizable. These assumptions are satisfied by all linear mechanical systems.

Since we know how to reconstruct $\boldsymbol{x}(t)$ from $\boldsymbol{y}(t)$ (using the matrix V as in (7.18)), and we know the form of the solutions of the simple equations (7.22), we can now write the most general solution to our original system (7.16) in two different forms

$$\begin{aligned} y_i(t) &= \mathsf{y}_i \cos\big(\omega_i t + \theta_i\big), \qquad\qquad i = 1, \ldots, N, \\ \Rightarrow \qquad \begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix} &= \begin{bmatrix} v_{11}\ \mathsf{y}_1 \cos\big(\omega_1 t + \theta_1\big) + \cdots + v_{1n}\ \mathsf{y}_N \cos\big(\omega_N t + \theta_N\big) \\ \vdots \\ v_{n1}\ \mathsf{y}_1 \cos\big(\omega_1 t + \theta_1\big) + \cdots + v_{nn}\ \mathsf{y}_N \cos\big(\omega_N t + \theta_N\big) \end{bmatrix} \end{aligned} \tag{7.23}$$

$$\Rightarrow \qquad \begin{bmatrix} \\ \boldsymbol{x}(t) \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{v}_1 \\ \\ \end{bmatrix} \mathsf{y}_1 \cos\big(\omega_1 t + \theta_1\big) \ + \cdots + \ \begin{bmatrix} \\ \boldsymbol{v}_N \\ \\ \end{bmatrix} \mathsf{y}_N \cos\big(\omega_N t + \theta_N\big) \tag{7.24}$$

The form (7.23) says that the response $x_i(t)$ for each mechanical degree of freedom is generally *a linear combination* of N normal modes of the form $\mathsf{y}_i \cos(\omega_i t + \theta_i)$, i.e. *a mixture* of sinusoids of frequencies $\omega_1, \ldots, \omega_N$. These are the "natural frequencies" of the system. Since we have N mechanical degrees of freedom, we also have N *natural frequencies*.

The form (7.24) gives another perspective which can be interpreted as revealing the "shape of vibrations". If we think of $\boldsymbol{x}(t)$ as a vector in N-dimensional space, it is a linear combination of n basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$, each defining a particular direction. Equation (7.24) says that *the vector $\boldsymbol{x}(t)$ vibrates in each direction $\boldsymbol{v}_i$ with frequency ω_i*. Another look at Figure 7.5b will help clarify this point of view.

7.3 Initial Conditions and Pure Normal Mode Responses

As Equation (7.23) demonstrates, the response of a system will typically contain a linear combination of all normal modes. The exact mixture of those normal modes depends on the initial conditions. An interesting question is whether there are initial conditions that produce responses that are made up of *only one of the normal modes* and none of the others (i.e. a pure normal mode response)? Specifically, which initial conditions produce a response where only one frequency (say ω_k) appears in (7.23)-(7.24), with the terms corresponding to all other frequencies ω_i, $i \neq k$ being zero. As seen earlier in the special solutions (7.17), the eigenvectors play a role in answering this question.

To find the exact relations between the initial conditions of $\boldsymbol{x}(t)$ and those of $\boldsymbol{y}(t)$ we simply use the matrix relation $\boldsymbol{x}(t) = V\boldsymbol{y}(t)$ which implies (since V does not depend on time)

$$\boldsymbol{x}(0) \ = \ V\ \boldsymbol{y}(0), \qquad \dot{\boldsymbol{x}}(0) \ = \ V\ \dot{\boldsymbol{y}}(0).$$

These relations imply that choosing $\boldsymbol{x}(0)$ or $\dot{\boldsymbol{x}}(0)$ as scalings of one of the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ is equivalent to choosing $\boldsymbol{y}(0)$ or $\dot{\boldsymbol{y}}(0)$ as scalings of one of the canonical basis vectors $\boldsymbol{e}_k$, i.e.

$$\boldsymbol{x}(0) \ = \ V\boldsymbol{y}(0) \ = \ V\left(\alpha \boldsymbol{e}_k\right) \ = \ \left[\begin{array}{c:c:c:c} & & & \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_N & \\ & & & \end{array}\right] \begin{bmatrix} \mathbb{0} \\ \alpha \\ \mathbb{0} \end{bmatrix} \leftarrow k\text{'th component} \qquad \Rightarrow \qquad \boldsymbol{x}(0) = \alpha \begin{bmatrix} \\ \boldsymbol{v}_k \\ \\ \end{bmatrix},$$

where $\mathbb{0}$ is a vector of all zeros of the appropriate size. Similarly, choosing $\dot{\boldsymbol{y}}(0) = \beta \boldsymbol{e}_k$ implies that $\dot{\boldsymbol{x}}(0) = \beta \boldsymbol{v}_k$.

The implications of $\boldsymbol{y}(0) = \alpha \boldsymbol{e}_k$ and $\dot{\boldsymbol{y}}(0) = \beta \boldsymbol{e}_k$ are the following

$$\ddot{y}_i(t) = -\omega_i^2\ y(t), \qquad\qquad y_i(0) = 0,\ \dot{y}_i(0) = 0, \qquad i \neq k \tag{7.25}$$

$$\ddot{y}_k(t) = -\omega_k^2\ y(t), \qquad\qquad y_k(0) = \alpha,\ \dot{y}_k(0) = \beta. \tag{7.26}$$

This means that the solutions of all normal modes equations for $i \neq k$ will be identically zero for all time, i.e. $y_i(t) = 0$, $t \geq 0$ for $i \neq k$. Only the normal mode $y_k(t)$ will be non-zero. Recall from Theorem 3.1 that the solution of the 1-DOF system (7.26) is given by

$$y_k(t) \ = \ \mathsf{y}\ \cos\big(\omega_k t + \theta\big), \qquad\qquad \mathsf{y} = \sqrt{\alpha^2 + \tfrac{\beta^2}{\omega_k^2}}, \qquad \theta = -\tan^{-1}\tfrac{\beta}{\alpha\ \omega_k}.$$

In this case the solution vector $\boldsymbol{y}(t)$ would have all other components zero (7.25), i.e. it would be

$$\boldsymbol{y}(t) \;=\; \begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} \;=\; \begin{bmatrix} \mathbb{0} \\ y_k(t) \\ \mathbb{0} \end{bmatrix} \;=\; \begin{bmatrix} \mathbb{0} \\ 1 \\ \mathbb{0} \end{bmatrix} y_k(t) \;=\; \begin{bmatrix} \\ \boldsymbol{e}_k \\ \\ \end{bmatrix} \; \mathsf{y} \cos\big(\omega_k t + \theta\big).$$

This in turn means that in the original coordinates $\boldsymbol{x}(t)$ the solution is

$$\boldsymbol{x}(t) \;=\; V\boldsymbol{y}(t) \;=\; V\boldsymbol{e}_k \; y_k(t) \;=\; \boldsymbol{v}_k \; y_k(t),$$

i.e. it is an oscillatory solution which is always in the direction of the eigenvector $\boldsymbol{v}_k$ for all $t \geq 0$. This is summarized in the next statement.

Lemma 7.6. *Consider an N-degree of freedom, undamped, linear mechanical system of the form*

$$\mathsf{M}\,\ddot{\boldsymbol{x}}(t) \;+\; \mathsf{K}\,\boldsymbol{x}(t) \;=\; 0,$$

and let $\boldsymbol{v}_k$ be an eigenvector *of* $\mathsf{M}^{-1}\mathsf{K}$ *with* eigenvalue $\lambda_k > 0$. *If the initial conditions vectors are in the direction of $\boldsymbol{v}_k$, then the solution vector $\boldsymbol{x}(t)$ will remain in the direction of $\boldsymbol{v}_k$ for all t*

$$\begin{bmatrix} \\ \boldsymbol{x}(0) \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{v}_k \\ \\ \end{bmatrix} \alpha, \quad \textit{and} \quad \begin{bmatrix} \\ \dot{\boldsymbol{x}}(0) \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{v}_k \\ \\ \end{bmatrix} \beta \qquad \Rightarrow \qquad \begin{bmatrix} \\ \boldsymbol{x}(t) \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{v}_k \\ \\ \end{bmatrix} \mathsf{a} \;\cos\big(\omega_k t + \theta\big), \quad (7.27)$$

where $\omega_k := \sqrt{\lambda_k}$ is the k'th mode of vibration *(aka the k'th* natural frequency*), and the amplitude* a *and phase θ are determined by initial conditions as follows*

$$\mathsf{a} = \sqrt{\alpha^2 + \tfrac{\beta^2}{\omega_k^2}}, \qquad \theta = -\tan^{-1}\tfrac{\beta}{\alpha\;\omega_k}.$$

The vector $\boldsymbol{v}_k$ is referred to as that mode's shape*. For such initial conditions, the response (7.27) is called a* pure mode response*.*

In other words, choosing the initial conditions to be along an eigenvector $\boldsymbol{v}_k$ of $\big(\mathsf{M}^{-1}\mathsf{K}\big)$ will produce a vibration in only that direction in n-dimensional space with a single frequency $\omega_k = \sqrt{\lambda_k}$ associated with that eigenvector. Since the $N \times N$ matrix $\mathsf{M}^{-1}\mathsf{K}$ has n eigenvalues and eigenvectors, an N-DOF system always has N modes of vibration with possibly N different vibration natural frequencies. Note that in the special case $\dot{\boldsymbol{x}}(0) = 0$ (i.e. $\beta = 0$) when initial velocities are zero, the phase $\theta = 0$ and the amplitude $\mathsf{a} = \alpha$ is simply the initial condition amplitude.

Example 7.7. (System I). Consider first the System I described earlier

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 + k_c & -k_c \\ -k_c & k_2 + k_c \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = 0,$$

and assume for simplicity that $m_1 = m_2 = 1$ and that all spring constants are equal $k_1 = k_2 = k_c = k$. We then have

$$\big(M^{-1}K\big) \;=\; \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix} \;=\; k \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

The eigenvalues of this matrix can be easily found (manually or by symbolic calculations)[4] to be

$$\lambda_1 = k, \;\; \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \lambda_2 = 3k, \;\; \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \Rightarrow \qquad V = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \; \Lambda = \begin{bmatrix} k & 0 \\ 0 & 3k \end{bmatrix}. \quad (7.28)$$

[4] Note that in this case the matrix is of the form kM, where M is a numerical matrix whose eigenvalues can be calculated numerically. As can be easily verified, for a matrix of the form kM, the eigenvectors are the same as the eigenvectors of M, and the eigenvalues of kM are the eigenvalues of M multiplied by k.

Thus the normal mode frequencies are

$$\omega_1 = \sqrt{\lambda_1} = \sqrt{k}, \qquad \omega_2 = \sqrt{\lambda_2} = \sqrt{3k}. \tag{7.29}$$

The initial conditions that would produce those two pure modes are the eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, e.g.

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} \dot{x}_1(0) \\ \dot{x}_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \boldsymbol{x}(t) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cos\left(\sqrt{k}\ t\right),$$
$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ \text{-}1 \end{bmatrix}, \quad \begin{bmatrix} \dot{x}_1(0) \\ \dot{x}_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \boldsymbol{x}(t) = \begin{bmatrix} 1 \\ \text{-}1 \end{bmatrix} \cos\left(\sqrt{3k}\ t\right),$$

and similarly for the initial conditions on the derivatives (though with different amplitudes and phases). The first initial condition above corresponds to frequency $\sqrt{k}$ and a motion where $x_1(t)$, $x_2(t)$ move "together" left and right, i.e. "unison" motion. The second corresponds to a higher frequency $\sqrt{3k}$, but the masses move in opposite directions, or in "differential" motion. These are illustrated in Figure 7.5.

Example 7.8. (System II). This system has the equations

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 & \text{-}k_1 \\ \text{-}k_1 & k_1 + k_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = 0.$$

Assume for simplicity that $m_1 = m_2 = 1$ and $k_1 = k_2 = k$. We then have

$$\left(M^{\text{-}1}K\right) \ = \ \begin{bmatrix} k & \text{-}k \\ \text{-}k & 2k \end{bmatrix} \ = \ k \begin{bmatrix} 1 & \text{-}1 \\ \text{-}1 & 2 \end{bmatrix}.$$

The eigenvalues of this matrix can be easily found (manually or by symbolic calculations) to be

$$\lambda_1 = \tfrac{3-\sqrt{5}}{2}\, k, \quad \boldsymbol{v}_1 = \begin{bmatrix} 1+\sqrt{5} \\ 2 \end{bmatrix} \approx \begin{bmatrix} 3.24 \\ 2 \end{bmatrix}, \qquad \lambda_2 = \tfrac{3+\sqrt{5}}{2}\, k, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1-\sqrt{5} \\ 2 \end{bmatrix} \approx \begin{bmatrix} \text{-}1.24 \\ 2 \end{bmatrix}. \tag{7.30}$$

We again see two normal modes with a lower and higher frequency modes at

$$\omega_1 = \sqrt{\tfrac{3-\sqrt{5}}{2}\, k} \ \approx \ 0.62\ \sqrt{k}, \qquad \omega_2 = \sqrt{\tfrac{3+\sqrt{5}}{2}\, k} \ \approx \ 1.62\ \sqrt{k}. \tag{7.31}$$

We can again use $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ as the initial conditions that would produce those pure modes, but a physical interpretation of these initial conditions is harder to imagine compared to System (I). None the less, the eigenvector and eigenvalue calculations always produce the normal modes regardless of physical insight.

Pure mode motions for this system are illustrated in Figure 7.6. Note that they again represent differential and unison motion respectively, though unlike System (I), each displacement has a different amplitude. This is due to the asymmetry between the two masses m_1 and m_2, unlike the completely symmetric case of System (I). Note in particular the solution forms

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 3.24 \\ 2 \end{bmatrix} \qquad \Rightarrow \qquad \boldsymbol{x}(t) = \begin{bmatrix} 3.24 \\ 2 \end{bmatrix} \cos\left(0.62\sqrt{k}\ t\right),$$
$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} \text{-}1.24 \\ 2 \end{bmatrix} \qquad \Rightarrow \qquad \boldsymbol{x}(t) = \begin{bmatrix} \text{-}1.24 \\ 2 \end{bmatrix} \cos\left(1.62\sqrt{k}\ t\right),$$

which can be clearly seen in Figure 7.5.

Finally note that System (III) with no external forcing (i.e. $u(t) = 0$) and no damping is exactly the same as System (II), and therefore has the same normal modes. The differences between Systems (II) and (III) appear when we consider the effect of external inputs on the response, such as in the expressions for the frequency response. This is the subject of the next Chapter.

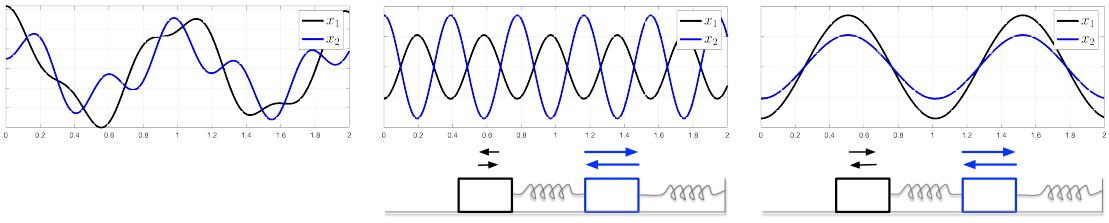


Figure 7.6: Trajectories of masses' positions for System (II) from three different sets of initial conditions. $m_1 = m_2 = 1$ and $k_1 = k_2 = 100$ in these simulations. (*Left*) Trajectories from a "generic" initial conditions appear to contain two different frequencies in each mass' trajectory. (*Middle*) Trajectories from the special initial conditions $x(0) = -\boldsymbol{v}_2$ obtained from the 2nd eigenvector $\boldsymbol{v}_2$ appears to correspond to differential motion. The trajectories appear to be a "pure normal mode" with frequency $\omega = \frac{2\pi}{T} \approx \frac{2\pi}{0.39} \approx 16$ rad/s, which matches the theoretically predicted value $\omega_2 \approx 1.6\sqrt{k} \approx 16$ rad/s. (*Right*) Trajectories from the initial conditions $x(0) = \boldsymbol{v}_1$ obtained from the other eigenvector. This again appears to be a pure normal mode with frequency $\omega = \frac{2\pi}{T} \approx \frac{2\pi}{1} \approx 6.3$ rad/s, which approximately matches the theoretically predicted value for this mode of $\omega_1 = 0.62\sqrt{k} = 6.2$. The two masses move together in sync, but with different amplitudes, which makes it easy to distinguish them on this plot.

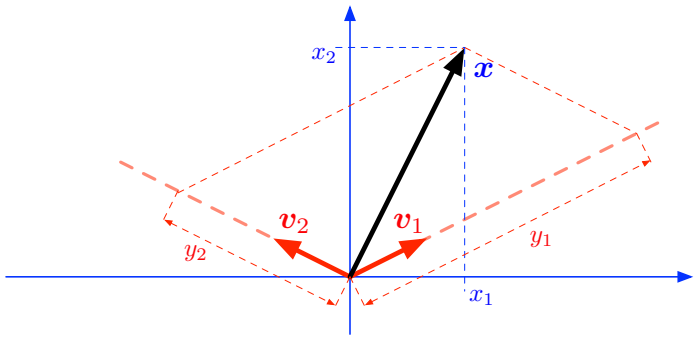


Figure 7.7: Expressing a vector $\boldsymbol{x}$ in terms of a different basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.

Appendix

7.A Change of Bases and Diagonalization

Consider an $n \times n$ matrix V and two n-vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ related by the matrix-vector product $\boldsymbol{x} = V\boldsymbol{y}$. You have already seen the basic definition of a matrix-vector product as

$$\boldsymbol{x} = V\boldsymbol{y} \qquad \Leftrightarrow \qquad x_i = \sum_{j=1}^{n} v_{ij}\ y_j, \qquad i = 1, \dots, n, \tag{7.32}$$

where v_{ij} is the ij'th entry of the matrix V, and x_i and y_j are the i'th and j'th components of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. This definition is very algebraic. However, there is some beautiful geometry underlying this definition which we will now discuss.

Change of Bases

The product (7.32) can have several geometric interpretations. The first one we consider is as a *basis expansion*. First, recall that the components $x_1, \dots, x_n$ of a vector $\boldsymbol{x}$ are geometrically interpreted as the Cartesian coordinates of the vector in an orthonormal coordinate system as shown in Figure 7.7 for the special case of 2-vectors (i.e. vectors in the plane). Any set of linearly independent vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ is a *basis* in n-dimensional space, and we can always express any given vector $\boldsymbol{x}$ in that basis by writing is as a linear combination of the basis vectors

$$\boldsymbol{x} \ = \ y_1\boldsymbol{v}_1 + \cdots + y_n\boldsymbol{v}_n, \tag{7.33}$$

where the scalars $y_1, \dots, y_n$ are the new coordinates of $\boldsymbol{x}$ in this new basis. See Figure 7.7 for a geometrical illustration. Now we ask the following question: given a vector $\boldsymbol{x}$ and a *new basis* $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$, how do we compute the coefficients $y_1, \dots, y_n$ of the expansion (7.33) in the new basis?

The key idea is to form a matrix V whose columns are the basis vectors, and observe that a matrix-vector product can be interpreted differently from (7.32) as follows

$$\begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ V \\ \\ \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{y} \\ \\ \end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \overset{1}{=} \underbrace{\begin{bmatrix} \\ \boldsymbol{v}_1 \\ \\ \end{bmatrix} y_1 + \ \cdots \ + \begin{bmatrix} \\ \boldsymbol{v}_n \\ \\ \end{bmatrix} y_n}_{\substack{\text{linear combination of the} \\ \text{vectors } \boldsymbol{v}_1, \dots, \boldsymbol{v}_n}}, \tag{7.34}$$

The equality $\overset{1}{=}$ follows from the matrix-vector product definition (7.32). For example, the first component y_1 multiplies only the first column of V, and so on. The last expression in (7.34) expresses $\boldsymbol{x}$ as a *linear combination* of the vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$, with the scalars $y_1, \dots, y_n$ as coefficients in this

linear combination, i.e. a *basis expansion* of $\boldsymbol{x}$ in the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$. Although this might constitute a "mental shift" from the way you normally think about a matrix-vector product, it is a very useful interpretation as you will shortly see. The coefficients $y_1, \ldots, y_n$ can be calculated all at once by forming the inverse V^{-1} to obtain

$$\boldsymbol{x} = V\boldsymbol{y} \qquad \Rightarrow \qquad \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \\ V^{-1} \\ \\ \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix}. \tag{7.35}$$

You may have seen other algorithms for computing the coefficients of a basis expansion. They are all equivalent to the above, which is very compactly expressed using a matrix inverse, and a matrix-vector product.

Effect of Change of Bases on Matrices

Given an $n \times n$ matrix A, it can be considered as a *mapping* between vectors as

$$\boldsymbol{x'} \;=\; A\,\boldsymbol{x} \qquad \Leftrightarrow \qquad \boldsymbol{x} \;\mapsto\; \boldsymbol{x'} = A\boldsymbol{x}.$$

That is, A maps the vector $\boldsymbol{x}$ to the vector $\boldsymbol{x'} = A\boldsymbol{x}$ by matrix-vector multiplication. We now ask another question: Given a new basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, *how does A map the basis coefficients of $\boldsymbol{x}$ to those of $\boldsymbol{x'}$?*

This question is easy to answer. Using the previous material, we can find the basis coefficients of $\boldsymbol{x}$ and $\boldsymbol{x'}$ using the matrix V (made from the basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ as its columns)

$$\boldsymbol{y} \;:=\; V^{-1}\,\boldsymbol{x}, \qquad\qquad \boldsymbol{y'} \;:=\; V^{-1}\,\boldsymbol{x'},$$

where the expansion coefficients of $\boldsymbol{x}$ and $\boldsymbol{x'}$ are the components of the vectors $\boldsymbol{y}$ and $\boldsymbol{y'}$ respectively. Next we find the mapping from $\boldsymbol{y}$ to $\boldsymbol{y'}$ by

$$\begin{aligned} \boldsymbol{y'} \;&=\; V^{-1}\,\boldsymbol{x'} && (\boldsymbol{y'} \text{ contains the coefficients in the new basis. Apply (7.35)}) \\ &=\; V^{-1}\,A\boldsymbol{x} && (\text{since } \boldsymbol{x'} = A\boldsymbol{x}) \\ &=\; V^{-1}A\,V\boldsymbol{y} && (\text{since } \boldsymbol{x} = V\boldsymbol{y}) \\ \Rightarrow \qquad \boldsymbol{y'} \;&=\; \left(V^{-1}AV\right)\,\boldsymbol{y}. \end{aligned}$$

Note that $\left(V^{-1}AV\right)$ is another $n \times n$ matrix. It is the representation of the original mapping A in the new basis. We can now think of a *transformation of matrices* $A \mapsto \left(V^{-1}AV\right)$, which is called a *similarity transformation*. Let's summarize what we have just done precisely as a theorem.

Theorem 7.9. *Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a set of n linearly independent n-vectors, i.e. a basis of the space of n-vectors. Form the non-singular $n \times n$ matrix V whose columns are the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Then*

1. *For any vector $\boldsymbol{x}$, the coefficients $y_1, \ldots, y_n$ of its expansion in the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ are given by*

$$\boldsymbol{y} \;=\; \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \\ V^{-1} \\ \\ \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} \qquad \Leftrightarrow \qquad \boldsymbol{x} = \begin{bmatrix} \\ \boldsymbol{v}_1 \\ \\ \end{bmatrix} y_1 + \cdots + \begin{bmatrix} \\ \boldsymbol{v}_n \\ \\ \end{bmatrix} y_n.$$

2. *For any (possibly singular) $n \times n$ matrix A, the transformation $\boldsymbol{x} \mapsto \boldsymbol{x'}$ with $\boldsymbol{x'} = A\boldsymbol{x}$ can be expressed in terms of their basis expansion coefficients $\boldsymbol{y} := V^{-1}\boldsymbol{x}$ and $\boldsymbol{y'} := V^{-1}\boldsymbol{x'}$ as*

$$\boldsymbol{y'} = \left(V^{-1}AV\right)\boldsymbol{y} \qquad \Leftrightarrow \qquad \boldsymbol{x'} = A\,\boldsymbol{x}$$

One way to phrase the above theorem is that the *basis transformation* on vectors $\boldsymbol{x} \mapsto \boldsymbol{y} = V^{-1}\boldsymbol{x}$ induces a *similarity transformation* $A \mapsto \big(V^{-1}AV\big)$ on matrices. Another way to say it is that $\big(V^{-1}AV\big)$ is the new representation of the matrix A in the new basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$. One major motivation for bases changes is that a matrix A with complicated structure may be converted to a matrix $A' = \big(V^{-1}AV\big)$ with much simpler structure. The process of *diagonalizing* a matrix through a similarity transformation (i.e. a change of basis) illustrated next is a prime example of this simplification.

Diagonalization of a Matrix using its Eigenvectors and Eigenvalues

Consider an $n \times n$ matrix A that has n linearly independent eigenvectors[5] $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. This means that we have n eigenvalue/eigenvector relations of the form

$$A\boldsymbol{v}_i \;=\; \lambda_i \boldsymbol{v}_i, \qquad\qquad i = 1, \ldots, n.$$

It is an elementary, but powerful observation that these n matrix-vector relations can be rewritten as the following single matrix equation

$$\begin{bmatrix} A\boldsymbol{v}_1 & \cdots & A\boldsymbol{v}_n \end{bmatrix} \;=\; \begin{bmatrix} \lambda_1\boldsymbol{v}_1 & \cdots & \lambda_n\boldsymbol{v}_n \end{bmatrix} \tag{7.36}$$

$$\Updownarrow$$

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} \;=\; \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \tag{7.37}$$

$$\Updownarrow$$

$$AV \;=\; V\Lambda, \tag{7.38}$$

where Λ is the diagonal matrix made up of the eigenvalues of A arranged in the same order as their corresponding eigenvectors in the columns of V. The fact that (7.37) is equivalent to (7.36) follows from the definition of matrix-matrix products. The notation shown above is called *partitioned matrix* notation, and is a very compact way of expressing the above relations without exhibiting the details of matrix-matrix products.

Finally, observe that by multiplying by V^{-1} from the left or the right, equation (7.38) implies that

$$A' \;:=\; V^{-1}AV \;=\; \Lambda \qquad \Leftrightarrow \qquad A \;=\; V\Lambda V^{-1} \tag{7.39}$$

Thus $A \mapsto V^{-1}AV = \Lambda$, where V is made up of the eigenvectors of A, is precisely the similarity transformation that diagonalizes A. The resulting diagonal matrix Λ has all the eigenvalues of A along the diagonal (and zero entries otherwise).

Diagonal matrices are the simplest matrices to work with. For systems of equations (either algebraic or differential equations), if the underlying matrix is diagonal, then the system of equations is completely *uncoupled*, and they can therefore be solved as scalar equations independently of each other.

[5] Not all $n \times n$ matrices have n linearly independent eigenvectors. In general an $n \times n$ matrix may have anywhere from 1 to n such eigenvectors. This is not an issue for us since the matrices we consider in this handout all have as many linearly independent eigenvectors as their dimension.

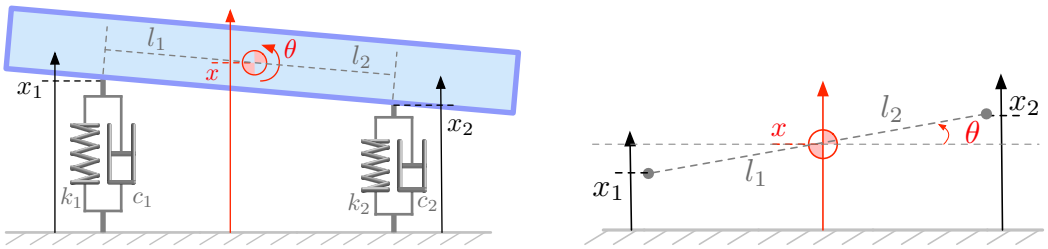


Figure 7.8: The sprung beam with damping and possibly unequal lengths l_1, l_2.

7.B Eigenvalues/vectors of Mass and Stiffness Matrices

K is the sum of a positive-semi-definite matrix and possibly a diagonal with positive entries, so always has non-negative real eigenvalues.

$$\text{PE} \;=\; x^*\mathsf{K}x \;=\; \sum_{i=1}^{N} \tfrac{1}{2} k_i\, x_j^2 \;+\; \sum_{(i,j)\in\mathcal{E}} \tfrac{1}{4} k_{ij}\,(x_i - x_j)^2$$

TBC

7.C More Models

7.C.1 The General Sprung Beam Model

We derive the differential equations for a more general sprung beam than the one considered in the introduction. The model is shown in Figure 7.8 with damping elements and possibly unequal lengths l_1 and l_2. The equations in this case are

$$\begin{aligned} m\,\ddot{x} \;&=\; -k_1x_1 - k_2x_2 - c_1\dot{x}_1 - c_2\dot{x}_2, \\ \mathrm{J}\,\ddot{\theta} \;&=\; l_1\,(k_1x_1 + c_1\dot{x}_1) \;-\; l_2\,(k_2x_2 + c_2\dot{x}_2)\,. \end{aligned} \tag{7.40}$$

The kinematics shown in Figure 7.8 imply

$$\left.\begin{aligned} x_1 \;&\approx\; x - l_1\sin\theta \approx x - l_1\theta \\ x_2 \;&\approx\; x + l_2\sin\theta \approx x + l_2\theta \end{aligned}\right\} \quad\Rightarrow\quad \left\{\begin{aligned} m\,\ddot{x} \;&=\; -k_1x_1 - k_2x_2 - c_1\dot{x}_1 - c_2\dot{x}_2 \\ &=\; -(k_1+k_2)\,x + (k_1l_1 - k_2l_2)\,\theta \\ &\quad\; -(c_1+c_2)\,\dot{x} + (c_1l_1 - c_2l_2)\,\dot{\theta} \\ \mathrm{J}\,\ddot{\theta} \;&=\; l_1\,(k_1x_1 + c_1\dot{x}_1) \;-\; l_2\,(k_2x_2 + c_2\dot{x}_2) \\ &=\; (k_1l_1 - k_2l_2)\,x - (k_1l_1^2 + k_2l_2^2)\,\theta \\ &\quad\; (c_1l_1 - c_2l_2)\,\dot{x} - (c_1l_1^2 + c_2l_2^2)\,\dot{\theta} \end{aligned}\right. \tag{7.41}$$

These equations are rewritten as vector differential equations as follows

$$\begin{bmatrix} m\,\ddot{x} + (c_1+c_2)\,\dot{x} - (c_1l_1 - c_2l_2)\,\dot{\theta} + (k_1+k_2)\,x - (k_1l_1 - k_2l_2)\,\theta \\ \mathrm{J}\,\ddot{\theta} - (c_1l_1 - c_2l_2)\,\dot{x} + (c_1l_1^2 + c_2l_2^2)\,\dot{\theta} - (k_1l_1 - k_2l_2)\,x + (k_1l_1^2 + k_2l_2^2)\,\theta \end{bmatrix} \;=\; \tag{7.42}$$

$$\Leftrightarrow \quad \begin{bmatrix} m & 0 \\ 0 & \mathrm{J} \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} + \begin{bmatrix} c_1+c_2 & c_2l_2 - c_1l_1 \\ c_2l_2 - c_1l_1 & c_1l_1^2 + c_2l_2^2 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} k_1+k_2 & k_2l_2 - k_1l_1 \\ k_2l_2 - k_1l_1 & k_1l_1^2 + k_2l_2^2 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} \;=\; \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{7.43}$$

The alternate model using the coordinates (x_1, x_2) is obtained from the transformation

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;=\; \begin{bmatrix} 1 & -l_1 \\ 1 & l_2 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} \qquad\Rightarrow\qquad \begin{bmatrix} x \\ \theta \end{bmatrix} \;=\; \frac{1}{l_1+l_2} \begin{bmatrix} l_2 & l_1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;=\; \begin{bmatrix} (l_2x_1 + l_1x_2)/(l_1+l_2) \\ (x_2 - x_1)/(l_1+l_2) \end{bmatrix}$$

Substituting these expressions for (x, θ) in (7.41) gives

$$\begin{aligned} m\,(l_2\ddot{x}_1 + l_1\ddot{x}_2)/(l_1+l_2) &= -k_1x_1 - k_2x_2 - c_1\dot{x}_1 - c_2\dot{x}_2, \\ \mathrm{J}\,(-\ddot{x}_1 + \ddot{x}_2)/(l_1+l_2) &= l_1\,(k_1x_1 + c_1\dot{x}_1) \;-\; l_2\,(k_2x_2 + c_2\dot{x}_2)\,, \end{aligned}$$

which can be reorganized into the following single vector differential equation.

$$\frac{1}{l_1+l_2}\begin{bmatrix} ml_2 & ml_1 \\ -\mathrm{J} & \mathrm{J} \end{bmatrix}\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} + \begin{bmatrix} c_1 & c_2 \\ c_1l_1 & -c_2l_2 \end{bmatrix}\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} + \begin{bmatrix} k_1 & k_2 \\ k_1l_1 & -k_2l_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0. \tag{7.44}$$

As can be seen from the above, there is no easily-recognizable set of coordinates that renders the equations in a decoupled form unless $l_1 = l_2$, $c_1 = c_2$ and $k_1 = k_2$. Such assumptions may not be realistic in all problems.

Chapter 8

Forced n-DOF Systems: The Frequency Response Matrix

Forced n-DOF systems can be analyzed in the frequency domain using techniques similar to those for 1-DOF systems. The main tool is again phasor representations and the frequency response. For n-DOF systems, the mechanical coordinates form vectors, and multiple forcing inputs are also described mathematically as a vector of inputs. Phasors representing outputs and inputs become vectors of phasors, and the input-output relation between the phasor vectors becomes the frequency response matrix. *Starting from the matrix coefficients of vector differential equations, formulas for the frequency response matrix are given in terms of those matrix coefficients. It thus becomes important to keep track of matrix algebra, and in particular the non-commutativity of matrix products in the analysis. In comparison to 1-DOF systems, n-DOF systems exhibit new vibrations phenomena such as multiple resonances and zeros of certain frequency responses. The latter represents phenomena of vibration absorption that are used in devices such as tuned mass dampers.*

8.1 The Frequency Response Matrix

Phasor and frequency response analysis can be applied to the vector equation

$$\mathsf{M}\,\ddot{\boldsymbol{x}}(t) \;+\; \mathsf{C}\,\dot{\boldsymbol{x}}(t) \;+\; \mathsf{K}\,\boldsymbol{x}(t) \;=\; \mathsf{B}_2\,\ddot{\boldsymbol{u}}(t) \;+\; \mathsf{B}_1\,\dot{\boldsymbol{u}}(t) \;+\; \mathsf{B}_0\,\boldsymbol{u}(t) \tag{8.1}$$

in much the same manner as for scalar equations. Care must be taken with matrix algebra though. Phasor analysis of (8.1) enables us to determine the *frequency response* from any component of the input *vector* $\boldsymbol{u}(t)$ to any component of the response *vector* $\boldsymbol{x}(t)$. We will be extra careful with notation in this section to avoid confusion. Phasors and complex numbers in general will be denoted with a "hat" in sans serif font, e.g. $\hat{\mathsf{x}}_i$ will be a scalar phasor which is the i'th component of a vector of phasors $\hat{\mathbf{x}}$.

To see how phasor analysis applies to (8.1), let $\boldsymbol{u}(t)$ be a vector where each component is a sinusoid of some amplitude and phase, but all with the *same frequency*. We express each component in terms of its corresponding phasor to obtain a *vector of phasors* as follows

$$\begin{aligned}\boldsymbol{u}(t) \;&=\; \begin{bmatrix}\mathsf{u}_1\cos\big(\omega t+\phi_1\big)\\ \vdots \\ \mathsf{u}_n\cos\big(\omega t+\phi_n\big)\end{bmatrix} \;=\; \begin{bmatrix}\mathbb{R}\big(\mathsf{u}_1\,e^{j(\omega t+\phi_1)}\big)\\ \vdots \\ \mathbb{R}\big(\mathsf{u}_n\,e^{j(\omega t+\phi_n)}\big)\end{bmatrix} \;=\; \mathbb{R}\left(\begin{bmatrix}\mathsf{u}_1\,e^{j(\omega t+\phi_1)}\\ \vdots \\ \mathsf{u}_n\,e^{j(\omega t+\phi_n)}\end{bmatrix}\right)\\ &=\; \mathbb{R}\left(\begin{bmatrix}\mathsf{u}_1\,e^{j\phi_1}\\ \vdots \\ \mathsf{u}_n\,e^{j\phi_n}\end{bmatrix}e^{j\omega t}\right) \;=:\; \mathbb{R}\left(\begin{bmatrix}\hat{\mathsf{u}}_1\\ \vdots \\ \hat{\mathsf{u}}_n\end{bmatrix}e^{j\omega t}\right) \;=:\; \mathbb{R}\big(\hat{\mathbf{u}}\,e^{j\omega t}\big)\,,\end{aligned} \tag{8.2}$$

where $\hat{\mathbf{u}}$ is the *phasor vector* defined above. Note that the amplitudes and phases u_i, ϕ_i of each component are encoded in the magnitudes and phases of the complex numbers $\hat{\mathsf{u}}_i$, i.e. $|\hat{\mathsf{u}}_i| = \mathsf{u}$ and $\angle\hat{\mathsf{u}}_i = \phi_i$, all of which make up the phasor vector $\hat{\mathbf{u}}$.

Since the vector of functions $\boldsymbol{u}(t)$ is the real part of the vector of complex functions $\hat{\mathbf{u}}\, e^{j\omega t}$, it can also be written as the sum of a complex function and its conjugate as

$$\boldsymbol{u}(t) = \begin{bmatrix} \mathsf{u}_1 \cos\left(\omega t + \phi_1\right) \\ \vdots \\ \mathsf{u}_n \cos\left(\omega t + \phi_n\right) \end{bmatrix} = \mathbb{R}\big(\hat{\mathbf{u}}\, e^{j\omega t}\big) \qquad \Leftrightarrow \qquad \boldsymbol{u}(t) \;=\; \tfrac{1}{2}\left(\hat{\mathbf{u}}\, e^{j\omega t} + \hat{\mathbf{u}}^*\, e^{j\omega t}\right),$$

where $\hat{\mathbf{u}}^*$ is the vector whose components are the complex conjugates of the components of $\hat{\mathbf{u}}$. The superposition property of linear equations implies that the steady state response to $\boldsymbol{u}(t)$ is (half) the sum of the responses to $\hat{\mathbf{u}}\, e^{j\omega t}$ and $\hat{\mathbf{u}}^*\, e^{j\omega t}$ respectively.

We will show next that the steady-state response to an input of the form $\hat{\mathbf{u}}\, e^{j\omega t}$ is of the form $\hat{\mathbf{x}}\, e^{j\omega t}$, and obtain the *frequency response matrix* which gives the relation between the vector phasors $\hat{\mathbf{u}}$ and $\hat{\mathbf{x}}$. Substituting the forms $\boldsymbol{u}(t) = \hat{\mathbf{u}}\, e^{j\omega t}$ and $\boldsymbol{x}(t) = \hat{\mathbf{x}}\, e^{j\omega t}$ in the differential equation (8.1) (for notational simplicity in the following we assume $\mathsf{B}_1 = 0$, and $\mathsf{B}_2 = 0$), and recalling that the derivative of a vector function of t is simply the vector of derivatives of each component

$$\begin{bmatrix} \quad \mathsf{M} \quad \end{bmatrix} \frac{d^2}{dt^2} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} e^{j\omega t} + \begin{bmatrix} \quad \mathsf{C} \quad \end{bmatrix} \frac{d}{dt} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} e^{j\omega t} + \begin{bmatrix} \quad \mathsf{K} \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} e^{j\omega t} = \begin{bmatrix} \quad \mathsf{B}_0 \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{u}}_1 \\ \vdots \\ \hat{\mathsf{u}}_n \end{bmatrix} e^{j\omega t}$$

$$\Rightarrow \begin{bmatrix} \quad \mathsf{M} \quad \end{bmatrix} (j\omega)^2 \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} e^{j\omega t} + \begin{bmatrix} \quad \mathsf{C} \quad \end{bmatrix} (j\omega) \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} e^{j\omega t} + \begin{bmatrix} \quad \mathsf{K} \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} e^{j\omega t} = \begin{bmatrix} \quad \mathsf{B}_0 \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{u}}_1 \\ \vdots \\ \hat{\mathsf{u}}_n \end{bmatrix} e^{j\omega t}$$

Now since $e^{j\omega t}$ is a common factor to all terms on both sides of the equation, we divide by it to obtain the purely algebraic relation

$$(j\omega)^2 \begin{bmatrix} \quad \mathsf{M} \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} + (j\omega) \begin{bmatrix} \quad \mathsf{C} \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} + \begin{bmatrix} \quad \mathsf{K} \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} = \begin{bmatrix} \quad \mathsf{B}_0 \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{u}}_1 \\ \vdots \\ \hat{\mathsf{u}}_n \end{bmatrix}$$

$$\Rightarrow \qquad \left((j\omega)^2 \begin{bmatrix} \quad \mathsf{M} \quad \end{bmatrix} + (j\omega) \begin{bmatrix} \quad \mathsf{C} \quad \end{bmatrix} + \begin{bmatrix} \quad \mathsf{K} \quad \end{bmatrix} \right) \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \vdots \\ \hat{\mathsf{x}}_n \end{bmatrix} = \begin{bmatrix} \quad \mathsf{B}_0 \quad \end{bmatrix} \begin{bmatrix} \hat{\mathsf{u}}_1 \\ \vdots \\ \hat{\mathsf{u}}_n \end{bmatrix}. \tag{8.3}$$

Now that the relation is clear, we rewrite (8.3) in compact matrix notation, which then shows clearly how to solve for $\hat{\mathbf{x}}$ from $\hat{\mathbf{u}}$ using a matrix inversion

$$\left(-\omega^2\, \mathsf{M} + j\omega\, \mathsf{C} + \mathsf{K}\right) \hat{\mathbf{x}} = \mathsf{B}_0\, \hat{\mathbf{u}} \qquad \Rightarrow \qquad \hat{\mathbf{x}} = \left(-\omega^2\, \mathsf{M} + j\omega\, \mathsf{C} + \mathsf{K}\right)^{-1} \mathsf{B}_0\, \hat{\mathbf{u}} \;=:\; \mathbf{H}(\omega)\, \hat{\mathbf{u}}. \tag{8.4}$$

The frequency response $\mathbf{H}(\omega)$ defined by this formula is now a complex *matrix* for each frequency ω. The reader should compare this with the complex frequency response formula we obtained earlier for a single degree of freedom system (where $\hat{\mathbf{x}}$ and $\hat{\mathbf{u}}$ were complex *scalars*, and we had $\mathsf{B}_0 = 1$)

$$\mathsf{H}(\omega) \;=\; \frac{\hat{\mathsf{x}}}{\hat{\mathsf{u}}} \;=\; \frac{1}{-m\,\omega^2 \;+\; jc\,\omega \;+\; k}.$$

Since we are now working with matrices, it does not make sense to do something like $1/\mathsf{M}$ when M is a matrix. The corresponding operation is to take a matrix inverse as in (8.4). Finally, a repetition of the arguments above show that the frequency response of the most general system (8.1) with possibly non-zero B_1 and B_2 gives the general frequency response formula. We state this as a theorem.

Theorem 8.1. *Consider the vector input-output differential equation*

$$\mathsf{M}\, \ddot{\boldsymbol{x}}(t) \;+\; \mathsf{C}\, \dot{\boldsymbol{x}}(t) \;+\; \mathsf{K}\, \boldsymbol{x}(t) \;=\; \mathsf{B}_2\, \ddot{\boldsymbol{u}}(t) \;+\; \mathsf{B}_1\, \dot{\boldsymbol{u}}(t) \;+\; \mathsf{B}_0\, \boldsymbol{u}(t),$$

and the associated frequency response matrix *function of* ω

$$\boxed{\mathbf{H}(\omega) \;=\; \left(-\omega^2\, \mathsf{M} + j\omega\, \mathsf{C} + \mathsf{K}\right)^{-1} \left(-\omega^2\, \mathsf{B}_2 + j\omega\, \mathsf{B}_1 + \mathsf{B}_0\right).} \tag{8.5}$$

1. *If the input is a complex harmonic vector of the form* $\boldsymbol{u}(t) = \hat{\mathbf{u}}\, e^{j\omega t}$, *the steady-state response is also a complex harmonic vector* $\boldsymbol{x}(t) = \hat{\mathbf{x}}\, e^{j\omega t}$ *with the same frequency* ω. *The complex phasor vectors* $\hat{\mathbf{u}}$ *and* $\hat{\mathbf{x}}$ *are related by the frequency response matrix as*

$$\hat{\mathbf{x}} \;=\; \mathbf{H}(\omega)\; \hat{\mathbf{u}}. \tag{8.6}$$

2. *If the input components are sinusoids of one frequency* ω, *then so are the output components, and the two are also related using the frequency response matrix by*

$$\boldsymbol{u}(t) = \begin{bmatrix} \mathsf{u}_1 \cos(\omega t + \phi_1) \\ \vdots \\ \mathsf{u}_n \cos(\omega t + \phi_n) \end{bmatrix} \;\Rightarrow\; \boldsymbol{x}(t) = \begin{bmatrix} \mathsf{x}_1 \cos(\omega t + \theta_1) \\ \vdots \\ \mathsf{x}_n \cos(\omega t + \theta_n) \end{bmatrix}, \quad \begin{bmatrix} \mathsf{x}_1 e^{j\theta_1} \\ \vdots \\ \mathsf{x}_n e^{j\theta_n} \end{bmatrix} = \begin{bmatrix} \\ \mathbf{H}(\omega) \\ \\ \end{bmatrix} \begin{bmatrix} \mathsf{u}_1 e^{j\phi_1} \\ \vdots \\ \mathsf{x}_n e^{j\phi_n} \end{bmatrix}. \tag{8.7}$$

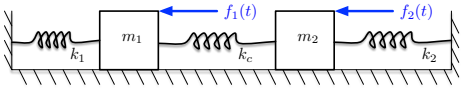
Note how the formula (8.4) derived earlier is a special case of the more general formula (8.5). Note also how the relation (8.7) between the amplitudes and phases of the input and output is the same as that between the phasors (8.6). The fact that *2* follows from *1* is a consequence of the superposition property and the representation of sinusoids as the real part of complex exponentials

$$\begin{aligned} \boldsymbol{u}(t) = \tfrac{1}{2}\left(\hat{\mathbf{u}}\, e^{j\omega t} + \hat{\mathbf{u}}^*\, e^{-j\omega t}\right) \qquad \Rightarrow \qquad \boldsymbol{x}(t) &= \tfrac{1}{2}\left(\mathbf{H}(\omega)\,\hat{\mathbf{u}}\, e^{j\omega t} + \mathbf{H}(-\omega)\,\hat{\mathbf{u}}^*\, e^{-j\omega t}\right) \\ &= \tfrac{1}{2}\left(\mathbf{H}(\omega)\,\hat{\mathbf{u}}\, e^{j\omega t} + \mathbf{H}^*(\omega)\,\hat{\mathbf{u}}^*\, e^{-j\omega t}\right) \\ &= \tfrac{1}{2}\left(\hat{\mathbf{x}}\, e^{j\omega t} + \hat{\mathbf{x}}^*\, e^{-j\omega t}\right), \qquad \hat{\mathbf{x}} \;=\; \mathbf{H}(\omega)\; \hat{\mathbf{u}}, \end{aligned}$$

where the second equality follows from the fact that $\mathbf{H}(-\omega) = \mathbf{H}^*(\omega)$ as can be verified directly from (8.5). Thus the amplitudes and phases x_i, θ_i of the response's components are obtained from the amplitudes and phases u_i, ϕ_i of the input using the elegant complex matrix-vector relation $\hat{\mathbf{x}} = \mathbf{H}(\omega)\hat{\mathbf{u}}$.

We now apply the frequency response matrix formula (8.5) to the three examples discussed earlier to get some insight into the meaning of the frequency response matrix $\mathbf{H}(\omega)$.

Example 8.2. (Frequency Response of System I)



Recall the diagram of System (I) reproduced here. We make the simplifying assumptions that $m_1 = m_2 = 1$ and $k_1 = k_2 = k_c = k$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}.$$

Applying formula (8.5) to find the frequency response ($\mathsf{B}_0 = \mathrm{I}$ in this case, so we don't write it out explicitly)

$$\begin{aligned} \mathbf{H}(\omega) \;=\; \left(-\omega^2\,\mathsf{M} + j\omega\,\mathsf{C} + \mathsf{K}\right)^{-1} &= \left(-\omega^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix}\right)^{-1} = \begin{bmatrix} 2k-\omega^2 & -k \\ -k & 2k-\omega^2 \end{bmatrix}^{-1} \\ &= \frac{1}{\omega^4 - 4k\omega^2 + 3k^2} \begin{bmatrix} 2k-\omega^2 & k \\ k & 2k-\omega^2 \end{bmatrix}. \end{aligned}$$

Examine the meaning of this by writing it out as the relation between the components of the forcing and the components of the response

$$\hat{\mathbf{x}} \;=\; \mathbf{H}(\omega)\; \hat{\mathbf{f}} \qquad \Leftrightarrow \qquad \begin{bmatrix} \hat{\mathsf{x}}_1 \\ \hat{\mathsf{x}}_2 \end{bmatrix} = \begin{bmatrix} \frac{2k-\omega^2}{\omega^4-4k\omega^2+3k^2} & \frac{k}{\omega^4-4k\omega^2+3k^2} \\ \frac{k}{\omega^4-4k\omega^2+3k^2} & \frac{2k-\omega^2}{\omega^4-4k\omega^2+3k^2} \end{bmatrix} \begin{bmatrix} \hat{\mathsf{f}}_1 \\ \hat{\mathsf{f}}_2 \end{bmatrix}. \tag{8.8}$$

Note that all four entries of this matrix have the same denominator. The roots of the denominator are the frequencies at which the frequency responses become infinite, these are the *resonance frequencies*.

It's instructive to calculate them. The denominator is a 4'th order polynomial, and therefore has 4 roots. Because this polynomial has only even-order terms, finding the roots reduces to finding the roots of a 2'nd order polynomial as follows

$$\omega^4 - 4k\ \omega^2 + 3k^2 = 0 \qquad \Leftrightarrow \qquad s^2 - 4k\ s + 3k^2 = 0, \qquad s = w^2.$$

The roots of the 2'nd order polynomial are obtained from the quadratic formula

$$s_{1,2} \ = \ 2k \pm \sqrt{4k^2 - 3k^2} \ = \ 2k \pm k \qquad \Rightarrow \qquad \omega_{1,2} = \sqrt{s_{1,2}} = \sqrt{k} \text{ and } \sqrt{3k}, \tag{8.9}$$

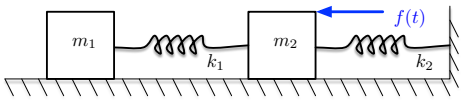
where we have ignored the negative roots since they correspond to negative frequencies.

The two resonance frequencies calculated in (8.9) are precisely the frequencies of the normal modes calculated previously in (7.29) using an eigenvalue calculation. This is not an accident, but a manifestation of a fact that *the frequency response resonances are the same as the frequencies of the normal modes of a system*. This will be shown to be true in general in Section 8.2.

There is a lot of symmetry in the frequency response matrix (8.8). The (1,1) and the (2,2) entries of the matrix are equal. Those represent the frequency response from f_1 to x_1 (the position of m_1) and the frequency response from f_2 to x_2. The fact that they are equal means that the dynamical effect of f_1 on x_1 is exactly equivalent to the dynamical effect of f_2 on x_2. Furthermore, the (1,2) and (2,1) entries are also equal. Those represent the effect of f_2 on x_1 and the effect of f_1 on x_2 respectively. This is again due to the symmetry in this problem. Note that if we change any of the assumptions on equality between m_1, m_2 or on k_1, k_2 and k_{12}, this symmetry will disappear.

Note that in the previous example, all entries of the frequency response matrix (8.8) have no imaginary parts. This is because there is no damping in this example. Examining the formula (8.5) for the general form of the frequency response, we see that if $\mathsf{C} = 0$ and $\mathsf{B}_1 = 0$ (i.e. all the coefficients of the first derivative terms are zero), the matrix $\mathbf{H}(\omega)$ will have no imaginary parts.

Example 8.3. (Frequency Response of System II)



Consider System (II) with the simplifying assumptions that $m_1 = m_2 = m$, $k_1 = k_2 = k$, and no damping $c_1 = c_2 = 0$

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k & -k \\ -k & 2k \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} f(t) \qquad \Leftrightarrow \qquad \mathsf{M}\ \ddot{\boldsymbol{x}}(t) \ + \ \mathsf{K}\ \boldsymbol{x}(t) \ = \ \mathsf{B}_0\ \boldsymbol{u}(t).$$

Applying formula (8.5) to find the frequency response

$$\begin{aligned} \mathbf{H}(\omega) \ &= \ \left(-\omega^2\, \mathsf{M} + \mathsf{K}\right)^{-1} \mathsf{B}_0 \ = \left(-\omega^2\, m \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + k \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\right)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \tfrac{1}{k} \left(-\tfrac{\omega^2}{k/m} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\right)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \tfrac{1}{k} \begin{bmatrix} 1-\Omega^2 & -1 \\ -1 & 2-\Omega^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \left(\Omega := \tfrac{\omega}{\sqrt{k/m}}\right) \\ &= \ \frac{1}{k}\, \frac{1}{\Omega^4 - 3\Omega^2 + 1} \begin{bmatrix} 2-\Omega^2 & 1 \\ 1 & 1-\Omega^2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{k} \begin{bmatrix} \frac{1}{\Omega^4 - 3\Omega^2 + 1} \\ \frac{1-\Omega^2}{\Omega^4 - 3\Omega^2 + 1} \end{bmatrix}. \end{aligned}$$

Note how in the second line, multiplication by the 2×1 matrix B_0 "picks out" only the second column of the 2×2 matrix.

Since in this case there is only one input (one forcing term) the frequency response is a 2×1 matrix. Each entry is the frequency response from f to the displacement of each mass

$$\begin{bmatrix} \hat{\mathsf{x}}_1 \\ \\ \hat{\mathsf{x}}_2 \end{bmatrix} \ = \ \begin{bmatrix} \frac{1/k}{\Omega^4 - 3\Omega^2 + 1} \\ \\ \frac{(1-\Omega^2)/k}{\Omega^4 - 3\Omega^2 + 1} \end{bmatrix} \hat{\mathsf{f}} \ =: \ \begin{bmatrix} \mathsf{H}_1(\Omega) \\ \\ \mathsf{H}_2(\Omega) \end{bmatrix} \hat{\mathsf{f}}. \tag{8.10}$$

Note again that the denominators of both components of the frequency response are the same. The resonance frequencies are calculated by finding the roots of the denominator polynomial. This is again a

4'th order polynomial with only even terms, and the calculation can be reduced to that of a 2'nd order polynomial

$$\Omega^4 - 3\Omega^2 + 1 = 0 \qquad \Leftrightarrow \qquad s^2 - 3s + 1 = 0, \qquad s := \Omega^2.$$

The roots of the 2'nd order polynomial are obtained from the quadratic formula

$$s_{1,2} \;=\; \tfrac{1}{2}\left(3 \pm \sqrt{9-4}\right) \;=\; \tfrac{3\pm\sqrt{5}}{2} \qquad \Rightarrow \qquad \Omega_{1,2} = \sqrt{s_{1,2}} = \pm\sqrt{\tfrac{3\pm\sqrt{5}}{2}}.$$

The two positive frequencies are

$$\Omega_1 \;\approx\; 0.62, \qquad \Omega_2 \;\approx\; 1.62. \tag{8.11}$$

Note again that those are the same frequencies of the normal modes calculated in (7.30). As already stated, this is a general fact.

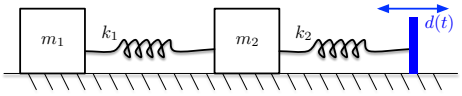
The two frequency responses $\mathsf{H}_1(\Omega)$ and $\mathsf{H}_2(\Omega)$ are plotted in Figure 8.1a. The two resonances Ω_1 and Ω_2 calculated above are clearly visible in the plot. In addition, the frequency response $\mathsf{H}_2(\Omega)$ has an interesting feature that we have not seen before. It becomes exactly zero at a particular frequency

$$\mathsf{H}_2(\Omega) \;=\; \frac{1}{k}\,\frac{1-\Omega^2}{\Omega^4 - 3\Omega^2 + 1} \qquad \Rightarrow \qquad \mathsf{H}_2(\Omega)\Big|_{\Omega=1} \;=\; 0, \tag{8.12}$$

because the *numerator* polynomial has a root at $\Omega_z := 1$ (which corresponds to $\omega_z = \sqrt{k/m}$)[1]. This has a very interesting interpretation, namely that forcing with $f(t)$ at that frequency will result in *zero displacement* in steady state of $x_2(t)$. In other words, *at that frequency $x_2(t)$ is "shielded" from the forcing $f(t)$*! Figure 8.1b illustrates the time response of $x_1(t)$ and $x_2(t)$ when $f(t)$ is at frequency Ω_z. Note how $x_1(t)$ appears to "absorb" all the vibrations caused by $f(t)$, while $x_2(t)$ is shielded from the effects of $f(t)$. This is an example of a "vibration absorber", sometime called a *tuned mass damper*. Figure 8.1c gives a physical interpretation of this vibration absorption phenomenon. At this frequency, the motion of the m_1 is "synchronized" with the forcing so that the force m_1 produces on m_2 exactly cancels the external force f.

In general, if a frequency response is exactly zero at a particular frequency, that frequency is termed a *zero of the frequency response* (thus the notation ω_z). In vibrations problems, this corresponds to a vibration absorption phenomenon. In Electrical Engineering, ω_z is called a "notch frequency", usually used in some kind of filter (sometimes referred to as a "band-stop" filter) to suppress oscillations at that frequency.

Example 8.4. (Frequency Response of System III)



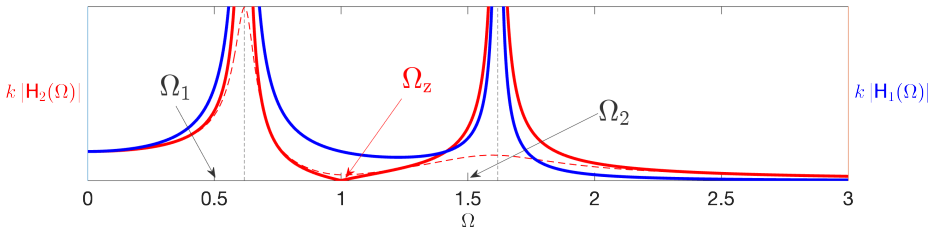
This system also exhibits a vibration absorption phenomenon similar to the previous example. Consider the simplified equations with no damping

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}\begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1+k_2 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ k_2 \end{bmatrix} d(t).$$

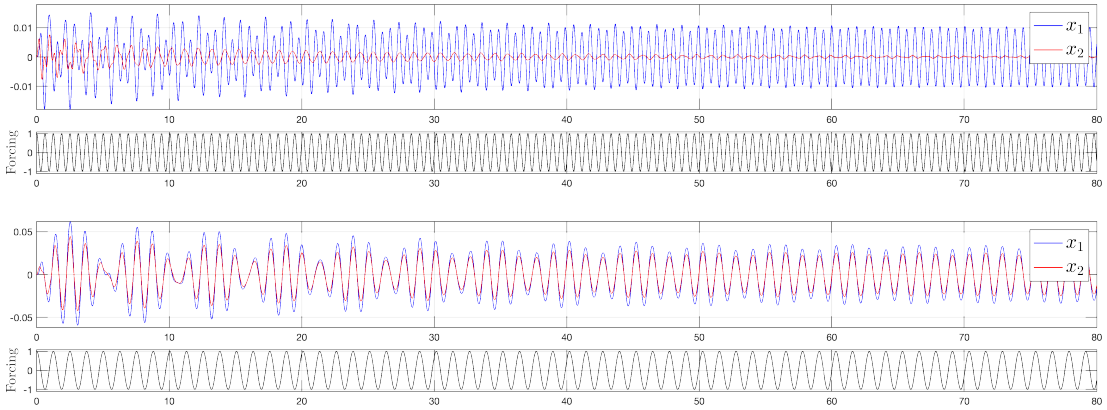
Note that these equations are essentially the same as those of System II, so it is no surprise that the dynamical behavior will be similar. Applying formula (8.5) to this system

$$\begin{aligned} \mathbf{H}(\omega) \;&=\; \left(-\omega^2\,\mathsf{M} + \mathsf{K}\right)^{-1}\mathsf{B}_0 \;=\; \left(-\omega^2\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} + \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1+k_2 \end{bmatrix}\right)^{-1}\begin{bmatrix} 0 \\ k_2 \end{bmatrix} \\ &=\; \begin{bmatrix} \dfrac{k_1 k_2}{m_1 m_2\;\omega^4 \;-\; (k_1 m_1 + k_1 m_2 + k_2 m_1)\;\omega^2 \;+\; k_1 k_2} \\[2ex] \dfrac{k_2\left(k_1 - m_1\omega^2\right)}{m_1 m_2\;\omega^4 \;-\; (k_1 m_1 + k_1 m_2 + k_2 m_1)\;\omega^2 \;+\; k_1 k_2} \end{bmatrix} \;=:\; \begin{bmatrix} \mathsf{H}_1(\omega) \\ \mathsf{H}_2(\omega) \end{bmatrix} \end{aligned}$$

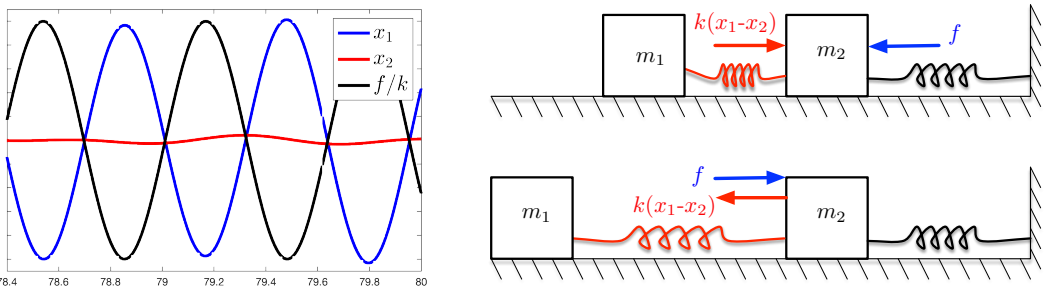
[1] The reader should note that while in a 1-DOF MSD system, $\sqrt{k/m}$ corresponds to the natural frequency (normally denoted by ω_n, we cannot call $\sqrt{k/m}$ a "natural frequency" in this two-mass system. The two natural frequencies of this system are $\omega_1 \approx 0.62\sqrt{k/m}$ and $\omega_2 \approx 1.62\sqrt{k/m}$ as derived in (8.11). In this case, $\sqrt{k/m}$ plays a different role as the vibration absorption frequency.



(a) The amplitude frequency responses of System II. $|\mathsf{H}_1(\omega)|$ (solid red curve) is the response from forcing f to displacement x_1 of m_1, and similarly for $|\mathsf{H}_2(\omega)|$. The resonances of both responses are the same frequencies ω_1 and ω_2 since those are determined by the denominators of the respective frequency responses, which are the same for H_1 and H_2 as can be seen from (8.10). The frequency response H_2 has a "zero" at a particular frequency $\omega_z = \sqrt{k}$, which is the "vibration absorption" frequency. If the input is oscillating at this frequency, then the steady state value of $x_2(t)$ is expected to be zero as verified below. The dashed red curve is the frequency response of H_1 when some amount of damping is introduced between m_1 and m_2. While damping reduces the resonance peaks, it also reduces the quality of vibration absorption at the frequency ω_z.



(b) The time responses of System II when the input $f(t)$ is oscillating at the vibration absorption frequency $\omega_z = \sqrt{k}$ (*top*), and at $\omega = 0.5\sqrt{k}$ (*bottom*). Both simulations include a slight amount of damping ($\zeta \approx 0.01$), which explains the long time needed to settle to steady state). Note how when the forcing is at ω_z, the response of x_2 is almost zero in steady state, and thus m_1 absorbs all the forcing in its vibrations. The fact that the steady state amplitude of x_2 is not exactly zero is due to the slight damping. At other forcing frequencies (e.g. bottom figure), the vibrations of m_1 and m_2 are comparable in amplitude.



(c) A physical interpretation of the vibration absorption phenomenon. At forcing frequency ω_z, the motion of m_1 is "almost synchronized" with the forcing so that they effectively cancel each other's effects on m_2. When the forcing is positive, x_1 is negative (see plot on left) and the coupling spring (between m_1 and m_2) is compressed, producing a negative force which counteracts the applied force f as depicted in the diagram on the right. When the forcing is negative, the opposite occurs.

Figure 8.1: Illustrations of the frequency responses of System II showing *resonances* at ω_1, ω_2 and a *zero* ω_z in the response from forcing to x_2. The time responses illustrate the implications of having a zero in the frequency response as a type of "vibration absorption".

The frequency response $\mathsf{H}_2(\omega)$ from d to x_2 has a zero at the frequency $\omega_z = \sqrt{k_1/m_1}$. This is again a vibration absorption phenomenon. A physical interpretation can be given by imagining that when the input $d(t)$ is oscillating at that frequency, mass m_2 is stationary, and only mass m_1 vibrates. Since m_2 is stationary, the vibrations of m_1 are determined by its mass m_1 and spring k_1, which explains the frequency $\omega_z = \sqrt{k_1/m_1}$. This is a very similar interpretation to that given in Figure 8.1c.

Vibration absorption is a common technique for vibration suppression used in so-called *tuned mass dampers*. Examples of these include "Stockbridge dampers" installed on long cables and transmission lines. Some skyscrapers use tuned mass dampers which are laterally vibrating large masses installed near the top to reduce the vibrations due to dynamic wind loading.

8.2 Relation between Normal Modes and Resonances

Frequency response resonances are a property of the response of a system to "external excitation", where the excitation can be an externally applied force or displacement. In the language of systems analysis, these resonances are *input-output (i.e. external) properties* of a system. Normal modes on the other hand arise from non-equilibrium initial conditions without any external excitation, so they may be considered as *internal properties* of a system. None the less, there is direct connection between the input-output and the internal properties. In particular, for undamped systems, we will show in this section that the normal mode frequencies are precisely the resonance frequencies in the frequency response in a system with no damping. This correspondence also holds as a good approximation in lightly damped systems.

add the simpler derivation using change of basis already introduced before

Consider an undamped system with $\mathsf{C} = 0$. The frequency response formula (8.5) becomes

$$\mathbf{H}(\omega) \;=\; \left(-\omega^2\,\mathsf{M} + \mathsf{K}\right)^{-1} \mathsf{B}_0,$$

where we assume for simplicity that the input enters without differentiation, i.e. $\mathsf{B}_1 = 0$ and $\mathsf{B}_2 = 0$. This formula can be "massaged" so that the diagonalization $\mathsf{M}^{-1}\mathsf{K} = V\Lambda V^{-1}$ is applied as follows[2]

$$\begin{aligned}
\mathbf{H}(\omega) &= \left(-\omega^2\,\mathsf{M} + \mathsf{K}\right)^{-1} \mathsf{B}_0 \\
&= \left(\mathsf{M}\left(-\omega^2\,\mathrm{I} + \mathsf{M}^{-1}\mathsf{K}\right)\right)^{-1} \mathsf{B}_0 && \text{(I is the } n\times n \text{ identity matrix)} \\
&= \left(\mathsf{M}\left(-\omega^2\,\mathrm{I} + V\Lambda V^{-1}\right)\right)^{-1} \mathsf{B}_0 && \text{(substitute the diagonalization of } M^{-1}K) \\
&= \left(\mathsf{M}\,V\left(-\omega^2\,\mathrm{I} + \Lambda\right)V^{-1}\right)^{-1} \mathsf{B}_0 && \text{(“extract” } V \text{ and } V^{-1}\text{, and note } V^{-1}\mathrm{I}V = \mathrm{I}) \\
&= V\left(-\omega^2\,\mathrm{I} + \Lambda\right)^{-1} V^{-1}\mathsf{M}^{-1}\mathsf{B}_0 && \text{(using } (XYZ)^{-1} = Z^{-1}Y^{-1}X^{-1}) \\
&= V\begin{bmatrix} -\omega^2+\lambda_1 & & \\ & \ddots & \\ & & -\omega^2+\lambda_N \end{bmatrix}^{-1} D && \left(\begin{array}{r}(-\omega^2\,\mathrm{I}+\Lambda) \text{ is a diagonal matrix} \\ V^{-1}\mathsf{M}^{-1}\mathsf{B}_0 \text{ is some constant matrix, call it } D\end{array}\right) \\
&= V\begin{bmatrix} \frac{1}{\lambda_1-\omega^2} & & \\ & \ddots & \\ & & \frac{1}{\lambda_N-\omega^2} \end{bmatrix} D. && \text{(the inverse of a diagonal matrix is also diagonal)}
\end{aligned}$$

The exact values of the entries of the vector D are not important for the analysis next.

Now examine more closely the meaning of the last expression above for the response phasor vector $\hat{\mathbf{x}}$. First recall that the normal mode frequencies ω_i are related to the eigenvalues λ_i of the matrix $\mathsf{M}^{-1}\mathsf{K}$ by $\omega_i^2 = \lambda_i$. Assume for simplicity that there is a single external input f, and therefore its phasor $\hat{\mathsf{f}}$ is a scalar rather than a vector. In this case the matrix D is $n \times 1$ and the relation

[2] The matrix manipulations here are equivalent to changing to new coordinates z using $x = Vz$, finding the decoupled frequency responses, and then changing back to the original coordinates x.

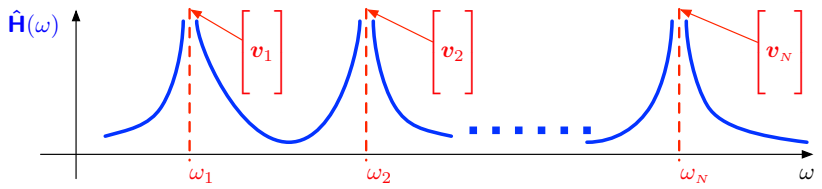


Figure 8.2: A depiction of the frequency response of an n-degree of freedom, undamped (or lightly damped) mechanical system. This depiction is for a system with a single forcing input. The resonance frequencies $\omega_1, \dots, \omega_n$ are the same as the "internal" normal modes. Furthermore, when the forcing is at a frequency close to a resonance $\omega \approx \omega_i$, then the response phasor vector is nearly aligned (i.e. $\hat{\mathbf{x}} = \mathbf{H}(\omega)\hat{\mathsf{f}} \approx \boldsymbol{v}_i \alpha$) with the eigenvector $\boldsymbol{v}_i$ of that normal mode as follows from (8.13). This determines the "vector shape" of the response near each resonance in the sense that the response as a vector function of time is $\boldsymbol{x}(t) \approx \boldsymbol{v}_i \ \alpha \cos\big(\omega t + \theta\big)$. A version of this diagram for the specific example of System II is shown in Figure 8.3.

between $\hat{\mathbf{x}}$ and $\hat{\mathsf{f}}$ is

$$\hat{\mathbf{x}} \ = \ \mathbf{H}(\omega)\ \hat{\mathsf{f}} \ = \ V \begin{bmatrix} \frac{1}{\omega_1^2 - \omega^2} & & \\ & \ddots & \\ & & \frac{1}{\omega_N^2 - \omega^2} \end{bmatrix} D\ \hat{\mathsf{f}} \ = \ \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_N \end{bmatrix} \begin{bmatrix} \frac{1}{\omega_1^2 - \omega^2} & & \\ & \ddots & \\ & & \frac{1}{\omega_N^2 - \omega^2} \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix} \hat{\mathsf{f}}$$

$$\begin{bmatrix} \\ \hat{\mathbf{x}} \\ \\ \end{bmatrix} \ = \ \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_N \end{bmatrix} \begin{bmatrix} \frac{d_1 \hat{\mathsf{f}}}{\omega_1^2 - \omega^2} \\ \vdots \\ \frac{d_n \hat{\mathsf{f}}}{\omega_N^2 - \omega^2} \end{bmatrix} \ = \ \begin{bmatrix} \\ \boldsymbol{v}_1 \\ \\ \end{bmatrix} \frac{d_1 \hat{\mathsf{f}}}{\omega_1^2 - \omega^2} \ + \ \cdots \ + \ \begin{bmatrix} \\ \boldsymbol{v}_N \\ \\ \end{bmatrix} \frac{d_N \hat{\mathsf{f}}}{\omega_N^2 - \omega^2}. \tag{8.13}$$

This last expression says a lot! As the frequency approaches each normal mode $\omega \to \omega_i := \sqrt{\lambda_i}$, the corresponding term $\frac{d_i \hat{\mathsf{f}}}{\omega_i^2 - \omega^2}$ goes to infinity, i.e. a resonance in the frequency response[3]. Thus we see that provided none of coefficients d_i are zero (which holds generically), *the internal normal mode frequencies are the same as the external resonance frequencies*. This is depicted in Figure 8.2.

There is another important observation from (8.13) about the "shape" of vibrations near resonances. Near each ω_i, the dominant term multiplies the eigenvector $\boldsymbol{v}_i$. This means that the phasor vector $\hat{\mathbf{x}}$ is essentially aligned with $\boldsymbol{v}_i$. This says that as a vector, $\boldsymbol{x}(t)$ should be aligned with $\boldsymbol{v}_i$ for each t, i.e.

$$\begin{array}{rl} & f(t) \ = \ \mathsf{f} \ \cos\big(\omega t + \phi\big), \quad \text{and } \omega \approx \omega_i \\ \Rightarrow \quad & \hat{\mathbf{x}} \ \approx \ \boldsymbol{v}_i \ \alpha \end{array} \qquad \Rightarrow \qquad \begin{bmatrix} \\ \boldsymbol{x}(t) \\ \\ \end{bmatrix} \ \approx \ \begin{bmatrix} \\ \boldsymbol{v}_i \\ \\ \end{bmatrix} \alpha \ \mathsf{f} \cos\big(\omega t + \phi\big), \tag{8.14}$$

where $\alpha := \frac{d_i}{\omega_i^2 - \omega^2}$ is a scalar which depends on the exact value of ω, i.e. when $\omega \approx \omega_i$, the "shape of the vibrations" will be a temporal oscillation (with frequency ω_i) of the vector $\boldsymbol{v}_i$.

Example 8.5. In System II, the normal modes were calculated (in (7.30) and (7.31)) to be (in units of normalized frequency $\Omega := \omega/\sqrt{km}$)

$$\Omega_1 \approx \ 0.62, \quad \boldsymbol{v}_1 \approx \begin{bmatrix} 3.24 \\ 2 \end{bmatrix}, \qquad \Omega_2 \approx 1.62, \quad \boldsymbol{v}_2 \approx \begin{bmatrix} -1.24 \\ 2 \end{bmatrix}. \tag{8.15}$$

This means that at the first resonance ω_1, the two masses will move "in phase" (i.e. in the same direction) with their amplitudes having the ratio $\mathsf{x}_1/\mathsf{x}_2 \approx 3.24/2 \approx 1.62$, while at the second resonance ω_2, the two masses will move "completely out of phase" (i.e. with 180° degree phase difference, which means moving in opposite directions) with their amplitudes having the ratio $\mathsf{x}_1/\mathsf{x}_2 \approx 1.24/2 \approx 0.62$. Figure 8.3 shows the frequency response magnitude together with (steady state) time responses for forcing near the two resonance frequencies. Note that these responses look the same as the (unforced) responses from the special initial conditions $\boldsymbol{x}(0) = \boldsymbol{v}_1$ and $\boldsymbol{x}(0) = \boldsymbol{v}_2$ shown in Figure 7.6.

[3] If there is any damping in the system, the frequency response will have a "peak" at ω_i rather than going to infinity.

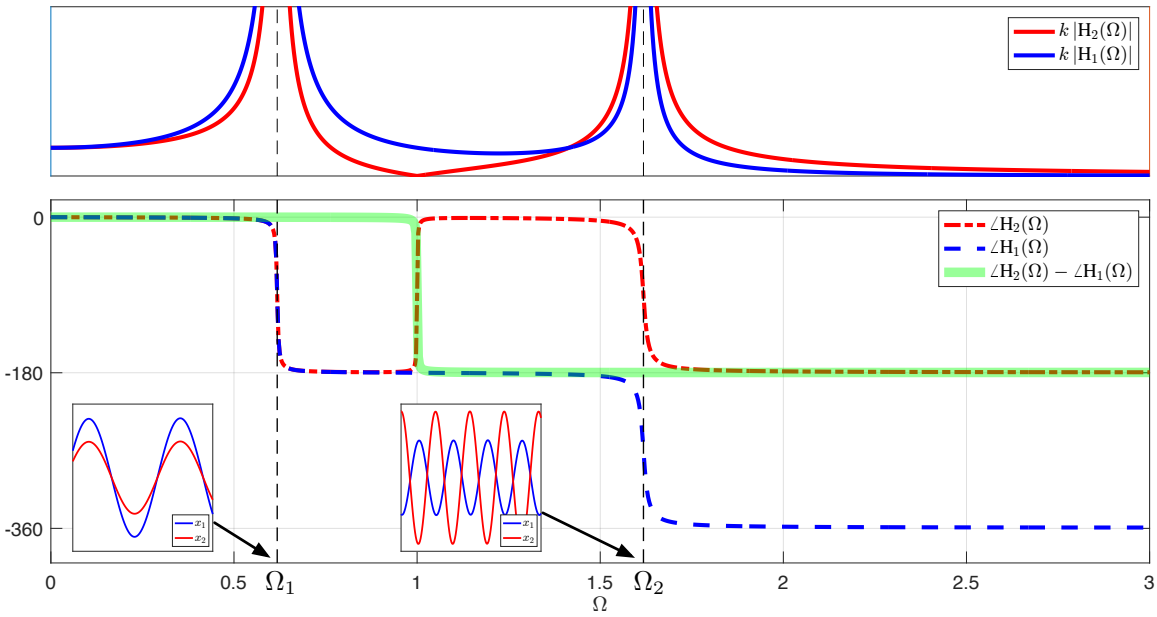


Figure 8.3: Magnitude (*top*) and phase (*bottom*) of System II. The insets show the time responses when forcing is near each of the two resonance frequencies Ω_1 and Ω_2. These responses are the same as the respective free vibrations starting from eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ respectively as initial conditions (Figure 7.6). For forcing at Ω_1, the masses move in-phase, while they move out-of-phase when forcing is at Ω_2. The phase plots $\angle\mathsf{H}_1$, $\angle\mathsf{H}_2$ and their difference $\angle\mathsf{H}_1 - \angle\mathsf{H}_2$ are shown. Note that $\mathsf{H}_i(\Omega)$ is the phase difference between *forcing input* and position of mass m_i. On the other hand, the green curve $\angle\mathsf{H}_1(\Omega) - \angle\mathsf{H}_2(\Omega)$ shows the phase difference *between the two masses*. We see clearly that below $\Omega < 1$, m_1, m_2 are in phase (unison motion), while for $\Omega > 1$, they are 180° out of phase (differential motion). The plots were generated with very small damping ($\zeta = 0.01$).

Figure 8.3 shows additional information that can be obtained from examining the phase plots $\angle\mathsf{H}_1(\Omega)$ and $\angle\mathsf{H}_1(\Omega)$. Those two plots indicate the phase difference between forcing input and each mass' position. The difference $\angle\mathsf{H}_1(\Omega) - \angle\mathsf{H}_1(\Omega)$ is also shown. This indicates the phase difference between the two masses positions (irrespective of their phase difference relative to forcing). We see that for $\Omega < 1$, $\angle\mathsf{H}_1(\Omega) - \angle\mathsf{H}_2(\Omega) \approx 0$, i.e. the two masses move in phase, which is consistent with the first resonance (at $\Omega \approx 0.62 < 1$) corresponding to unison motion. On the other hand, for $\Omega > 1$, $\angle\mathsf{H}_1(\Omega) - \angle\mathsf{H}_2(\Omega) \approx 180°$, i.e. the two masses move completely out of phase[4], which is consistent with the second resonance (at $\Omega \approx 1.62 > 1$) corresponding to differential motion.

Figure 8.3 illustrates how phase plots can reveal useful information about the "shape of vibrations". To have a full picture of harmonically forced vibrations in n-DOF systems, both magnitude and phase plots need to be considered. We will see further examples of this when considering large-scale n-DOF systems in Chapter 9.

8.2.1 Hidden Modes

In the pervious section we made the statement that "generically" the internal normal mode frequencies are the same as the external resonance frequencies. The precise statement is that *the external resonance frequencies can only be a subset of the internal mode frequencies*. This means there are no external resonances that did not come from internal normal modes. However, there could be internal modes that do not appear as an input-output resonance in the frequency response. We call such frequencies "hidden modes". They only arise in very specific geometries or systems with symmetries. In this section we give such examples to add insight into the relationship between internal and external vibration phenomena.

[4]Recall that a phase difference of $-180°$ is the same as a phase difference of $+180°$.

Consider the three mass system shown in Figure 8.4a. The differential equations can be derived as follows[5]

$$\begin{aligned} m_1\,\ddot{x}_1 &= -k_1x_1 - k_{12}(x_1-x_2) + f_1 &&= -(k_1+k_{12})\,x_1 + k_{12}\,x_2 + f_1,\\ m_2\,\ddot{x}_2 &= k_{12}(x_1-x_2) - k_{23}(x_2-x_3) + f_2 &&= k_{12}\,x_1 - (k_{12}+k_{23})\,x_2 + k_{23}\,x_3 + f_2,\\ m_3\,\ddot{x}_2 &= k_{23}(x_2-x_3) - k_3x_3 &&= k_{23}\,x_2 - (k_{23}+k_3)\,x_3. \end{aligned}$$

Those equation can be put in the standard matrix-vector form (7.12)

$$\begin{bmatrix} m_1 & 0 & 0\\ 0 & m_2 & 0\\ 0 & 0 & m_3 \end{bmatrix}\begin{bmatrix} \ddot{x}_1\\ \ddot{x}_2\\ \ddot{x}_3 \end{bmatrix} + \begin{bmatrix} k_1+k_{12} & -k_{12} & 0\\ -k_{12} & k_{12}+k_{23} & -k_{23}\\ 0 & -k_{23} & k_{23}+k_3 \end{bmatrix}\begin{bmatrix} x_1\\ x_2\\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0\\ 0 & 1\\ 0 & 0 \end{bmatrix}\begin{bmatrix} f_1\\ f_2 \end{bmatrix}$$

We first analyze this system under the simplifying assumption that all masses are equal $m_1 = m_2 = m_3 = m$, and all spring constants are equal $k_1 = k_{12} = k_{23} = k_3 = k$. In this case $\mathsf{M}^{-1}\mathsf{K}$ becomes

$$\mathsf{M}^{-1}\mathsf{K} = \frac{k}{m}\begin{bmatrix} 2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2 \end{bmatrix} =: \frac{k}{m}\,\Delta,$$

where Δ is the matrix of constants defined above. Since the matrix $\mathsf{M}^{-1}\mathsf{K}$ is a scalar multiple of Δ, the *eigenvectors* of $\mathsf{M}^{-1}\mathsf{K}$ are the same as those of Δ, and the *eigenvalues* are $\frac{k}{m}$ times those of Δ, all of which can be computed numerically. The results are best summarized by defining a "normalized frequency" $\Omega := \omega/\omega_{\rm n}$ where $\omega_{\rm n} = \sqrt{k/m}$ (the natural frequency of a single mass-spring system). The computed normal modes are

$$\lambda_1(\Delta) \approx 0.59,\quad \Omega_1 \approx 0.77,\quad \boldsymbol{v}_1 = \begin{bmatrix} 1\\ \sqrt{2}\\ 1 \end{bmatrix},\qquad \lambda_2(\Delta) = 2,\quad \Omega_2 = \sqrt{2},\quad \boldsymbol{v}_2 = \begin{bmatrix} 1\\ 0\\ -1 \end{bmatrix},\qquad \lambda_3(\Delta) \approx 3.4,\quad \Omega_3 \approx 1.8,\quad \boldsymbol{v}_3 = \begin{bmatrix} -1\\ \sqrt{2}\\ -1 \end{bmatrix}$$

Note that the eigenvalues listed are those of the constant matrix Δ (e.g. $\lambda_1(\Delta)$ is the first eigenvalues of Δ), while the actual normal modes of the system are obtained from $\omega_i = \Omega_i\omega_{\rm n} = \Omega_i\sqrt{k/m}$. The physical interpretations of the three modes of motion are depicted in Figure 8.4b.

The frequency response from forcing inputs f_1, f_2 to masses positions has some peculiar behavior in this system. It is calculated from formula (8.5) as

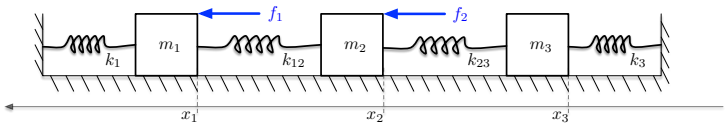
$$\begin{aligned} \mathbf{H}(\omega) &= \big(-\omega^2\,\mathsf{M}+\mathsf{K}\big)^{-1}\,\mathsf{B}_0 = \big(-\omega^2\,m\,\mathrm{I}+k\,\Delta\big)^{-1}\,\mathsf{B}_0\\ &= \frac{1}{k}\big(-\Omega^2\,\mathrm{I}+\Delta\big)^{-1}\,\mathsf{B}_0 \qquad \Big(\Omega := \omega/\sqrt{k/m} \;\Leftrightarrow\; \omega^2 m = \Omega^2 k\Big)\\ &= \frac{1}{k}\left(-\Omega^2\begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & 0\\ 0 & 1\\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Using symbolic calculations, the frequency response matrix (in terms of Ω) is computed as

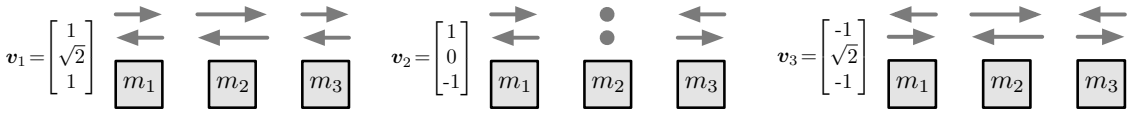
$$\mathbf{H}(\Omega) = \frac{1}{k}\begin{bmatrix} \frac{-(\Omega^4-4\Omega^2+3)}{(\Omega^4-4\Omega^2+2)(\Omega^2-2)} & \frac{1}{(\Omega^2-2)^2}\\ \frac{1}{(\Omega^4-4\Omega^2+2)} & \frac{-1}{(\Omega^2-2)}\\ \frac{-1}{(\Omega^4-4\Omega^2+2)(\Omega^2-2)} & \frac{1}{(\Omega^2-2)^2} \end{bmatrix} =: \begin{bmatrix} \mathsf{H}_{11} & \mathsf{H}_{12}\\ \mathsf{H}_{21} & \mathsf{H}_{22}\\ \mathsf{H}_{31} & \mathsf{H}_{32} \end{bmatrix} \quad\Leftrightarrow\quad \begin{bmatrix} \hat{\mathsf{x}}_1\\ \hat{\mathsf{x}}_2\\ \hat{\mathsf{x}}_3 \end{bmatrix} = \begin{bmatrix} \mathsf{H}_{11} & \mathsf{H}_{12}\\ \mathsf{H}_{21} & \mathsf{H}_{22}\\ \mathsf{H}_{31} & \mathsf{H}_{32} \end{bmatrix}\begin{bmatrix} \hat{\mathsf{f}}_1\\ \\ \hat{\mathsf{f}}_2 \end{bmatrix},$$

where the dependence on Ω is suppressed in the notation for the individual matrix entries H_{ij} for simplicity of notation. There is a lot to unpack in this frequency response relation, and we list some below as individual remarks.

[5] As usual, the coordinates x_i are chosen so that $x_1 = 0$, $x_2 = 0$ and $x_3 = 0$ corresponding to all springs being unstretched.



(a) The three mass system with two forcing inputs.

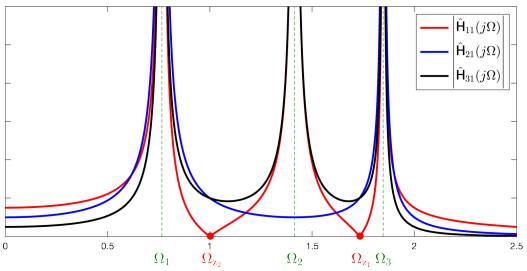


Mode 1: unison motion. m_1 and m_3 oscillate together with same amplitude. m_2 oscillates in tandem with a larger amplitude

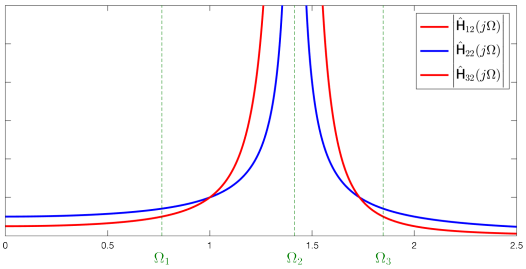
Mode 2: m_2 stationary. m_1 and m_3 oscillate in opposite directions

Mode 3: m_1 and m_3 oscillate in tandem, while m_2 oscillates in the opposite direction with a larger amplitude

(b) The three normal modes of the system at frequencies $\Omega_1 \approx 0.77$, $\Omega_2 = \sqrt{2}$, and $\Omega_3 \approx 1.8$. The eigenvectors are shown together with depictions of each corresponding mode of motion.



(c) Frequency responses from f_1 to the three mass positions. All three normal modes appear as resonances in the responses to x_1 and x_3. However, mode Ω_2 is "hidden" in the response H_{21}. This means that when f_1 is at frequency Ω_2, x_1 and x_3 will move according to mode 2 above, and x_2 will remain relatively stationary.



(d) Frequency responses from f_2 to the three mass positions. Modes Ω_1 and Ω_3 are "hidden" and only mode Ω_2 appears as a resonance to all mass positions. Note how H_{22} (the response to x_2) has lower growth than the other two responses. This is again a manifestation of mode 2 above, where the center mass remains relatively stationary compared to the other two at that mode.

Figure 8.4: Analysis of the three mass system with two different kinds of forcing in the case where all masses and spring constants are equal respectively. The middle figure depicts the normal modes, while the bottom figure shows the various frequency responses. Some of the frequency responses have a subset of the normal modes "hidden", i.e. they do not appear as resonances. This phenomenon is due to the symmetry in the problem. The plots are shown in terms of the normalized frequency $\Omega := \omega/\omega_n := \omega/\sqrt{k/m}$.

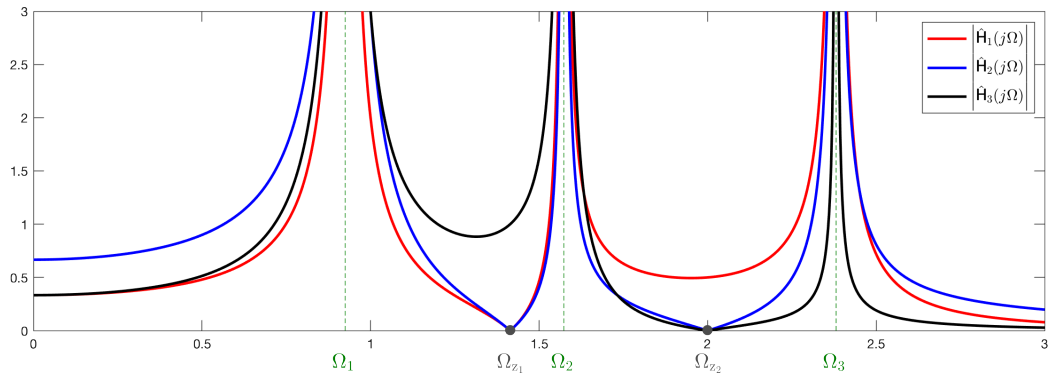


Figure 8.5: The three components of the frequency response of the three-mass system as a function of the normalized frequency $\Omega = \omega/\sqrt{k/m}$. Each represents the frequency response from forcing to the respective mass position x_i. All three responses have their resonances at all three normal mode frequencies. There are also two zeros, each for a subset of the responses, and those represent two different phenomena of vibration absorption.

- The second column of $\mathbf{H}(\omega)$ corresponds to the effect of f_2 on the three masses with f_1 set to zero. This is a very special and symmetric situation since the force acts on the center mass, and all masses and spring constants are respectively equal. Note that even though we have three masses, and thus three normal modes, there is only one resonance at $\Omega = \sqrt{2}$ for all the entries of the second column. To see the behavior of the masses around this resonance

$$\begin{bmatrix} \frac{1}{(\Omega^2-2)^2} \\ \frac{-1}{(\Omega^2-2)} \\ \frac{1}{(\Omega^2-2)^2} \end{bmatrix} = \frac{1}{(\Omega^2-2)^2} \begin{bmatrix} 1 \\ -(\Omega^2-2) \\ 1 \end{bmatrix} \sim \frac{1}{(\Omega^2-2)^2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{(\Omega^2-2)^2} \begin{bmatrix} \boldsymbol{v}_2 \end{bmatrix}, \qquad (\text{as } \Omega \to 2)$$

  where $\sim$ refers to the "leading order" behavior as $\Omega \to \sqrt{2}$.

  Thus the only resonance from f_2 corresponds approximately to vibrations of normal mode $\boldsymbol{v}_2$. More precisely, as $\Omega \to \sqrt{2}$, m_2 vibrates with amplitude proportional to $1/(\Omega^2-2)$, but m_1 and m_3 vibrate with the much larger amplitudes $1/(\Omega^2-2)^2$. Thus near that resonance the motion resembles that of the 2nd mode depicted in Figure 8.4b.

The roots of the denominator are the frequencies where the response goes to infinity (i.e. resonance), and the positive ones are exactly the frequencies Ω_1, Ω_2 and Ω_3 calculated earlier as the normal modes. Figure 8.5 illustrates the magnitudes of each component of $\mathsf{H}(\omega)$ separately. It is clear that each goes to infinity at each of the resonances as expected. An interesting feature of this system is that it has *two zeros* Ω_{z_1} and Ω_{z_2} corresponding to two different scenarios of vibration absorption. The first, Ω_{z_1} corresponds to $\mathsf{H}_1(\Omega_{z_1}) = 0$ and $\mathsf{H}_2(\Omega_{z_1}) = 0$, i.e. m_1 and m_2 not moving and all the vibrations being absorbed by m_3. The second corresponds to $\mathsf{H}_2(\Omega_{z_1}) = 0$ and $\mathsf{H}_3(\Omega_{z_1}) = 0$, i.e. m_2 and m_3 not moving and all the vibrations being absorbed by m_1. The reader should try to think of a physical interpretation of these phenomena similar to that given in Figure 8.1c.

$$\mathbf{H}(\omega) = \left(-\omega^2 m \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + k \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{k} \left(-\frac{\omega^2}{k/m} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \frac{1}{k} \frac{1}{\Omega^6 - 6\Omega^4 + 10\Omega^2 - 4} \begin{bmatrix} 2\left(2-\Omega^2\right) \\ \left(2-\Omega^2\right)\left(\Omega^4-4\right) \\ \left(4-\Omega^2\right) \end{bmatrix}. \qquad \left(\Omega := \omega/\omega_{\text{n}} = \frac{\omega}{\sqrt{k/m}}\right)$$

Chapter 9

Vibrations in Large-Scale Systems

Normal mode and frequency response analysis for n-DOF systems can be applied to systems of any order. This chapter gives examples where n can be quite large either due to the system comprising many masses, or when the n-DOF dynamics arise from spatial discretizations of continuum models such as those of beam or wave phenomena. The first example involves a chain of a large number of masses, which has similar behavior to the wave equation in one dimension. The second is the wave in one spatial dimension which has an infinite number of normal modes. After a finite-difference discretization (in space), the equations resemble those of an n-DOF system, which can then be analyzed using the tools developed so far. The third example is the Euler-Bernoulli beam equation, which is analyzed through a similar finite-difference discretization. In all the examples, the relations between the normal modes and the frequency response is explored.

Introduction

The first example presented is a system with N masses where N can be arbitrary. In the other examples presented here of the wave and beam equations, the number of degrees of freedom N is the size of the discretization grid used in finite-difference approximations of the underlying Partial Differential Equations (PDEs). Spatial discretization of a PDE over a spatial grid of size N yields N coupled ODEs representing an N-DOF system. Thus N can be as large as needed for a sufficiently accurate approximation of the PDE.

The wave and beam equations are termed "continuum" systems since their mechanical degrees of freedom form an infinite continuum. Spatial discretizations of the underlying PDEs provide numerical schemes for converting such systems to finite DOF system, which can then be analyzed using the tools developed so far, and this is the approach taken in this chapter. Chapter 10 will present analytical methods for analysis of some continuum systems without the need for discretization methods.

9.1 A Mass Chain

The first example is a "chain" of N masses connected by springs and moving in one dimension as depicted in Figure 9.1. In the free-vibrations setting, the two springs at the boundaries are attached to rigid wall anchors (top of Figure 9.1), while in the forced-vibrations setting (bottom figure), the left wall is replaced by a moving boundary which moves with an externally determined displacement $u(t)$. The forced-vibrations setting is used to explore the frequency response of the system from that input.

For simplicity of exposition, the springs are assumed identical. Mass chain models with the two ends either free or connected to rigid anchors represent idealized dynamics of several real systems. Those include many vehicles connected in series such as the cars of a long locomotive, or the dynamics of a tightly packed "string" of vehicles on a single lane of a highway. In the latter case the spring connections are "virtual", i.e. they represents the behavior of the drivers of vehicles (whether humans

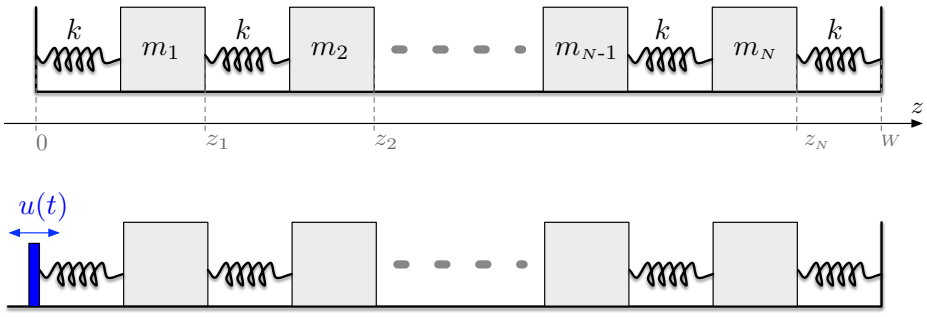


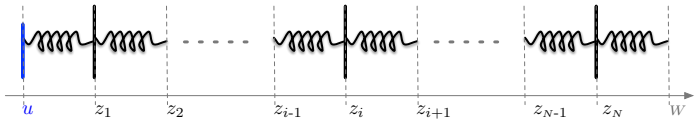
Figure 9.1: A chain of N masses connected with identical springs. The mass' positions $\{z_i\}$ are measured on a single coordinate axis. The analysis is however simplified using an alternative set of coordinates $x_i := z_i - iL$, which measure the "deviation" of the i'th mass position from its equilibrium position iL (where L is the equilibrium length of each spring, and the masses are assumed to have zero width). The mass chain is an idealized model of large interacting subsystems arranged in a one-dimensional "lattice". This idealization is useful for some traffic models such as connected locomotive cars and a tight chain of vehicles traveling on a highway lane. The bottom figure depicts a mass chain driven by the position $u(t)$ of the platform on the left.

or automated systems) in terms of their response to inter-vehicle gaps' deviation from desired value. The control force law in this case can be modeled as a "virtual spring". At the other extreme end of spatial scales, an extremely large mass chain is a so-called "harmonic solids" model of atomic lattices where the masses represent atoms and the springs represents bonds between them. The vibrational modes of such a crystal lattice are called "phonons" in quantum mechanical models of crystal vibrations.

Let the position of the i'th mass be denoted by $z_i(t)$. The positions $\{z_i\}$ are measured using one coordinate system as shown. A change of coordinates will simplify the equations of motions by "removing" constant terms which are not relevant to vibrational behavior as shown next. For simplicity assume each mass to have zero width. The equilibrium length of each spring is L, and at static equilibrium all springs are uncompressed which means that the distance between the two vertical walls must be

$$W \;=\; (N+1)\,L.$$

Now we write Newton's 2nd law for each of the masses in the interior region $2 \leq i \leq N-1$, and then for the two masses at the left and right end respectively as shown in the figure below (suppressing



the dependence on t in the equations below for simplicity of notation)

$$\begin{aligned}
m_i\,\ddot{z}_i &= -k\left(z_i - z_{i-1} - L\right) + k\left(z_{i+1} - z_i - L\right) && (i = 2,..,N\text{-}1)\\
m_1\,\ddot{z}_1 &= -k\left(z_1 - u - L\right) + k\left(z_2 - z_1 - L\right) && (\text{left end})\\
m_N\,\ddot{z}_n &= -k\left(z_N - z_{N-1} - L\right) + k\left(W - z_N - L\right) && (\text{right end})
\end{aligned}$$

These equations contain the constant terms W and L, which can be removed by defining new coordinates which are the *deviations from equilibrium positions*, i.e.

$$\begin{aligned}
x_i \;:=\; z_i - i\,L \qquad &\Leftrightarrow \qquad z_i \;=\; x_i + i\,L, && (i = 1,2,\ldots,N)\\
\Rightarrow \quad x_i - x_{i-1} \;&=\; \big(z_i - iL\big) - \big(z_{i-1} - (i-1)L\big) \;=\; \big(z_i - z_{i-1} - L\big)\\
x_1 \;&=\; z_1 - L\\
x_N \;&=\; z_N - NL \;=\; z_N - W + L && (\text{using } W = NL + L)
\end{aligned}$$

In terms of the new variables $\{x_i(t)\}$, the equations become simpler with no constant terms

$$\begin{aligned} m_i\,\ddot{x}_i &= -k\,(x_i - x_{i-1}) + k\,(x_{i+1} - x_i) &&= -k\,(-x_{i-1} + 2x_i - x_{i+1}) && (i = 2,\ldots,N\text{-}1)\\ m_1\,\ddot{x}_1 &= -k\,(x_1 - u) - k\,(x_2 + x_1) &&= -k\,(2x_1 - x_2) && (\text{left end})\\ m_N\,\ddot{x}_N &= -k\,(x_N - x_{N-1}) - k\,x_N &&= -k\,(-x_{N-1} - 2x_N) && (\text{right end}) \end{aligned}$$

Of course once these equations are solved, we can reconstruct the original position variables using the relation

$$z_i(t) \;=\; x_i(t) \;+\; i\,L, \qquad\qquad i = 1, 2, \ldots, N. \tag{9.1}$$

The differential equations for the x_i variables are all mutually coupled. They have to be solved all together. The best way to understand their structure is to write them all together as a *single vector* differential equation as follows

$$\begin{bmatrix} m_1 & & & & \\ & m_2 & & & \\ & & \ddots & & \\ & & & m_{N\text{-}1} & \\ & & & & m_N \end{bmatrix} \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \\ \vdots \\ \ddot{x}_{N\text{-}1}(t) \\ \ddot{x}_N(t) \end{bmatrix} + k \underbrace{\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}}_{\mathcal{L}_N} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{N\text{-}1}(t) \\ x_N(t) \end{bmatrix} = k \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} u(t) \tag{9.2}$$

$$\Leftrightarrow \qquad \mathsf{M}\,\ddot{\boldsymbol{x}}(t) + \mathsf{K}\,\boldsymbol{x}(t) \;=\; \mathsf{B}_0\,u(t)$$

where the mass M, stiffness $\mathsf{K} := k\,\mathcal{L}_N$, and B_0 matrices are as shown. The numerical matrix $\mathcal{L}_N$ defined above is the Laplacian[1] of size N, and therefore denoted by $\mathcal{L}_N$. While M is diagonal, the stiffness matrix K is *tri-diagonal*; a structure reflecting the "nearest neighbor" interaction between the masses.

Free Vibrations of the Mass Chain

For analysis of free (unforced) vibrations of the mass chain, we simply set $u(t) = 0$, i.e. the right hand side of (9.2) is zero. Note that $u(t) = 0$ is equivalent to attaching the leftmost spring to a rigid wall. This system with N masses moving in only one coordinate has N degrees of freedom. The normal modes of vibration are given by the eigenvalues of $\mathsf{M}^{-1}\mathsf{K}$, and the normal mode shapes by the eigenvectors. If we make the simplifying assumptions that all masses are equal $m_i = m,\ \ i = 1, \ldots, N$, then we have

$$\mathsf{M}^{-1}\mathsf{K} \;=\; \tfrac{k}{m}\,\mathcal{L}_N.$$

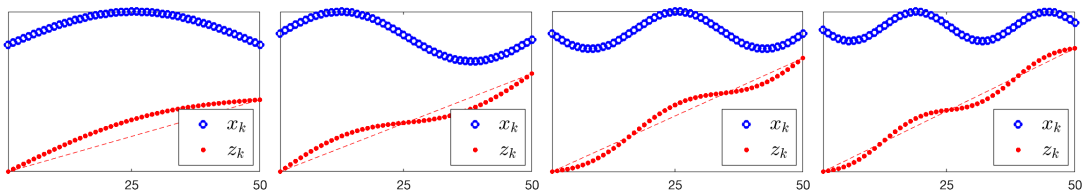
This means that $\mathsf{M}^{-1}\mathsf{K}$ is a scalar multiple of the numerical matrix $\mathcal{L}_N$. Thus the eigenvalues of $\mathsf{M}^{-1}\mathsf{K}$ are the eigenvalues of $\mathcal{L}_N$ multiplied by the scalar factor k/m, while the eigenvectors of $\mathsf{M}^{-1}\mathsf{K}$ are exactly the eigenvectors of $\mathcal{L}_N$. Both the eigenvectors and eigenvalues of $\mathcal{L}_N$ can be calculated numerical or analytically[2]. In particular, it can be shown that for large N, the first few eigenvalues of $\mathcal{L}_N$ are well approximated by

$$\lambda_l(\mathcal{L}_N) \approx \left(\tfrac{\pi}{N}\,l\right)^2 \quad \Rightarrow \quad \lambda_l\left(\tfrac{k}{m}\mathcal{L}_N\right) \approx \tfrac{k}{m}\left(\tfrac{\pi}{N}\,l\right)^2 \quad \Rightarrow \quad \omega_l \approx \left(\tfrac{\pi}{N}\sqrt{\tfrac{k}{m}}\right) l, \quad l = 1, 2, \ldots, \tag{9.3}$$

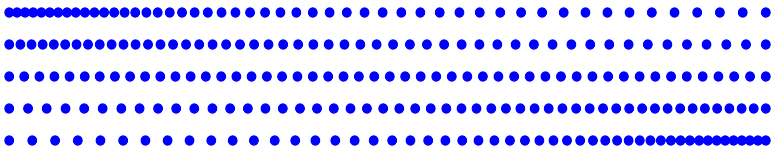
where $\lambda_l(\mathcal{L}_N)$ denotes the l'th eigenvalue of $\mathcal{L}_N$, and ω_l is the l'th natural frequency of the N-mass chain system. The formulas for the eigenvalues and natural frequencies are a relatively good approximation for the first 20% of natural frequencies (when arranged in increasing order), which

[1] This name originates from the fact that $-\mathcal{L}_N$ is the finite-difference discretization of the second derivative operator $\frac{\partial^2}{\partial x^2}$, traditionally referred to as the "Laplacian".

[2] The matrix $\mathcal{L}_N$ is a tridiganoal so-called *Toeplitz matrix*, for which the eigenvalues and vectors can be derived analytically.



(a) Mode shapes of the first four modes of a mass chain system of size 50. The blue circles represent the position deviation variables $\{x_k\}$, while the red dots represent the absolute mass positions $\{z_k\}$ obtained from the deviations by (9.1). The red dashed lines indicate the equilibrium positions of the masses. Click for Animation.



(b) A depiction of the first normal mode of the mass chain using five snapshots over 1/2 of the period. A "compression wave" starts from one end and propagates to the other end. The wave then reverses and the process repeats periodically. Click for Animation.

(c) The second normal mode begins with a compression wave either in the middle or at both edges. The compression then propagates from one configuration to the other repeatedly. The diagram depicts 1/2 of a period. Click for animation. See also the 3rd mode animation, the 4th mode animation, as well as the animation of the first 4 modes together.

Figure 9.2: Illustrations of some of the normal modes of the mass chain system computed with 50 masses. The modes are usually ordered by increasing values of their corresponding natural frequencies. The most important modes are those with the lowest natural frequencies since in any real system, damping tends to suppress higher frequency oscillations more than lower frequency ones.

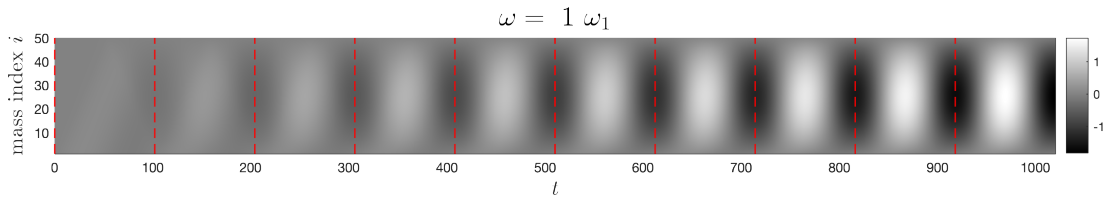
are typically the most important modes of vibration[3]. For comparison, the first few actual natural frequencies computed from the eigenvalues of $\mathcal{L}_N$ for $N = 100$ are

l	1	2	3	4
$\omega_l \Big/ \left(\frac{\pi}{N}\sqrt{\frac{k}{m}}\right)$	0.99	1.98	2.97	3.96

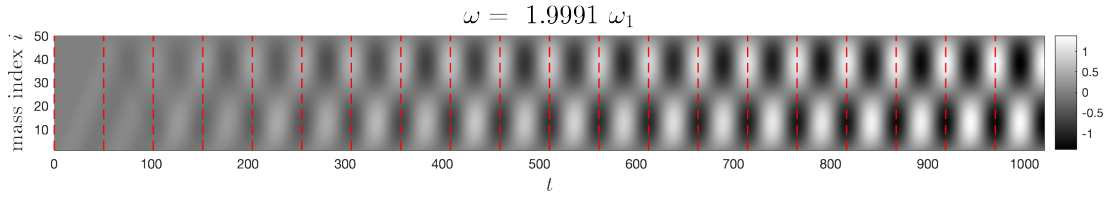
which should be compared with the approximation (9.3). We note that this approximation becomes exact in the limit of a large number of masses $N \to \infty$.

Figure 9.2 depicts a visualization of the first few normal modes (ordered in terms of oscillation frequency) calculated from the *eigenvectors* of $\mathcal{L}_N$. Those modes have the structure of "compression waves". The first mode is a compression wave oscillating back and forth between the left and right end, while the second mode is a compression wave oscillating back and forth between the edges of the chain towards the center. Higher frequency modes also look like compression waves, but with more minimum and maximum compression regions. Figure 9.2 shows the modes for $N = 50$. For such a large system, the eigenvectors are best visualized as graphs over the mass index running from 1 to 50. These are shown in the top panel of the figure using both the original coordinates $\{z_i\}$ and the deviations from equilibrium positions coordinates $\{x_i\}$. Animations of the first four modes of vibrations are also shown in that figure.

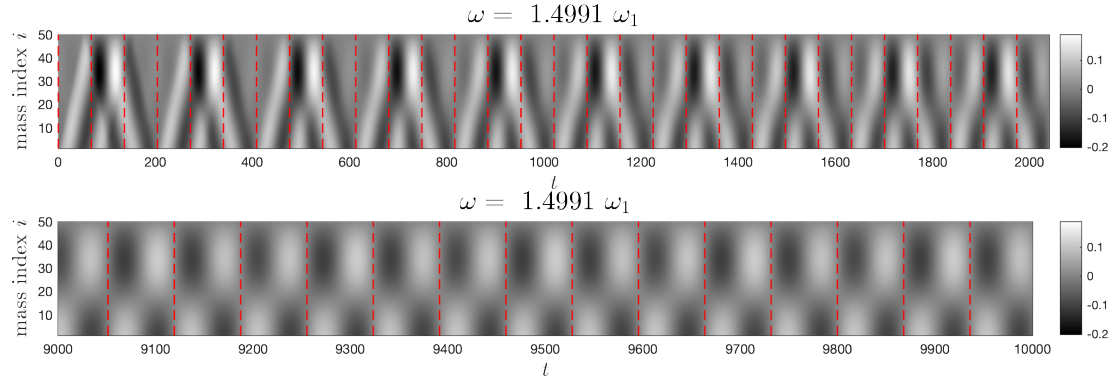
[3]In many real systems, damping tends to be proportional to natural frequencies, i.e. higher damping for higher frequencies. Thus only the first few modes of vibrations are important in such systems since the higher frequency modes are typically so heavily damped that they have little contribution to the total superposition of modes.



(a) The input $u(t)$ is a sinusoid with the same frequency ω as the first normal mode frequency ω_1. Note how the response grows with the same "shape" as the first mode depicted in Figure 9.2a. Note also how the response's period matches that of the input's period.



(b) The input is at $\omega = \omega_2 \approx 2\omega_1$, which is the 2nd normal mode's frequency. The response shape matches that of the 2nd mode depicted in Figure 9.2a, and the response frequency matches that of the input.



(c) The input frequency is at $\omega \approx 1.7\omega_1$, which is in between the 1st and 2nd mode frequencies. The response shape is a more complicated combination of the modes, and initially appears (*top panel*) not to have a well defined frequency. In steady state (*bottom panel*), the response has the same frequency as the input, but its shape is a combination of primarily the 1st and 2nd mode shapes. Note also how the maximum amplitude is about an order of magnitude smaller than the two resonant responses above.

Figure 9.3: Time responses of an $N = 50$ mass chain subject to purely harmonic excitation of frequency ω at the input $u(t)$. The pcolor plots show the deviation positions ($x_i(t)$ variables) as a function of time, with amplitude scales indicated by the color bar gray scale. Three cases are shown with ω at the 1st and 2nd normal mode frequencies, as well as a frequency in between those two modes. The amplitude of $u(t)$ is the same in all three cases for comparison of the responses' amplitudes. The gaps between the red dashed lines indicate one period of the forcing input $u(t)$.

Forced Vibrations of the Mass Chain

Now assume the mass chain is driven by an input $u(t)$ as depicted in the bottom part of Figure 9.1. Before analyzing the frequency response, we consider examples of the time response of this system when $u(t)$ is a pure harmonic excitation, i.e. when $u(t)$ is a pure sinusoid of frequency ω. In order to simulate this system we require a small amount of damping in order to have a steady state response[4] It is not difficult to see that if we place dampers c in parallel with each spring in Figure 9.1, the resulting differential equation becomes

$$\mathsf{M}\,\ddot{\boldsymbol{x}}(t) \;+\; c\,\mathcal{L}_N\,\dot{\boldsymbol{x}}(t) \;+\; k\,\mathcal{L}_N\,\boldsymbol{x}(t) \;=\; \mathsf{B}_0 u(t). \tag{9.4}$$

The fact that the damping matrix is $c\,\mathcal{L}_N$ follows from the dampers being in parallel with the springs, for which the stiffness matrix is $k\,\mathcal{L}_N$.

[4]When there is no damping, and the input is at the (normal mode) resonance frequencies of the system, the response grows unboundedly in time, i.e. there is no "steady state".

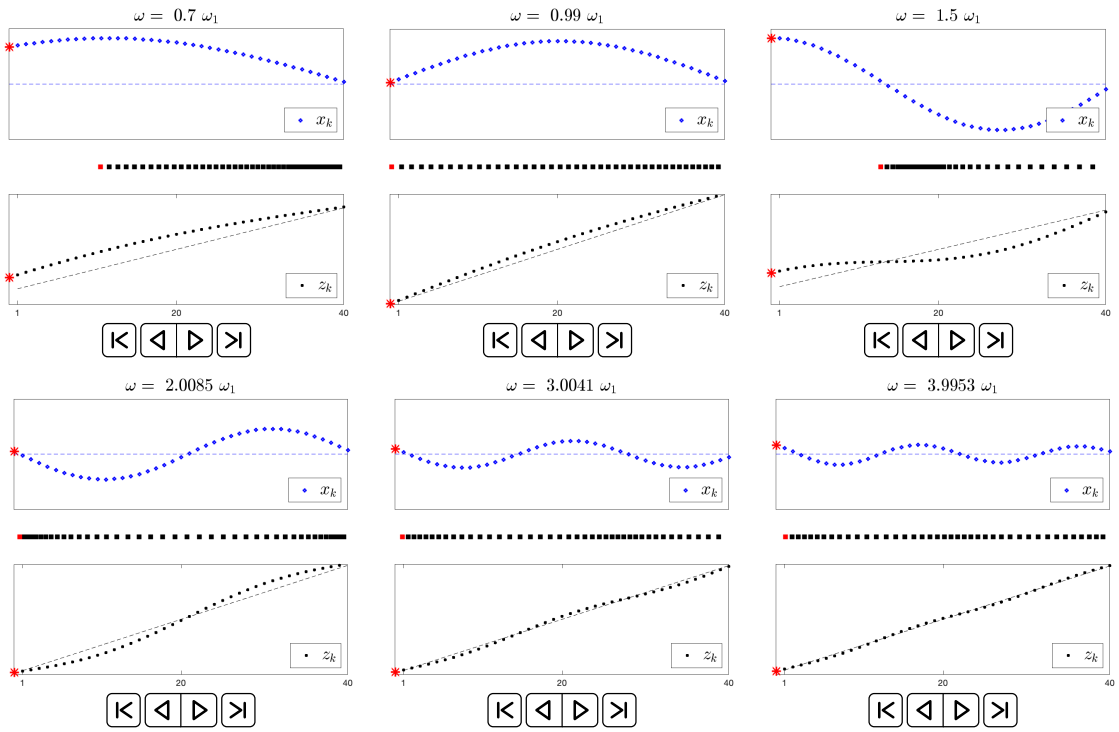


Figure 9.4: Animations of the steady state response for forcing at various frequencies. The red marker indicates the input displacement $u(t)$, and the blue and black markers indicate the masses' positions in terms of the deviations' positions $x_k(t)$, and absolute positions $z_k(t)$.

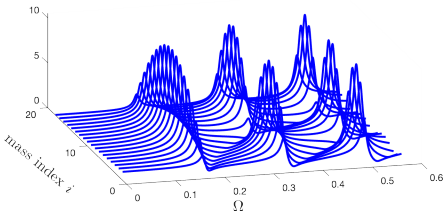
The system (9.4) is simulated with a small amount of damping c and the results are illustrated in Figure 9.3. The first and second cases correspond to the input frequency set to the 1st and 2nd normal mode frequencies. In those cases, the responses build up significantly in amplitude and the "shape" of the responses matches very closely to the corresponding normal mode shapes depicted in Figure 9.2a. Thus we have clear resonance phenomenon in those two cases. In addition, when the input frequency is close to a normal mode frequency, the system's response appears to "select" primarily that mode as the shape of the vibrations (regardless of initial conditions). This is to be expected from the analysis in Section 8.2 of the relation between normal modes and resonances.

In the third case, the excitation frequency ω is in between ω_1 and ω_2. During the initial transients, this response does not appear to have a well-defined frequency. In steady state its frequency is the same as the input's frequency as predicted by the theory, and it appears to have a more complex shape that is a combination of the first two modes. In addition, the response's amplitudes in this case are about one order of magnitude smaller than the two resonant cases. Animations of the response to forcing at various frequencies are shown in Figure 9.4.

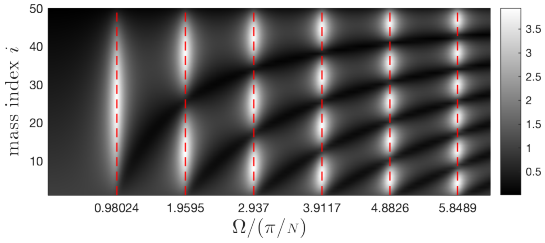
The phenomena observed in the time response experiments just described can be fully understood and quantified using the frequency response matrix of this system. Recall formula (8.5) of Theorem 8.1. The frequency response $\mathbf{H}(\omega)$ in this case is an $N \times 1$ matrix since there is a single input u, and N outputs x_i, $i = 1, \ldots, N$ representing the position of each mass in the chain. Applying formula (8.5) to (9.4) with all masses $m_i = m$, springs, and dampers assumed identical gives

$$\mathbf{H}(\omega) \;=\; \left(-\omega^2\, m\, \mathrm{I} + j\omega\; c\, \mathcal{L}_N + k\, \mathcal{L}_N\right)^{-1} \mathsf{B}_0,$$

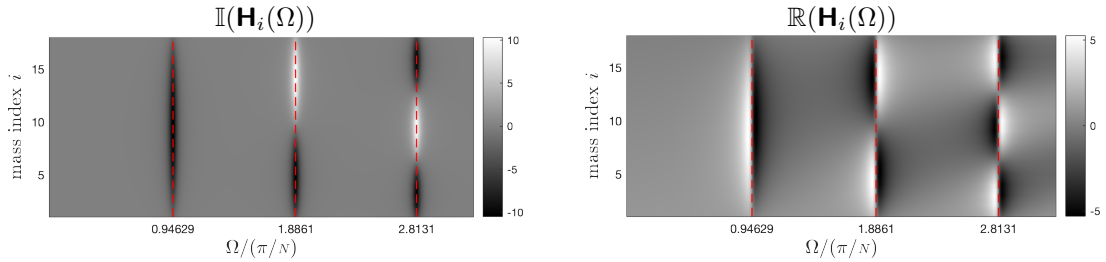
where I is the $N \times N$ identity matrix, and $\mathsf{B}_0 = k\, \boldsymbol{e}_1$ is the matrix given in (9.2). Now rewrite this



(a) The magnitude frequency responses (for $N = 18$) showing the first 3 resonances. The mode shapes shown in Figure 9.2a are clearly visible here in terms of the peaks and nodes of each mode shape.



(b) A pseudo-color plot of all the magnitude frequency responses $|\mathbf{H}_i(\Omega)|$, $i = 1, .., N$ (for $N = 50$). The frequency axis is scaled as $\Omega/(\pi/N)$, and the resonances are expected at integer values as predicted by (9.3). Actual resonance frequencies are indicated by the dashed red lines. Note again the peaks at each resonance frequency, as well as the node locations at all frequencies.



(c) Real and imaginary part of the frequency response. Note that color bar scale. The relative signs of the mode shapes near each resonance are visible in these plots. The reader should compare to the normal mode shapes shown in Figure 9.2a. For example, the 2nd mode has two peaks of opposite signs, while the 3rd mode has the two "outer" peaks of the same sign, which is opposite to the sign of the "middle" peak.

Figure 9.5: Various views of the $N \times 1$ frequency response matrix of the forced mass chain. The input is the leftmost displacement, and the outputs are the positions of each mass.

equation in terms of the normalized frequency $\Omega := \omega/\sqrt{k/m}$ and the damping ratio $\zeta := 2\frac{c}{\sqrt{km}}$

$$\mathbf{H}(\Omega) = \left(-\omega^2\, m\, \mathrm{I} + j\omega\; c\, \mathcal{L}_N + k\, \mathcal{L}_N\right)^{-1} k\, \boldsymbol{e}_1 \;=\; \left(-\frac{\omega^2}{k/m}\, \mathrm{I} + \mathcal{L}_N + j\; 2\, \frac{c}{2\sqrt{mk}}\, \frac{\omega}{\sqrt{k/m}}\, \mathcal{L}_N\right)^{-1} \boldsymbol{e}_1$$
$$= \;\left(\left(\mathcal{L}_N - \Omega^2\, \mathrm{I}\right) + j\; 2\zeta\, \Omega\; \mathcal{L}_N\right)^{-1} \boldsymbol{e}_1. \tag{9.5}$$

It is helpful to compare this expression for the single Mass-Spring-Damper case which reads

$$\mathsf{H}(\Omega) = \frac{1}{(1 - \Omega^2) + j\; 2\zeta\, \Omega}. \tag{9.6}$$

It is useful to think of the frequency response (9.5) as the *matrix version* of the single Mass-Spring-Damper frequency response (9.6) for a "networked" system of identical masses, springs and dampers. Note the appearance of the matrix $\mathcal{L}_N$, which describes the interconnection, or network structure of the system. Any other system of identical elements connected in another manner, e.g. a 2-dimensional array of masses, would have the same expression as (9.5), but with a different matrix $\mathcal{L}_N$ encoding the different network structure.

For the mass-chain system of Figure 9.1 with $N = 50$, the numerically computed frequency response matrix $\mathbf{H}(\Omega)$ is illustrated in Figure 9.5. Various ways of visualizing this frequency response are presented since the 50×1 matrix $\mathbf{H}(\Omega)$ is complex-valued. Its absolute value, real and imaginary parts are shown. Resonances occur at the normal mode frequencies, and the "shape" of the frequency response near the resonances resembles the shape of the normal modes of vibrations as expected.

Aside from showing the shape of vibrations at resonances, the frequency response plot 9.5b also shows the shape of vibrations at other frequencies. Examine the steady state vibrations in 9.3c when $\omega \approx 1.5\,\omega_1$, and compare with the frequency response in 9.5b at that same frequency. You will see

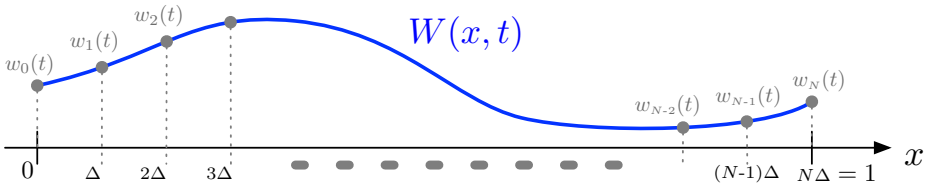


Figure 9.6: Spatial discretization of a spatially-distributed and time-varying field $W(x,t)$ over the domain $0 \le x \le 1$. The spatial domain is approximated by a grid of $N+1$ points $x_i := i\Delta$, $i = 0, 1, 2, \ldots, N$. Define the time-varying value of $W(x,t)$ evaluated at the i'th grid point to be the function of time $w_i(t) := W(i\Delta, t)$. The $N+1$ functions $\{w_i(t)\}_{i=0}^{N}$ are then collected together as the components of a vector-valued function $\boldsymbol{w}(t)$. Provided N is sufficiently large, the vector signal $\boldsymbol{w}(t)$ will be a good approximation to the field $W(x,t)$. Dynamic spatial discretization techniques amount to finding the ODEs that govern $\boldsymbol{w}(t)$ from approximations of the PDE that governs $W(x,t)$.

that the frequency response plot predicts the shape of the steady state vibrations in this case as well.

9.2 The Wave Equation

The wave equation in a finite, one-dimensional medium $0 \le x \le L$ is

$$\frac{\partial^2}{\partial t^2} W(x,t) \;=\; \nu^2 \frac{\partial^2}{\partial x^2} W(x,t), \qquad\qquad 0 \le x \le L, \tag{9.7}$$

where the boundary conditions have not been explicitly stated yet. This is a model for the vibrations of a string, or the transmission of acoustic vibrations in a one dimensional medium. The constant ν is called the wave speed since if the medium is infinite $-\infty < x < \infty$ then it easy to verify by differentiation that this equation has the two "d'Alembert" solutions

$$W_1(x,t) \;=\; W(x - \nu t, 0), \qquad W_2(x,t) \;=\; W(x + \nu t, 0). \tag{9.8}$$

Thus an initial spatial profile $W(.,0)$ is advected right as $W_1(.,)$ and left as $W_2(.,t)$, which are right and left moving with velocity ν respectively. However, when the spatial domain is finite, "reflections" occur at the boundaries and the solutions (9.8) are no longer valid. This makes the wave equation behave in more interesting ways over finite domains. In this chapter we take a spatial discretization approach to the wave equation, while an analytical treatment is presented in Chapter 10. The discretization approach will also show the analogies of the wave equation to the mass chain system studied earlier.

The wave equation (9.7) can be approximated by a finite DOF system by using a spatial finite-difference discretization as follows. Define $N+1$ functions of time $\{w_i(t)\}$, which are the values of $W(x,t)$ on a uniform spatial grid of size $N+1$ as shown in Figure 9.6 (from now on we assume $L = 1$)

$$w_i(t) \;:=\; W\left(i\Delta, t\right), \qquad\qquad i = 0, 1, \ldots, N. \qquad \Delta := 1/N.$$

The time derivatives of the functions $w_i(t)$ are the partial derivatives of $W(x,t)$ at the grid points

$$\frac{d^2}{dt^2} w_i(t) \;=\; \frac{\partial^2}{\partial t^2} W(x,t)\Big|_{x=i\Delta} \;=\; \frac{\partial^2}{\partial t^2} W(i\Delta, t). \tag{9.9}$$

This approximates the left hand side of (9.7). To approximate the right hand side, we need to approximate the 2nd *spatial derivative* $\frac{\partial^2}{\partial x^2} W(i\Delta, t)$ at the grid points. This can be done using the following finite-difference scheme (with 2nd order accuracy) implemented with the stencil

$$\begin{aligned} \frac{\partial^2}{\partial x^2} W\big(i\Delta, t\big) &\approx \frac{1}{\Delta^2}\Big(W\big((i-1)\Delta, t\big) - 2\, W\big(i\Delta, t\big) + W\big((i+1)\Delta, t\big)\Big) \\ &= \frac{1}{\Delta^2}\Big(w_{i-1}(t) - 2\, w_i(t) + w_{i+1}(t)\Big), \qquad i = 1, \ldots, N\text{-}1. \end{aligned} \tag{9.10}$$

Note that we cannot use this stencil at $i = 0$ or $i = N$ since i-1 and $N + 1$ are not point on the grid. The values of $w_0(t)$ and $w_N(t)$ will have to be determined using boundary conditions.

Before discussing boundary conditions, note that the original PDE (9.7) can now be approximated by using the expression (9.9) for the time derivatives, (9.10) for the spatial derivatives, and substituting in the wave equation (9.7). This results in the set of ODEs for $\{w_k(t)\}$ written here in vector form as

$$\frac{d^2}{dt^2}\begin{bmatrix} \vdots \\ w_{i\text{-}1}(t) \\ w_i(t) \\ w_{i+1}(t) \\ \vdots \end{bmatrix} \approx \frac{\nu^2}{\Delta^2}\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & 0 & 1 & -2 & 1 & 0 & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}\begin{bmatrix} \vdots \\ w_{i\text{-}1}(t) \\ w_i(t) \\ w_{i+1}(t) \\ \vdots \end{bmatrix}, \qquad i = 1, \ldots, N\text{-}1. \tag{9.11}$$

Note the structure of the matrix here which is "constant along diagonals", with the entries on those diagonals being exactly the coefficients of the stencil (9.10). The equation above only holds at the "interior" grid points. To find the correct equations near the boundaries, we have to incorporate boundary conditions. Next we show how that is done for two types of boundary conditions imposed on either the function W itself or on its first spatial derivative.

Boundary conditions can be imposed on either end $\bar{x} = 0, 1$ and on either the value of $W(\bar{x}, t)$ or the spatial derivative $\frac{\partial}{\partial x}W(\bar{x}, t)$ for all time as the following examples show.

$$\begin{aligned} W(0,t) &= 0 && \text{(left end is kept at zero)} \\ \tfrac{\partial}{\partial x}W(0,t) &= 0 && \text{(derivative at left end is kept at zero)} \\ W(0,t) &= u(t) && \text{(left end is actuated with } u(t)\text{)} \end{aligned}$$

An example of the actuated boundary condition is when the left end of a string is connected to an actuator that moves with the command $u(t)$. To approximate boundary conditions using the samples $\{w_i(t)\}$ of the function $W(x, t)$, the unidirectional finite-difference stencils can be used[5]

$$\begin{aligned} W(0,t) &\approx w_0(t) \\ W(1,t) &\approx w_N(t) \\ \tfrac{\partial}{\partial x}W(0,t) &\approx \tfrac{1}{\Delta}\big(w_1(t) - w_0(t)\big) && \text{(forward finite difference)} \\ \tfrac{\partial}{\partial x}W(1,t) &\approx \tfrac{1}{\Delta}\big(w_N(t) - w_{N\text{-}1}(t)\big) && \text{(backward finite difference)} \end{aligned}$$

Consider a wave equation that is "actuated" at both left and right ends according to time-varying inputs $u_\mathrm{l}(t)$ and $u_\mathrm{r}(t)$ respectively

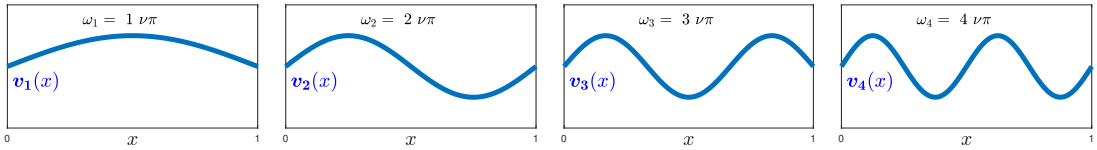
$$W(0,t) = w_0(t) = u_\mathrm{l}(t), \qquad W(1,t) = w_N(t) = u_\mathrm{r}(t).$$

The discretization of the wave equation with these boundary conditions becomes

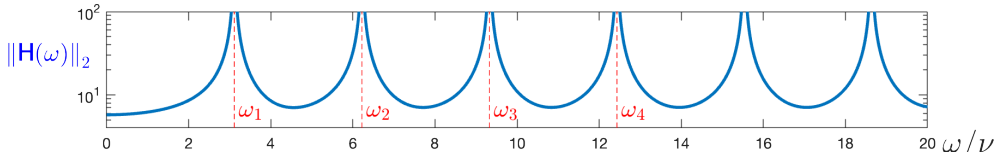
$$\begin{aligned} \ddot{w}_i(t) &= \tfrac{\nu^2}{\Delta^2}\Big(w_{i\text{-}1}(t)\text{-}2\,w_i(t) + w_{i+1}(t)\Big), && i = 1, \ldots, (N\text{-}1) && \text{(at interior points)} \\ \ddot{w}_1(t) &= \tfrac{\nu^2}{\Delta^2}\Big(w_0(t)\text{-}2\,w_1(t) + w_2(t)\Big) = \tfrac{\nu^2}{\Delta^2}\Big(\text{-}2\,w_1(t) + w_2(t)\Big) + \tfrac{\nu^2}{\Delta^2}\,u_\mathrm{l}(t) && \left(\begin{smallmatrix}\text{at left end} \\ w_0(t) = u_\mathrm{l}(t)\end{smallmatrix}\right) \\ \ddot{w}_{N\text{-}1}(t) &= \tfrac{\nu^2}{\Delta^2}\Big(w_{N\text{-}2}(t)\text{-}2\,w_{N\text{-}1}(t) + w_N(t)\Big) = \tfrac{\nu^2}{\Delta^2}\Big(w_{N\text{-}2}(t)\text{-}2\,w_{N\text{-}1}(t)\Big) + \tfrac{\nu^2}{\Delta^2}\,u_\mathrm{r}(t) && \left(\begin{smallmatrix}\text{at right end} \\ w_N(t) = u_\mathrm{r}(t)\end{smallmatrix}\right) \end{aligned}$$

Note that there is no need for the differential equations for $w_0(t)$ or $w_N(t)$ since those variables are determined by the external inputs $u_\mathrm{l}(t)$ and $u_\mathrm{r}(t)$. Thus we have $N - 1$ differential equations for the evolution of $w_i(t)$ at the $N - 1$ interior points of a grid of size $N + 1$. The differential equations just derived can be collected into a single vector ODE with the input vector $u(t) := \big(u_\mathrm{l}(t), u_\mathrm{r}(t)\big)$ as

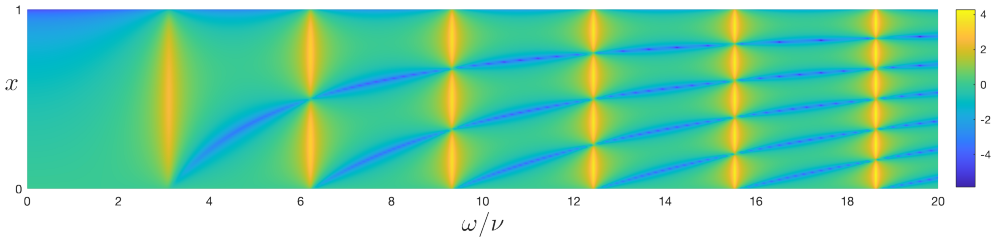
[5] Although the stencil (9.10) has 2nd order accuracy, the boundary stencils presented here have only 1st order accuracy, implying that the overall scheme will only have 1st order accuracy.



(a) The first four modes and their natural frequencies.



(b) The aggregate frequency response of the wave equation showing the regularly spaced resonances on a semi-log plot. The four natural frequencies depicted in red correspond to the first four modes shown above.



(c) The spatially-resolved frequency response as a color plot. The color at each point in this plot represents $\log\left(|\mathsf{H}_x(\omega)|\right)$, the frequency response at a particular location x when the input is oscillating at frequency ω. The shapes of each of the resonant modes is clearly seen in this plot. In particular, the nodes of oscillation are marked in blue. Note that since the actuation is at the $x = 0$ end, the plot shows asymmetry with respect to the position x.

Figure 9.7: The first few modes of the wave equation together with two different ways of visualizing its frequency response.

follows

$$\frac{d^2}{dt^2}\begin{bmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_{N\text{-}2}(t) \\ w_{N\text{-}1}(t) \end{bmatrix} = \underbrace{\frac{\nu^2}{\Delta^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix}}_{\mathcal{L}_N} \underbrace{\begin{bmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_{N\text{-}2}(t) \\ w_{N\text{-}1}(t) \end{bmatrix}}_{\boldsymbol{w}(t)} + \underbrace{\frac{\nu^2}{\Delta^2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathsf{B}_0} \underbrace{\begin{bmatrix} u_l(t) \\ u_r(t) \end{bmatrix}}_{\boldsymbol{u}(t)}$$

$$\Leftrightarrow \qquad \ddot{\boldsymbol{w}}(t) \;=\; -\,\mathsf{K}\,\boldsymbol{w}(t) \;+\; \mathsf{B}_0\,\boldsymbol{u}(t) \tag{9.12}$$

Note the reappearance of the Laplacian matrix $\mathcal{L}_N$ and the similarity of the stiffness matrix $\mathsf{K} := -\frac{\nu^2}{\Delta^2}\,\mathcal{L}_N$ above to that in the mass-chain example (9.2).

9.2.1 Normal Modes and Frequency Response

The normal modes of the wave equation are obtained from the eigenvalues and eigenvectors of $\mathcal{L}_N$. Recall the approximate expression (9.3) for the eigenvalues which then gives the natural frequencies

$$\lambda_l(\mathcal{L}_N) \approx \left(\pi\,\frac{l}{N}\right)^2 \qquad \Rightarrow \qquad \lambda_l\left(\tfrac{\nu^2}{\Delta^2}\mathcal{L}_N\right) \approx \frac{\nu^2}{\Delta^2}\left(\pi\,\frac{l}{N}\right)^2 = N^2\nu^2\left(\pi\,\frac{l}{N}\right)^2 = \nu^2\pi^2\,l^2$$
$$\Rightarrow \qquad \omega_l = \nu\pi\,l, \qquad\qquad l = 1, 2, \ldots. \tag{9.13}$$

This expression turns out to be exact for the wave equation which has an infinite number of modes corresponding to all positive integers l as will be shown in Chapter 10 (note the disappearance of

the approximation order N from the above expression). The eigenvectors of K are the same as the eigenvectors of $\mathcal{L}_N$ since the scaling with $\frac{\nu^2}{\Delta^2}$ scales only the eigenvalues and not the eigenvectors. The first four eigenvectors are depicted in Figure 9.7a together with the corresponding natural frequencies.

The input is the actuation $u(t)$ which is the left boundary condition. For the wave equation modeling the vibrations of a string, this physically corresponds to vibrating the left end of the string *vertically* according to the signal $u(t)$. The "aggregate" frequency response of (9.12) is defined as

$$\|\mathsf{H}(\omega)\|_2 \;:=\; \sqrt{|\mathsf{H}_1(\omega)|^2 + \cdots + |\mathsf{H}_{N\text{-}1}(\omega)|^2},$$

where $\mathsf{H}_i(\omega)$ is the frequency response from input u to the displacement $w_i(t)$. Figure 9.7b shows a portion of this frequency response where the regularly spaced resonance frequencies $\{\omega_l\}_{l=1}^{\infty}$ in (9.13) are clearly seen.

Figure 9.7c shows more details of the frequency response. If $\mathsf{H}_i(\omega)$ is thought of as $\mathsf{H}_x(\omega)$ where $x \in [0,1]$ is the location corresponding to the index i, then $\mathsf{H}_x(\omega)$ represents the frequency response from the input to the wave at a particular location x. This can be visualized as the color map shown in the figure where the colors corresponding to the value of $\log\big(|\mathsf{H}_x(\omega)|\big)$. The two axes are then the normalized frequency ω/ν and the spatial location x. The respective mode shapes can now be clearly seen at each resonance frequency.

9.3 Euler-Bernoulli Beam

The dynamic Euler-Bernoulli equation for an unloaded, undamped beam of length L is

$$\frac{\partial^2}{\partial t^2}\, W(x,t) \;=\; -\,\alpha \frac{\partial^4}{\partial x^4}\, W(x,t), \qquad\qquad 0 \le x \le L, \qquad\qquad \alpha := \frac{EI}{\rho}, \tag{9.14}$$

where $W(x,t)$ is the beam's deflection at location x and time t. E, I, ρ are the Young's modulus, area moment of inertia, and the linear mass density of the beam respectively. The boundary conditions on this equation are determined by how the beam is supported (e.g. clamped at one end and free at the other, pinned at both ends, etc.).

This equation can be discretized to yield a vector ODE in a similar manner to what was done with the wave equation. Assume a beam length of $L = 1$ and define the values of $W(x,t)$ at $N+1$ grid points by

$$w_k(t) \;:=\; W\left(k\Delta, t\right), \qquad\qquad k = 0, 1, \ldots, N. \qquad\qquad \Delta := 1/N,$$

As before, the time derivatives of $w_k(t)$ are the partial derivatives of $W(x,t)$ at the grid points

$$\frac{d^2}{dt^2} w_k(t) \;=\; \frac{\partial^2}{\partial t^2} W(x,t)\Big|_{x=k\Delta}. \tag{9.15}$$

The fourth order spatial derivative $\frac{\partial^4}{\partial x^4} W(x,t)$ can be approximated with the 2nd order accuracy stencil

$$\begin{aligned} \tfrac{\partial^4}{\partial x^4} W\big(k\Delta, t\big) &\approx \tfrac{1}{\Delta^4}\Big(W\big((k\text{-}2)\Delta, t\big) - 4W\big((k\text{-}1)\Delta, t\big) + 6W\big(k\Delta, t\big) - 4W\big((k\text{+}1)\Delta, t\big) + W\big((k\text{+}2)\Delta, t\big)\Big) \\ &= \tfrac{1}{\Delta^4}\Big(w_{k\text{-}2}(t) - 4\; w_{k\text{-}1}(t) + 6\; w_k(t) - 4\; w_{k+1}(t) + w_{k+2}(t)\Big), \end{aligned} \tag{9.16}$$

which gives the following form for the vector ODE at the interior points

$$\frac{d^2}{dt^2}\begin{bmatrix} \vdots \\ w_{k\text{-}1}(t) \\ w_k(t) \\ w_{k+1}(t) \\ \vdots \end{bmatrix} \approx \frac{\text{-}\alpha}{\Delta^4}\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & 0 & 1 & \text{-}4 & 6 & \text{-}4 & 1 & 0 & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}\begin{bmatrix} \vdots \\ w_{k\text{-}1}(t) \\ w_k(t) \\ w_{k+1}(t) \\ \vdots \end{bmatrix}. \tag{9.17}$$

Note the "constant along diagonals" structure of the stiffness matrix above.

9.3.1 Boundary Conditions

There are a variety of boundary conditions that can be imposed on the beam at either end. Those correspond to the beam being "pinned", "actuated", "clamped" or "free", and any combination of those at either end. Mathematically, those physical constraints amount to constraints on W and its derivatives at the respective ends. For example, the four different possible constraints at the left end $x = 0$ are

$$\begin{aligned}
0 &= W(0,t), \text{ and } 0 = \tfrac{\partial}{\partial x} W(0,t) && \text{(left end } \textit{clamped}\text{)} \\
u(t) &= W(0,t), \text{ and } 0 = \tfrac{\partial}{\partial x} W(0,t) && \text{(left end } \textit{actuated} \text{ vertically by } u(t)\text{)} \\
0 &= W(0,t), \text{ and } 0 = \tfrac{\partial^2}{\partial x^2} W(0,t) && \text{(left end } \textit{pinned}\text{)} \\
0 &= \tfrac{\partial^2}{\partial x^2} W(0,t), \text{ and } 0 = \tfrac{\partial^3}{\partial x^3} W(0,t) && \text{(left end } \textit{free}\text{)}
\end{aligned}$$

and similarly at the other end. Note that these constraints have to hold for all $t \geq 0$.

To incorporate the boundary conditions into the dynamical equations (9.17), we can use *unidirectional* finite-difference stencils at either end. Those stencils are listed below

$$\begin{aligned}
W(0,t) &= w_0(t), & W(L,t) &= w_N(t) \\
\tfrac{\partial}{\partial x} W(0,t) &\approx \tfrac{1}{\Delta} \left(\text{-}w_0(t) + w_1(t) \right) & \tfrac{\partial}{\partial x} W(L,t) &\approx \tfrac{1}{\Delta} \left(\text{-}w_{N\text{-}1}(t) + w_N(t) \right) \\
\tfrac{\partial^2}{\partial x^2} W(0,t) &\approx \tfrac{1}{\Delta^2} \left(w_0(t) \text{-} 2\, w_1(t) + w_2(t) \right) & \tfrac{\partial^2}{\partial x^2} W(L,t) &\approx \tfrac{1}{\Delta^2} \left(w_{N\text{-}2}(t) \text{-} 2\, w_{N\text{-}1}(t) + w_N(t) \right) \\
\tfrac{\partial^3}{\partial x^3} W(0,t) &\approx \tfrac{1}{\Delta^3} \left(\text{-}w_0(t) + 3\, w_1(t) \text{-} 3\, w_2(t) + w_3(t) \right) & & \\
\tfrac{\partial^3}{\partial x^3} W(L,t) &\approx \tfrac{1}{\Delta^3} \left(\text{-}w_{N\text{-}3}(t) + 3\, w_{N\text{-}2}(t) \text{-} 3\, w_{N\text{-}1}(t) + w_N(t) \right) & &
\end{aligned} \tag{9.18}$$

The stencils above are the forward and backwards finite difference stencils with first order accuracy for simplicity. Higher order accuracy stencils can also be used, but give more complicated expressions. For one dimensional problems like this one, this additional complexity is not warranted as accuracy can always be increased by increasing the grid size while keeping computational complexity within feasible bounds.

Usually two boundary conditions are needed at each end. The differential equation (9.17) has $N + 1$ functions of time. A boundary condition using any of the stencils (9.18) represents one algebraic constraint between those functions of time, which then allows for the elimination of one of the functions. Applying four boundary conditions then eliminates four of the functions leaving an unconstrained differential equation of dimension $(N + 1) - 4 = N - 3$. The next two examples illustrate this procedure.

Example 9.1. [Pinned-Pinned Beam]

If both ends are pinned, then the boundary conditions are

$$\begin{aligned}
u(t) &= W(0,t), & 0 &= W(L,t), \\
0 &= \tfrac{\partial^2}{\partial x^2} W(0,t), & 0 &= \tfrac{\partial^2}{\partial x^2} W(L,t)
\end{aligned}$$

The approximation of these conditions using the stencils (9.18) at the left and right end respectively

$$\begin{aligned}
& 0 = w_0(t), && 0 = w_N(t) \\
& 0 = w_0(t) \text{-} 2\, w_1(t) + w_2(t) && 0 = w_{N\text{-}2}(t) \text{-} 2\, w_{N\text{-}1}(t) + w_N(t) \\
\Rightarrow \quad & w_1(t) = w_2(t)/2, && w_{N\text{-}1}(t) = w_{N\text{-}2}(t)/2,
\end{aligned}$$

which allow for elimination of w_0, w_1, $w_{N\text{-}1}$ and w_N. Substituting this into the ODEs (9.17)

$$\begin{aligned}
\ddot{w}_2 &= (\text{-}\alpha/\Delta^4)\ \big[w_0 - 4\ w_1 + 6\ w_2 - 4\ w_3 + w_4\big] &&= (\text{-}\alpha/\Delta^4)\ \big[4\ w_2 - 4\ w_3 + w_4\big],\\
\ddot{w}_3 &= (\text{-}\alpha/\Delta^4)\ \big[w_1 - 4\ w_2 + 6\ w_3 - 4\ w_4 + w_5\big] &&= (\text{-}\alpha/\Delta^4)\ \big[-3.5\ w_2 + 6\ w_3 - 4\ w_4 + w_5\big],\\
\ddot{w}_{N\text{-}3} &= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}5} - 4\ w_{N\text{-}4} + 6\ w_{N\text{-}3} - 4\ w_{N\text{-}2} + w_{N\text{-}1}\big]\\
&= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}5} - 4\ w_{N\text{-}4} + 6\ w_{N\text{-}3} - 3.5\ w_{N\text{-}2}\big]\\
\ddot{w}_{N\text{-}2} &= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}4} - 4\ w_{N\text{-}3} + 6\ w_{N\text{-}2} - 4\ w_{N\text{-}1} + w_N\big]\\
&= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}4} - 4\ w_{N\text{-}3} + 4\ w_{N\text{-}2}\big]
\end{aligned}$$

The final set of ODEs for the pinned-pinned beam are then

$$\ddot{\boldsymbol{w}}(t) = \frac{d^2}{dt^2}\begin{bmatrix} w_2\\ w_3\\ w_4\\ \vdots\\ w_{N\text{-}4}\\ w_{N\text{-}3}\\ w_{N\text{-}2}\end{bmatrix} = \text{-}\alpha\ \tfrac{1}{\Delta^4}\begin{bmatrix} 4 & \text{-}4 & 1 & & & & & \\ \text{-}3.5 & 6 & \text{-}4 & 1 & & & & \\ 1 & \text{-}4 & 6 & \text{-}4 & 1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & \text{-}4 & 6 & \text{-}4 & 1 & \\ & & & 1 & \text{-}4 & 6 & \text{-}3.5 & \\ & & & & 1 & \text{-}4 & 4 & \end{bmatrix}\begin{bmatrix} w_2\\ w_3\\ w_4\\ \vdots\\ w_{N\text{-}4}\\ w_{N\text{-}3}\\ w_{N\text{-}2}\end{bmatrix} =: -\alpha\ K\ \boldsymbol{w}(t) \tag{9.19}$$

where the stiffness matrix K is defined as above, and we collected all the function $\{w_k(t)\}$ as components of the vector-valued function $\boldsymbol{w}(t)$. Note that except for the first and last two rows respectively, the matrix entries are the same as in (9.17). This is because of the use of two boundary conditions at each end to "adjust" the differential equations so that boundary conditions are enforced.

Example 9.2. [Actuated, Clamped End with other End Free]
If the left end ($x = 0$) is clamped to an actuation device (e.g. a shaker) that can oscillate vertically as $u(t)$, and the right end ($x = L$) is free, then the boundary conditions are

$$\begin{aligned}
u(t) &= W(0,t), & 0 &= \tfrac{\partial^2}{\partial x^2} W(L,t),\\
0 &= \tfrac{\partial}{\partial x} W(0,t), & 0 &= \tfrac{\partial^3}{\partial x^3} W(L,t).
\end{aligned}$$

Approximating these conditions using the stencils (9.18) at the left end gives

$$\begin{aligned}
u(t) &= W(0,t) \quad \Rightarrow \quad u(t) = w_0(t),\\
0 &= \tfrac{\partial}{\partial x} W(0,t) \quad \Rightarrow \quad 0 = \tfrac{1}{\Delta}\left(\text{-}w_0(t) + w_1(t)\right) \quad \Rightarrow \quad w_1(t) = w_0(t) = u(t).
\end{aligned} \tag{9.20}$$

Thus we can eliminate $w_0(t)$ and $w_1(t)$ and rewrite them in terms of the input $u(t)$. Similarly, at the right end the homogenous boundary conditions for the 2nd and 3rd derivatives can be used to eliminate $w_N(t)$ and $w_{N\text{-}1}(t)$ be rewriting them in terms of $w_{N\text{-}3}(t)$ and $w_{N\text{-}2}(t)$ using the corresponding stencils from (9.18)

$$\left.\begin{aligned} 0 &= w_{N\text{-}2}(t) \text{-} 2\ w_{N\text{-}1}(t) + w_N(t)\\ 0 &= \text{-}w_{N\text{-}3}(t) + 3\ w_{N\text{-}2}(t) \text{-} 3\ w_{N\text{-}1}(t) + w_N(t)\end{aligned}\right\} \quad \Rightarrow \quad \begin{bmatrix} 2 & -1\\ 3 & -1\end{bmatrix}\begin{bmatrix} w_{N\text{-}1}\\ w_N\end{bmatrix} = \begin{bmatrix} 0 & 1\\ -1 & 3\end{bmatrix}\begin{bmatrix} w_{N\text{-}3}\\ w_{N\text{-}2}\end{bmatrix}$$

$$\Rightarrow \quad \begin{bmatrix} w_{N\text{-}1}\\ w_N\end{bmatrix} = \begin{bmatrix} -1 & 2\\ -2 & 3\end{bmatrix}\begin{bmatrix} w_{N\text{-}3}\\ w_{N\text{-}2}\end{bmatrix}. \tag{9.21}$$

Now a complete set of ODEs incorporating (9.17) and the boundary conditions can be written for the remaining functions $w_2, \ldots, w_{N\text{-}2}$. The differential equations for w_2 and w_3 are

$$\begin{aligned}
\ddot{w}_2 &= (\text{-}\alpha/\Delta^4)\ \big[w_0 - 4\ w_1 + 6\ w_2 - 4\ w_3 + w_4\big] &&= (\text{-}\alpha/\Delta^4)\ \big[6\ w_2 - 4\ w_3 + w_4\ -\ 3\ u\big],\\
\ddot{w}_3 &= (\text{-}\alpha/\Delta^4)\ \big[w_1 - 4\ w_2 + 6\ w_3 - 4\ w_4 + w_5\big] &&= (\text{-}\alpha/\Delta^4)\ \big[-4\ w_2 + 6\ w_3 - 4\ w_4 + w_5\ +\ u\big],
\end{aligned}$$

where we have used (9.20) to substitute for w_0 and w_1. At the other end, we use (9.21)

$$\begin{aligned}
\ddot{w}_{N\text{-}2} &= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}4} - 4\ w_{N\text{-}3} + 6\ w_{N\text{-}2} - 4\ w_{N\text{-}1} + w_N\big]\\
&= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}4} - 4\ w_{N\text{-}3} + 6\ w_{N\text{-}2} - 4\big(\text{-}w_{N\text{-}3} + 2w_{N\text{-}2}\big) + \big(\text{-}2w_{N\text{-}3} + 3w_{N\text{-}2}\big)\big]\\
&= (\text{-}\alpha/\Delta^4)\ \big[w_{N\text{-}4} - 2\ w_{N\text{-}3} + w_{N\text{-}2}\big].
\end{aligned}$$

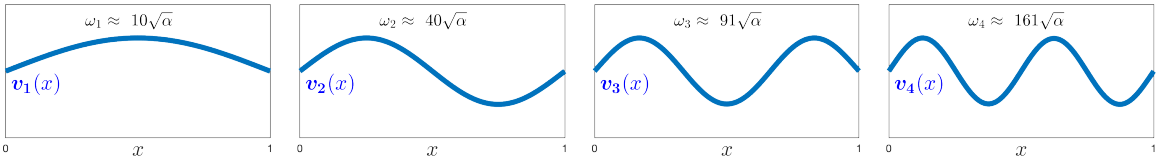


Figure 9.8: The first 4 normal modes of vibration of the Euler-Bernoulli beam with pinned-pinned boundary conditions. These were computed numerically from a discretization with $N = 100$ of the stiffness matrix K in (9.19). The eigenvectors $\boldsymbol{v}_i(x)$, $i = 1, 2, 3, 4$ are plotted against the spatial variable $0 \leq x \leq 1$ by assuming that the values $\boldsymbol{v}_i(k\Delta)$ of the function at the grid points are the components of the numerically computed eigenvector. The mode frequencies ω_i are obatined from the numerically computed eigenvalues of K. Note that the natural frequencies behave approximately like $\omega_i \sim 10\ i^2\ \sqrt{\alpha}$.

A similar calculation for $\ddot{w}_{N\text{-}3}$ gives

$$\ddot{w}_{N\text{-}3} \ = \ \left(\text{-}\alpha/\Delta^4\right) \ \left[w_{N\text{-}5} - 4\ w_{N\text{-}4} + 5\ w_{N\text{-}3} - 2\ w_{N\text{-}2}\right].$$

Now putting all of the above together we finally arrive at the required set of ODEs written in vector form

$$\begin{aligned}
\frac{d^2}{dt^2}\begin{bmatrix} w_2 \\ w_3 \\ w_4 \\ \vdots \\ w_{N\text{-}4} \\ w_{N\text{-}3} \\ w_{N\text{-}2} \end{bmatrix} &= \frac{\text{-}\alpha}{\Delta^4}\begin{bmatrix} 6 & \text{-}4 & 1 & & & & \\ \text{-}4 & 6 & \text{-}4 & 1 & & & \\ 1 & \text{-}4 & 6 & \text{-}4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & \text{-}4 & 6 & \text{-}4 & 1 \\ & & & 1 & \text{-}4 & 5 & \text{-}2 \\ & & & & 1 & \text{-}2 & 1 \end{bmatrix}\begin{bmatrix} w_2 \\ w_3 \\ w_4 \\ \vdots \\ w_{N\text{-}4} \\ w_{N\text{-}3} \\ w_{N\text{-}2} \end{bmatrix} + \frac{\text{-}\alpha}{\Delta^4}\begin{bmatrix} \text{-}3 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix} u \\
\Rightarrow \qquad \ddot{\boldsymbol{w}}(t) \ &= \ -K\ \boldsymbol{w}(t) \ + \ B_0\ u(t)
\end{aligned} \tag{9.22}$$

where the matrices K and B_0 are defined as above, and we collected all the function $\{w_k(t)\}$ as components of the vector-valued function $\boldsymbol{w}(t)$. Note that this system is of the form (7.12) with $M = I$, and $C = 0$, but with an input $u(t)$ that enters the equations through the matrix B_0. Recall that the input is the vertical position of the left end that is actuated with an external device.

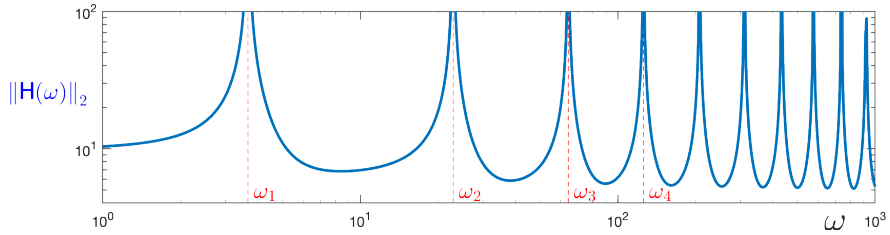
Note that except for the first and last two rows in (9.22), the matrix entries are the same as in (9.17) and in (9.19). The first and last two rows here are different from those in the pinned-pinned case (9.19) since we are enforcing a different set of boundary conditions.

The systems (9.19) and (9.22) are both of the form (7.12) (with $M = I$, $C = 0$, and an input in the latter case). We can therefore apply all the previous material on normal modes and frequency responses to analyze this system as we show next.

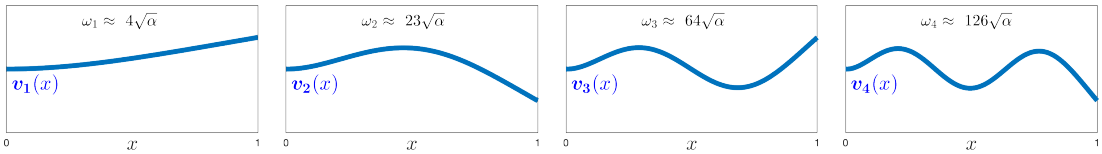
9.3.2 Normal Modes and Frequency Response

Equations (9.19) and (9.22) are a spatially discretized versions of the dynamic Euler-Bernoulli model (9.14) with two different boundary conditions assumptions. The PDE (9.14) has an infinite number of vibration modes, while the discretized models have $N - 3$ modes (the number of variables $\{w_k\}$ that remain after using the boundary condition to eliminate 4 of them). The larger the grid size N, the better of an approximation the discretized model will be. In the results shown next, $N = 100$ is used.

Example 9.3. [Normal modes of free vibrations of pinned-pinned beam] The normal modes are obtained from the eigenvalues and eigenvectors of the stiffness matrix K in (9.19) (since we've made the simplify assumption that all masses are equal to one $m_i = 1$, i.e. $M = I$, and therefore $M^{-1}K = K$). While it is possible to calculate analytical expressions for the eigenvalues and eigenvectors as will be shown in Chapter 10, the quantities displayed in Figure 9.8 are computed numerically from the discretization with $N = 100$, and are very close to the actual true quantities.



(a) The magnitude frequency response showing the first 10 of an infinite number of resonances. This frequency response is shown on a log-log plot. The frequencies ω_i scale like i^2, which makes gaps between successive frequencies progressively larger and harder to visualize on a regular plot. The first 4 modes are indicated by dashed lines on the plot, and their corresponding "mode shapes" are shown below.



(b) The first 4 modes of oscillation. The eigenvectors $\boldsymbol{v}_i$ of the matrix K in (9.17) are the normal modes. Since the vector indices represent points on the x grid, the length $N+1$ vectors are plotted here as a function of the spatial location $0 \leq x \leq 1$. These modes are the "shape" of the vibrations that occur at the respective resonances (i.e. mode $\boldsymbol{v}_i$ oscillates in time with frequency ω_i. Note the difference in mode shapes and natural frequencies between the pinned-pinned case of Figure 9.8 and the clamped-free case shown here.

Figure 9.9: The actuated, clamped-free Euler-Bernoulli beam frequency response showing the first 10 resonances, as well as the "mode shapes" of the first 4 normal modes. The length of the beam is assumed here as $L = 1$, and the coefficient $\alpha = 1$.

Example 9.4. [Modes and frequency response of actuated, clamed-free beam] To find the frequency response, use formula (8.5) with the matrices K and B_0 from (9.22) (note that in this case $M = I$ and $C = 0$)

$$\mathsf{H}(\omega) \;=\; \left(-\omega^2 I + K\right)^{-1} B_0.$$

For each ω, $\mathsf{H}(\omega)$ is an $(N-3)\times 1$ matrix, so to plot the frequency response, we plot its "norm" defined as the square root of the sum of squares of the components

$$\|\mathsf{H}(\omega)\|_2 \;:=\; \left(|\mathsf{H}_3(\omega)|^2 + \cdots + |\mathsf{H}_{N-2}(\omega)|^2\right)^{1/2}.$$

A plot of this magnitude frequency response is shown in Figure 9.9a. Many resonances are clearly seen. For the original PDE model, there's actually an infinite number of resonances of progressively increasing frequencies. The plot only shows the first 10, which should be well approximated with $N = 100$.

Recall that a system excited with a frequency near any one its normal modes ω_i, will oscillate with a "shape" determined by the i'th eigenvector $\boldsymbol{v}_i$ of the matrix $M^{-1}K$ (which is just K in this case). The first 4 mode shapes are shown in Figure 9.9b by simply plotting the components of each eigenvector $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$. Recall that the components of the vector $\boldsymbol{w}(t)$ are approximations to the function $W(x,t)$ at the grid points. Therefore the plots of the eigenvectors can be considered as approximations to functions over the domain $0 \leq x \leq 1$. These "mode shapes" will be the shape of the oscillation when the beam is excited at frequency ω_i. More precisely, we can say that for the original PDE

$$u(t) = \alpha \cos\left(\omega t + \phi\right), \text{ and } \omega \approx \omega_i \qquad \Rightarrow \qquad W(x,t) \;\approx\; \beta\; v_i(x)\; \cos(\omega t + \theta),$$

where $v_i(x)$ is the eigenvector of K (considered as a function over $[0,1]$). Note the separation of $W(x,t)$ into a product of a *spatial mode shape* $v_i(x)$ and a *temporal oscillation* $\cos(\omega t + \theta)$. This only occurs near resonances. Away from resonances, the spatial shape of oscillations will usually be "a mixture" of many normal modes.

Chapter 10

Vibrations in Continua

Vibrations of strings, membranes, beams and plates are described by Partial Differential Equations (PDEs). The mechanical degrees of freedom are represented by displacements which are functions of both time and a spatial variable. There is an infinite, continuum number of mechanical degrees of freedom, and such systems are therefore referred to as continuum systems. The mathematical tools used for analysis of continuum systems are conceptually similar to those of N-DOF, except that instead of matrices, the objects to analyze are spatial differential operators. Those operators typically have an infinite number of eigenvalues with associated eigenfunctions that play the same role eigenvectors do in n-DOF systems.

Introduction and the Big Picture

The main theme of this introduction is to highlight the similarities between the analysis methods of continuum vibrations and those of n-DOF systems. The key idea is an abstraction where spatial differential operators are thought of in a similar manner to matrices, i.e. as linear operators. Eigenvectors of matrices can then be abstracted to more general objects which are the eigen*functions* of linear differential operators. To explain how this works, consider the following PDEs treated in this chapter in comparison with the standard form for (an undamped) n-DOF system

$$\text{vector ODE for } n\text{-DOF system with } M=I \;:\quad \frac{d^2}{dt^2}\,\boldsymbol{w}(t) \;=-K\;\boldsymbol{w}(t) \tag{10.1}$$

$$\text{1D wave equation for strings} \;:\quad \frac{\partial^2}{\partial t^2}\,w(x,t) \;=\; \nu^2\frac{\partial^2}{\partial x^2}\,w(x,t) \tag{10.2}$$

$$\text{2D wave equation for membranes} \;:\quad \frac{\partial^2}{\partial t^2}\,w(x,y,t) \;=\; \nu^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\,w(x,y,t) \tag{10.3}$$

$$\text{Euler-Bernoulli equation for beams} \;:\quad \frac{\partial^2}{\partial t^2}\,w(x,t) \;=-\alpha\,\frac{\partial^4}{\partial x^4}\,w(x,t) \tag{10.4}$$

$$\text{2D Euler-Bernoulli equation for plates} \;:\quad \frac{\partial^2}{\partial t^2}\,w(x,y,t) \;=-\alpha\left(\frac{\partial^4}{\partial x^4}+2\,\frac{\partial^2}{\partial x^2}\frac{\partial^2}{\partial y^2}+\frac{\partial^4}{\partial y^4}\right)w(x,y,t) \tag{10.5}$$

The four PDEs are all of the form

$$\frac{\partial^2}{\partial t^2}\,w(.,t) \;=\; -\,\mathcal{K}\,w(.,t), \tag{10.6}$$

where $\mathcal{K}$ is some *spatial* PDE operator like the ones in the four examples above. This form corresponds to placing the second-order time derivative on one side of the equation, and all the spatial derivatives on the other side. Notice the suppression of the spatial coordinate in the notation $w(.,t)$. This means that we consider $\mathcal{K}\,w(.,t)$ to be an operation on the whole of the function $w(.,t)$ at each

time. This slight abstraction is a very important "mental shift" which is developed carefully in the next section.

For the n-DOF system (10.1), the normal modes of vibration are obtained from the eigenvalues/vectors of the stiffness matrix K. In continuum systems, normal modes of vibrations are obtained from the eigenvalues and *eigenfunctions* of the "stiffness" operation $\mathcal{K}$. Abstractly a function $v(x)$ of a spatial variable is an eigenfunction of a differential operation $\mathcal{K}$ if for some (possibly complex) number λ

$$\mathcal{K}v(x) \;=\; \lambda\, v(x).$$

Finding such eigenvalues λ and corresponding eigenfunctions $v(x)$ is called an *eigenvalue problem*. For example, for the 1D and 2D wave equations respectively, the eigenvalue problems are to find numbers λ and functions v that satisfy

$$-\nu^2 \frac{\partial^2}{\partial x^2} v(x) \;=\; \lambda\, v(x), \tag{10.7}$$

$$-\nu^2 \frac{\partial^2}{\partial x^2} v(x,y) \;-\; \nu^2 \frac{\partial^2}{\partial y^2} v(x,y) \;=\; \lambda\, v(x,y). \tag{10.8}$$

Note the absence of the time variable in these problems. They are differential equations that involve only the spatial variables. Therefore (10.7) is actually an ODE in the single spatial variable x, while (10.8) is a PDE in the two spatial variables x, y.

Once an eigenfunction is found, it is easy to show that if a system of the form (10.6) starts from an initial condition which is an eigenfunction of the operation $\mathcal{K}$, it will oscillate in time as a multiple of that spatial eigenfunction. For example, for the 1D wave equation, if the initial condition is a multiple of an eigenfunction satisfying (10.7)

$$w(x,0) \;=\; \alpha\, v(x), \qquad \frac{\partial}{\partial t} w(x,0) \;=\; \beta\, v(x),$$

then the solution for all time will be of the form

$$w(x,t) \;=\; v(x) \;\; \mathsf{a} \;\cos(\omega t + \theta), \qquad\qquad \omega = \sqrt{\lambda},$$

where the temporal oscillation frequency ω is the square root of the eigenvalue λ *associated with* the eigenfunction $v(x)$. The amplitude a and phase θ of the temporal oscillation is determined by the details of the initial conditions (i.e. α and β). The fact that this is a solution of the PDE is easily shown by direct verification. If $v(x)$ satisfies (10.7), then

$$\nu^2 \frac{\partial^2}{\partial x^2} v(x) \;=\; \lambda\, v(x) \qquad \Rightarrow \qquad \begin{array}{r} \frac{\partial^2}{\partial t^2} \left(v(x) \;\; \mathsf{a} \cos(\omega t + \theta) \right) \;=\; -\omega^2 \left(v(x) \;\; \mathsf{a} \cos(\omega t + \theta) \right) \\ \nu^2 \frac{\partial^2}{\partial x^2} \left(v(x) \;\; \mathsf{a} \cos(\omega t + \theta) \right) \;=\; -\lambda \;\; \left(v(x) \;\; \mathsf{a} \cos(\omega t + \theta) \right) \end{array}$$

10.1 Eigenvalues and Eigenfunctions of Differential Operators

Differentiation of a function is an operation on functions. Going forward, we will be very careful with notation so as to emphasize this point. Up to now, we have been using notation like $v(x)$ to refer to a function of an independent variable x. From now on, we adopt the notation that v or $v(.)$ refers to the whole function as an object, while $v(x)$ is the *single number which is the value of the function v at the point x*. Thus $v(x)$ is a number, while v or $v(.)$ is the whole function. This notational distinction is important to avoid confusion going forward.

Differentiation is an *operation* that takes functions to functions. We want to think of this in analogy with a matrix being an operation that takes vectors to vectors. To emphasize this analogy, the operation of differentiation with respect to an independent variable x will be denoted as follows

$$\left(\boldsymbol{\partial}_{\mathbf{x}} v\right)(x) \;:=\; \frac{\partial v}{\partial x}(x).$$

Parse through this notation carefully. v is a function, and the operator $\boldsymbol{\partial_x}$ transforms the function v to the function $\boldsymbol{\partial_x} v$. This transformation is given by the recipe above, i.e. the value $(\boldsymbol{\partial_x} v)(x)$ of the function $\boldsymbol{\partial_x} v$ at the point x is given by the partial derivative of v with respect to x. While at first this notation may seem unnecessarily pedantic, it is important for understanding normal mode analysis of PDEs in analogy with that for ODEs.

Using the notation above, we can express the fact that differentiation is a linear operation on functions, i.e. given any two functions v_1 and v_2, and any linear combination $\alpha v_1 + \beta v_2$ of these functions

$$\begin{aligned} & & \boldsymbol{\partial_x}(\alpha v_1 + \beta v_2) &= \alpha\, \boldsymbol{\partial_x} v_1 + \beta\, \boldsymbol{\partial_x} v_2 && \text{(equality as functions)} \\ \Leftrightarrow \quad & \text{for all } x, & \big(\boldsymbol{\partial_x}(\alpha v_1 + \beta v_2)\big)(x) &= \alpha\, \big(\boldsymbol{\partial_x} v_1\big)(x) + \beta\, \big(\boldsymbol{\partial_x} v_2\big)(x) && \text{(equality at each point } x) \\ \Leftrightarrow \quad & \text{for all } x, & \frac{\partial}{\partial x}\big(\alpha v_1(x) + \beta v_2(x)\big) &= \alpha\, \frac{\partial v_1}{\partial x}(x) \;+\; \beta\, \frac{\partial v_2}{\partial x}(x) && \text{(previous notation)} \end{aligned}$$

Thus differentiation is a linear operation in the same way that matrices are linear operations on vectors

$$K\big(\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2\big) \;=\; \alpha\, K\boldsymbol{v}_1 + \beta\, K\boldsymbol{v}_2,$$

where α, β are scalars, $\boldsymbol{v}_1, \boldsymbol{v}_2$ are n-vectors, and K is an $n \times n$ matrix. This linearity property is the key to generalizing the concepts of eigenvalues and eigenvectors from matrices to differential operators.

Let's take this analogy a little further. Given an $n \times n$ matrix K, we view it as an operation on vectors; given an n-vector $\boldsymbol{v}$, the matrix-vector product $K\boldsymbol{v}$ produces another n-vector. Repeating this operation say m times

$$\underbrace{K \;\cdots\; K}_{m \text{ times}} \boldsymbol{v} \;=\; K^m \boldsymbol{v},$$

is equivalent to acting with the matrix K^m. Similarly, differentiating a function $v(.)$ m times is expressed as

$$\frac{\partial^m v}{\partial x^m}(x) \;=\; \underbrace{\frac{\partial}{\partial x} \;\cdots\; \frac{\partial}{\partial x}}_{m \text{ times}} v(x) \qquad \Leftrightarrow \qquad \big(\boldsymbol{\partial_x^m} v\big)(x) \;=\; \frac{\partial^m v}{\partial x^m}(x).$$

Similarly, mixed partial derivatives can be written as "compositions" of differentiation operations

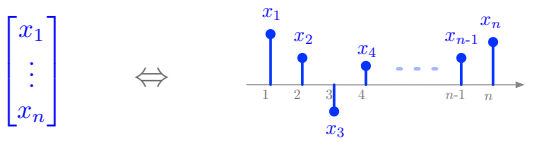
$$\big(\boldsymbol{\partial_x}\boldsymbol{\partial_y} v\big)(x,y) \;=\; \frac{\partial^2 v}{\partial x \partial y}(x,y).$$

Here v is a function of two independent variables x, y (i.e. function over a 2D domain), and the operation $\boldsymbol{\partial_x}\boldsymbol{\partial_y} v$ produces another function of the two independent variables x, y. This is an example of the composition of two linear operators that *commute* since

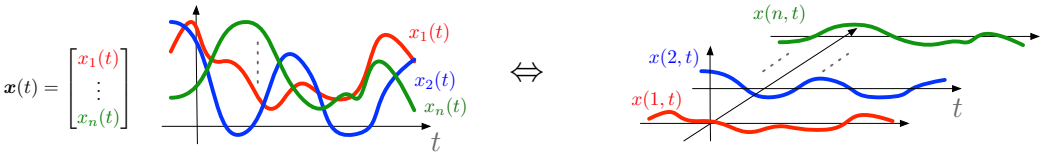
$$\text{for any function } v, \quad \frac{\partial^2 v}{\partial x \partial y}(x,y) = \frac{\partial^2 v}{\partial y \partial x}(x,y) \qquad \Leftrightarrow \qquad \boldsymbol{\partial_x}\boldsymbol{\partial_y} = \boldsymbol{\partial_y}\boldsymbol{\partial_x}.$$

The first statement above follows from multlivariable calculus; the two successive operations of partial differentiation with respect to one of the variables followed by differentiation with respect to the other can be performed in either order. This is succinctly stated by $\boldsymbol{\partial_x}\boldsymbol{\partial_y} = \boldsymbol{\partial_y}\boldsymbol{\partial_x}$, i.e. the two operators $\boldsymbol{\partial_x}$ and $\boldsymbol{\partial_y}$ commute. Note that in general linear operators do not commute in the same manner that not all matrices commute. However, when two operators do commute, then so do all their powers. For example, for the mixed partial derivatives this implies

$$\boldsymbol{\partial_x^2}\boldsymbol{\partial_y^2} \;=\; \boldsymbol{\partial_x} \underbrace{\boldsymbol{\partial_x}\boldsymbol{\partial_y}}_{\text{switch}} \boldsymbol{\partial_y} \;=\; \boldsymbol{\partial_x}\boldsymbol{\partial_y} \underbrace{\boldsymbol{\partial_x}\boldsymbol{\partial_y}}_{\text{switch}} \;=\; \underbrace{\boldsymbol{\partial_x}\boldsymbol{\partial_y}}_{\text{switch}} \boldsymbol{\partial_y}\boldsymbol{\partial_x} \;=\; \boldsymbol{\partial_y} \underbrace{\boldsymbol{\partial_x}\boldsymbol{\partial_y}}_{\text{switch}} \boldsymbol{\partial_x} \;=\; \boldsymbol{\partial_y}\boldsymbol{\partial_y}\boldsymbol{\partial_x}\boldsymbol{\partial_x} \;=\; \boldsymbol{\partial_y^2}\boldsymbol{\partial_x^2}$$



(a) A vector as a function



(b) A vector as a function

Figure 10.1: Vectors as functions and functions as vectors

Similar arguments show that all higher powers of those two operators commute. A nice application of this fact allows us to view the two-dimensional PDE for plate vibrations as a generalization of the one-dimensional Euler-Bernoulli beam equation as follows. First, note that the commutativity shown above implies the following

$$\begin{aligned} \left(\partial_{\mathbf{x}}^2 + \partial_{\mathbf{y}}^2\right)^2 &= \left(\partial_{\mathbf{x}}^2 + \partial_{\mathbf{y}}^2\right)\left(\partial_{\mathbf{x}}^2 + \partial_{\mathbf{y}}^2\right) = \partial_{\mathbf{x}}^2\partial_{\mathbf{x}}^2 + \underbrace{\partial_{\mathbf{x}}^2\partial_{\mathbf{y}}^2 + \partial_{\mathbf{y}}^2\partial_{\mathbf{x}}^2}_{\text{(equal because } \partial_{\mathbf{x}}^2\partial_{\mathbf{y}}^2 = \partial_{\mathbf{y}}^2\partial_{\mathbf{x}}^2)} + \partial_{\mathbf{y}}^2\partial_{\mathbf{y}}^2 \\ &= \partial_{\mathbf{x}}^4 + 2\,\partial_{\mathbf{x}}^2\partial_{\mathbf{y}}^2 + \partial_{\mathbf{y}}^4 \end{aligned} \tag{10.9}$$

This last differential operator is the one appearing on the right hand side of PDE (10.5) for plate vibrations. In one dimension, second derivative operator $\partial_{\mathbf{x}}^2$ is called the *Laplacian*. In two dimensions, the Laplacian is $\partial_{\mathbf{x}}^2 + \partial_{\mathbf{y}}^2$. Thus in the Euler-Bernoulli beam equation (10.4), the right hand side is given by the square $\left(\partial_{\mathbf{x}}^2\right)^2 = \partial_{\mathbf{x}}^4$ of the one-dimensional Laplacian, while the right hand side of the plate vibrations PDE (10.5) is also given by the square of the Laplacian, but in this case we use the two-dimensional Laplacian (10.9). This is why we call the plate vibrations PDE the "2D Euler-Bernoulli equation".

Functions as Continuum versions of Vectors

The final ingredient in this new way of thinking about PDEs is to write fields like $w(x,t)$ so as to view the dependence on x differently from the dependence on t.

$$\boldsymbol{w}(t) \;:=\; w(.,t), \qquad \big(\boldsymbol{w}(t)\big)(x) \;:=\; w(x,t)$$

We now rewrite the equations (10.2)-(10.5) in this operator notation

$$\begin{aligned} \tfrac{d^2}{dt^2}\,\boldsymbol{w}(t) &= -K\,\boldsymbol{w}(t) && \text{(vector ODE for } n\text{-DOF system with } M = I) \\ \partial_{\mathbf{t}}^2\boldsymbol{w}(t) &= \nu^2\partial_{\mathbf{x}}^2\,\boldsymbol{w}(t) && \text{(1D wave equation for strings)} \\ \partial_{\mathbf{t}}^2\boldsymbol{w}(t) &= \nu^2\left(\partial_{\mathbf{x}}^2 + \partial_{\mathbf{y}}^2\right)\,\boldsymbol{w}(t) && \text{(2D wave equation for membranes)} \\ \partial_{\mathbf{t}}^2\boldsymbol{w}(t) &= -\alpha\,\partial_{\mathbf{x}}^4\,\boldsymbol{w}(t) && \text{(Euler-Bernoulli equation for beams)} \\ \partial_{\mathbf{t}}^2\boldsymbol{w}(t) &= -\alpha\left(\partial_{\mathbf{x}}^2 + \partial_{\mathbf{y}}^2\right)^2\boldsymbol{w}(t) && \text{(2D Euler-Bernoulli equation for plates)} \end{aligned}$$

Continuum Vibration Modes are Eigenfunctions

In n-DOF systems, the modes of vibrations are given by the eigenvalues and eigenvectors of the matrix $M^{-1}K$, i.e. the matrix that appears on the right hand side of the vector differential equation.

Similar statements will apply to continuum systems, but we will consider the eigenvalue problem for the spatial differential operators that appear on the right hand side of the PDE.

Consider any PDE written in the form

$$\boldsymbol{\partial}_{\mathbf{t}}^2 \boldsymbol{w}(t) \;=\; -\mathcal{K}\,\boldsymbol{w}(t) \qquad \Leftrightarrow \qquad \boldsymbol{\partial}_{\mathbf{t}}^2 w(x,t) \;=\; -\left(\mathcal{K} w(.,t)\right)(x,t), \tag{10.10}$$

where $\mathcal{K}$ is a *spatial* differential operator. A function $v(.)$ is called an eigenfunction of a linear operator if for some (possibly complex) number λ

$$\mathcal{K} v \;=\; \lambda\, v \qquad \Leftrightarrow \qquad \left(\mathcal{K} v\right)(x) \;=\; \lambda\, v(x), \tag{10.11}$$

where λ is called the eigenvalue corresponding to this eigenfunction. Just like n-DOF where initial conditions which are eigenvectors produce solutions which are pure modes, the situation is the same here. Let the initial condition of (10.10) be an eigenfunction like (10.11), it follows that the solution must be

$$\begin{array}{r} w(x,0) \;=\; \alpha\; v(x) \\ \partial_t w(x,0) \;=\; \beta\; v(x) \end{array} \qquad \Rightarrow \qquad w(x,t) \;=\; v(x)\;\; \mathsf{w}\cos(\omega t + \theta), \qquad\qquad \omega = \sqrt{\lambda}.$$

Note the "separation form" of the solution as a *product* of a function of x and a function of t. The solution oscillates in time with a frequency $\omega = \sqrt{\lambda}$ determined by the eigenvalue, but the "spatial shape" of the solution is always a multiple of eigenfunction $v(.)$. The amplitude w and phase θ of the temporal oscillations are determined by the initial condition parameters α and β.

For example, consider the 1D Laplacian $\boldsymbol{\partial}_{\mathbf{x}}^2$. A function $v(.)$ is an an eigenfunction of this operator if for some (possibly complex) number λ

$$\left(\boldsymbol{\partial}_{\mathbf{x}}^2 v\right)(x) \;=\; \lambda\, v(x) \qquad \Leftrightarrow \qquad \frac{d^2}{dx^2} v(x) \;=\; \lambda\, v(x). \tag{10.12}$$

In this case λ is called the eigenvalue corresponding to this eigenfunction.

To see what eigenfunctions imply for the dynamics of PDEs, consider the 1D wave equation with an initial condition which is an eigenfunction (10.12) of $\boldsymbol{\partial}_{\mathbf{x}}^2$

$$\boldsymbol{\partial}_{\mathbf{t}}^2 w(x,t) \;=$$

We will briefly describe this more abstract approach using the Euler-Bernoulli equation. Recall the dynamics (9.14)

$$\frac{\partial^2}{\partial t^2}\, W(x,t) \;=\; -\,\bar{s}\boldsymbol{\partial}_{\mathbf{x}}^4\, W(x,t), \qquad\qquad 0 \le x \le L. \tag{10.13}$$

The spatial derivative can be thought of as a linear operator $\mathcal{K} := \bar{s}\boldsymbol{\partial}_{\mathbf{x}}^4$ which acts on functions defined over the spatial domain $0 \le x \le L$. A function $v(x)$ on that domain is an *eigenfunction* of $\mathcal{K}$ with eigenvalue λ if

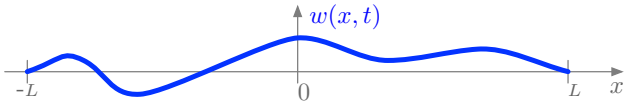
$$\mathcal{K}\, v \;=\; \lambda\, v \qquad\qquad \Leftrightarrow \qquad\qquad \bar{s}\boldsymbol{\partial}_{\mathbf{x}}^4\, v(x) \;=\; \lambda\, v(x), \qquad 0 \le x \le L. \tag{10.14}$$

Notice that the statement $\mathcal{K}v = \lambda v$ is the same as if $\mathcal{K}$ were a matrix, and v were a vector. The new feature now is that $\mathcal{K} = -\bar{s}\boldsymbol{\partial}_{\mathbf{x}}^4$ is a linear operator (that operates on *functions* rather than vectors), and v is a function.

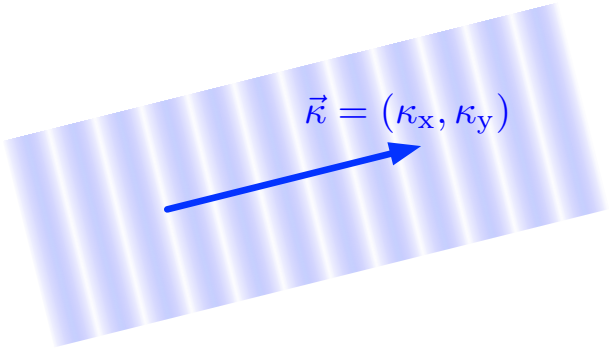
The statement on the right in (10.26) is just the statement $\mathcal{K}v = \lambda v$ in detail. This statement is actually a differential equation. To see this, assume again a beam clamped at $x = 0$ (with no actuation in this case), and free at $x = L$. The (purely spatial) differential equation for determining the eigenfunctions becomes

$$\bar{s}\; v^{(4)}(x) \;-\; \lambda\, v(x) \;=\; 0, \qquad\qquad \begin{array}{rr} v(0) \;=\; 0, & v^{(2)}(L) \;=\; 0 \\ v^{(1)}(0) \;=\; 0, & v^{(3)}(L) \;=\; 0 \end{array}. \tag{10.15}$$

This is a 4'th order differential equation for $v(x)$ with a "free parameter" λ. It is called a "two point boundary value problem" since the boundary conditions are given at two points, namely $x = 0$ and $x = L$. Note that time does not play a role in this equation, the same way that when solving for normal modes, we solve an eigenvalue problem for a matrix $K\boldsymbol{v} = \lambda \boldsymbol{v}$, and this eigenvalue problem does not involve time.



(a) Wave motion of a string are modeled using a function of spatial location x and time t. $w(x,t)$ describes the time-varying transverse displacement of the string. The fact that the string is pinned at both ends is expressed by the boundary conditions $w(\pm L, t) = 0$ for all $t \geq 0$.



(b) In 2D, an acoustic plane wave is a pressure fluctuation of the form $p(x,y,t) = \psi\left((\kappa_x x + \kappa_y y) - \nu t\right)$. It is a wave traveling in the direction of the wavenumber vector $\vec{\kappa} = (\kappa_x, \kappa_y)$ with speed ν. A plane wave in 2D (or 3D) can be represented as a one-dimensional wave in the spatial coordinate $z := \kappa_x x + \kappa_y y$, while being constant in an orthogonal coordinate.

Figure 10.2: The one-dimensional wave equation can describe either (a) the vibrations of a string, or (b) a plane wave. Here an acoustic wave of pressure fluctuations is depicted, but the same analysis is applicable electromagnetic or elastic plane waves.

10.2 String Vibrations and Acoustics: The Wave Equation

The one-dimensional wave equation for a quantity $w(x,t)$ that varies in both time t and a one-dimensional medium with spatial coordinate x is

$$\boldsymbol{\partial_t^2} w(x,t) \;=\; \nu^2(x)\; \boldsymbol{\partial_x^2} w(x,t), \tag{10.16}$$

where the coefficient function $\nu(x)$ describes certain material properties of the medium. When ν is a constant in x, it corresponds to the *wave speed* as described in Section 9.2. The wave equation is encountered in a wide variety of physical phenomena such as string vibrations, and "plane waves" in elasticity, acoustics and electromagnetics. Appendices 10.A.1 and 10.A.2 detail the derivations of the respective wave equations for string vibrations and acoustic waves respectively. We now summarize how each of the two phenomena are modeled with an equation of the form (10.16).

The dynamics of a string of length $2L$ pinned at both ends $x = \pm L$ illustrated in Figure 10.2a are given by the one-dimensional wave equation with the following boundary conditions

$$\boldsymbol{\partial_t^2}\; w(x,t) \;=\; \nu^2(x)\; \boldsymbol{\partial_x^2} w(x,t), \qquad w(\text{-}L,t) = 0, \qquad w(L,t) = 0. \tag{10.17}$$

The function $\nu(x)$ depends on the material properties of the string, which in analogy with $\sqrt{k/m}$ is the square root of ratio of local tension over local (linear) density of the string. For a string of uniform density and tension, ν is independent of x, which is the case we will consider in this chapter.

In a non-moving gas, small pressure $p(x,t)$ and velocity $v(x,t)$ fluctuations forming a plane wave as shown in Figure 10.2b can be described by the pair of one-dimensional, coupled PDEs (10.53) which we recall here

$$\begin{aligned} \boldsymbol{\partial_t} v(x,t) \;&=\; \tfrac{1}{\bar\rho}\; \boldsymbol{\partial_x} p(x,t), \\ \boldsymbol{\partial_t} p(x,t) \;&=\; \gamma \bar p\; \boldsymbol{\partial_x} v(x,t). \end{aligned} \tag{10.18}$$

In these equations $v(x,t)$ and $p(x,t)$ are air particles' velocity and pressure fluctuations respectively, and $\bar\rho$ and $\bar p$ are the nominal constant density and pressure about which the fluctuations are occurring.

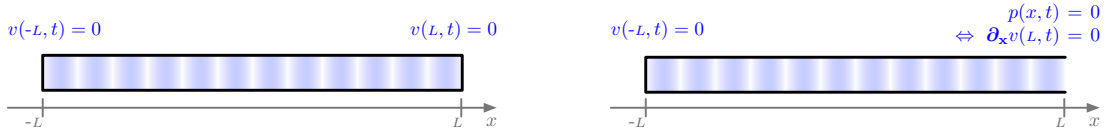


Figure 10.3: Acoustic boundary conditions in a high aspect ratio tube. At closed ends, the no-slip condition results in zero velocity boundary conditions. At open ends, more complicated "radiative" boundary conditions are typically needed. However, for high aspect ratio tubes, a reasonable approximation is to assume zero pressure fluctuations at the open end. Due the structure of the equations, a zero pressure fluctuation boundary condition can equivalently be represented as zero velocity-derivative condition.

$\gamma := 1 + \frac{R}{c_v}$ is the heat capacity ratio, which is a dimensionless number equal to 1.4 for diatomic gases like air. Although the equations (10.18) at first look do not resemble the wave equation (10.16), the pressure and velocity fields individually obey wave equations. To see that, differentiate the first equation with respect to t and the second one with respect to x

$$\begin{aligned}
&\partial_t v(x,t) = \tfrac{1}{\bar\rho}\,\partial_x p(x,t) && \Rightarrow && \partial_t^2 v(x,t) = \tfrac{1}{\bar\rho}\,\partial_t\partial_x p(x,t) \\
&\partial_t p(x,t) = \gamma\bar p\,\partial_x v(x,t) && \Rightarrow && \partial_x\partial_t p(x,t) = \gamma\bar p\,\partial_x^2 v(x,t) \\
& && \Rightarrow && \partial_t^2 v(x,t) = \tfrac{\gamma\bar p}{\bar\rho}\,\partial_x^2 v(x,t) && (\text{since } \partial_x\partial_t = \partial_t\partial_x)
\end{aligned}$$

This is a wave equation with wave speed $\nu = \sqrt{\gamma\bar p/\bar\rho}$. Alternatively, if we differentiate the first equation in (10.18) with respect to x and the second one with respect to t

$$\begin{aligned}
&\partial_t v(x,t) = \tfrac{1}{\bar\rho}\,\partial_x p(x,t) && \Rightarrow && \partial_x\partial_t v(x,t) = \tfrac{1}{\bar\rho}\,\partial_x^2 p(x,t) \\
&\partial_t p(x,t) = \gamma\bar p\,\partial_x v(x,t) && \Rightarrow && \partial_t^2 p(x,t) = \gamma\bar p\,\partial_t\partial_x v(x,t) \\
& && \Rightarrow && \partial_t^2 p(x,t) = \tfrac{\gamma\bar p}{\bar\rho}\,\partial_x^2 p(x,t). && (\text{since } \partial_x\partial_t = \partial_t\partial_x)
\end{aligned}$$

Thus pressure also obeys a wave equation with the same wave speed as velocity wave equation[1].

The boundary conditions for pressure and velocity depend on the geometry. For example, to model acoustic waves in narrow tubes as shown in Figure 10.3, the one-dimensional model is a reasonable approximation. If one end of the tube is closed, then particle velocity at that end must be zero for all time by the no-slip condition. Thus the acoustic dynamics in a narrow tube with both ends closed would be described by the wave equation (for velocity) with zero boundary conditions

$$\partial_t^2 v(x,t) \;=\; \tfrac{\gamma\bar p}{\bar\rho}\,\partial_x^2 v(x,t) \qquad v(-L,t) = 0, \qquad v(L,t) = 0, \tag{10.19}$$

which resembles the equation for string vibrations (10.17).

On the other hand, if one end is open, then a "radiative" boundary condition[2] needs to be imposed. For tubes with very high aspect ratio, a reasonable approximation is that pressure fluctuations at the open end are zero. We can apply those boundary conditions regardless of whether we take the wave equation for velocity or the equivalent alternative wave equation for pressure. This because when the equations relating velocity and pressure (10.18) are applied at a particular point $\bar x$ for all time

$$\left.\begin{aligned}
&\text{for } t \geq 0, \quad v(\bar x,t) = 0 \;\Rightarrow\; \partial_t v(\bar x,t) = 0 \\
&\text{and} \qquad \partial_t v(\bar x,t) = \tfrac{1}{\bar\rho}\,\partial_x p(x,t)
\end{aligned}\right\} \quad \Rightarrow \quad \partial_x p(\bar x,t) = 0$$

$$\left.\begin{aligned}
&\text{for } t \geq 0, \quad p(\bar x,t) = 0 \;\Rightarrow\; \partial_t p(\bar x,t) = 0 \\
&\text{and} \qquad \partial_t p(\bar x,t) = \gamma\bar p\,\partial_x v(x,t)
\end{aligned}\right\} \quad \Rightarrow \quad \partial_x v(\bar x,t) = 0$$

Thus we can replace a zero velocity boundary condition with a zero derivative pressure boundary condition and vice versa.

[1] Standardized atmospheric pressure at sea level is 1 atm (atmosphere) := 101,235 Pa. The standardized air density at sea level is approximately 1.2 Kg/m^2. This leads to a speed of sound of $\sqrt{\gamma\bar p/\bar\rho} \approx 344$ m/s.

[2] A "radiative" or "Robin" boundary condition specifies a linear relation between pressure and velocity at the open end $v(L,t) = \alpha p(L,t)$, where the parameter α is related to the "acoustic impedance" at the tube opening.

For example, for a tube closed at the left end $x = \text{-}L$ and open at the right end $x = L$, we can use either of wave equations with their following respective boundary conditions

$$\begin{aligned} \boldsymbol{\partial}_{\mathbf{t}}^2 v(x,t) &= \tfrac{\gamma\bar{p}}{\bar{\rho}}\, \boldsymbol{\partial}_{\mathbf{x}}^2 v(x,t), & v(\text{-}L,t) = 0, \qquad & \boldsymbol{\partial}_{\mathbf{x}} v(L,t) = 0, \\ \boldsymbol{\partial}_{\mathbf{t}}^2 p(x,t) &= \tfrac{\gamma\bar{p}}{\bar{\rho}}\, \boldsymbol{\partial}_{\mathbf{x}}^2 p(x,t), & \boldsymbol{\partial}_{\mathbf{x}} p(\text{-}L,t) = 0, \qquad & p(L,t) = 0. \end{aligned}$$

10.2.1 Normal Modes: Eigenvalues and Eigenfunctions

The normal modes of string vibrations are determined by the eigenvalues of the one-dimensional PDE operator $\nu^2(x)\, \partial_x^2$ together with the boundary conditions listed above. For strings with homogenous material properties, $\nu(x)$ is constant in x (which from now on we write as simply ν), and therefore the operator has constant coefficients, which greatly simplifies its analysis. The normal modes of vibrations are obtained from eigenfunctions which satisfy[3] equations of the form

$$\partial_x^2\, v(x) \;=\; \lambda\, v(x), \qquad\qquad v(\text{-}L) = 0, \quad v(L) = 0, \tag{10.20}$$

for some (possibly complex) number λ. Note that there is no time dependence in this eigenvalue problem, and therefore it is actually an ODE in the independent variable x

$$\frac{d^2}{dx^2} v(x) \;=\; \lambda\, v(x), \qquad v(\pm L) \;=\; 0.$$

This is a second order ODE with constant coefficients. It may not have solutions that satisfy the boundary conditions for all λ. We first characterize the numbers λ for which there exists solutions, and then construct those solutions. The characteristic roots are given as the roots $\bar{s}$ of the polynomial $(s^2 - \lambda) = 0$. The form of the solutions depends on whether $\lambda = \bar{s}^2$ is positive, zero, or negative as follows

$$v(x) \;=\; \begin{cases} \mathsf{v}_1\, e^{\bar{s}x} \;+\; \mathsf{v}_2\, e^{-\bar{s}x} & \lambda > 0, \;\; \bar{s} = \left|\sqrt{\lambda}\right|, \\ \mathsf{v}_1\, x \;+\; \mathsf{v}_2, & \lambda = 0, \\ \mathsf{v}\, e^{j\bar{s}x} \;+\; \mathsf{v}^*\, e^{-j\bar{s}x} & \lambda < 0, \;\; \bar{s} = \left|\sqrt{|\lambda|}\right|. \end{cases} \tag{10.21}$$

For each of the three cases, the constraint of satisfying the boundary conditions $v(\pm L) = 0$ can be expressed as a system of two linear equations (for the coefficients v_1, v_2 or v)

$$\overbrace{\begin{bmatrix} e^{\bar{s}L} & e^{-\bar{s}L} \\ e^{-\bar{s}L} & e^{\bar{s}L} \end{bmatrix}}^{\lambda>0} \begin{bmatrix} \mathsf{v}_1 \\ \mathsf{v}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \overbrace{\begin{bmatrix} L & 1 \\ -L & 1 \end{bmatrix}}^{\lambda=0} \begin{bmatrix} \mathsf{v}_1 \\ \mathsf{v}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \overbrace{\begin{bmatrix} e^{j\bar{s}L} & e^{-j\bar{s}L} \\ e^{-j\bar{s}L} & e^{j\bar{s}L} \end{bmatrix}}^{\lambda<0} \begin{bmatrix} \mathsf{v} \\ \mathsf{v}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus non-zero coefficients that satisfy the boundary conditions exist iff the corresponding 2×2 matrix is singular. This condition is most easily expressed in this case via their respective determinants

$$\text{non-trivial solutions exists for } \begin{cases} \lambda > 0 & \text{if } e^{2\bar{s}L} - e^{-2\bar{s}L} \;=\; \sinh(2\bar{s}L) \;=\; 0, \\ \lambda = 0 & \text{if } L + L \;=\; 2L \;=\; 0, \\ \lambda < 0 & \text{if } e^{2j\bar{s}L} - e^{-2j\bar{s}L} \;=\; \frac{1}{2j}\sin(2\bar{s}L) \;=\; 0. \end{cases}$$

Clearly the case $\lambda = 0$ is inadmissible. The case $\lambda > 0$ is also inadmissible since $\bar{s} > 0$, and $\sinh(2\bar{s}L) = 0$ only at $\bar{s} = 0$. This leaves the third case for which the condition becomes

$$\bar{s} > 0, \quad \sin(2L\bar{s}) = 0 \qquad \Rightarrow \qquad \bar{s} = \frac{\pi}{L} n, \;\; n = 1, 2, \ldots \qquad \Rightarrow \qquad \lambda_n = -\frac{\pi^2}{L^2} n^2, \;\; n = 1, 2, \ldots,$$

where the last equality recalls that $\lambda_n = -\bar{s}^2$ for the case $\lambda < 0$.

[3] Recall that the eigenvalues of a scaling $\nu^2\partial_x^2$ of the operator are simply ν^2 times the eigenvalues of ∂_x^2. On the other hand, the eigen*functions* of the two operators are the same.

Having calculated the eigenvalues, we now find the coefficient v to obtain the eigenfunctions. The general form for the case $\lambda < 0$ is that of a sinusoid of frequency $\bar{s}_n := n\pi/L$. Any such sinusoid can be written (ignoring the amplitude for now) as a sine with some phase β

$$v_n(x) \;=\; \sin\big(n(\pi/L)x + \beta\big). \tag{10.22}$$

Using the boundary conditions again

$$0 = v_n(\pm L) = \sin\left(n\frac{\pi}{L}(\pm L) + \beta\right) \qquad \Rightarrow \qquad \pm n\pi + \beta = k\ 2\pi \qquad \Rightarrow \qquad \beta = \pi(n + 2k),$$

where k is any integer. Note however that we can take $k = 0$ since an additional phase of kL for the sine function can only change its sign (i.e. $\sin(\gamma + kL) = \pm\sin(\gamma)$), which does not effect the condition above. Similarly, we can take $\beta = +Ln/2$ since if $\sin(\gamma + L/2) = 0$, the so is $\sin(\gamma - L/2) = 0$ (and similarly for $n > 1$). We finally conclude that the eigenfunctions and eigenvalues are all of the form

$$v_n(x) = \;\sin\left(\frac{n}{2}(x + L)\right), \qquad \omega_n = \frac{n}{2}, \quad n = 1, 2, \ldots. \tag{10.23}$$

10.3 The Euler-Bernoulli Beam

The dynamic Euler-Bernoulli equation for an unloaded, undamped beam is

$$\boldsymbol{\partial}_{\mathbf{t}}^2 W(x,t) \;=\; -\,\alpha\ \boldsymbol{\partial}_{\mathbf{x}}^4 W(x,t), \qquad 0 \le x \le L, \qquad \alpha := \frac{EI}{\rho}, \tag{10.24}$$

where $W(x,t)$ is the beam's deflection at location x and time t. E, I, ρ are the Young's modulus, area moment of inertia, and the linear mass density of the beam respectively. The boundary conditions on this equation are determined by how the beam is supported (e.g. clamped at one end and free at the other, pinned at both ends, etc.).

The results presented earlier relied on a finite-difference discretization of the spatial derivatives in the Euler-Bernoulli equation. There is an alternative, more analytical approach that lends additional insight compared to the numerical method. This approach relies on an abstraction where the spatial derivative $\boldsymbol{\partial}_{\mathbf{x}}^4$ is considered as an infinite dimensional version of a matrix, referred to as a *linear operator*. While matrices operate on vectors to produce other vectors, linear operators act on *functions* to produce other functions. The linear algebra concepts are very analogous though, linear operators have eigenvalues, but possibly an infinite number of them, and they have *eigenfunctions* as their version of eigenvectors.

We will briefly describe this more abstract approach using the Euler-Bernoulli equation. Recall the dynamics (9.14)

$$\boldsymbol{\partial}_{\mathbf{t}}^2 W(x,t) \;=\; -\,\alpha\ \boldsymbol{\partial}_{\mathbf{x}}^4 W(x,t), \qquad 0 \le x \le L. \tag{10.25}$$

The spatial derivative can be thought of as a linear operator $\mathcal{K} := \alpha\boldsymbol{\partial}_{\mathbf{x}}^4$ which acts on functions defined over the spatial domain $0 \le x \le L$. A function $v(x)$ on that domain is an *eigenfunction* of $\mathcal{K}$ with eigenvalue λ if

$$\mathcal{K}\ v \;=\; \lambda\ v \qquad\qquad \Leftrightarrow \qquad\qquad \alpha\boldsymbol{\partial}_{\mathbf{x}}^4\ v(x) \;=\; \lambda\ v(x), \qquad 0 \le x \le L. \tag{10.26}$$

Notice that the statement $\mathcal{K}v = \lambda v$ is the same as if $\mathcal{K}$ were a matrix, and v were a vector. The new feature now is that $\mathcal{K} = -\alpha\boldsymbol{\partial}_{\mathbf{x}}^4$ is a linear operator (that operates on *functions* rather than vectors), and v is a function.

The statement on the right in (10.26) is just the statement $\mathcal{K}v = \lambda v$ in detail. This statement is actually a differential equation. To see this, assume again a beam clamped at $x = 0$ (with no actuation in this case), and free at $x = L$. The (purely spatial) differential equation for determining the eigenfunctions becomes

$$\bar{s}\ v^{(4)}(x) \;-\; \lambda\ v(x) \;=\; 0, \qquad \begin{matrix} v(0) = 0, & v^{(2)}(L) = 0 \\ v^{(1)}(0) = 0, & v^{(3)}(L) = 0 \end{matrix}. \tag{10.27}$$

This is a 4'th order differential equation for $v(x)$ with a "free parameter" λ. It is called a "two point boundary value problem" since the boundary conditions are given at two points, namely $x = 0$ and $x = L$. Note that time does not play a role in this equation, the same way that when solving for normal modes, we solve an eigenvalue problem for a matrix $K\boldsymbol{v} = \lambda \boldsymbol{v}$, and this eigenvalue problem does not involve time.

The differential equation (10.27) can be solved with standard methods for constant coefficient ODEs. We will not show the details here. The main feature is that the parameter λ is not given, but it is to be determined. One can show that this equation has solutions that satisfy the boundary conditions only for values of λ which satisfy the equation

$$\cos\left(\lambda^{1/4}\right)\cosh\left(\lambda^{1/4}\right) \;+\; 1 \;=\; 0, \tag{10.28}$$

where for simplicity we assumed $\bar{s} = 1$ and $L = 1$. The equation (10.28) has an infinite set of solutions at discrete values $\{\lambda_i\}_{i=1}^{\infty}$. The normal modes are then obtained from

$$\omega_i \;=\; \sqrt{\lambda_i}, \qquad i = 1, 2, \ldots.$$

Equation (10.28) does not have analytical solutions, but good estimates can be obtained for large λ since $\cosh\left(\lambda^{1/4}\right)$ grows exponentially. Therefore

$$\begin{aligned} \lambda \gg 1 \qquad \Rightarrow \qquad \cos\left(\lambda^{1/4}\right) = \frac{-1}{\cosh\left(\lambda^{1/4}\right)} \;\approx\; 0 \qquad \Rightarrow \qquad \lambda_i^{1/4} \;\approx\; iL - \frac{L}{2} \\ \Rightarrow \qquad \omega_i \;=\; \sqrt{\lambda_i} \;\approx\; \left(iL - \frac{L}{2}\right)^2. \end{aligned} \tag{10.29}$$

Thus for large ω_i we see that the normal modes behave like $\omega_i \sim i^2$. For lower values of $\omega = \sqrt{\lambda}$, the zeros of equation (10.28) can be calculated numerically, and the table below gives the values for the first 4 normal modes. We see that the approximation (10.29) is quite reasonable even for ω_2.

	ω_1	ω_2	ω_3	ω_4
actual values	3.516	22.03	61.70	121.0
$\left(iL - \frac{L}{2}\right)^2$	2.467	22.21	61.69	120.9

Finally, the reader should note that the values of $\omega_1, \omega_2, \omega_3, \omega_4$ calculated from the eigenvalue problem here are precisely the resonances seen in Figure 9.9a as expected from the theory.

For each computed eigenvalue λ_i, one can substitute that number in the differential equation (10.27) and solve it for the i'th eigenfunction $v_i(x)$. This will produce the exact mode shapes shown in Figure 9.9b.

10.4 Vibrations of Thin Membranes: The 2D Wave Equation

The vibrations of a thin membranes are described by a function $w(x, y, t)$ of two spatial variables x, y and time t, where $w(x, y, t)$ is the instantaneous transverse displacement of the membrane. The vibration dynamics of such thin membranes with isotropic and homogenous material properties are governed by 2 dimensional version of the wave equation

$$\partial_t^2\; w(x, y, t) \;=\; \nu^2\left(\partial_x^2 w(x, y, t) + \partial_y^2 w(x, y, t)\right) \;=\; \nu^2\left(\partial_x^2 + \partial_y^2\right) w(x, y, t) \tag{10.30}$$

where ν is a constant the depends on material properties. The boundary conditions depend on the shape of the membrane described as a domain in 2D. This equation is the two dimensional analog of the one dimensional wave equation (10.17) describing the vibrations of a string. The main difference

is that spatial derivatives are now in terms of the 2D Laplacian $\partial_x^2 + \partial_y^2$ instead of the 1D Laplacian in (10.17).

In this section we only consider free vibrations of a membrane that is pinned at the boundary γ of a domain $\Gamma \subset \mathbb{R}^2$. Unlike the 1D case, the boundary conditions are now imposed on a closed curve γ in 2D rather than two end points. A good analogy to keep in mind is with the membrane of a drum, which is forced to not vibrate at the edges of the drum. The equation with its boundary conditions becomes

$$\partial_t^2\ w(x,y,t). \ = \ \nu^2 \left(\partial_x^2 + \partial_y^2\right) w(x,y,t), \qquad w(x,y,t)\Big|_{(x,y)\in\gamma} = 0,\ t \geq 0. \tag{10.31}$$

The normal modes of plate vibrations are determined by the eigenvalues of the two dimensional PDE operator $\partial_x^2 + \partial_y^2$ together with the applicable boundary conditions. For membranes with homogenous material properties, this has constant coefficients, which greatly simplifies its analysis. The normal modes of vibrations are obtained from eigenfunctions of this operator[4]

$$\partial_x^2 v(x,y) \ + \ \partial_y^2 v(x,y) \ = \ \lambda\ v(x,y). \tag{10.32}$$

The eigenfunctions are all of the form

$$v(x,y) \ = \ e^{j\kappa_{\rm x} x}\ e^{j\kappa_{\rm y} y} \ = \ e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)}, \tag{10.33}$$

where $\kappa_{\rm x}$ and $\kappa_{\rm y}$ are the *wavenumbers* that will be determined by applying the boundary conditions. Substituting the form (10.33) into the eigenfunction equation (10.32) we arrive at an algebraic equation for the eigenvalues

$$\begin{aligned} \lambda\ e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} \ &= \ \left(\partial_x^2 + \partial_y^2\right) e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} \ = \ \left((j\kappa_{\rm x})^2 + (j\kappa_{\rm y})^2\right) e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} \\ \Rightarrow \qquad \lambda \ &= \ -\left(\kappa_{\rm x}^2 + \kappa_{\rm y}^2\right). \end{aligned} \tag{10.34}$$

The allowable values of λ will then come from discovering the allowable values of $\kappa_{\rm x}$ and $\kappa_{\rm y}$ that are compatible with boundary conditions.

Example 10.1. [Square membrane] This is the simplest case to analyze. Let the membrane be of size $2L \times 2L$, and choose a centered coordinate system so that the boundaries are at $x = \pm L$ and $y = \pm L$. The pinned boundary conditions are then

$$v(\pm L, y) = 0, \quad y \in [-L, L], \qquad\qquad v(x, \pm L) = 0, \quad x \in [-L, L]. \tag{10.35}$$

For real eigenfunctions, we need function of the form (10.33) together with its complex conjugate, i.e.

$$v(x,y) \ = \ \mathsf{v}\ e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} + \mathsf{v}^*\ e^{-j(\kappa_{\rm x} x + \kappa_{\rm y} y)}.$$

Imposing the boundary conditions (10.35) on this form gives

$$0 = v(\pm L, y) \ = \ \mathsf{v}\ e^{j(\pm\kappa_{\rm x} L + \kappa_{\rm y} y)} + \mathsf{v}^*\ e^{-j(\pm\kappa_{\rm x} x + \kappa_{\rm y} y)}, \qquad y \in [-L, L]$$

10.5 Vibrations of Plates: The 2D Euler-Bernoulli Equation

The vibrations of a thin plate are described by a function $w(x,y,t)$ of two spatial variables x, y and time t, where $w(x,y,t)$ is the instantaneous transverse displacement of the plate. The vibration dynamics of such thin plates with isotropic and homogenous material properties are governed by the following PDE

$$\partial_t^2\ w(x,y,t) \ = \ -\,\bar{s}\left(\partial_x^4 w(x,y,t) + 2\ \partial_x^2\partial_y^2 w(x,y,t) + \partial_y^4 w(x,y,t)\right), \tag{10.36}$$

[4] Recall that the eigenvalues of a scaling $\nu^2(\partial_x^2+\partial_y^2)$ of the operator are simply ν^2 times the eigenvalues of $(\partial_x^2+\partial_y^2)$. On the other hand, the eigen*functions* of the two operators are the same.

where $\bar{s}$ is a constant the depends on material properties. The boundary conditions depend on the shape of the plate, described as a domain in 2D, and on which parts of its boundary the plate maybe pinned, clamped or free.

Although at first the PDE (10.36) appears quite different from the Euler-Bernoulli beam PDE (10.24), it is in fact just the two (spatial) dimensional version of the one spatial dimensional beam equation. To see that note that the right hand side of each PDE is the square of the Laplacian operator in 1 and 2 dimensions respectively

$$\begin{aligned} \Delta &:= \partial_x^2 \qquad \Rightarrow \qquad \Delta^2 = \partial_x^4 && \text{(in 1 dimension)} \\ \Delta &:= \partial_x^2 + \partial_y^2 \qquad \Rightarrow \qquad \Delta^2 = \left(\partial_x^2 + \partial_y^2\right)^2 = \partial_x^4 + 2\,\partial_x^2\partial_y^2 + \partial_y^4 && \text{(in 2 dimensions)} \end{aligned}$$

Thus the beam and plate PDEs can be written in a formally similar way as

$$\begin{aligned} \partial_t\, W(x,t) &= -\,\bar{s}\,\Delta^2 W(x,t) && \text{(beam equation)} \\ \partial_t\, W(x,y,t) &= -\,\bar{s}\,\Delta^2 W(x,y,t), && \text{(plate equation)} \end{aligned}$$

where the Laplacian Δ is defined as above depending on whether the problem is in 1 or 2 dimensions.

The normal modes of plate vibrations are determined by the eigenvalues of the two dimensional PDE operator Δ^2 together with the applicable boundary conditions. For plates with homogenous material properties, the PDE operator $\Delta^2 = \partial_x^4 + 2\,\partial_x^2\partial_y^2 + \partial_y^4$ has constant coefficients, which greatly simplifies its analysis. The normal modes of vibrations are obtained from eigenfunctions of this operator

$$\Delta^2 v(x,y) = \lambda\, v(x,y) \qquad \Leftrightarrow \qquad \partial_x^4 v(x,y) + 2\,\partial_x^2\partial_y^2 v(x,y) + \partial_y^4 v(x,y) = \lambda\, v(x,y). \tag{10.37}$$

The eigenfunctions are all of the form

$$v(x,y) = e^{j\kappa_{\rm x} x}\, e^{j\kappa_{\rm y} y} = e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)}, \tag{10.38}$$

where $\kappa_{\rm x}$ and $\kappa_{\rm y}$ are the *wavenumbers* that will be determined by applying the boundary conditions. Substituting the form (10.38) into the eigenfunction equation (10.37) we arrive at an equation that the eigenvalue λ must satisfy

$$\begin{aligned} \lambda\, e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} &= \left(\partial_x^4 + 2\,\partial_x^2\partial_y^2 + \partial_y^4\right) e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} = \left((j\kappa_{\rm x})^4 + 2\,(j\kappa_{\rm x})^2(j\kappa_{\rm y})^2 + (j\kappa_{\rm y})^4\right) e^{j(\kappa_{\rm x} x + \kappa_{\rm y} y)} \\ \Rightarrow \qquad \lambda &= \kappa_{\rm x}^4 + 2\,\kappa_{\rm x}^2\kappa_{\rm y}^2 + \kappa_{\rm y}^4 = \left(\kappa_{\rm x}^2 + \kappa_{\rm y}^2\right)^2. \end{aligned} \tag{10.39}$$

The allowable values of λ will then come from discovering the allowable values of $\kappa_{\rm x}$ and $\kappa_{\rm y}$ that are compatible with boundary conditions.

Example 10.2. [Square plate with free boundaries] This is the simplest case to analyze. Let the plate be of size $2L \times 2L$, and choose a centered coordinate system so that the boundaries are at $x = \pm L$ and $y = \pm L$. The free boundary conditions are then

$$\begin{aligned} &\partial_x^2 v(\pm L, y) = 0, \qquad y \in [-L, L], && \partial_y^2 v(x, \pm L) = 0, \qquad x \in [-L, L], \\ &\partial_x^3 v(\pm L, y) = 0, \qquad y \in [-L, L], && \partial_y^3 v(x, \pm L) = 0, \qquad x \in [-L, L]. \end{aligned}$$

TBC

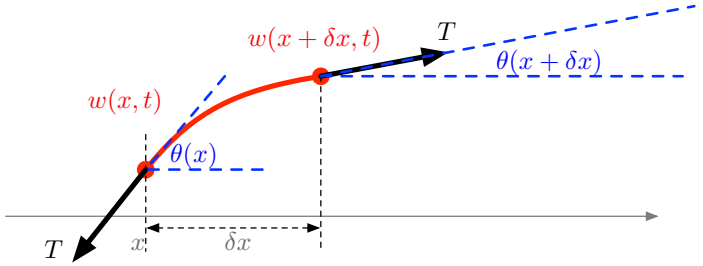


Figure 10.4: The transverse motion of a string segment (shown in red) are determined by the net vertical force on it. The tension $T(x)$ is assumed constant in magnitude throughout the string. However, the angle $\theta(x)$ at which the tension acts varies and is determined by the slope $\boldsymbol{\partial_x} w(x,t)$. The difference between the vertical components of $T(x)$ and $T(x+\delta x)$ is determined by the difference between $\theta(x)$ and $\theta(x+\delta x)$, which in turn is determined by the *curvature* $\boldsymbol{\partial_x^2} w(x,t)$ of the string. In the configuration shown, the string segment has negative curvature, and therefore it is accelerating downwards.

Appendix

10.A Physical Dynamics Described by the 1D Wave Equation

10.A.1 String Vibrations

To derive the differential equation for string vibrations, take a small segment of the string as shown in Figure 10.4 and analyze the forces on it. The assumption we make is that the internal tension force T in the string is constant in magnitude throughout. However, the angle $\theta(x)$ at which the tension forces act on a string element does depend on the instantaneous "slope" $\boldsymbol{\partial_x} w(x,t)$ of the string. Examining Figure 10.4 shows that the net vertical force on the string element shown is determined by the difference between the *vertical* components of the opposing tension forces. This difference is in turn determined by the "curvature" $\boldsymbol{\partial_x^2} w(x,t)$ of the string as we show next.

Let $\theta(x)$ be the slope of the function $w(x,t)$ at any one instant of time, i.e. the angle the vector tangent to $w(x,t)$ makes with the horizontal axis. The mass of a string element of length δx is given by $\delta x\ \rho(x)$, where $\rho(x)$ is the string's linear density. Newton's 2nd law for a small string segment relates the vertical acceleration $\boldsymbol{\partial_t^2} w(x,t)$ to the sum of vertical forces

$$\begin{aligned}
\Big(\delta x\ \rho(x)\Big)\ \boldsymbol{\partial_t^2} w(x,t) \ &= \ T\ \sin\Big(\theta(x+\delta x)\Big)\ -\ T\ \sin\Big(\theta(x)\Big) && \text{(sum of vertical forces)} \\
&\approx \ T\ \theta(x+\delta x)\ -\ T\ \theta(x) && \text{(using the small angle approximation } \sin(\phi)\approx\phi\text{)} \\
&= \ T\ \Big(\theta(x+\delta x)\ -\ \theta(x)\Big) && (10.40)
\end{aligned}$$

To rewrite this equation in terms of $w(x,t)$, we need to express this angle difference in terms of w or its spatial derivatives. First observe that the slope is related to the derivative $\boldsymbol{\partial_x} w(x,t)$ by using the small angle approximation again $\theta \approx \tan(\theta)$ and

$$\theta(x)\ \approx\ \tan\big(\theta(x)\big)\ =\ \boldsymbol{\partial_x} w(x,t),$$

where the last equality follows from the definition of the derivative. The slope at $x+\delta x$ is also

$$\theta(x+\delta x)\ \approx\ \boldsymbol{\partial_x} w(x+\delta x,t).$$

The above two expressions for the angles can now be used to express the angle difference in (10.40) in terms of the curvature $\boldsymbol{\partial_x^2} w(x,t)$

$$\begin{aligned}
\theta(x+\delta x)\ -\ \theta(x)\ &\approx\ \boldsymbol{\partial_x} w(x+\delta x,t)\ -\ \boldsymbol{\partial_x} w(x,t) \\
&\approx\ \Big(\boldsymbol{\partial_x} w(x,t) + \boldsymbol{\partial_x^2} w(x,t)\ \delta x\Big)\ -\ \boldsymbol{\partial_x} w(x,t) \\
& \qquad\qquad \text{(first-order approximation of } \boldsymbol{\partial_x} w(x+\delta x,t)\text{)} \\
&=\ \boldsymbol{\partial_x^2} w(x,t)\ \delta x
\end{aligned}$$

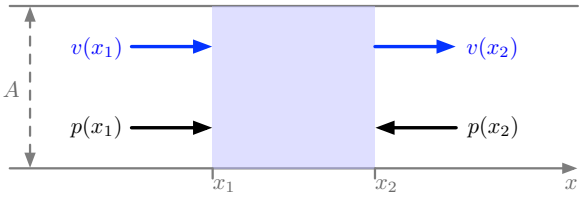


Figure 10.5: Depicting a control volume for the dynamics of a one-dimensional gas, or equivalently, a 3D gas where velocities, pressures and densities vary in only one dimension, depicted here as the x coordinate. $v(x_1)$ and $v(x_2)$ depict the velocities in and out of the volume respectively. $p(x_1)$ and $p(x_2)$ are the pressure forces exerted onto the volume by the remainder of the gas. These pressure forces are always positive and act inwards towards the volume. The density $\rho(x)$ (not shown) is also assumed to vary with coordinate x.

Finally we can substitute this last expression in (10.40) to get a PDE that is purely in terms of $w(x,t)$ and does not involve $\theta(x)$

$$\Big(\delta x\ \rho(x)\Big)\ \boldsymbol{\partial_t^2} w(x,t)\ \approx\ T\ \boldsymbol{\partial_x^2} w(x,t)\ \delta x. \tag{10.41}$$

Dividing through by δx and taking a limit $\delta x \to 0$ produces an exact equation[5] which is the wave equation

$$\boldsymbol{\partial_t^2} w(x,t)\ =\ \tfrac{T}{\rho(x)}\ \boldsymbol{\partial_x^2} w(x,t).$$

Note that the constant $\nu^2 = T/\rho(x)$, where the tension T plays the role of "stiffness per unit length" while the density $\rho(x)$ is the mass per unit length. Thus the wave equation can be considered to be the continuum analog of a Mass-Spring system.

10.A.2 One-dimensional Gas Dynamics and Acoustics

Figure 10.5 depicts a one-dimensional compressible gas medium, or equivalently a three-dimensional medium (with cross-sectional area A) where all fields are constant in the cross-sectional directions, e.g as in the propagation of plane waves. All fields are then simply functions of a single spatial coordinate x. The picture depicts a thin volume slice (a "control volume" in the interval $[x_1, x_2]$) in which we keep track of mass, momentum and energy transport. The fields relevant to gas dynamics are

- $v(x,t)$: the velocity field
- $\rho(x,t)$: the mass density field
- $p(x,t)$: the pressure field (the isotropic force exerted on the fluid at any point)
- $T(x,t)$: the temperature field
- $U(x,t)$: specific (per unit mass) internal energy. It accounts for energy stored in molecular motions and vibrations. For "calorically perfect" gasses, it is a function of only temperature $U(x,t) = c_v T(x,t)$, where c_v is the specific heat capacity.

Three physical laws that govern the behavior of these fields can be derived.

Mass Conservation

The mass flow rate rightwards through the boundary at x_i is equal to $v(x_i,t)\rho(x_i,t)A$. The total mass inside the control volume *increases* at a rate equal to the total mass flow rate *inwards* at those

[5] Note that we took first order approximations with δx and $\delta\theta \sim \delta x$ assumed small. This means that terms of order $(\delta x)^2$ and higher powers were ignored. If we kept track of them, they would appear in Equation (10.41). However, after dividing through by δx and taking the limit $\delta x \to 0$, those terms would limit to zero.

boundaries

$$\underbrace{\frac{d}{dt}\left(\int_{x_1}^{x_2} \rho(x,t)A\; dx\right)}_{\substack{\text{time rate of change of}\\ \text{total mass in control volume}}} \;=\; \underbrace{v(x_1,t)\;\rho(x_1,t)A}_{\text{mass flow rate in}} \;-\; \underbrace{v(x_2,t)\;\rho(x_2,t)A}_{\text{mass flow rate out}}.$$

Dividing by A and $\Delta_x := x_2 - x_1$, and taking the small-volume limit as $\Delta_x \to 0$ results in the continuity (aka transport) equation

$$\boldsymbol{\partial_t}\rho(x,t) \;=\; -\,\boldsymbol{\partial_x}\Big(v(x,t)\;\rho(x,t)\Big). \tag{10.42}$$

Forces and the Momentum Equation

The time rate of change of momentum of a control volume of fluid is equal to the sum of forces acting on it, plus the momentum flux due to material flow across the boundaries (now suppressing the independent variables and the area A for notational simplicity)

$$\frac{d}{dt}\quad\underbrace{\left(\int_{x_1}^{x_2} v\rho\; dx\right)}_{\substack{\text{total momentum of}\\ \text{fluid in control volume}}} \;=\; \underbrace{\left(v^2\rho\Big|_{x_1} - v^2\rho\Big|_{x_2}\right)}_{\substack{\text{momentum influx}\\ \text{through boundaries}}} + \underbrace{\left(p\Big|_{x_1} - p\Big|_{x_2}\right)}_{\substack{\text{total force at}\\ \text{boundaries}}}.$$

The first term on the right states that a mass flow rate of $v\rho A$ across a boundary adds momentum at the rate of $v\,(v\rho A) \;=\; v^2\rho A$. The second term is the total force at the boundary due to the isotropic pressure force (which always acts inwards on the volume). Taking the small-volume limit again results in the momentum equation

$$\boldsymbol{\partial_t}\Big(v(x,t)\;\rho(x,t)\Big) \;=\; -\,\boldsymbol{\partial_x}\Big(v^2(x,t)\rho(x,t)\Big) \;-\; \boldsymbol{\partial_x}p(x,t). \tag{10.43}$$

An equivalent (and more standard) form of this equation expressing acceleration is obtained if the continuity equation (10.42) is used to eliminate the $\boldsymbol{\partial_t}\rho$ terms (see Section 10.A.3)

$$\boldsymbol{\partial_t}v(x,t) \;=\; -\,v(x,t)\;\boldsymbol{\partial_x}v(x,t) \;-\; \frac{1}{\rho(x,t)}\;\boldsymbol{\partial_x}p(x,t). \tag{10.44}$$

The Energy Equation and Thermodynamic Effects

A volume of gas has both internal and kinetic energies. Its total *energy density* at location x can then be expressed as $\rho(x,t)U(x,t) + \frac{1}{2}\rho(x,t)v^2(x,t)$. An intuitive interpretation of internal energy density ρU is that it is a kind of potential energy stored in molecular motions and vibrations. The time rate of change of total energy in the control volume is accounted for by

$$\underbrace{\frac{d}{dt}\left(\int_{x_1}^{x_2}\left(\rho U + \tfrac{1}{2}\rho v^2\right)\; dx\right)}_{\text{time rate of change of energy}} \;=\; \underbrace{v\left(\rho U + \tfrac{1}{2}\rho v^2\right)\Big|_{x2}^{x1}}_{\substack{\text{power added by}\\ \text{material influx}}} + \underbrace{pv\Big|_{x2}^{x1}}_{\substack{\text{work rate of}\\ \text{pressure forces}}}. \tag{10.45}$$

The first term on the right represents the energy flux due to material entering and leaving the control volume with velocity v, while the second term is the rate of work done by the pressure forces at the boundaries. Taking the small volume limit leads to the PDE

$$\boldsymbol{\partial_t}\left(\rho U + \tfrac{1}{2}\rho v^2\right) \;=\; -\,\boldsymbol{\partial_x}\left(v\left(\rho U + \tfrac{1}{2}\rho v^2\right) \;+\; pv\right). \tag{10.46}$$

As shown in Section 10.A.3, this equation can be simplified and converted to a more standard form by using the continuity and momentum equations to substitute for the $\boldsymbol{\partial_t}\left(\frac{1}{2}\rho v^2\right)$ term, which then leads to

$$\boldsymbol{\partial_t}\left(\rho U\right) \;=\; -\,\boldsymbol{\partial_x}\left(v\rho U\right) \;-\; p\;\boldsymbol{\partial_x}v \tag{10.47}$$

Equivalent, and more useful forms of this equation are obtained by expressing the internal energy U in terms of physically observable quantities. This can be done using either pressure or temperature, and we choose pressure here since the resulting equations will have a more convenient form. To express internal energy in terms of pressure, first note that for an ideal gas $p = R\ \rho T$. Combining this with $U = c_v T$ yields

$$\rho U \ = \ c_v\ \rho T \ = \ \frac{c_v}{R}\ p. \tag{10.48}$$

This provides some useful intuition. Recall that internal energy density ρU can be understood as a kind of potential energy stored in molecular motions and vibrations. Equation (10.48) states that local pressure is directly proportional to this potential energy density for ideal gases. Rewriting Equation (10.47) using the substitution (10.48) for ρU we obtain the energy equation in terms of the pressure field

$$\boldsymbol{\partial_t} p \ = \ -\gamma\ p\ \boldsymbol{\partial_x} v \ - \ v\ \boldsymbol{\partial_x} p, \tag{10.49}$$

where $\gamma := 1 + \frac{R}{c_v}$ is the heat capacity ratio (a dimensionless number equal to 1.4 for diatomic gases like air).

Finally, the three equations of mass conservation (10.42), momentum (10.44) and energy (10.49) can now be combined into a single vector PDE that has the following form

$$\boxed{\boldsymbol{\partial_t} \begin{bmatrix} \rho \\ v \\ p \end{bmatrix} = \begin{bmatrix} -v\ \boldsymbol{\partial_x}\rho - \rho\ \boldsymbol{\partial_x} v \\ -v\ \boldsymbol{\partial_x} v \ - \ \frac{1}{\rho}\ \boldsymbol{\partial_x} p \\ -\gamma\ p\ \boldsymbol{\partial_x} v \ - \ v\ \boldsymbol{\partial_x} p \end{bmatrix} \ = \ - \begin{bmatrix} v & \rho & 0 \\ 0 & v & 1/\rho \\ 0 & \gamma p & v \end{bmatrix} \boldsymbol{\partial_x} \begin{bmatrix} \rho \\ v \\ p \end{bmatrix}.} \tag{10.50}$$

These are the *Euler Equations of Gas Dynamics* that describe compressible gas flows which have negligible thermal conductivity within the gas. Although this vector PDE is nonlinear, it has the following very special structure of

$$\boldsymbol{\partial_t} \Psi(x,t) \ = \ F\big(\Psi(x,t)\big)\ \boldsymbol{\partial_x} \Psi(x,t), \tag{10.51}$$

where $\Psi(x,t)$ is a vector-valued field and $F(.)$ is a function that depends on $\Psi(x,t)$, but not on any of its derivatives.

Linearization

Acoustic phenomena are described by the linearization of (10.50) around small oscillations. For a PDE of the form (10.51), consider a nominal solution $\bar{\Psi}(x)$ which is constant in x as well, i.e.

$$0 \ = \ \boldsymbol{\partial_x} \bar{\Psi}(x),$$

and fluctuations $\tilde{\Psi}(x,t)$ around the nominal, i.e. consider a general solution decomposed as

$$\Psi(x,t) \ = \ \bar{\Psi}(x) \ + \ \tilde{\Psi}(x,t),$$

Plugging this into (10.51) and using a Taylor series expansion of $F(.)$

$$\begin{aligned} \boldsymbol{\partial_t} \Big(\bar{\Psi}(x) \ + \ \tilde{\Psi}(x,t)\Big) \ &= \ F\Big(\bar{\Psi}(x) \ + \ \tilde{\Psi}(x,t)\Big)\ \boldsymbol{\partial_x} \Big(\bar{\Psi}(x) \ + \ \tilde{\Psi}(x,t)\Big) \\ \boldsymbol{\partial_t} \tilde{\Psi}(x,t) \ &= \ \Big(F\big(\bar{\Psi}(x)\big) \ + \ \frac{\partial F}{\partial \Psi} \tilde{\Psi}(x,t) + \cdots\Big) \Big(\boldsymbol{\partial_x} \bar{\Psi}(x) \ + \ \boldsymbol{\partial_x} \tilde{\Psi}(x,t)\Big) \\ \boldsymbol{\partial_t} \tilde{\Psi}(x,t) \ &= \ F\big(\bar{\Psi}(x)\big)\ \boldsymbol{\partial_x} \tilde{\Psi}(x,t) \ + \ \cdots, \end{aligned} \tag{10.52}$$

where $\ldots$ indicate terms of order 2 or higher in $\tilde{\Psi}$ which do not contribute to the linearized dynamics.

Now consider (10.50) with a nominal condition of constant density $\bar{\rho}$, velocity $\bar{v}$ and pressure $\bar{p}$. Using (10.52) we see that the density, velocity and pressure fluctuations $\tilde{\rho}, \tilde{v}, \tilde{p}$ have the following linearized dynamics

$$\boldsymbol{\partial_t} \begin{bmatrix} \tilde{\rho} \\ \tilde{v} \\ \tilde{p} \end{bmatrix} = - \begin{bmatrix} \bar{v} & \bar{\rho} & 0 \\ 0 & \bar{v} & 1/\bar{\rho} \\ 0 & \gamma \bar{p} & \bar{v} \end{bmatrix} \boldsymbol{\partial_x} \begin{bmatrix} \tilde{\rho} \\ \tilde{v} \\ \tilde{p} \end{bmatrix}.$$

This equation shows an interesting feature. Since $\bar{\rho}, \bar{v}, \bar{p}$ are constants known ahead of time, the differential equations for $\tilde{v}$ and $\tilde{p}$ are independent of $\tilde{\rho}$ (they only depend on the nominal density $\bar{\rho}$. Those two equations can then be considered independently of the evolution of $\tilde{\rho}$ as

$$\boldsymbol{\partial_t} \begin{bmatrix} \tilde{v} \\ \tilde{p} \end{bmatrix} = - \begin{bmatrix} \bar{v} & 1/\bar{\rho} \\ \gamma \bar{p} & \bar{v} \end{bmatrix} \boldsymbol{\partial_x} \begin{bmatrix} \tilde{v} \\ \tilde{p} \end{bmatrix}.$$

In particular, for the propagation of sound waves in a still medium (i.e. $\bar{v} = 0$), the equations take on a particularly simple form

$$\boldsymbol{\partial_t} \begin{bmatrix} \tilde{v} \\ \tilde{p} \end{bmatrix} = - \begin{bmatrix} 0 & 1/\bar{\rho} \\ \gamma \bar{p} & 0 \end{bmatrix} \boldsymbol{\partial_x} \begin{bmatrix} \tilde{v} \\ \tilde{p} \end{bmatrix}. \tag{10.53}$$

10.A.3 Simplifications Leading to Equations (10.44) and (10.47)

To avoid notational complexity in the following derivations, we adopt here the subscript notation for partial derivatives, i.e. $\boldsymbol{\partial_t} v = v_t$ and $\boldsymbol{\partial_x} v = v_x$ and so on.

From Equation (10.43) to Equation (10.44)

Starting from Equation (10.43) and then using Equation (10.42) follows

$$\begin{aligned}
& & (v\rho)_t &= -\left(v^2\rho\right)_x - p_x & & \\
& \Rightarrow & v_t\rho \;+\; v\rho_t &= -2vv_x\rho - v^2\rho_x - p_x & & \text{(using product and chain rule)} \\
\Rightarrow & & v_t\rho \;+\; v\left(-v\rho_x - v_x\rho\right) &= -2vv_x\rho - v^2\rho_x - p_x & & \text{(using (10.42) for } \rho_t\text{)} \\
& \Rightarrow & v_t\rho &= -vv_x\rho - p_x & & \\
& \Rightarrow & v_t &= -vv_x \;-\; \tfrac{1}{\rho}\left(p_x + \beta v\right), & &
\end{aligned}$$

From Equation (10.46) to Equation (10.47)

We start from Equation (10.46) and use the continuity (10.42) and momentum (10.44) equations (4th line below) as follows

$$\begin{aligned}
\left(\rho U + \frac{\rho v^2}{2}\right)_t &= -\left(v\left(\rho U + \frac{\rho v^2}{2}\right) + pv\right)_x \\
(\rho U)_t \;+\; \tfrac{1}{2}\left(\rho v^2\right)_t &= -\left(v\rho U\right)_x - \tfrac{1}{2}\left(\rho v^3\right)_x - (pv)_x \\
(\rho U)_t \;+\; \tfrac{1}{2}\left(\rho_t v^2 \;+\; 2\rho v v_t\right) &= -\left(v\rho U\right)_x - \tfrac{1}{2}\left(\rho_x v^3 + 3\rho v^2 v_x\right) - \left(p_x v + p v_x\right) \\
(\rho U)_t + \tfrac{1}{2}\left(\left(-v_x\rho - v\rho_x\right)v^2 + 2\rho v\left(-vv_x \;-\; \tfrac{1}{\rho}p_x\right)\right) &= -\left(v\rho U\right)_x - \tfrac{1}{2}\left(\rho_x v^3 + 3\rho v^2 v_x\right) - \left(p_x v + p v_x\right) \\
(\rho U)_t - \tfrac{1}{2}\left(v^2 v_x\rho + v^3\rho_x\right) - v^2 v_x\rho - p_x v &= -\left(v\rho U\right)_x - \tfrac{1}{2}\left(\rho_x v^3 + 3\rho v^2 v_x\right) - \left(p_x v + p v_x\right) \\
(\rho U)_t - \tfrac{3}{2}v^2 v_x\rho - \tfrac{1}{2}v^3\rho_x - p_x v &= -\left(v\rho U\right)_x - \tfrac{1}{2}\rho_x v^3 - \tfrac{3}{2}\rho v^2 v_x - \left(p_x v + p v_x\right) \\
(\rho U)_t &= -\left(v\rho U\right)_x - p v_x
\end{aligned}$$

Bibliography