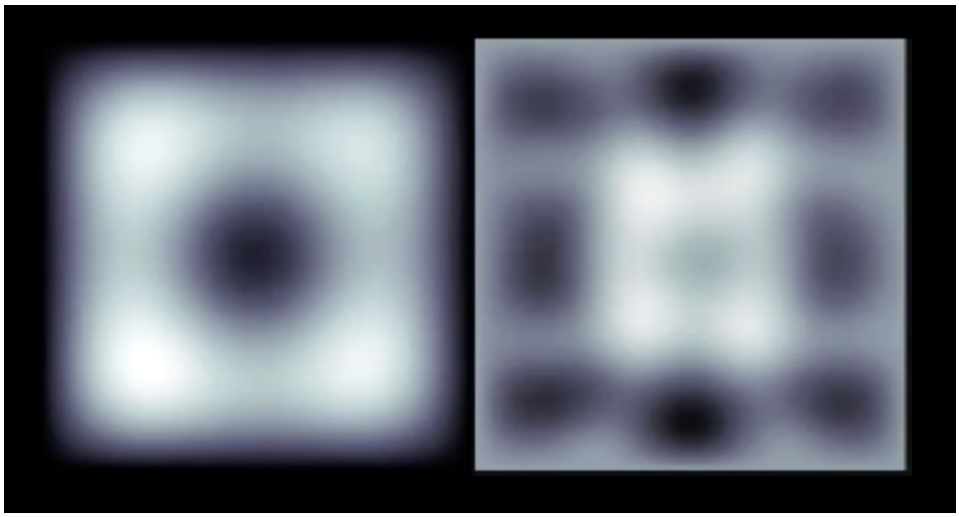
Computational Acoustics

Lecture Notes for the Course
B4936 - COMPUTATIONAL ACOUSTICS M - 6 CFU

Michele Ducceschi
michele.ducceschi@unibo.it
mdphys.org



University of Bologna
2024-2025
Secondo Ciclo

Copyright Notice: The entire manuscript is Copyright by Michele Ducceschi. All Rights Reserved. The notes may not be copied or duplicated in whole or part by any means without express prior agreement in writing or unless specifically noted.

Contents

Chapter 1

Introduction

Acoustics is a large research topic concerned with modelling, measuring, analysing, and simulating sound waves propagating in elastic media. Examples include elastic waves in cables and strings, rods, bars, tubes, membranes, plates and rooms, [1]. The nature of sound propagation varies case by case, and physical models change accordingly. Cables, strings, rods and cylindrical tubes display the simplest kind of wave propagation, which, in a first approximation, can be described in terms of the equation by D'Alembert [2]; see also [3]. This is a one-dimensional type of wave equation for which extensive analytical tools exist. Most notably, the d'Alembert equation admits an analytic solution in the form of two travelling wavefronts that never change shape or size. Some computational models for the wave equation exploit such property to achieve an extremely efficient simulation, such as *digital waveguides*, [4, 5]. The mathematical validity of the simple wave equation is limited to lossless, non-dispersive systems in one dimension. When losses and wave dispersion are included, the shape of the initial wavefronts changes over time, and the d'Alembert solution is no longer valid. More involved acoustic systems not described by the d'Alembert equation also require different solution techniques. For these reasons, it is worth introducing the simulation methods that can be applied to all such problems, starting from general principles.

Numerical simulation is now commonplace among scientists in the field. Commercially available software has enabled the study of systems that cannot be approached analytically, and many computational tools and frameworks exist. *Finite elements*, a kind of numerical method in which the solution is computed as a superposition of elementary functions with compact support, occupy a central role in many branches of computational physics, including acoustics [6, 7]. In this book, however, we will develop the method of *finite differences*, which are conceptually very different, see e.g. [8, 9]. Instead of expanding the solution in terms of elementary functions, finite differences discretise the differential operators directly at specific locations along the domain, called the *nodes*. The continuous solutions are then represented by an approximate grid function defined at the nodes. When the problem is time-dependent, as is often the case in acoustics, the grid functions are updated in time, which is also discretised. Time updating requires storing the grid function's values from previous time steps. There are multiple reasons why one would want to learn the method of finite differences:

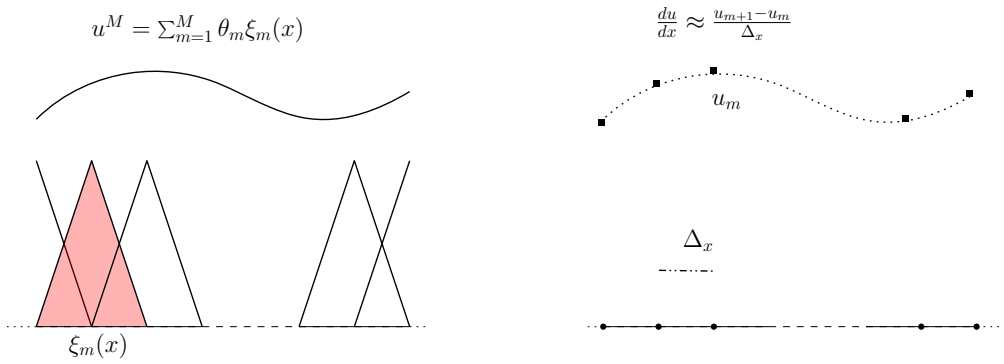


Figure 1.1: Finite elements (left) and finite differences (right). In the FEM paradigm, a solution is discredited using a finite sum of simple test functions with a compact support (the elements). In the FD paradigm, differential operators are discredited on a grid of points.

- They are conceptually simple. The basic workings of finite differences can be understood using high-school calculus.
- Typical wave phenomena in acoustics are described by smooth functions of space and time, for which derivatives exist and are smooth. The related models can be written in their "strong" form, and the differential operators can be discretised directly.
- Finite differences are central to computational acoustics in classical and modern applications. Examples of recent developments of the finite difference method in acoustics include the simulation of large acoustic spaces (see, e.g., [10, 11, 12, 13]) and nonlinear systems such as strings and plates (see, e.g., [14, 15, 16, 17]).
- No matter which computational tool one adopts to approximate the spatial part of a differential problem, time integration is performed using finite differences.
- Recent applications have overcome traditional limitations imposed by the method, particularly in discretising non-cartesian domains and complex boundary conditions, [18, 19].

Figure 1.1 summarises the finite element and finite difference paradigms.

1.1 The finite difference method

The working principle of finite differences is straightforward: one approximates derivatives of functions to turn a differential problem into an algebraic one. The finite difference method can be applied to discretise both the temporal and the spatial domains. However, the two present specific difficulties and must be approached using tailored discretisation techniques. A typical problem in time discretisation concerns the growth of high-frequency oscillations, an issue known at least since the seminal work by Courant, Friedrichs and Lewy

[20]. When such growth is unbounded, numerical instability ensues as the numerical solution diverges from the true solution. This unwanted numerical behaviour may be avoided by respecting appropriate *stability conditions*, using various analytical techniques such as frequency domain methods [21] or energy methods [22, 23]. On the other hand, difficulties arising in spatial discretisation include the correct realisation of the boundary geometry and the setting of appropriate numerical boundary conditions.

The temporal and spatial problems become closely intertwined in discretising partial differential equations. Stability conditions yield the smallest grid spacing for a given time step or the largest time step for a given mesh size. In some other cases, the schemes are unconditionally stable, allowing one to select any time step and grid spacing combination. Not all choices will return a meaningful output as other numerical artefacts arise, such as frequency warping effects and other kinds of numerical distortion severely affecting any musical application [24]. Designing stable schemes with the least amount of numerical artefacts is the core objective of any numerical simulation.

1.2 Boundary Value Problems

Because of the inherent difficulties in discretising the temporal and spatial domains, it is worth approaching the two problems separately. In a typical numerical workflow, the spatial part is discretised first. This turns the original differential problem into an initial-value problem for a system of coupled time-dependent equations, which may be updated in time using a time-stepping scheme. Thus, it is worth approaching the study of finite differences beginning with the theory of boundary value problems (BVPs). In acoustics, typical boundary-value problems include the *Laplace equation* and the *biharmonic equation*, with or without a source*. Without a source, solutions to such problems may be found in one dimension by direct integration. Such solutions depend on a few constants of integration, which are fixed by imposing appropriate boundary conditions. Since the performance of the numerical approximations can be assessed against the closed-form solution obtained via direct integration, approaching these simple problems numerically allows one to understand the working principles of the finite difference method. Concepts such as the order of accuracy and the order of convergence may be understood through such examples [8]. More complex problems in one spatial dimension arise when the medium properties are space-dependent. Examples include acoustic tubes with a variable cross-section or cables and rods with non-uniform material and geometric properties. In these cases, analytic solutions, if any, are much harder to obtain, and a numerical approach is necessary. Non-uniform mesh sizes can be adapted to account for such spatial variability [25].

Solutions to BVPs in more than one space dimension are seldom available in closed form. Notable exceptions include problems defined on a cartesian domain with simple boundary conditions. Problems defined over non-cartesian domains or presenting more complex boundary conditions can only be approached numerically using extensions of the techniques developed for the one-dimensional case. Particular care must be taken to ensure that the boundary geometry is realised appropriately, keeping the overall accuracy of the discretisation unchanged compared to the interior points.

*the Laplace equation with a source is known as the *Poisson equation*

1.3 Eigenvalue and Frequency Domain Problems

When the problem is time-dependent, but certain assumptions on the system's time evolution hold, a reduction to some form of BVP is usually possible. The most notable example in acoustics is a time-dependent system's unforced, undamped, steady state [26]. The solutions to this problem are time-harmonic and, hence, the BVP is an *eigenvalue problem* (in the case of the Laplace operator, the defining equation is known as the *Helmholtz equation*). Time-harmonic solutions also arise when a system is forced at a specific frequency. In this case, if a single point in space can approximate the spatial extent of the source, the solution to the spatial differential problem is a *Green's function*. Such descriptions are commonplace in acoustics, where problems are often linear and time-invariant and can be solved equivalently in the time and frequency domains. The frequency-domain approach is sometimes preferable since closed-form solutions are easier to obtain, and most acoustic wave propagation properties in media are naturally described as a function of frequency, such as decay times, absorption coefficients, and so on. Additionally, frequency-domain approaches serve as the basis for many inverse modelling techniques, such as material parameter estimation.

1.4 Modal Methods

Applications in the frequency domain are restricted to steady-state or time-harmonic problems. However, simulating transient responses requires an appropriate update of the acoustic equations over time. Once spatial semi-discretisation is carried out through finite differences or finite elements, the resulting system of ordinary differential equations may be advanced in time using a suitable time-stepping scheme. Time is often discretised using a finite-difference approach, typically employing a constant sample rate throughout the simulation. Many approaches, collectively known as *time-domain* methods, operate directly on the spatially discretised equations without further transformation. Examples range from room acoustics [27, 19, 12], to musical instrument simulation [28, 29, 30, 24, 31]. Time-domain methods have numerous applications in acoustics. Finite-difference time-domain (FDTD) techniques gained prominence in the 1990s, largely due to the pioneering work of Chaigne and colleagues; see, e.g., [32, 33, 34], and have since become a key simulation method in acoustics, see [24]. Finite-element approaches in the time domain have experienced a similar trajectory, thanks to influential work by Hughes [6, 35], Joly and associates [36, 37, 38], and others. Besides time-domain and frequency-domain approaches, hybrid methods exist. Modal methods are one such prominent example. In this approach, spatial discretisation is first completed, after which the resulting semi-discretised system is projected onto a modal basis to derive the modal equations. These modal equations are then advanced in time, and the solution is reconstructed at specific output points using a reduced modal sum, see e.g. [39, 40, 41, 42, 43]. This approach entails additional steps before time integration compared to purely time-domain methods. These preliminary steps are often performed offline and may become a computational bottleneck for larger systems. However, once the modal equations are obtained, the system reduces to a set of parallel oscillators that can be updated highly efficiently over time. Additionally, exact time integrators exist for linear problems, effectively eliminating artefacts from time discretisation errors [44, 45]. This modal approach may thus be preferable for certain acoustic systems

and will be further illustrated in this lecture. Of course, synthesis methods exist that employ a variety of numerical techniques, such as the boundary conditions and body radiativity via modal methods and wave propagation via time-domain methods, see e.g. [46, 47].

1.5 Book Outline

This lecture develops the method of finite differences from first principles. Chapter 2 derives approximations of derivatives in one dimension, demonstrating how difference operators may be constructed using straightforward secondary school calculus. Interpolation, a somewhat more advanced topic, is subsequently introduced, enabling the construction of more general difference operators through the coefficients of interpolating polynomials. This approach proves particularly effective in constructing difference operators for functions sampled at non-uniform intervals. Grids and grid functions are subsequently introduced, building upon the preceding discussion.

Chapter 3 presents the first application of the finite difference method via the solution of simple boundary value problems (BVPs) in one dimension. These problems, which admit analytical solutions through direct integration, introduce concepts such as the order of convergence of a difference scheme. Both uniform and non-uniform grids are employed in the solutions, thereby disproving the common misconception that difference schemes are restricted to domains with constant mesh sizes. Eigenvalue problems in acoustics are then explored as a special case of a BVP, specifically in the context of the one-dimensional wave equation and the Euler-Bernoulli beam with variable thickness.

The foundational ideas in vibration modelling are introduced in Chapter 4. Vibration models are expressed in the time domain, and relevant analysis techniques are introduced. These include frequency-domain methods such as the Laplace and Fourier transforms, phase-space analysis and energy methods. The latter has the advantage of generalising well to the nonlinear case. The chapter focuses mainly on lumped systems described by a single degree of freedom, such as the harmonic oscillator and several generalisations, including damping, sources and nonlinearities. The chapter ends with a discussion on systems with multiple degrees of freedom.

Chapter 6 introduces the time discretisation methods. A review of the most common time difference operators is given analogously to the spatial case discussed earlier. Importantly, discrete-time versions of the frequency-domain and energy methods discussed earlier are given, including z transform techniques and time difference operator identities.

Chapter 7 introduces an application to the oscillator equations. Discrete-time counterparts of the frequency-domain and energy techniques are discussed. The convergence of time-stepping schemes in simple cases is discussed using formal arguments and proofs.

Chapter 2

Finite Difference Approximations

Suppose one wants to derive an approximate value for the derivative of a function of one variable. Let the function be $u = u(x) : \mathcal{I} \subseteq \mathbb{R} \to \mathbb{R}$, where $\mathcal{I}$ is a *closed interval* in $\mathbb{R}$, and let the function be sufficiently smooth, such that one may compute derivatives u', u'', ... and these are continuous. Furthermore, consider the small parameter $\Delta_x > 0$. The derivative of $u(x)$ at $x_0 \in \mathcal{I}$ may be defined in terms of the limit of a difference:

$$u'(x_0) = \lim_{\Delta_x \to 0} \frac{u(x_0 + \Delta_x) - u(x_0)}{\Delta_x}. \tag{2.1}$$

Of course, this definition is not unique.

Definition 2.0.1. The *identity*, *forward shift* and *backward shift* operators are defined as, respectively:

$$1u(x) = u(x), \quad e_{x+}u(x) = u(x + \Delta_x), \quad e_{x-}u(x) = u(x - \Delta_x) \tag{2.2}$$

□

Definition 2.0.2. The *forward*, *backward*, *centred* difference operators are defined as, respectively:

$$\delta_{x+} = \frac{e_{x+} - 1}{\Delta_x}, \quad \delta_{x-} = \frac{1 - e_{x-}}{\Delta_x}, \quad \delta_{x\cdot} = \frac{e_{x+} - e_{x-}}{2\Delta_x}. \tag{2.3}$$

□

These definitions can be used interchangeably to define the continuous derivative, such that one has:

$$u'(x_0) = \lim_{\Delta_x \to 0} \delta_{x+}u(x_0) = \lim_{\Delta_x \to 0} \delta_{x-}u(x_0) = \lim_{\Delta_x \to 0} \delta_{x\cdot}u(x_0). \tag{2.4}$$

Applying the definition of the difference operators to the function $u(x)$ yields the expres-

sions:

$$\delta_{x+}u(x) = \frac{u(x+\Delta_x)-u(x)}{\Delta_x}, \tag{2.5a}$$

$$\delta_{x-}u(x) = \frac{u(x)-u(x-\Delta_x)}{\Delta_x}, \tag{2.5b}$$

$$\delta_{x\cdot}u(x) = \frac{u(x+\Delta_x)-u(x-\Delta_x)}{2\Delta_x}. \tag{2.5c}$$

When Δ_x is finite, the difference operators yield an approximate value of the derivative, and an error is introduced.

2.1 Truncation errors

The error introduced by the difference operators is obtained via Taylor series arguments. Since the function $u(x)$ is smooth, one has:

$$u(x_0+\Delta_x) \approx u(x_0)+\Delta_x u'(x_0)+\frac{\Delta_x^2}{2}u''(x_0)+\frac{\Delta_x{}^3}{6}u'''(x_0), \tag{2.6a}$$

$$u(x_0-\Delta_x) \approx u(x_0)-\Delta_x u'(x_0)+\frac{\Delta_x^2}{2}u''(x_0)-\frac{\Delta_x{}^3}{6}u'''(x_0). \tag{2.6b}$$

These can be used to infer the order of the approximation of the difference operators:

$$\delta_{x+}u(x_0) \approx u'(x_0)+\frac{\Delta_x}{2}u''(x_0) = u'(x_0)+\mathcal{O}(\Delta_x), \tag{2.7a}$$

$$\delta_{x-}u(x_0) \approx u'(x_0)-\frac{\Delta_x}{2}u''(x_0) = u'(x_0)+\mathcal{O}(\Delta_x), \tag{2.7b}$$

$$\delta_{x\cdot}u(x_0) \approx u'(x_0)+\frac{\Delta_x^2}{6}u'''(x_0) = u'(x_0)+\mathcal{O}(\Delta_x^2). \tag{2.7c}$$

Note that the odd powers drop out of the Taylor series for $\delta_{x\cdot}$, a property typical of centred operators.

Definition 2.1.1. The *truncation error* is defined as:

$$E_{\delta_\circ} := \delta_\circ u(x_0)-u'(x_0), \tag{2.8}$$

where $\delta_\circ$ is any of the difference operators defined in (2.3).

Definition 2.1.2. The *order of accuracy* p of a difference operator is defined as the exponent of the leading term in the Taylor series of the error E:

$$E_{\delta_\circ} \approx C\Delta_x^p, \rightarrow p = \log(\Delta_x)^{-1}\left(\log(|E_{\delta_\circ}|)-\log(|C|)\right), \tag{2.9}$$

where $\delta_\circ$ is any of the difference operators defined in (2.3) □

Thus, from (2.7), the forward and backward differences are *first-order* accurate, whereas the centred difference is *second-order* accurate.

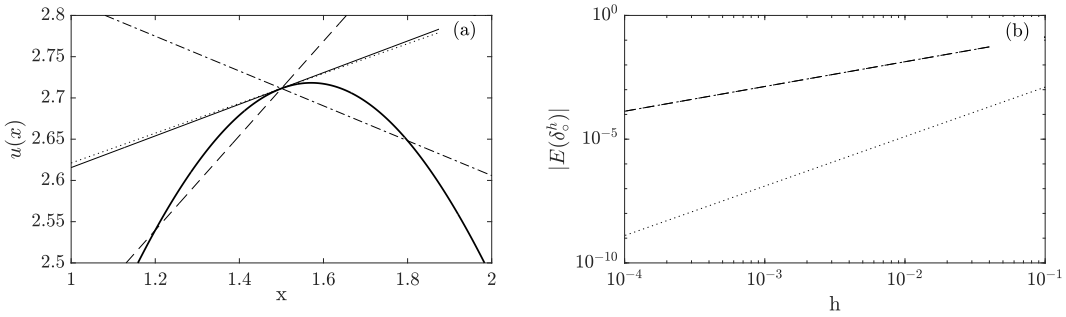


Figure 2.1: Approximation of the derivative of $u(x) = e^{\sin(x)}$ using finite differences. (a): slopes obtained using δ_{x+} (dashed), δ_{x-} (dash-dotted), $\delta_{x\cdot}$ (dotted), and exact (solid). In all cases, $x_0 = 1.5$, $\Delta_x = 0.3$. The continuous, thick line is $u(x)$. (b): log-log plot of the absolute value of the truncation error for the difference operators, with the same line style as panel (a).

Δ_x	$E_{\delta_{x+}}$	$E_{\delta_{x-}}$	$E_{\delta_{x\cdot}}$
0.1000	-0.135	0.133	-0.00127
0.0178	-0.024	0.0239	-4.03e-5
0.0032	-0.00426	0.00425	-1.27e-6
0.0006	-7.57e-4	7.57e-4	-4.03e-8
0.0001	-1.35e-4	1.35e-4	-1.27e-9

Table 2.1: Truncation errors.

Example 2.1.1. To understand the behaviour of the difference operators, the derivatives of $e^{\sin(x)}$ at $x_0 = 1.5$ are computed using the definitions (2.3). Figure 2.1 reports the slopes and the error trends of the finite difference operators, under various values of the small parameter Δ_x. The error values are reported in Table 2.1. One important aspect concerns the power expansions (2.7): the derivatives of $u(x)$ appearing in them do not depend on Δ_x, and are, therefore, constants when the expansion point x_0 is kept fixed. Thus, one should expect the errors in Table 2.1 to have the form (2.7), with the expansion coefficients given by the appropriate derivatives of $u(x)$. This is, indeed, the case. Running a polynomial fit on the columns of Table 2.1 yields:

$$E_{\delta_{x+}} \approx -1.3540\,\Delta_x, \quad E_{\delta_{x-}} \approx 1.3280\,\Delta_x, \quad E_{\delta_{x\cdot}} \approx -0.1269\,\Delta_x^2,$$

and note that:

$$\frac{u''(x_0)}{2} \approx -1.3456, \quad \frac{u'''(x_0)}{6} \approx -0.1275, \tag{2.10}$$

Hence, the numerical error trends are in good agreement with the corresponding analytical expressions. Since the errors are defined as powers, according to definition (2.9), log plots such as panel (b) of Figure 2.1 can be employed, so that the order of accuracy p appears as the slope of the error lines □

Before proceeding, it is worth illustrating the Taylor series of the backward and forward

difference operators when applied at $x_0 \pm \Delta_x/2$. Formally:

$$\begin{aligned} \delta_{x+} u\left(x_0 - \frac{\Delta_x}{2}\right) &= \delta_{x-} u\left(x_0 + \frac{\Delta_x}{2}\right) \\ &= \frac{u\left(x_0 + \frac{\Delta_x}{2}\right) - u\left(x_0 - \frac{\Delta_x}{2}\right)}{\Delta_x}. \end{aligned} \tag{2.11}$$

Note that these operators are centred around x_0 and, hence, one can expect the odd powers to drop out of the Taylor series, yielding higher-accurate operators. This is, indeed, the case. One has:

$$\begin{aligned} \delta_{x+} u\left(x_0 - \frac{\Delta_x}{2}\right) &= \delta_{x-} u\left(x_0 + \frac{\Delta_x}{2}\right) \\ &\approx u'(x_0) + \frac{\Delta_x^2}{8} u'''(x_0) = u'(x_0) + \mathcal{O}(\Delta_x^2). \end{aligned}$$

2.2 Polynomial interpolation

The discussion in the previous section suggests building the difference operators using the values of a function at arbitrary locations, provided the correct coefficients are applied at the sampled points of the function. In essence, the problem of finding appropriate difference coefficients amounts to finding an appropriate interpolating function [8]. See, e.g. [48, 49] for an introduction to interpolation.

2.2.1 Vandermonde method

One may find such weights using the Taylor series approach. Suppose to have sampled the function $u(x)$ at x_0, $x_0 + \Delta_x$, $x_0 + 2\Delta_x$. The sample points are equally spaced in this case, but the technique described below would work no matter where the points are located. In this example, the difference operator is defined as the following:

$$\delta_{x\star} u(x_0) := c_0\, u(x_0) + c_1\, u(x_0 + \Delta_x) + c_2\, u(x_0 + 2\Delta_x), \tag{2.12}$$

for unknown weight coefficients c_0, c_1, c_2. One may expand $u(x_0 + \Delta_x)$ as suggested in (2.6), and use an analogous expansion for $u(x_0 + 2\Delta_x)$. Inserting the expansions in (2.12), one obtains:

$$\delta_{x\star} u(x_0) \approx (c_0 + c_1 + c_2)\, u(x_0) + (c_1 + 2c_2)\Delta_x\, u'(x_0) + \frac{(c_1 + 4c_2)\Delta_x^2}{2}\, u''(x_0).$$

For this to approximate the first derivative, one requires:

$$c_0 + c_1 + c_2 = 0, \quad (c_1 + 2c_2)\Delta_x = 1, \quad c_1 + 4c_2 = 0, \tag{2.13}$$

yielding

$$c_0 = -\frac{3}{2\Delta_x}, \quad c_1 = \frac{2}{\Delta_x}, \quad c_2 = -\frac{1}{2\Delta_x}. \tag{2.14}$$

Notice that the coefficients are proportional to Δ_x^{-1}, and hence one can expect $\delta_{x\star}$ to be second-order accurate, that is:

$$\delta_{x\star}u(x_0) = u'(x_0) + \mathcal{O}(\Delta_x^2). \tag{2.15}$$

This approach can be extended further to derive a general formula for constructing difference operators. Suppose to have sampled the function $u(x)$ at $M+1$ points $x_0, ..., x_M$ (not necessarily equally spaced) and to be wanting to compute the coefficients of an $M+1$-point difference operator acting at $\bar{x}$ (this may or may not be equal to any of the sampling points x_m). One constructs a polynomial $p(x)$ of degree M as:

$$p(x) = \alpha_0 + \alpha_1(x-\bar{x}) + \alpha_2(x-\bar{x})^2 + ... + \alpha_M(x-\bar{x})^M, \tag{2.16}$$

and imposes $p(x_0) = u(x_0)$, $p(x_1) = u(x_1)$, ..., $p(x_M) = u(x_M)$. This can be expressed conveniently via the *Vandermonde matrix* as:

$$\begin{bmatrix} 1 & (x_0-\bar{x}) & (x_0-\bar{x})^2 & \cdots & (x_0-\bar{x})^M \\ 1 & (x_1-\bar{x}) & (x_1-\bar{x})^2 & \cdots & (x_1-\bar{x})^M \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (x_M-\bar{x}) & (x_M-\bar{x})^2 & \cdots & (x_M-\bar{x})^M \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_M \end{bmatrix} = \begin{bmatrix} u(x_0) \\ u(x_1) \\ \vdots \\ u(x_M) \end{bmatrix}. \tag{2.17}$$

Solving this system returns the coefficients α_m as a function of the sampled points $u(x_m)$. Note that the system is well-defined when the sample points are distinct. Then, the analytic expression of the $M+1$-point difference operator acting at $\bar{x}$ is recovered by computing $p'(\bar{x})$, that is, α_1.

Example 2.2.1. The coefficients c_0, c_1, c_2 in (2.12) are computed using the Vandermonde matrix method. To that end, note that $\bar{x} = x_0$, and that $x_1 - x_0 = \Delta_x$, $x_2 - x_0 = 2\Delta_x$. Thus:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & \Delta_x & \Delta_x^2 \\ 1 & 2\Delta_x & 4\Delta_x^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} u(x_0) \\ u(x_0+\Delta_x) \\ u(x_0+2\Delta_x) \end{bmatrix}. \tag{2.18}$$

Solving the system, one obtains:

$$\alpha_1 = -\frac{3u(x_0)}{2\Delta_x} + \frac{2u(x_0+\Delta_x)}{\Delta_x} - \frac{u(x_0+2\Delta_x)}{2\Delta_x}, \tag{2.19}$$

that is, the same as (2.12) with coefficients (2.14) □

2.2.2 Lagrange method

In some cases, particularly for large M, the Vandermonde matrix becomes poorly conditioned, and small perturbations in the matrix entries produce large errors in the interpolating coefficients α_m. In practice, large values of $M+1$ are never needed to build difference operators. However, it is worth illustrating an alternative way of building the interpolating polynomial $p(x)$ through Lagrange polynomials. Applications of Lagrange interpolants are much wider than merely deriving finite difference coefficients and are thus worth discussing here. As before, assume to have sampled the function $u(x)$ at $M+1$ points $x_0, x_1, ..., x_M$.

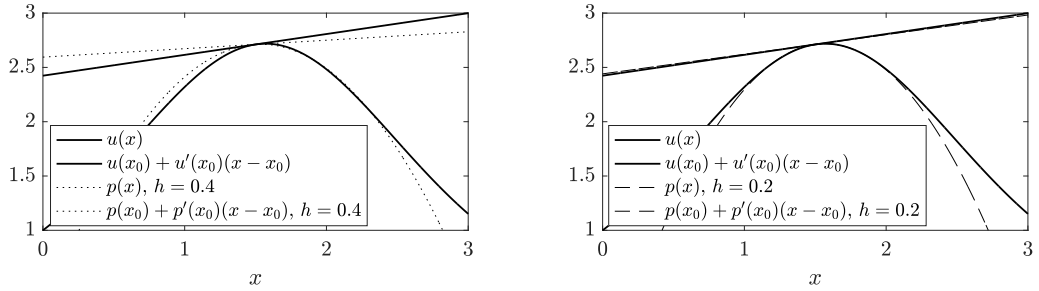


Figure 2.2: The function $u(x) = e^{\sin(x)}$, the interpolating polynomials $p(x)$ obtained via the Vandermonde method, using the sampled points $u(x_0)$, $u(x_0 + \Delta_x)$, $u(x_0 + 2\Delta_x)$, with $\Delta_x = 0.4$ (left) and $\Delta_x = 0.2$ (right) and $x_0 = 1.5$. The slopes at x_0 are also plotted.

Definition 2.2.1. The Lagrange basis functions $l_m(x)$, $m = 0, ..., M$ are defined as:

$$\begin{aligned} l_m(x) &:= \frac{(x-x_0)...(x-x_{m-1})(x-x_{m+1})...(x-x_M)}{(x_m-x_0)...(x_m-x_{m-1})(x_m-x_{m+1})...(x_m-x_M)} \\ &= \frac{\prod_{n\neq m}(x-x_n)}{\prod_{n\neq m}(x_m-x_n)}, \end{aligned}$$

and the associated Lagrange polynomial is:

$$p(x) = \sum_{m=0}^{M} u(x_m)\, l_m(x). \tag{2.20}$$

Note that $p(x_m) = u(x_m)\, \forall m$, which is the required condition for interpolation □

Compared to the Vandemorde matrix method, the Lagrange interpolant is built explicitly and not via the solution of a linear system. This allows the construction of polynomials of large degrees without much difficulty. Furthermore, if the function $u(x)$ is sampled further via $u(x_{M+1})$, $u(x_{M+2})$, ..., the first $M + 1$ Lagrange basis functions can be computed by adjusting the old ones: one only needs to multiply the numerator and the denominator of each $l_m(x)$ by the factors associated with the new sample points. As before, the analytic expression of the $M + 1$-point difference operator acting at $\bar{x}$ is recovered by computing $p'(\bar{x})$.

Example 2.2.2. The coefficients c_0, c_1, c_2 in (2.12) are computed using Lagrange interpolation. In this case, one has:

$$l_0 = \frac{(x-x_1)(x-x_2)}{2\Delta_x^2}, \quad l_1 = -\frac{(x-x_0)(x-x_2)}{\Delta_x^2}, \quad l_2 = \frac{(x-x_0)(x-x_1)}{2\Delta_x^2}. \tag{2.21}$$

Using these to construct $p(x)$ as in (2.20), and computing $p'(x_0)$, one obtains again the coefficients (2.14) □

2.3 Approximation of higher derivatives

All the techniques illustrated in the case of the first derivative extend directly to the cases of higher derivatives.

subsubsectionSecond derivative The second derivative in one dimension is the simplest example of the *Laplace operator*, appearing in many physical laws modelling a variety of phenomena: elastic bending, wave propagation, and heat conduction, just to name a few. This operator, thus, deserves special treatment. In the continuous case, a definition of the second derivative is often given as:

$$\begin{aligned} u''(x_0) &= \lim_{\Delta_x \to 0} \frac{(\delta_{x+} - \delta_{x-})u(x_0)}{\Delta_x} \\ &= \lim_{\Delta_x \to 0} \frac{u(x_0 + \Delta_x) - 2u(x_0) + u(x_0 - \Delta_x)}{\Delta_x^2}. \end{aligned} \tag{2.22}$$

Definition 2.3.1. The *second difference* operator is defined as:

$$\delta_{xx} := \frac{\delta_{x+} - \delta_{x-}}{\Delta_x} = \frac{e_+ - 2 + e_-}{\Delta_x^2} \tag{2.23}$$

□

It is easy to obtain the Taylor series of this operator using (2.6). Thus, one has:

$$\begin{aligned} \delta_{xx}u(x_0) &= \frac{u(x_0 + \Delta_x) - 2u(x_0) + u(x_0 - \Delta_x)}{\Delta_x^2} \\ &\approx u''(x_0) + \frac{\Delta_x^2}{12}u''''(x_0) = u'' + \mathcal{O}(\Delta_x^2). \end{aligned}$$

Since δ_{xx} is a centred operator, the odd terms drop out of the series. Note as well that the error is second-order in the expansion parameter Δ_x. Various other definitions are possible, and obtainable using much of the same techniques as detailed in Section (2.2).

Example 2.3.1. The second derivative is approximated at x_0 via Lagrange interpolation, by sampling $u(x)$ at the points $x_{-2} = x_0 - 2\Delta_x$, $x_{-1} = x_0 - \Delta_x$, x_0, $x_1 = x_0 + \Delta_x$, $x_2 = x_0 + 2\Delta_x$. In this case, one has:

$$\begin{aligned} l_{-2} &= \frac{(x - x_{-1})(x - x_0)(x - x_1)(x - x_2)}{24\Delta_x^4}, \\ l_{-1} &= \frac{(x - x_{-2})(x - x_0)(x - x_1)(x - x_2)}{6\Delta_x^4}, \ \dots . \end{aligned}$$

These are used to construct $p(x)$ as in (2.20). To obtain the second difference operator, the second derivative is required, giving:

$$p''(x_0) = -\frac{u(x_0 - 2\Delta_x)}{12\Delta_x^2} + \frac{4u(x_0 - \Delta_x)}{3\Delta_x^2} - \frac{5u(x_0)}{2\Delta_x^2} + \frac{4u(x_0 + \Delta_x)}{3\Delta_x^2} - \frac{u(x_0 + 2\Delta_x)}{12\Delta_x^2}.$$

Computing the Taylor series, one has:

$$p''(x_0) = u''(x_0) + \mathcal{O}(\Delta_x^4), \tag{2.24}$$

and, hence, fourth-order accuracy is achieved. Whilst the definition of δ_{xx} was given in terms of the difference of δ_{x+} and δ_{x-}, note that the following identity holds, as one may show immediately:

$$\delta_{xx} = \delta_{x+}\delta_{x-} = \delta_{x-}\delta_{x+}. \tag{2.25}$$

Thus, higher difference operators may be constructed by *composition*. This idea will be exploited later when interpreting finite difference operators as matrices acting on grid functions: in that framework, the composition of the operators is realised by multiplying the corresponding matrices □

2.3.1 Third and fourth derivatives

One may go on and build difference operators for higher derivatives. The fourth derivative, also known as the *biharmonic operator*, appears in the models of the thin bar and is worth introducing here. One has:

$$\begin{aligned} u''''(x_0) &= \lim_{\Delta_x \to 0} \frac{u''(x_0 - \Delta_x) - 2u''(x_0) + u''(x_0 + \Delta_x)}{\Delta_x^2} = \\ &= \lim_{\Delta_x \to 0} \frac{u(x_0 - 2\Delta_x) - 4u(x_0 - \Delta_x) + 6u(x_0) - 4u(x_0 + \Delta_x) + u(x_0 + 2\Delta_x)}{\Delta_x^4}. \end{aligned}$$

Definition 2.3.2. The *fourth difference* operator is defined as:

$$\delta_{xxxx} := \delta_{xx}\delta_{xx} = \frac{e_{x-}^2 - 4e_{x-} + 6 - 4e_{x+} + e_{x+}^2}{\Delta_x^4} \tag{2.26}$$

□

Expanding the operator using a Taylor series, one obtains:

$$\delta_{xxxx}u(x_0) = u''''(x_0) + \mathcal{O}(\Delta_x^2), \tag{2.27}$$

that is, the operator is second-order accurate. Third derivatives appear when analysing the boundary conditions of the biharmonic operator. The composition of operators can be used such that $\delta_{xx}\delta_{x+}$, $\delta_{xx}\delta_{x-}$, and $\delta_{xx}\delta_{x\cdot}$ can all be employed to approximate the third derivative. The third difference operators will be denoted as follows:

$$\delta_{xxx+} := \delta_{xx}\delta_{x+}\,, \quad \delta_{xxx-} := \delta_{xx}\delta_{x-}\,, \quad \delta_{xxx\cdot} := \delta_{xx}\delta_{x\cdot}\,. \tag{2.28}$$

The order of the approximations may be computed via the Taylor series and left as an exercise for the reader. It is easy to show that both δ_{xxx+} and δ_{xxx-} are first-order accurate operators: they inherit this property directly from the first-order accuracy of the difference operators δ_{x+}, δ_{x-}. The centred operator $\delta_{xxx\cdot}$ is second-order accurate, as all the odd terms drop out of its Taylor series.

2.4 Grids and grid functions

The concepts introduced in the previous sections allow estimating the derivatives of a sampled, continuous function $u(x)$ defined over a closed interval $\mathcal{I} \subseteq \mathbb{R}$. The finite difference coefficients appearing in the definitions of the difference operators can be found by interpolating the continuous function at the available sample points x_m, yielding several difference operators with varying accuracy.

It is yet unclear how one may use such techniques to solve a differential problem where $u(x)$ and its derivatives appear. The method of finite differences is used to compute an *approximate* solution at specific locations in $\mathcal{I}$. Suppose that the interval is bounded so that $\mathcal{I} = \{x \,|\, 0 \leq x \leq L\}$. Just like several sample points were used above to interpolate polynomials, a set of discrete points belonging to $\mathcal{I}$ is used to compute an approximate solution to a differential problem.

Definition 2.4.1. One refers to a *grid* (sometimes also called a *mesh*) as the collection of *grid points* (also called the *nodes*) x_m and related *cells* $[x_m, x_{m+1}]$ such that:

$$0 = x_0 < x_1 < ... < x_m < x_{m+1} ... < x_M = L \tag{2.29}$$

□

Various kinds of grids can be defined, [25]:

- The simplest grid is certainly the *uniform grid*: here, all the grid points are separated by the *grid spacing* Δ_x, and one has $x_m = m\Delta_x$, $m \in [0, M] \subset \mathbb{N}$, and the number of grid intervals is $M = \Delta_x / L$. This type of grid is often used in the simulation of isotropic, homogeneous problems in which the properties of the medium are independent of space.
- The *smooth, non-uniform grid* is defined by the smooth function $y(x)$, and one has $y_m = y(x_m)$. The smooth function y maps $\mathcal{I}$ onto itself and is one-to-one. An example is given by $y = (L^2 - x^2)/L$.
- The *non-smooth, non-uniform grid* is a non-uniform grid where $y(x)$ is not smooth. An example is given by a grid whose distance between nodes alternates $\Delta_x, 2\Delta_x, \Delta_x, ...$, or where the grid intervals are sampled from a random distribution.

In all cases, the length between two nodes is bounded by a smallest and by a largest value, such that

$$C_{\text{min}}\, \Delta_x \leq x_{m+1} - x_m \leq C_{\text{max}}\, \Delta_x, \tag{2.30}$$

where Δ_x is a typical cell size (for instance, the mean value of all the cells).

Example 2.4.1. For the uniform grid, $C_{\text{min}} = C_{\text{max}} = 1\ \forall M$. For the grid defined by $y_{M-m} = (L^2 - x_m^2)/L$ with $L = 1.2$, $M = 120$, one has $\Delta_x = 0.01$, $C_{\text{min}} = 0.0083$, $C_{\text{max}} = 1.9917$ □

Another important grid is the *interleaved grid*, parametrised by a half-integer. So, all the grid points are still separated by a Δ_x, but they are found at $x_{m+\frac{1}{2}} = (m + \frac{1}{2})\Delta_x$, $m \in [1, M-1] \subset \mathbb{N}$

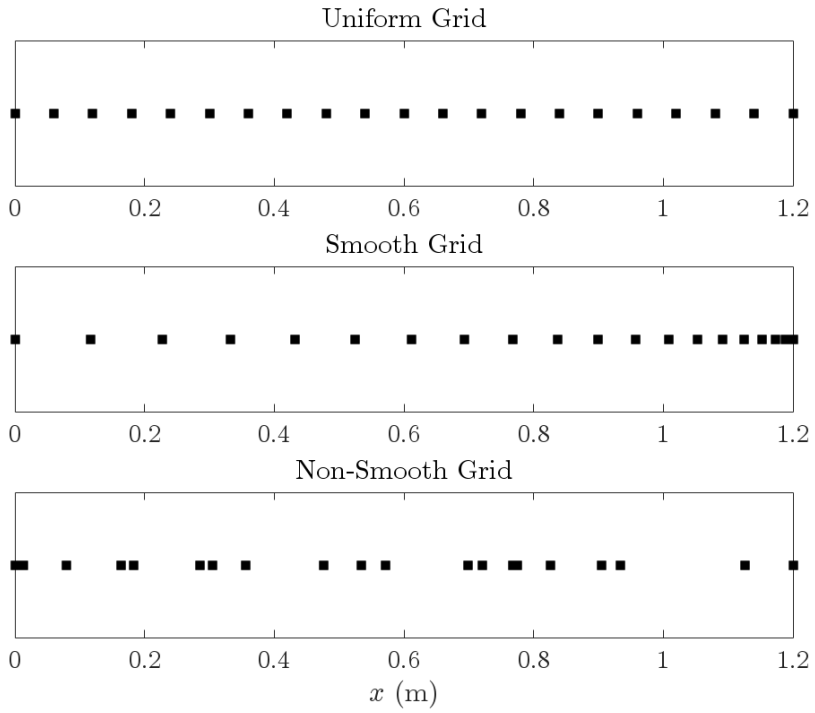


Figure 2.3: Examples of uniform, smooth and non-smooth grids, using $M = 20$ and $L = 1.2$. The smooth grid is obtained as $y_{M-m} = (L^2 - x_m^2)/L^2$.

Group	Definition
Identity	$1u_m = u_m$
Shift	$e_{x+}u_m = u_{m+1}$ (Forward) $e_{x-}u_m = u_{m-1}$ (Backward)
First Difference	$\delta_{x+}u_m = \Delta_x^{-1}(u_{m+1} - u_m)$ (Forward) $\delta_{x-}u_m = \Delta_x^{-1}(u_m - u_{m-1})$ (Backward) $\delta_{x\cdot}u_m = \frac{1}{2}\Delta_x^{-1}(u_{m+1} - u_{m-1})$ (Centred)
Second Difference (Laplacian)	$\delta_{xx}u_m = \Delta_x^{-2}(u_{m+1} - 2u_m + u_{m-1})$ (Centred)
Third Difference	$\delta_{xxx+}u_m = \Delta_x^{-3}(u_{m+2} - 3u_{m+1} + 3u_m - u_{m-1})$ (Forward) $\delta_{xxx-}u_m = \Delta_x^{-3}(u_{m+1} - 3u_m + 3u_{m-1} - u_{m-2})$ (Backward) $\delta_{xxx\cdot}u_m = \frac{1}{2}\Delta_x^{-3}(u_{m+2} - 2u_{m+1} + 2u_{m-1} - u_{m-2})$ (Centred)
Fourth Difference (Biharmonic)	$\delta_{xxxx}u_m = \Delta_x^{-4}(u_{m+2} - 4u_{m+1} + 6u_m - 4u_{m-1} + u_{m-2})$ (Centred)

Table 2.2: Summary of the spatial finite difference operators acting on a grid function u_m defined on a uniform grid with grid spacing Δ_x.

Definition 2.4.2. A *grid function* $\mathbf{u} : \mathfrak{I} \to \mathbb{R}^{M+1}$ is any $M + 1 \times 1$ vector defined at the grid points x_m. In practice, grid functions approximate the "true" solution of a differential problem at the grid locations. It will be convient to denote grid functions using indices, such that $u_m := (\mathbf{u})_m$ (the m-th component of the vector $\mathbf{u}$) □

2.4.1 Finite difference operators acting on grid functions

The difference operators defined above can be applied conveniently to grid functions. Though the formal definition of the operators changes slightly when applied to a grid function u_m as opposed to a continuous function $u(x)$, the interpretation is straightforward: the differences are now computed as differences of the *elements* of the vector $\mathbf{u}$. Table 2.2 reports the definitions of the spatial difference operators acting on the grid function u_m.

Definition 2.4.3. The finite difference method is used to solve a differential problem in $u(x)$ by computing a grid function $\mathbf{u}$ such that:

$$u_m = u(x_m) + \mathcal{O}(\Delta_x^q), \tag{2.31}$$

where q is the *order of convergence* of the method □

This definition implies that, as the grid spacing decreases towards zero, the grid function approximates the true solution exactly at the grid points. The definition (2.31) above is not entirely satisfying, as it implies a local convergence of one specific point in the domain. Refining a grid by making Δ_x smaller defines a sequence of grids $\mathfrak{I}^n$. The problem here is that the same grid point may not be present in all grids in the sequence, making the definition (2.31) impractical or ill-defined. It is more convenient, then, to define a form of global error over the whole grid, such as the *root mean square deviation*:

$$\text{RMSD}(\Delta_x) := \sqrt{\frac{\Delta_x}{L} \sum_{m=1}^{M-1} (u_m - u(x_m))^2}, \tag{2.32}$$

for which one also has $\text{RMSD}(\Delta_x) = \mathcal{O}(\Delta_x^q)$ when (2.31) holds.

The order of convergence q is closely related to the order of accuracy of the difference operators p, as defined in (2.9), though the two have different definitions and meanings. Solving a differential problem using p-th accurate difference operators often yields a p-th convergent finite difference scheme. In some cases, particularly when solving the model problem over non-uniform grids, the local truncation error may be as large as first-order for a given Δ_x, yet the RMSD has a higher convergence rate. This happens because the formal low order of accuracy is compensated by a rather small error at certain grid locations where the interval between grid points becomes much smaller than Δ_x. Determining the conditions for which the approximation (2.31) holds is the central problem of the finite difference method.

Chapter 3

Boundary Value Problems

The techniques illustrated in the preceding sections allow for the investigation of several problems of interest in acoustics. We shall begin from simple *boundary value problems*. These problems do not depend on time and form a subclass of the model problems encountered in acoustics, which usually involve spatial as well as temporal differential operators. The model equations are, thus, *partial differential equations*. Boundary value problems emerge as special cases when, for instance, the steady state of a system is considered, and the time dependence of the model problem may be simplified accordingly.

Definition 3.0.1. A boundary value problem (BVP) in one dimension is defined as:

$$f(u, u', u'', ..., u^{(n)}) = w(x), \tag{3.1}$$

where $w(x)$ is a source term, and n is the order of the BVP. To be complete, the BVP is supplied with n *boundary conditions* specifying the value of u and/or its derivatives up to the order $n-1$ at the domain's boundary □

3.1 The one-dimensional Poisson equation

It may be useful to start from the simple case of the one-dimensional Poisson equation, which, whilst not modelling an acoustic phenomenon as such, is useful as a test case for the finite difference techniques described in the previous section. This equation models the steady state of a diffusion-type system with a time-independent source, a problem encountered in fluid dynamics such as in the heat equation.

Definition 3.1.1. The one-dimensional Poisson equation with Dirichlet boundary conditions is defined as:

$$u''(x) = w(x), \quad u(0) = \mathfrak{u}_0,\ u(L) = \mathfrak{u}_L, \tag{3.2}$$

where $w(x)$ is a time-independent source, and $\mathfrak{u}_0, \mathfrak{u}_L$ are constants □

Clearly, an analytic solution can be obtained by directly integrating (3.2) twice and setting the two integration constants using the boundary conditions. The accuracy of the finite difference method can be assessed against such an analytic solution.

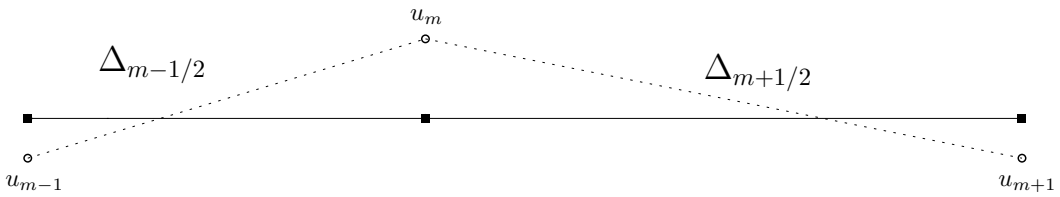


Figure 3.1: Three-point non-uniform mesh for the approximation of the Laplace operator.

Example 3.1.1. The Poisson equation for $w(x) = \tanh^2(x) - 1$, and with $\mathfrak{u}_0 = \mathfrak{u}_L = 0$ is solved via direct integration. In this case, the solution is

$$u(x) = L^{-1}x\log(\cosh(L)) - \log(\cosh(x)), \tag{3.3}$$

as one can verify after differentiating the solution twice □

Solving the Poisson equation involves approximating the Laplace operator. Here, approximations will be constructed on the three types of grids presented in Section 2.4 that is, the uniform grid, the non-uniform smooth grid and the non-uniform non-smooth grid. In all three cases, the Laplacian will be approximated using:

$$u''(x_m) \approx \frac{u(x_{m-1})}{\Delta_{m-1/2}\widetilde{\Delta}_m} - \frac{2u(x_m)}{\Delta_{m-1/2}\Delta_{m+1/2}} + \frac{u(x_{m+1})}{\Delta_{m+1/2}\widetilde{\Delta}_m} \tag{3.4}$$

where $\widetilde{\Delta}_m := (\Delta_{m-1/2} + \Delta_{m+1/2})/2$. The approximation (3.4) may be obtained using the Vandermonde or Lagrange polynomial method, as detailed in Section 2.2.

Assume to generate a mesh using $M+1$ (generally unequally spaced) sample points, as per (2.29). The boundary grid points u_0 and u_M need not be stored or updated since the boundary conditions fix their value. The Poisson equation defined by (3.2) is thus discretised as:

$$\frac{-2u_1}{\Delta_{1/2}\Delta_{3/2}} + \frac{u_2}{\Delta_{3/2}\widetilde{\Delta}_1} = w_1, \tag{3.5a}$$

$$\frac{-2u_2}{\Delta_{3/2}\Delta_{5/2}} + \frac{u_2}{\Delta_{5/2}\widetilde{\Delta}_2} = w_2, \tag{3.5b}$$

$$.... + ... = ...$$

$$\frac{-2u_{M-1}}{\Delta_{M-3/2}\Delta_{M-1/2}} + \frac{u_{M-2}}{\Delta_{M-3/2}\widetilde{\Delta}_{M-1}} = w_{M-1}, \tag{3.5c}$$

which may be cast more compactly as the $M-1 \times M-1$ system:

$$\mathbf{D}_{xx}\,\mathbf{u} = \mathbf{w} \tag{3.6}$$

with:

$$\mathbf{D}_{xx} := \begin{bmatrix} -\frac{2}{\Delta_{\frac{1}{2}}\Delta_{\frac{3}{2}}} & \frac{1}{\Delta_{\frac{3}{2}}\widetilde{\Delta}_1} & 0 & \dots & 0 \\ \frac{1}{\Delta_{\frac{3}{2}}\widetilde{\Delta}_2} & -\frac{2}{\Delta_{\frac{3}{2}}\Delta_{\frac{5}{2}}} & \frac{1}{\Delta_{\frac{5}{2}}\widetilde{\Delta}_2} & 0 & \vdots \\ 0 & \frac{1}{\Delta_{\frac{5}{2}}\widetilde{\Delta}_3} & -\frac{2}{\Delta_{\frac{5}{2}}\Delta_{\frac{7}{2}}} & \frac{1}{\Delta_{\frac{7}{2}}\widetilde{\Delta}_3} & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{1}{\Delta_{M-\frac{3}{2}}\widetilde{\Delta}_{M-1}} & -\frac{2}{\Delta_{M-\frac{1}{2}}\Delta_{M-\frac{3}{2}}} \end{bmatrix}$$

System (3.6) is tridiagonal: note that the rows of the matrix are linearly independent, and hence, the matrix is invertible, and a unique solution exists. One slightly worrying aspect is the absence of symmetry of the matrix for all but the special case of the uniform grid ($\Delta_{m-1/2} = \Delta_x \ \forall m$). This, however, may be remedied via a similarity transformation.

Definition 3.1.2. Two square matrices $\mathbf{A}, \mathbf{B}$ are called *similar* if there exist an invertible matrix $\mathbf{S}$ such that:

$$\mathbf{B} = \mathbf{S}\mathbf{A}\mathbf{S}^{-1} \tag{3.7}$$

□

Theorem 3.1.1. *Two similar matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *share the same eigenvalues. Furthermore, if* (3.7) *holds and* $\mathbf{a}$ *is an eigenvector of* $\mathbf{A}$ *with eigenvalue* λ*, then* $\mathbf{b} := \mathbf{S}\mathbf{a}$ *is an eigenvector of* $\mathbf{B}$ *with eigenvalue* λ*.*

The proof is immediate. By definition $\mathbf{A}\mathbf{a} = \lambda\mathbf{a}$. From here $\mathbf{S}\mathbf{A}\mathbf{S}^{-1}\mathbf{S}\mathbf{a} = \lambda\mathbf{S}\mathbf{a}$, from which the proof follows □

Example 3.1.2. The matrix $\mathbf{D}_{xx}$ is symmetrised using the similarity transformation $\mathbf{D}_{xx}^{sym} = \mathbf{S}\mathbf{D}_{xx}\mathbf{S}^{-1}$. To that end, consider the diagonal matrix:

$$[\mathbf{S}]_{m,m} := \prod_{j=2}^{m} \sqrt{\frac{\widetilde{\Delta}_j}{\widetilde{\Delta}_{j-1}}}, \ m = 2, ..., M-1, \tag{3.8}$$

and $[\mathbf{S}]_{1,1} = 1$. Then:

$$\mathbf{D}_{xx}^{sym} := \begin{bmatrix} -\frac{2}{\Delta_{\frac{1}{2}}\Delta_{\frac{3}{2}}} & \frac{1}{\Delta_{\frac{3}{2}}\sqrt{\widetilde{\Delta}_1\widetilde{\Delta}_2}} & 0 & \dots \\ \frac{1}{\Delta_{\frac{3}{2}}\sqrt{\widetilde{\Delta}_1\widetilde{\Delta}_2}} & -\frac{2}{\Delta_{\frac{3}{2}}\Delta_{\frac{5}{2}}} & \frac{1}{\Delta_{\frac{5}{2}}\sqrt{\widetilde{\Delta}_2\widetilde{\Delta}_3}} & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\Delta_{M-\frac{3}{2}}\sqrt{\widetilde{\Delta}_{M-1}\widetilde{\Delta}_{M-2}}} & -\frac{2}{\Delta_{M-\frac{1}{2}}\Delta_{M-\frac{3}{2}}} \end{bmatrix}$$

□

The numerical solutions on the three grids are plotted in Figure 3.2. A cursory inspection allows appreciating the qualitatively similar solutions computed on the three grids. The

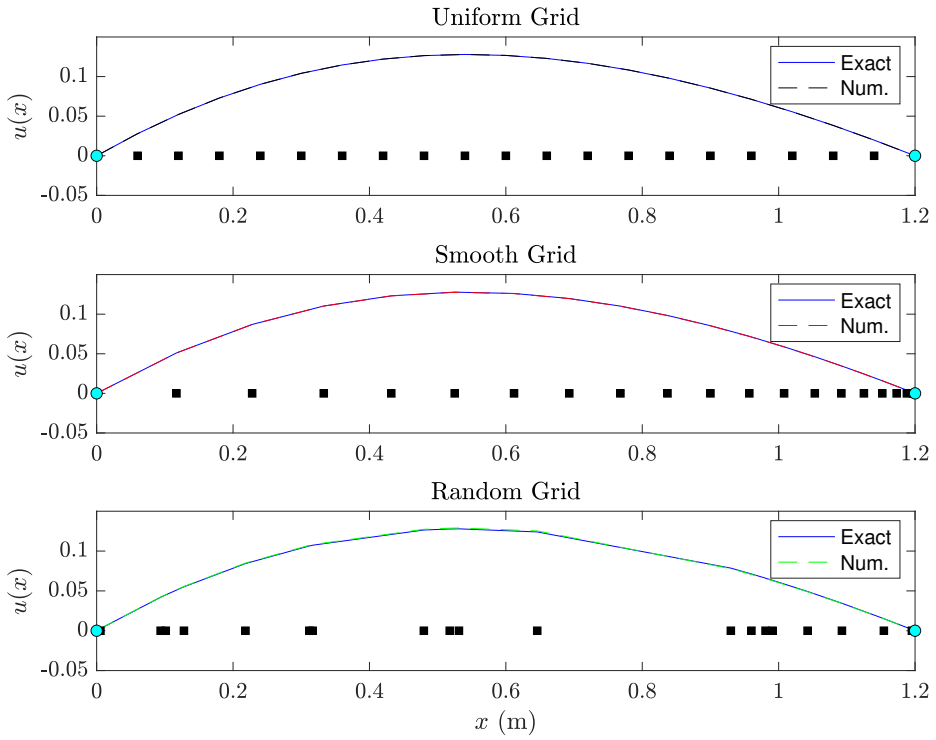


Figure 3.2: Numerical solutions of the Poisson equation (3.2) with $w(x) = \tanh^2(x) - 1$, using $L = 1.2$ and $M = 20$. The analytic solution (3.3) is also plotted at the grid points for reference.

solution computed on the random grid presents a somewhat less smooth behaviour when the interpolating spline is plotted through the solution points, but this is a consequence of the uneven distribution of the sample points rather than a degradation in the quality of the numerical approximation. This claim can be made more precise by plotting the RMSD defined in (2.32) as a function of $\Delta_x := L/M$, i.e. the grid spacing of the uniform grid. Such grid spacing can be used as a reference to compute the bounding constants C_{min}, C_{max} in (2.30), so that a cell's size never exceeds those bounds. Figure 3.3 plots the RMSD as a function of Δ_x. The uniform grid presents the smallest error of all three, followed by the smooth grid. However, the error behaviour is consistent for all three cases. In particular:

$$\text{RMSD}(\Delta_x) \approx \mathcal{O}(\Delta_x^2), \tag{3.9}$$

and, hence, the solutions computed on the three grids are second-order *convergent*. This is an important result since the local truncation error of the difference operator (3.4) is, in theory, only first-order in Δ_x for the non-uniform grids! But when the errors are computed globally, they compensate. This example also shows that one is better off using a uniform grid in this case, since not only $\mathbf{D}_{xx}$ is symmetric, but also Toeplitz. More importantly, the

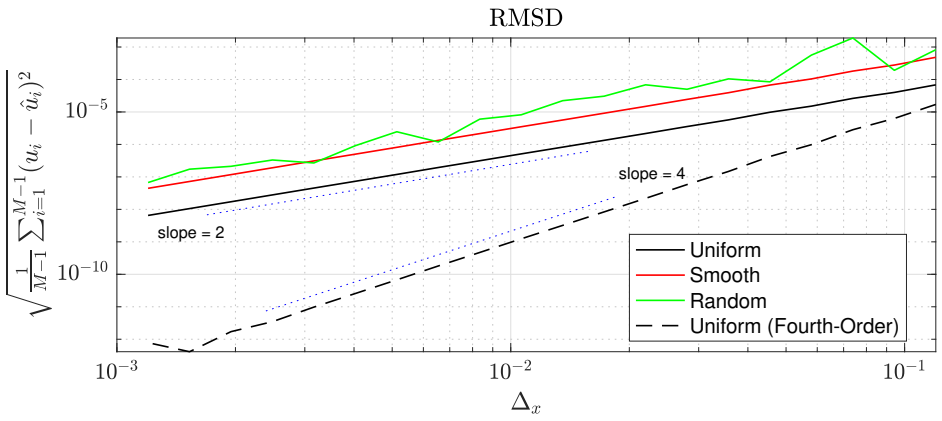


Figure 3.3: Root mean squared deviations (RMSD) as a function of $\Delta_x := L/M$. Blue dotted lines with slopes of two and four are also plotted for reference.

isotropy of the model problem (3.2) justifies the use of the uniform mesh on a physical basis. This makes the second-order convergence of the solutions computed on the non-uniform grids all the more remarkable. Using non-uniform grids is beneficial in simulating systems with varying spatial properties, as will be shown in later sections. Always in Figure 3.3, one sees the error trend of a solution computed on a uniform grid with a fourth-order accurate approximation of the Laplacian for interior points away from the boundaries, as per Example 2.3.1. Again, the expected error trend is recovered.

3.2 Eigenvalue Problems in One Dimension

Let us now turn the attention to *eigenvalue problems*. These crop up in various forms in the field of acoustics. Often, they originate after the temporal part of a model problem is transformed in the frequency domain via, e.g. a Fourier transform. Eigenvalue problems are a subclass of boundary value problems, as defined in (3.1), where f is linear, the source w is absent, and one is interested in finding the functions for which an action of the differential operator amounts to multiplying the function by a scalar.

Definition 3.2.1. For a linear spatial differential operator $\mathcal{L}$, the *eigenvalue problem* (EVP) is a subclass of the BVP defined in (3.1) such that:

$$\mathcal{L}\hat{u} = \lambda\hat{u}, \tag{3.10}$$

where $\hat{u}$ is called an *eigenfunction*, and λ is the corresponding *eigenvalue*. In most cases of interest in acoustics, the differential operator $\mathcal{L}$ is such that its eigenvalues are amenable to the squared resonant frequencies of the system, $\lambda = \omega^2$. Generally, the eigenvalue problem is unsolvable analytically, and one resorts to spatial discretisation of the problem to obtain a numerical eigenvalue problem of the form:

$$\mathbf{L}\hat{\mathbf{u}} = g(\lambda)\hat{\mathbf{u}}, \tag{3.11}$$

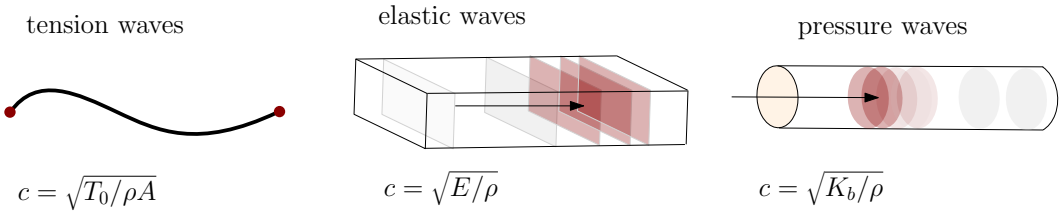


Figure 3.4: Examples of systems described by the one-dimensional wave equation: flexural waves in cables, longitudinal waves in bars, pressure waves in acoustics tubes.

where $\mathbf{L}$ is a sparse matrix discretising the differential operator $\mathcal{L}$, and $\hat{\mathbf{u}}$ is a corresponding eigenvector. $g(\lambda)$ is an approximation to the continuous eigenvalue λ, such that, usually, $g(\lambda) = \lambda + \mathcal{O}(\Delta_x^p)$, with $p \geq 1 \in \mathbb{N}$. The simplest examples are drawn from the simple wave equation, now described □

3.2.1 The wave equation

This simple equation describes a range of phenomena from tension waves in cables to longitudinal waves in elastic rods to acoustic pressure waves in tubes, taking the following forms:

$$\rho_c A \frac{\partial^2 u}{\partial t^2} = T_0 \frac{\partial^2 u}{\partial x^2}, \quad \rho_b \frac{\partial^2 v}{\partial t^2} = E \frac{\partial^2 v}{\partial x^2}, \quad \rho_a \frac{\partial^2 w}{\partial t^2} = K_a \frac{\partial^2 w}{\partial x^2}. \tag{3.12a}$$

In the above, u, v, w represent a cable's flexural displacement, a bar's longitudinal elongation, and air's longitudinal compression in a tube. ρ_c, ρ_b, ρ_a are the volume densities of the cable, the bar and the air, respectively; A is the cable's cross-section's area, T_0 is the applied tension; E is Young's modulus and K_a is the air's bulk modulus. From Figure 3.4, it results that the three equations can be written analogously as the one-dimensional wave equation with speed c.

Definition 3.2.2. The wave equation with speed c is given by:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}. \tag{3.13}$$

For tension waves in cables, u is a displacement, and we shall treat u as such in the forthcoming sections, knowing that the mathematics of the wave equation remains unchanged regardless of whichever physical system it describes □

Here, clearly, the state variable u is a function of both time and space, such that $u = u(t, x)$. It is convenient, for now, to consider temporal and spatial unbounded domains, such that $u = u(t, x) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. This is, of course, an abstraction since no physical system has existed from the beginning of time and is infinitely large, but this idealisation turns out to be useful for deriving a special relationship between the *frequency* and the *wavenumber* of a plane wave travelling through the system.

Definition 3.2.3. The *dispersion relation* for the wave equation is obtained as the relation $\omega = \omega(\gamma)$ after substituting the particular solution $u(t,x) = e^{j(\omega t+\gamma x)}$ (a plane wave) in (3.13), that is:

$$\omega = c\gamma, \tag{3.14}$$

where ω and γ are the radian temporal frequency and the wavenumber, respectively. These are real, positive numbers, i,e, $\omega, .\gamma \in \mathbb{R}_0^+$. These are related to the period of the plane wave $\mathfrak{t}$ and its wavelength $\mathfrak{l}$ by:

$$\mathfrak{t} = 2\pi\omega^{-1}, \quad \mathfrak{l} = 2\pi\gamma^{-1} \tag{3.15}$$

□

Definition 3.2.4. The *phase velocity* c_ϕ and the *group velocity* c_g are defined via the dispersion relation, as:

$$c_\phi := \frac{\omega}{\gamma}, \quad c_g := \frac{\mathrm{d}\omega}{\mathrm{d}\gamma}. \tag{3.16}$$

For the simple wave equation, both velocities are equal to c and, hence, independent of frequency. The wave equation is a *dispersionless* system □

To derive an eigenvalue problem, the wave equation must be supplied with appropriate boundary conditions. Let now the spatial domain be the closed, bounded interval $\mathcal{I} = \{x \,|\, x \in [0, L]\}$. Boundary conditions emerge naturally after an inspection of the energy of the continuous system (3.13).

Definition 3.2.5. For the wave equation, the total energy H is the sum of the kinetic and potential energies, defined as:

$$H_k := \frac{1}{2}\int_{\mathcal{I}} \left(\frac{\partial u}{\partial t}\right)^2 \mathrm{d}x, \quad H_p := \frac{c^2}{2}\int_{\mathcal{I}} \left(\frac{\partial u}{\partial x}\right)^2 \mathrm{d}x, \tag{3.17}$$

where the energies are here scaled by linear mass density in the case of tension wave in cables, and by the bar's and air's denisities in the case of compressional elastic waves in bars and acoustic tubes □

Taking the time derivative of the total energy $H = H_k + H_p$ and rearranging terms, one gets:

$$\int_{\mathcal{I}} \left(\frac{\partial u}{\partial t}\right)\left(\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2}\right)\mathrm{d}x = -c^2\frac{\partial u}{\partial t}\frac{\partial u}{\partial x}\Big|_0^L. \tag{3.18}$$

The left-hand side now expresses the weak form of the equation of motion. The boundary conditions are recovered by nullifying the right-hand side, i.e. when $\partial u/\partial t$ (Dirichlet condition) or $\partial u/\partial x$ (Neumann condition) vanishes at the boundary. Often, a Dirichlet condition is referred to as *fixed* (since u is constant at the boundary), and a Neumann condition as *free*. The latter expression refers to the fact that the boundary is *free of loads* and is, thus, able to move.

Definition 3.2.6. The *eigenvalue problem* for the wave equation is obtained after substituting a time-transformed solution $u(t,x) = \hat{u}(x)e^{j\omega t}$ in the wave equation (3.13) defined over

the closed, bounded interval $\mathcal{I}$, and imposing either a Dirichlet or a Neumann condition at $x = 0, L$. That is:

$$-c^2\hat{u}''(x) = \omega^2\hat{u}(x), \tag{3.19}$$

with either $\hat{u}$ or $\hat{u}'$ specified at the boundaries $x = \{0, L\}$. Considering the general definition of the eigenvalue problem (3.10), here $\mathcal{L} = -c^2\frac{\partial^2}{\partial x^2}$, and hence the eigenvalues of the wave equation are real and positive □

The boundary conditions of the Dirichlet and Neumann types are not the only ones available. They have the interpretation of realising a form of energy conservation for the wave equation. In some cases, particularly in acoustics, the boundary may itself store or dissipate energy. The boundary condition is often expressed as an *impedance* in such cases. For instance, the bridge of a guitar is not entirely fixed, and its mobility changes with frequency. Such more complicated boundary expressions will be considered here. See e.g. [47] for an example of impedance conditions in violin string vibration.

Solutions to the second-order differential problem (3.19) are obtained again through an appeal to complex exponentials. Using $\omega = c\gamma$ as per (3.14), consider the following particular solution:

$$\hat{u}(x) = A_+e^{j\gamma x} + A_-e^{-j\gamma x}, \tag{3.20}$$

for constant $A_\pm$. When substituted in (3.19), one obtains an identity and therefore $\hat{u}$ solves the eigenvalue problem. In general, $A_\pm \in \mathbb{C}$, but imposing the boundary conditions of Dirichlet or Neumann kind yields a set of real eigenfunctions. To show this, use $\hat{u}(0) = \hat{u}(L) = 0$ for Dirichlet, and $\hat{u}'(0) = \hat{u}'(L) = 0$ for Neumann. Two systems result:

$$\text{Dir: } \begin{bmatrix} 1 & 1 \\ e^{j\gamma L} & e^{-j\gamma L} \end{bmatrix} \begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ Neu: } \begin{bmatrix} 1 & -1 \\ e^{j\gamma L} & -e^{-j\gamma L} \end{bmatrix} \begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Non-trivial solutions are obtained by imposing the determinant of the matrices to be null, implying in both cases:

$$\sin\gamma L = 0 \implies \gamma = \gamma^{(p)} := p\pi/L,\ p \in \mathbb{N}. \tag{3.21}$$

The wavenumbers are now quantised by the positive integer p. Turning to the eigenfunctions, one may use the first row for both systems above, giving $A_+ = -A_-$ for the Dirichlet case and $A_+ = A_-$ for Neumann. Using these in (3.20), one gets:

$$\text{Dir: } \hat{u}^{(p)}(x) = A^{(p)}\sin(\gamma^{(p)}x), \quad \text{Neu: } \hat{u}^{(p)}(x) = A^{(p)}\cos(\gamma^{(p)}x), \tag{3.22}$$

where the constant of proportionality may be chosen to normalise the eigenfunctions in some manner. It is customary to choose:

$$\text{Dir: } A^{(p)} = \left(\int_{\mathcal{I}} \sin^2(\gamma^{(p)}x)\,\mathrm{d}x\right)^{-1/2}, \quad \text{Neu: } A^{(p)} = \left(\int_{\mathcal{I}} \cos^2(\gamma^{(p)}x)\,\mathrm{d}x\right)^{-1/2}$$

and note that, for the Dirchilet case, $p = 1, 2, 3, \ldots$ whereas for Neumann p may be zero, i.e. $p = 0, 1, 2, \ldots$. Under such choice for the normalisation constants, one has $\int_{\mathcal{I}}(\hat{u}^{(p)})^2\,\mathrm{d}x = 1$, that is, the eigenfunctions are *normalised*. Note as well that, regardless of the constant of normalisation, $\int_{\mathcal{I}} \hat{u}^{(p)}\hat{u}^{(q)}\,\mathrm{d}x = 0\ \forall p \neq q$, that is, the eigenfunctions are also *orthogonal*.

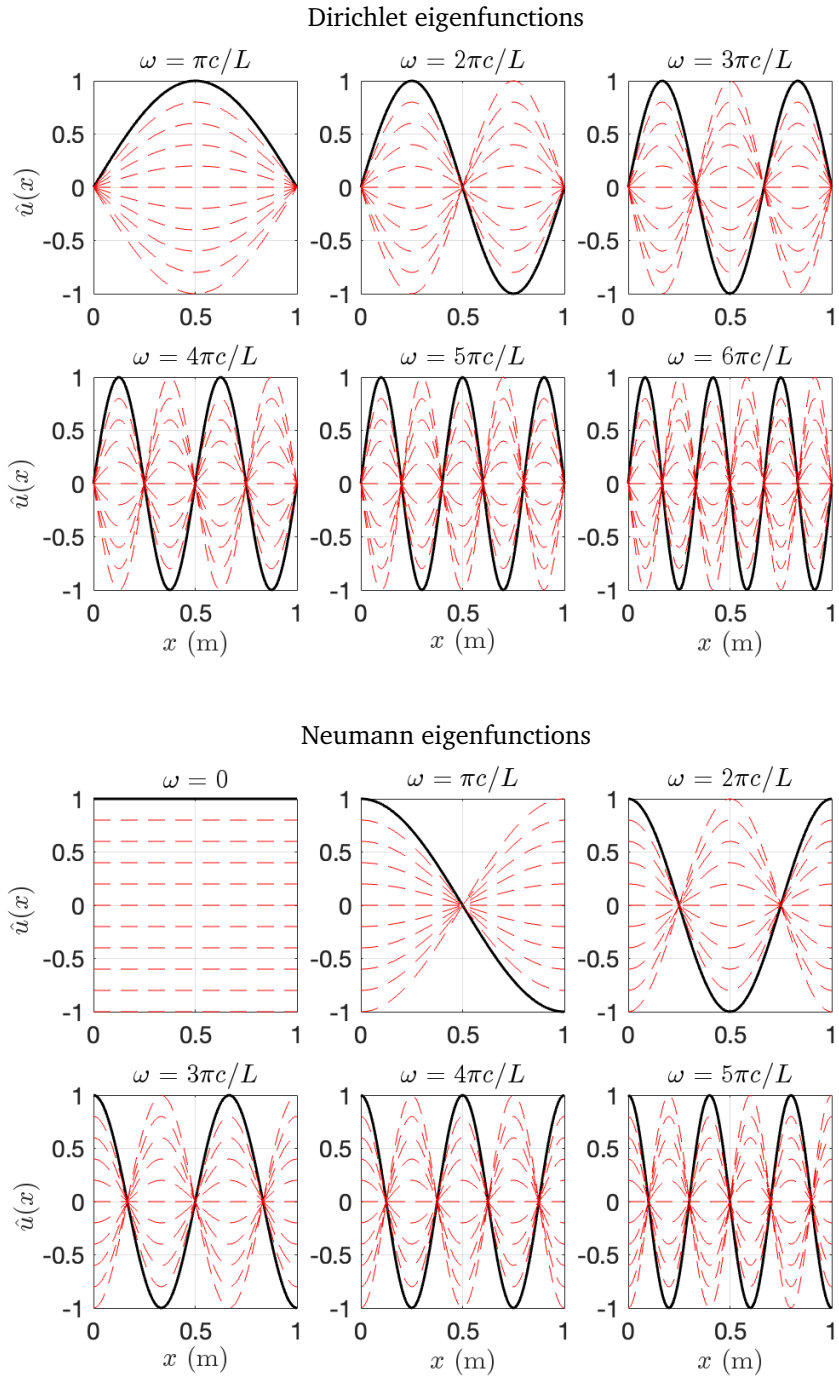


Figure 3.5: First six eigenfunctions for the eigenvalue problem (3.19) under a choice of Dirichlet or Neumann boundary conditions, as given.

subsubsectionThe discrete eigenvalue problem

A discrete eigenvalue problem is achieved simply by discretising the second derivative over a grid. It is convenient to begin the discussion using the uniform grid. Thus, consider a number $M := L/\Delta_x$ of grid subintervals, all of length Δ_x, yielding $M+1$ grid points, including the endpoints.

Then, apply the definition of the Laplacian from Table 2.2 to the grid function $e^{j\gamma m\Delta_x}$:

$$\delta_{xx}e^{j\gamma m\Delta_x} = \frac{e^{j\gamma\Delta_x} - 2 + e^{-j\gamma\Delta_x}}{\Delta_x^2}e^{j\gamma m\Delta_x}$$

which, using the half-angle trigonometric formulae, can be written as:

$$\delta_{xx}e^{j\gamma m\Delta_x} = -\frac{4}{\Delta_x^2}\sin^2\frac{\gamma\Delta_x}{2}e^{j\gamma m\Delta_x} = -\gamma^2\,\mathrm{sinc}^2\,\frac{\gamma\Delta_x}{2}e^{j\gamma m\Delta_x},$$

with $\mathrm{sinc}(x) := x^{-1}\sin(x)$. This shows that the complex exponential $e^{j\gamma m\Delta_x}$ is an eigenfunction of the discrete operator δ_{xx}, with eigenvalue $-\gamma^2\,\mathrm{sinc}^2\,\frac{\gamma\Delta_x}{2} = -\gamma^2 + \mathcal{O}(\Delta_x^2)$. The same eigenvalue is returned by applying δ_{xx} to $e^{-j\gamma m\Delta_x}$. Note that the function $\sin^2(x)$ is periodic with period π. Considering the positive half of the range of $[-\pi/2, \pi/2]$ (of length π), the useful range for γ becomes the following:

$$0 \le \gamma \le \pi/\Delta_x. \tag{3.23}$$

Notice that the sinc function "warps" the eigenvalues of the discrete wave equation compared to the eigenvalues of the continuous wave equation, and the effect is more and more pronounced γ approaches the upper limit in (3.23).

Definition 3.2.7. The *discrete eigenvalue problem* for the wave equation defined on a uniform grid with grid size Δ_x is given by:

$$-c^2\delta_{xx}\hat{u}_m = \omega^2\,\mathrm{sinc}^2\,\frac{\omega\Delta_x}{2c}\hat{u}_m, \tag{3.24}$$

where $\omega = c\gamma$ has the interpretation of an eigenfrequency of the continuous eigenvalue problem □

The eigenfunctions are given in this case by:

$$\hat{u}_m = A_+e^{j\gamma m\Delta_x} + A_-e^{-j\gamma m\Delta_x}. \tag{3.25}$$

As per the continuous case, the discrete eigenvalue problem (3.24) must be supplied with appropriate boundary conditions. Such conditions are given here as $\hat{u}_0 = \hat{u}_M = 0$ in the Dirichlet case, and $\delta_{x\cdot}\hat{u}_0 = \delta_{x\cdot}\hat{u}_M = 0$ in the Neumann case. Substituting such conditions in the expression (3.25), two systems result:

$$\text{Dir: } \begin{bmatrix} 1 & 1 \\ e^{j\gamma L} & e^{-j\gamma L} \end{bmatrix}\begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{Neu: } \begin{bmatrix} 1 & -1 \\ e^{j\gamma L} & -e^{-j\gamma L} \end{bmatrix}\begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where now $L = M\Delta_x$. Remarkably, nullifying the determinants results in the same equation as for the continuous case, (3.21). Thus:

$$\gamma^{(p)} := p\pi/L = p\pi/M\Delta_x \tag{3.26}$$

are the quantised solutions of the discrete eigenvalue problem. Solving for the eigenfunctions, similarly to what was done to obtain (3.22), one gets:

$$\text{Dir: } \hat{u}_m^{(p)}(x) = A^{(p)} \sin(\gamma^{(p)} m \Delta_x), \text{ Neu: } \hat{u}_m^{(p)}(x) = A^{(p)} \cos(\gamma^{(p)} m \Delta_x),$$

The constant of normalisation may, in this case, be chosen such that:

$$\text{Dir: } A^{(p)} = \left(\sum_{m=1}^{M-1} \sin^2(\gamma^{(p)} m \Delta_x) \right)^{-\frac{1}{2}}, \text{ Neu: } A^{(p)} = \left(\sum_{m=0}^{M} \cos^2(\gamma^{(p)} m \Delta_x) \right)^{-\frac{1}{2}}.$$

It may be useful in some cases to write the discrete eigenvalue problem (3.24) in matrix form. This is:

$$-c^2 \mathbf{D}_{xx} \hat{\mathbf{u}} = \omega^2 \operatorname{sinc}^2 \frac{\omega \Delta_x}{2c} \hat{\mathbf{u}}, \tag{3.27}$$

where the explicit form of the Laplacian under Dirichlet or Neumann conditions are:

$$\mathbf{D}_{xx}^{\text{Dir}} = \frac{1}{\Delta_x^2} \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix} \in \mathbb{R}^{M-1 \times M-1},$$

$$\mathbf{D}_{xx}^{\text{Neu}} = \frac{1}{\Delta_x^2} \begin{bmatrix} -2 & 2 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 2 & -2 \end{bmatrix} \in \mathbb{R}^{M+1 \times M+1}.$$

Note that the Dirichlet matrix has smaller dimensions: in this case, the endpoints u_0, u_M need not be stored or updated since the boundary conditions fix their value to zero at all times. Note also that the Neumann matrix is not symmetric, but its eigenvalues are real, as shown. The matrix may be diagonalised via a similarity transformation analogous to the one described in Definition 3.1.1.

Regardless, the Laplacian admits the following *eigenvalue decomposition*:

$$\mathbf{L} := -c^2 \mathbf{D}_{xx} = \hat{\mathbf{U}} \mathbf{\Omega}^2 \hat{\mathbf{U}}^{-1}, \tag{3.28}$$

where $\mathbf{\Omega}$ is diagonal and it contains the eigenvalues $(\omega^{(p)}) \operatorname{sinc} \frac{\omega^{(p)} \Delta_x}{2c}$, where $\omega^{(p)} = c\gamma^{(p)}$ is obtained from (3.26), and where $\hat{\mathbf{U}}$ is a matrix whose columns are the eigenvectors, i.e.

$$\hat{\mathbf{U}} := [\hat{\mathbf{u}}^{(1)}, \hat{\mathbf{u}}^{(2)}, ...]. \tag{3.29}$$

According to the boundary conditions, the eigenvectors $\hat{\mathbf{u}}^{(p)}$ are either the sampled sines or cosines. The form of $\hat{\mathbf{U}}$ is special for the Dirichlet case: the Laplacian is symmetric, meaning that $\hat{\mathbf{U}}$ is *orthonormal*. In practice, the inverse of $\hat{\mathbf{U}}$ in (3.28) is the transpose, and via the orthonormality property one has $\hat{\mathbf{U}}\hat{\mathbf{U}}^\intercal = \hat{\mathbf{U}}^\intercal \hat{\mathbf{U}} = \mathbf{I}$. The orthonormality condition holds

only if the normalisation constants $A^{(p)}$ are set as suggested above. If the eigenvectors are normalised differently, $\hat{\mathbf{U}}$ remains *orthogonal*, meaning that multiplication by its transpose gives rise to a diagonal matrix whose diagonal elements correspond to the squared norms of the sine functions. Figure 3.6 reports the first few eigenvectors of the Laplacian matrix for the two cases.

3.3 The Euler-Bernoulli equation for rods

A qualitatively different kind of wave propagation is displayed by flexural motion in bars. Bars in bending are subjected to elastic forces due to the shearing of the cross-section. When a bar is bent, a twisting moment is generated due to the deformation of the cross-section, resulting in a flexural elastic force.

Definition 3.3.1. The *Euler-Bernoulli* equation, describing flexural waves in bars, is given by:

$$\rho A \frac{\partial^2 u}{\partial t^2} = -\frac{\partial^2 \mu}{\partial x^2}, \quad \mu = EI\frac{\partial^2 u}{\partial x^2} \tag{3.30}$$

□

Here, $u = u(t, x)$ is the vertical displacement of an element of the bar, and $\mu = \mu(t, x)$ is the twisting moment originating from the internal stresses when the bar bends. Constants appear as: ρ, the volume density of the bar, and E, Young's modulus. In the following, $A = A(x)$ is the area of the cross-section, allowed here to vary; $I = I(x)$ is the corresponding area moment of inertia, defined as:

$$I(x) = \int_{A(x)} \zeta^2 \mathrm{d}A, \tag{3.31}$$

In the above, ζ is the distance of the area element $\mathrm{d}A$ from the axis of rotation. For a circular cross-section of radius r, $I = \pi r^4/4$, whereas for a rectangular cross-section of width b and height a, $I = a^3 b/12$. Notice that both A and I are solely geometric parameters with no dependence on the bar's material parameters. Bars with a spatially varying cross-section are ubiquitous in musical acoustics and found across a variety of percussion instruments such as xylophones and marimbas. Before proceeding with discretising (3.30), it is worth studying the behaviour of the elastic waves in the continuous case, as this allows selecting the mesh size according to wave speed considerations.

The dispersion relation for system (3.30) is obtained after substituting $u(x, t) = e^{j\omega t} e^{(j\gamma + r)x}$. As per (3.15), γ and ω represent a plane wave's spatial and temporal frequency. Here, however, a spatially dependent wave amplitude, scaled by e^{rx}, is assumed, accounting for the geometric variation of the cross-section. After substitution of the assumed solution in (3.30), equating the imaginary part to zero leads to:

$$0 = 2I\gamma^2 r + \frac{\mathrm{d}I}{\mathrm{d}x}\gamma^2 - 2Ir^3 - 3\frac{\mathrm{d}I}{\mathrm{d}x}r^2 - \frac{\mathrm{d}^2 I}{\mathrm{d}x^2}r \approx 2I\gamma^2 r + \frac{\mathrm{d}I}{\mathrm{d}x}\gamma^2, \tag{3.32}$$

where the approximation is valid for large enough γ such that $|r|\gamma^{-1} < 1$. Thus, one gets:

$$r \approx -\frac{1}{2I}\frac{\mathrm{d}I}{\mathrm{d}x}. \tag{3.33}$$

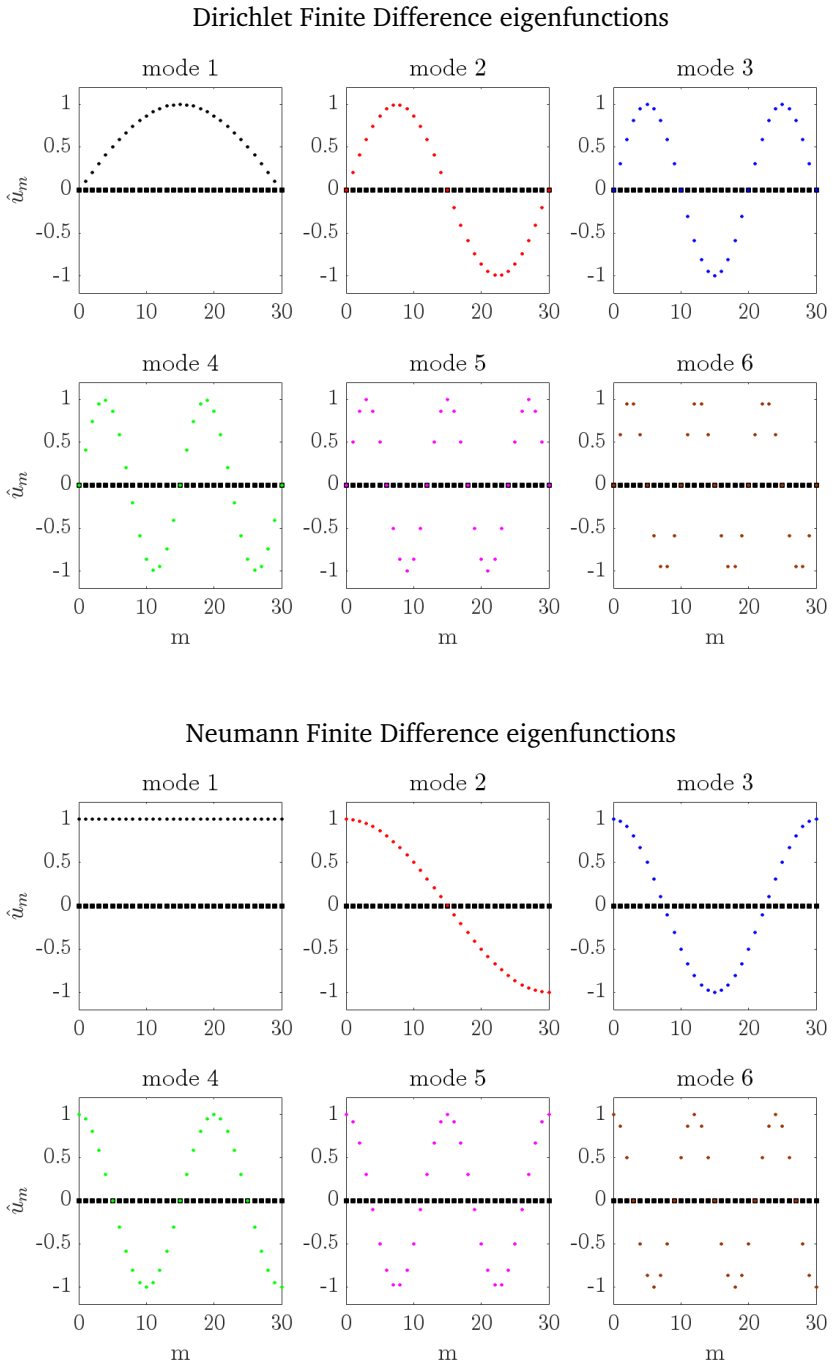


Figure 3.6: First six eigenfunctions for the eigenvalue problem (3.24) under a choice of Dirichlet or Neumann boundary conditions, as given, for $M = 30$.

Figure 3.7: Picture of the Majestic 5.0 Octave Rosewood Bar Concert Marimba. Copyright: Majestic Percussion. Link to video.

Note that $r > 0$ when $\mathrm{d}I/\mathrm{d}x < 0$ and, hence, the wave amplitude increases as the moment of inertia decreases. Substituting this value into the real part of the dispersion relation and solving for γ, one gets:

$$\gamma \approx \left(\frac{\rho A}{EI}\right)^{\frac{1}{4}} \omega^{\frac{1}{2}} + \frac{1}{4I}\frac{\mathrm{d}^2 I}{\mathrm{d}x^2}\left(\frac{EI}{\rho A}\right)^{\frac{1}{4}} \omega^{-\frac{1}{2}} \quad \text{for } |\omega| > 1, \tag{3.34}$$

neglecting higher-order corrections in $\mathrm{d}I/\mathrm{d}x$, $\mathrm{d}^2I/\mathrm{d}x^2$, which are, generally, small. This formula has a correction term proportional to $\omega^{-\frac{1}{2}}$ compared to the case of a bar with a constant cross-section. Using the mathematical expressions presented in Definition 3.2.4, the phase and group velocities for the bar with a varying cross-section are obtained from (3.34), as:

$$c_\phi(x,\omega) \approx \left(\frac{EI(x)}{\rho A(x)}\right)^{\frac{1}{4}} \omega^{\frac{1}{2}}, \quad c_g(x,\omega) \approx 2c_\phi(x,\omega), \tag{3.35}$$

and, hence, the bar is a *dispersive* system since the phase velocity depends on frequency. As a consequence, an initial wavefront travelling across the bar will gradually distort and flatten out as shorter wavelengths travel faster. The bar with varying cross-sections is further characterised by the phase and group velocities' spatial dependence through $A(x)$ and $I(x)$.

Example 3.3.1. A marimba's typical cross-section is considered here, as per Fig. 3.7. A rectangular cross-section is assumed throughout, for which:

$$I = \frac{a^3 b}{12}, \quad \frac{\mathrm{d}I}{\mathrm{d}x} = \frac{a^2 b}{4}\frac{\mathrm{d}a}{\mathrm{d}x}, \quad \frac{\mathrm{d}^2 I}{\mathrm{d}x^2} = \frac{ab}{2}\left(\frac{\mathrm{d}a}{\mathrm{d}x}\right)^2 + \frac{a^2 b}{4}\frac{\mathrm{d}^2 a}{\mathrm{d}x^2}, \tag{3.36}$$

where $a = a(x)$ is the thickness of the cross-section, and b is the width. Assume that the thickness varies as a power law:

$$a = a^{(min)} + \frac{2^p \Delta a}{l^p}|x|^p, \ (-l/2 \le x \le l/2), \tag{3.37}$$

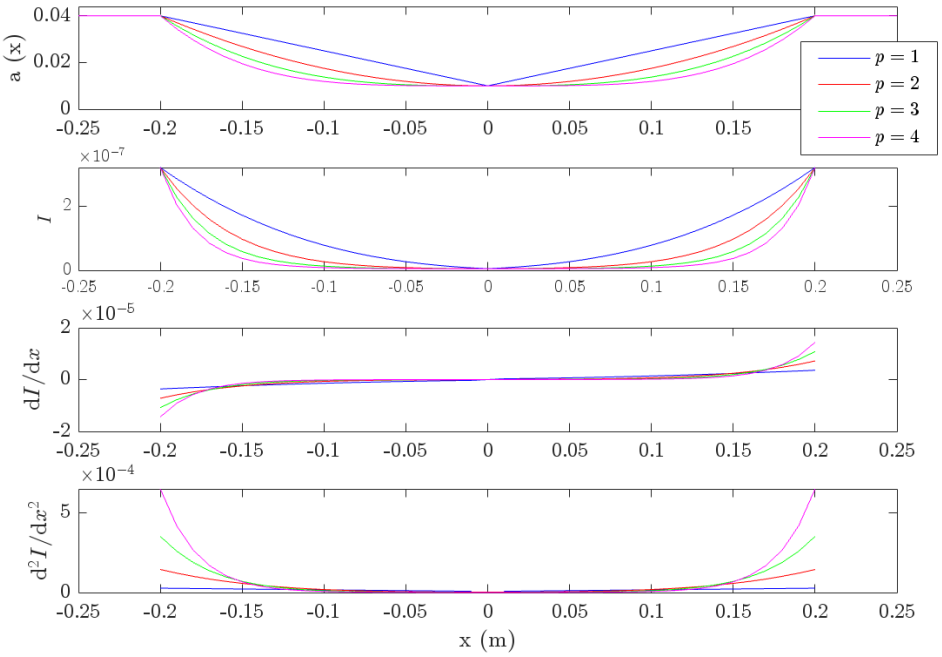


Figure 3.8: Thickness profiles and corresponding moment of inertia and derivatives. The thickness profile follows the power law (3.37), under four different values for the exponent p, as indicated. Here, $a^{(max)} = 4$ cm, $a^{(min)} = 1$ cm, $b = 6$ cm, $l = 40$ cm. Note that the moment of inertia and its derivatives are small, such that higher-order powers can safely be neglected when deriving results.

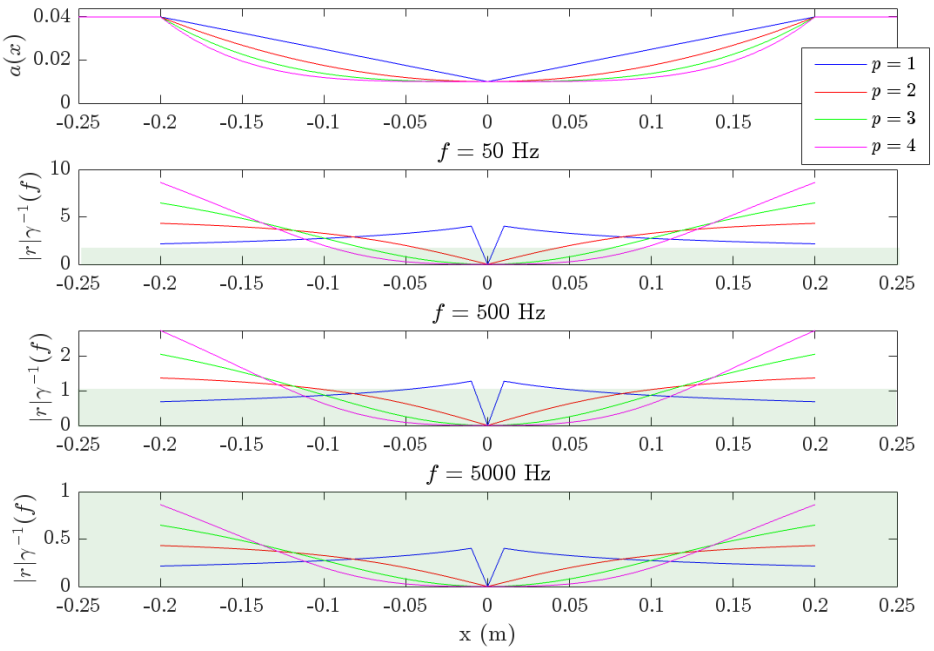


Figure 3.9: Thickness profiles (the same as Fig. 3.8) and corresponding amplitude parameter $|r|\gamma^{-1}$, as per (3.33). The shaded area in the bottom plot corresponds to $|r|\gamma^{-1} < 1$, for which approximation (3.33) is valid. The wavenumber is computed via (3.34) at three representative frequencies.

with $\Delta a := a^{(max)} - a^{(min)}$. In the following, take $\Delta a = 3$ cm, $l = 40$ cm, $a^{(min)} = 1$ cm, $b = 6$ cm, and $p \in \{1, 2, 3, 4\}$. Results are summarised in Fig. 3.8 where it is seen that the moment of inertia and its derivatives are small. Note, however, terms such as $\frac{1}{I}\frac{\mathrm{d}I}{\mathrm{d}x}$ are not as small □

Example 3.3.2. For the same marimba bars, the scaled amplitude parameter $|r|\gamma^{-1}$ is computed at various reference frequencies. The wavenumber is given by (3.34). Fig. 3.9 summarises the results. For larger frequencies, $|r|\gamma^{-1}$ becomes small along the whole length of the bar □

Definition 3.3.2. The energy H of a bent bar is obtained as the sum of its kinetic and potential energies H_k, H_p, defined as follows:

$$H_k := \int_{\mathcal{I}} \frac{\rho A(x)}{2} \left(\frac{\partial u}{\partial t}\right)^2 \mathrm{d}x, \quad H_p := \frac{1}{2}\int_{\mathcal{I}} \frac{\partial^2 u}{\partial x^2}\mu(x,t)\,\mathrm{d}x, \tag{3.38}$$

where the definition of the moment $\mu(x,t)$ is given in (3.30), and where $\mathcal{I} = \{x \,|\, 0 \le x \le L\}$ □

Taking the time derivative of $H = H_k + H_p$ one gets:

$$\int_{\mathcal{I}} \left(\frac{\rho A}{2} \frac{\partial^2 u}{\partial t^2} + \frac{\partial^2 \mu}{\partial x^2} \right) \mathrm{d}x = \frac{\partial u}{\partial t} q - \frac{\partial^2 u}{\partial t \partial x} \mu \Big|_0^L, \tag{3.39}$$

where the beam's shear force was defined as $q := \frac{\partial \mu}{\partial x}$. Nullifying the integrand above results in the strong form of the equation of motion (3.30); the right-hand side returns the boundary conditions. Here, two such conditions emerge at each end, a feature distinguishing the bar model compared to the simple wave equation. Conditions of classic type emerge as:

$$\text{free: } \mu = q = 0, \ \text{clamped: } u = \frac{\partial u}{\partial x} = 0, \ \text{simply-supported: } \mu = u = 0.$$

Free-end conditions correspond to the vanishing of the applied moment and its derivative, the net shear force; clamped conditions are purely geometrical with vanishing displacement and slope; simply-supported conditions are somewhat in between, having a null moment and displacement. Note that a fourth condition arises mathematically, namely $\partial u/\partial x = \partial \mu/\partial x = 0$, but this is rarely given in textbooks as it is seldom realised in practice.

3.3.1 The eigenvalue problem with constant thickness

In the simplest case of bars have a constant cross-section, the Euler-Bernoulli equation simplifies to:

$$\frac{\partial^2 u}{\partial t^2} = -\kappa^4 \frac{\partial^4 u}{\partial x^4}, \tag{3.40}$$

with $\kappa := (EI/\rho A)^{1/4}$. Let $\mathcal{I} := \{x | x \in [0, L]\}$ denote the domain occupied by the unstretched bar. The *eigenvalue problem* for the bar in bending is obtained after substituting the trial solution $u(x,t) = \hat{u}(x) e^{j\omega t}$, resulting in:

$$\kappa^4 \hat{u}''''(x) = \omega^2 \hat{u}(x), \tag{3.41}$$

This equation admits four distinct solutions in the form of exponentials, such that:

$$\hat{u}(x) = A_+ e^{j\gamma x} + A_- e^{-j\gamma x} + B_+ e^{\gamma x} + B_- e^{-\gamma x}, \tag{3.42}$$

with $\gamma^4 := \omega^2/\kappa^4$. The four constants A_+, A_-, B_+, B_- are determined by applying the boundary conditions at the left and right endpoints of the interval $\mathcal{I}$. Nullifying the determinant of the resulting system gives a transcendental equation whose solution returns the quantised wave numbers $\gamma^{(p)}$.

Example 3.3.3. The quantised wavenumbers for a bar simply-supported at both ends are obtained by imposing the boundary conditions $\hat{u} = \mathrm{d}^2\hat{u}/\mathrm{d}x^2 = 0$ at both ends. Imposing the condition at $x = 0$ results in $\hat{u} = 2jA_+ \sin(\gamma x) + 2B_+ \sinh(\gamma x)$. Further imposing the conditions at $x = L$ results in $0 = 2jA_+ \sin(\gamma L) + 2B_+ \sinh(\gamma L)$ and $0 = 2jA_+ \sin(\gamma L) - 2B_+ \sinh(\gamma L)$, which is only possible when $B_+ = 0$ and

$$\gamma^{(p)} = p\pi/L, \quad p \in \mathbb{N}. \tag{3.43}$$

Note that these are the same wavenumbers obtained for the simple wave equation, but now the corresponding frequencies are given by $\omega^{(p)} = (\gamma^{(p)})^2\kappa^2$. The corresponding eigenfunctions are, thus:

$$\hat{u}^{(p)}(x) = A^{(p)}\sin(\gamma^{(p)}x), \tag{3.44}$$

again, the same as the eigenfunctions of the simple wave equations under Dirichlet conditions □

Finding analytic expressions for the eigenvalues and frequencies becomes unwieldy for bars with a variable cross-section. When system (3.30) is transformed in time using a Fourier transform, as per examples 3.3.3, the following system results:

$$\omega^2\rho A(x)\hat{u}(x) = \hat{\mu}''(x), \quad \hat{\mu}(x) = EI(x)\hat{u}''(x) \tag{3.45}$$

The problem can be approached using a discrete eigenvalue problem defined over a grid.

3.3.2 The discrete eigenvalue problem

A formula analogous to (3.4) approximates the second derivatives over a non-uniform grid. The *discrete eigenvalue problem* for the Euler-Bernoulli bar with non-uniform cross-section is derived from:

$$\omega^2\rho A_m\hat{u}_m = \frac{\hat{\mu}_{m-1}}{\Delta_{m-1/2}\widetilde{\Delta}_m} - \frac{2\hat{\mu}_m}{\Delta_{m-1/2}\Delta_{m+1/2}} + \frac{\hat{\mu}_{m+1}}{\Delta_{m+1/2}\widetilde{\Delta}_m}, \tag{3.46a}$$

$$(EI_m)^{-1}\hat{\mu}_m = \frac{\hat{u}_{m-1}}{\Delta_{m-1/2}\widetilde{\Delta}_m} - \frac{2\hat{u}_m}{\Delta_{m-1/2}\Delta_{m+1/2}}\frac{\hat{u}_{m+1}}{\Delta_{m+1/2}\widetilde{\Delta}_m}, \tag{3.46b}$$

where, as per Section 3.1, $\Delta_{m-1/2}$ is the grid spacing between grid points x_{m-1} and x_m, and where $\widetilde{\Delta}_m := (\Delta_{m-1/2} + \Delta_{m+1/2})/2$. ω^2 is the eigenvalue associated with the eigenvector $\hat{\mathbf{u}}$ Here, $m \in [0, M] \subset \mathbb{N}$, defining the grid $\mathfrak{I} = \{x_j | x_j = \sum_{m=1}^{j}\Delta_{m-1/2}\}$.

Numerical boundary conditions must imposed at the bar's ends to account for the action of the difference operators on points near the boundary. Considering the marimba bars of Figs. 3.8 and 3.9, the cross-section is constant near the boundary, and the boundary conditions can be assumed to be of free type. Hence, the following are imposed:

$$\hat{\mu}_0 = 0, \quad \hat{\mu}_M = 0, \quad \hat{\mu}_{-1} = \hat{\mu}_1, \quad \hat{\mu}_{M+1} = \hat{\mu}_{M-1}, \tag{3.47}$$

discretising free-end conditions with constant bar cross-section near the boundary. Thus, $\hat{\mu}_0$ and $\hat{\mu}_M$ need not be stored or updated since their value is fixed at all times. Starting with (3.46a), and using the numerical conditions above, one has:

$$\begin{aligned}
\omega^2\rho A_0\hat{u}_0 &= \frac{2\hat{\mu}_1}{\Delta_{1/2}\widetilde{\Delta}_0} \\
\omega^2\rho A_1\hat{u}_1 &= -\frac{2\hat{\mu}_1}{\Delta_{1/2}\Delta_{3/2}} + \frac{\hat{\mu}_2}{\Delta_{3/2}\widetilde{\Delta}_1} \\
\omega^2\rho A_2\hat{u}_2 &= \frac{\hat{\mu}_1}{\Delta_{3/2}\widetilde{\Delta}_2} - \frac{2\hat{\mu}_2}{\Delta_{3/2}\Delta_{5/2}} + \frac{\hat{\mu}_3}{\Delta_{5/2}\widetilde{\Delta}_2} \\
&\cdots
\end{aligned}$$

It is convenient to collect the coefficients multiplying the elements of the vector $\hat{\boldsymbol{\mu}}$ in a matrix:

$$\mathbf{D}_{xx} := \begin{bmatrix} \frac{1}{\Delta_{\frac{1}{2}}\Delta_{\frac{1}{2}}} & & & & 0 \\ -\frac{2}{\Delta_{\frac{1}{2}}\Delta_{\frac{3}{2}}} & \frac{1}{\Delta_{\frac{3}{2}}\widetilde{\Delta}_1} & 0 & \dots & 0 \\ \frac{1}{\Delta_{\frac{3}{2}}\widetilde{\Delta}_2} & -\frac{2}{\Delta_{\frac{3}{2}}\Delta_{\frac{5}{2}}} & \frac{1}{\Delta_{\frac{5}{2}}\widetilde{\Delta}_2} & 0 & \vdots \\ 0 & \frac{1}{\Delta_{\frac{5}{2}}\widetilde{\Delta}_3} & -\frac{2}{\Delta_{\frac{5}{2}}\Delta_{\frac{7}{2}}} & \frac{1}{\Delta_{\frac{7}{2}}\widetilde{\Delta}_3} & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{1}{\Delta_{M-\frac{3}{2}}\widetilde{\Delta}_{M-1}} & -\frac{2}{\Delta_{M-\frac{1}{2}}\Delta_{M-\frac{3}{2}}} \\ 0 & & & & \frac{1}{\Delta_{M-\frac{1}{2}}\Delta_{M-\frac{1}{2}}} \end{bmatrix},$$

defining a rectangular matrix of size $(M+1)\times(M-1)$. Collecting coefficients on the right hand side of (3.46b) produces the matrix:

$$\mathbf{D}^{\star}_{xx} := \begin{bmatrix} \frac{1}{\Delta_{\frac{1}{2}}\widetilde{\Delta}_1} & -\frac{2}{\Delta_{\frac{1}{2}}\Delta_{\frac{3}{2}}} & \frac{1}{\Delta_{\frac{3}{2}}\widetilde{\Delta}_1} & \dots & 0 & 0 \\ 0 & \frac{1}{\Delta_{\frac{3}{2}}\widetilde{\Delta}_2} & -\frac{2}{\Delta_{\frac{3}{2}}\Delta_{\frac{5}{2}}} & \frac{1}{\Delta_{\frac{5}{2}}\widetilde{\Delta}_2} & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \dots & 0 & \frac{1}{\Delta_{M-\frac{3}{2}}\widetilde{\Delta}_{M-1}} & -\frac{2}{\Delta_{M-\frac{1}{2}}\Delta_{M-\frac{3}{2}}} & \frac{1}{\Delta_{M-\frac{1}{2}}\widetilde{\Delta}_{M-1}} \end{bmatrix},$$

of size $(M-1)\times(M+1)$. Before proceeding, note that in the case of a uniform grid $\Delta_{m-1/2} = \Delta_x \;\forall\, m$ one has $\mathbf{D}^{\star}_{xx} = \mathbf{D}^{\intercal}_{xx}$.

Definition 3.3.3. The discrete eigenvalue problem for bars with varying cross-section is written compactly as:

$$\omega^2\rho\mathbf{A}\hat{\mathbf{u}} = \mathbf{D}_{xx}\hat{\boldsymbol{\mu}}, \quad (E\mathbf{I})^{-1}\hat{\boldsymbol{\mu}} = \mathbf{D}^{\star}_{xx}\hat{\mathbf{u}}, \tag{3.48}$$

where $\mathbf{A}$ and $\mathbf{I}$ are diagonal matrices storing the values of the area and moment of inertia of the cross-section. The system may be reduced further and be expressed in terms of $\hat{\mathbf{u}}$ alone, as:

$$\widetilde{\kappa}^4(\widetilde{A}\mathbf{A}^{-1})\mathbf{D}_{xx}(\widetilde{I}^{-1}\mathbf{I})\mathbf{D}^{\star}_{xx}\hat{\mathbf{u}}(x) = \omega^2\hat{\mathbf{u}}(x), \text{ with } \widetilde{\kappa} := \left(\frac{E\widetilde{I}}{\rho\widetilde{A}}\right)^{\frac{1}{4}}, \tag{3.49}$$

and where $\widetilde{A}$, $\widetilde{I}$ are scaling factors (such as the average value of the area and moment of inertia along the bar or their largest or smallest values) □

One obvious question is how to select an appropriate grid to solve the numerical eigenvalue problem.

Definition 3.3.4. The grid's *points per wavelength* (ppw) at frequency ω are defined as:

$$\text{ppw} := \frac{2\pi c_\phi(x_m,\omega)}{\omega\widetilde{\Delta}_m}, \tag{3.50}$$

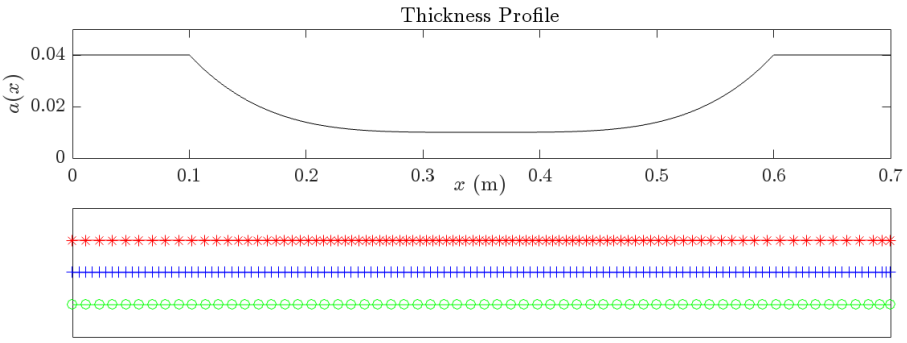


Figure 3.10: Top: thickness profile. Bottom: grid spacings using a coarse constant spacing (green); a fine constant spacing (blue); and adaptive spacing (red). The grid spacings are selected according to (3.50), using ppw = 50, and $\omega = 20000\pi$. For the constant grid spacings, the reference phase velocity is selected as the phase velocity near the bar's ends (yielding the largest grid spacing) and at the bar's centre (yielding the finest grid spacing). The adaptive mesh naturally maps the wave velocity to the corresponding grid spacing yielding a constant number of points per wavelength across the bar.

where $c_\phi(x_m, \omega)$ is the phase velocity at frequency ω and grid point x_m, as defined in (3.35) □

The idea of using "points-per-wavelength" as a control parameter for solving numerical eigenvalue problems is widespread in acoustics, particularly in room acoustics, [50, 51]. In order to control the finesse of the grid for the purpose of computing the eigenfrequencies and shapes of the bar, one may select a maximum largest frequency $\omega^{(max)}$ and an input ppw at such frequency, therefore yielding the mesh size Δ_m. Fig. 3.10 reports the non-uniform mesh obtained by sampling the interval $\mathcal{I}$ using definition (3.50). The same figure reports the grids obtained by sampling the interval using the smallest and largest grid spacing, corresponding to the largest and smallest phase velocity in the bar. Fig. 3.11 reports the convergence of the first three eigenfrequencies as a function of M. As expected, when plotted as a function of M, the uniform grids yield the same eigenfrequency values $\forall M$. The eigenfrequencies computed on the non-uniform grid follow a similar convergence trend. Whilst these notes do not cover the time domain in detail, the use of non-uniform meshes is crucial in the design of efficient time stepping schemes, maximising the output signal's bandwidth at any given input sample rate, as shown elegantly by [24, Ch. 7].

Example 3.3.4. In the context of marimba making, the process of removing material from the centre of the bars is called *tuning* or, more specifically, *undercutting*, [52]. This process involves carefully removing material from the underside of the wooden bars to achieve the desired pitch and harmonic overtones. The amount and location of the material removed affect the fundamental frequency and the relationship between the harmonics, allowing the instrument maker to fine-tune the bars for a resonant and balanced sound. Typically, the bar's centre influences the fundamental pitch, while adjustments near the nodal points can modify the harmonic relationships. In Fig. 3.12, three bar profiles are considered, following (3.37) and presenting three different $a^{(min)}$ values. The deformation of the eigenshapes is

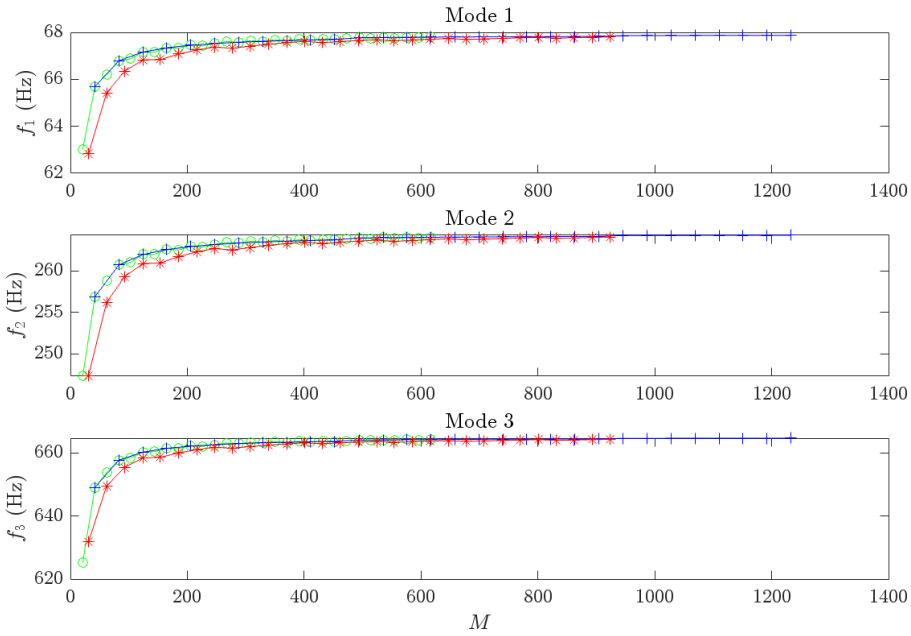


Figure 3.11: Convergence of the first three eigenfrequencies as a function of the number of grid subintervals M using a coarse constant spacing (green); a fine constant spacing (blue); and adaptive spacing (red). The three grid types yield a similar convergence profile.

	f_2/f_1	f_3/f_1
$a^{(min)} = 3$ cm	3.0	6.2
$a^{(min)} = 2$ cm	3.4	7.5
$a^{(min)} = 0.9$ cm	4.0	10.2

Table 3.1: Frequency relationships for the three bars of Fig. 3.12.

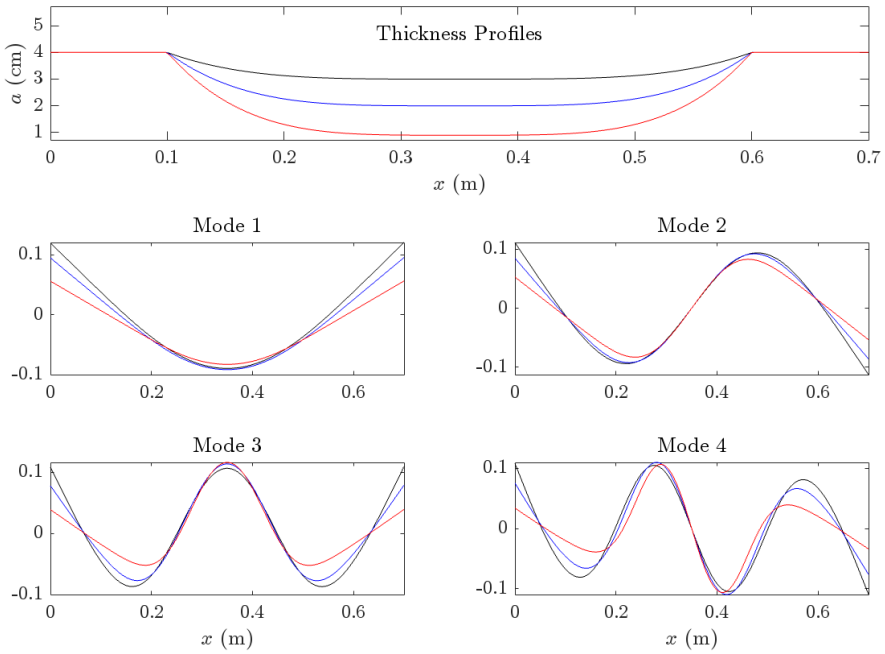


Figure 3.12: Bar profiles and first four eigenshapes. The thickness profiles follow the power law (3.37), under three different values for the smallest thickness $a^{(min)}$. Here, $a^{(max)} = 4$ cm, $p = 4$, $b = 6$ cm, $l = 40$ cm and $a^{(min)} = \{3$ cm (black), 2 cm (blue) and 0.9 cm (red)$\}$.

evident, with nodes moving toward the centre as the smallest thickness is decreased. Such wavelength shortening is unsurprising in light of the increased wave speed near the bar's centre. The relationships of the first two overtones to the fundamentals are given in Table 3.1, highlighting the effects of undercutting □

Chapter 4

Fundamentals of Vibration

Let us begin by studying the evolution of a simple vibrating lumped object, with a mass m and subjected to a force field originating from a potential,

$$F = -\frac{d\phi}{du}. \tag{4.1}$$

Here $\phi = \phi(u) : \mathbb{R} \to \mathbb{R} \in \mathcal{C}^1$ is the potential (assumed to be differentiable), and $u = u(t) : \mathbb{R}_0^+ \to \mathbb{R} \in \mathcal{C}^2$ is the displacement (assumed to be twice differentiable), measured from some convenient reference position. Here, $t \geq 0$ is time. Newton's law gives:

$$m\frac{d^2u}{dt^2} = -\frac{d\phi}{du}. \tag{4.2}$$

Equation (4.2) is an example of *ordinary differential equation* (ODE). The equation is *autonomous* (i.e., *time-invariant*), intended as the absence of explicit dependence on time t, except as via the argument of u. The equation is, in general, *nonlinear*. Linearity is expressed here as a superposition principle: if $u_1(t)$ and $u_2(t)$ are both solutions to (4.2), then, under linear conditions, $u_3(t) = a_1u_1(t) + a_2u_2(t)$ (with a_1, a_2 being constants) is also a solution. Since the second time derivative on the left-hand side of (4.2) is linear, linearity is obtained when

$$\frac{d\phi(u_3)}{du_3} = a_1\frac{d\phi(u_1)}{du_1} + a_2\frac{d\phi(u_2)}{du_2}, \tag{4.3}$$

which is solved for $\phi = cu^\alpha$, where c is a constant, and $\alpha \in \{0, 2\}$ (the case $\alpha = 0$ being the trivial case of zero force).

To be complete, (4.2) requires the specification of two *initial conditions*, usually given as the initial displacement and velocity, such that

$$u(t = 0) = u_0, \quad \frac{du}{dt}(t = 0) = v_0, \tag{4.4}$$

and since the independent variable is the time t, (4.2) plus (4.4) specify an *initial value problem* (IVP). When initial conditions are imposed, the system possesses one *unique solution* if appropriate conditions are satisfied. We shall not focus on such conditions here, as these are covered extensively in various textbooks on differential equations. We will assume that the equations encountered in this course admit one unique solution.

4.1 Energy analysis

Energy conservation is a fundamental principle of physics, bearing significant consequences in analysing the continuous systems and the numerical approximations used to simulate them. In the simple, one dimensional case (4.2), the work W done by a force pushing on m is:

$$W = \int_0^u F \, \mathrm{d}u, \tag{4.5}$$

where it is assumed that the initial position of the body, at the time $t = 0$, is 0. Using this definition in (4.2) gives

$$\int_0^u m \frac{d^2u}{dt^2} \, \mathrm{d}u = -\int_0^u \frac{d\phi}{du} \, \mathrm{d}u. \tag{4.6}$$

To integrate, one uses $du = \frac{du}{dt} dt$, giving

$$\int_0^t m \frac{d^2u}{dt^2} \frac{du}{dt} \, \mathrm{d}t = -\int_0^t \frac{d\phi}{du} \frac{du}{dt} \, \mathrm{d}t, \tag{4.7}$$

and, using simple identities, one gets

$$\int_0^t \frac{d}{dt} \left(\frac{m}{2} \left(\frac{du}{dt} \right)^2 + \phi \right) \mathrm{d}t = 0. \tag{4.8}$$

Since $u(t) \in \mathcal{C}^2$, $\phi(u) \in \mathcal{C}^1$, the equation is solved by taking

$$\frac{m}{2} \left(\frac{du}{dt} \right)^2 + \phi = H, \tag{4.9}$$

where H (a constant) is the system's total energy. It is convenient to identify the *kinetic* and *potential* components of the energy, given, respectively, by

$$H_k \triangleq \frac{m}{2} \left(\frac{du}{dt} \right)^2, \quad H_p \triangleq \phi. \tag{4.10}$$

The initial conditions determine the expression for H, and hence

$$H = \frac{m v_0^2}{2} + \phi(u_0) := H_0, \tag{4.11}$$

and the identity $H(t) = H_0$ holds $\forall t \geq 0$. Here, H_0 represents the expression of the energy computed using the initial data u_0, v_0.

4.2 Bounds on solution growth

Not only is energy conserved, but it is also *non-negative* when ϕ itself is non-negative, i.e. $H(t) \geq 0 \ \forall t$. This fact leads to the important result of *boundedness of the solutions*. In practice, since the kinetic and potential energies are *both* non-negative, one has

$$0 \leq H_k \leq H_0, \quad 0 \leq H_p \leq H_0. \tag{4.12}$$

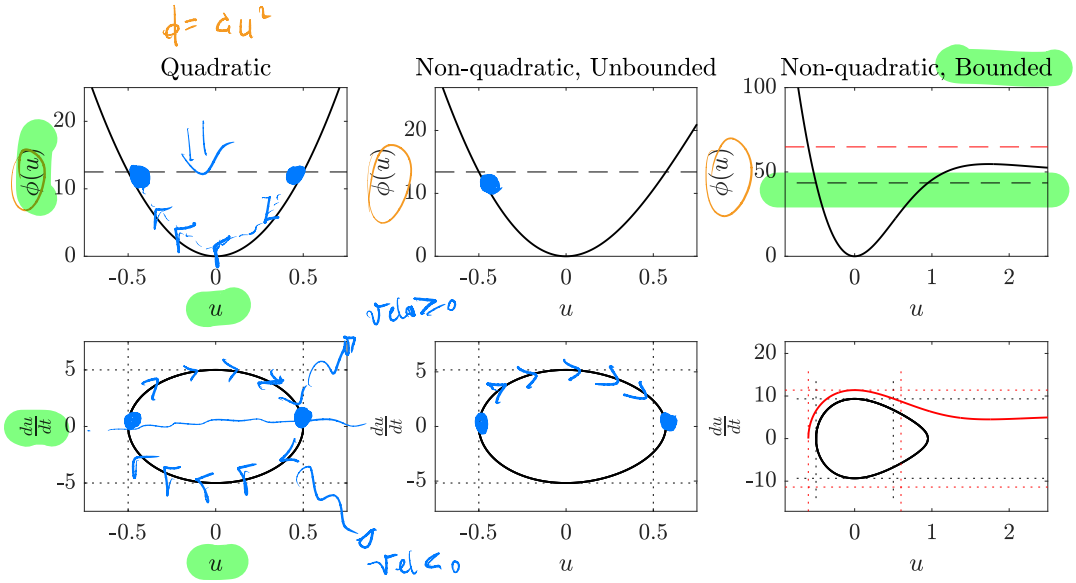


Figure 4.1: Potential functions and corresponding phase space orbits. The total energy is represented as dashed lines in the top plots. The dotted lines represent the bounds of the quadratic potential with the same total energy. The red orbit in the right panel represents unbounded motion in u, but note that the velocity remains bounded.

For the kinetic term, one has, in all cases,

$$\left|\frac{du(t)}{dt}\right| \leq \sqrt{2H_0/m} \ \forall t, \tag{4.13}$$

and thus, the velocity of the system is *always* bounded in terms of the initial energy. The displacement $u(t)$ itself may or may not be bounded. Consider the case of a quadratic potential, as in the left panel of Fig. 4.1: $\phi = \frac{Ku^2}{2}$. This case corresponds to the linear case, i.e. the simple harmonic oscillator with stiffness constant K, as will be discussed in some detail in Chapter 5. From (4.12), one has $|u| \leq \sqrt{2H_0/K}$. Hence, the displacement is bounded. When the potential is not quadratic, a nonlinear system is obtained. We shall consider the nonlinear oscillator later on. Here, qualitatively, two nonlinear potentials are shown in the centre and right panels of Fig. 4.1. Bounded motion is obtained whenever the potential attains infinity for large values of $|u|$. However, unbounded motion can be observed if some finite constant bounds the potential ϕ. To understand these claims, it can be useful to visualise the trajectories in the *phase plane*. This is a plane whose axes are $x = u(t)$ and $y = du(t)/dt$. For the quadratic case, trajectories are ellipses and symmetric about the x and y axes. Periodicity is evident by inspection of the phase portraits, which appear as closed loops. The energy components also oscillate periodically, in this case, at one single frequency. For the central panel of Figure 4.1, trajectories are now symmetric only about the u axis: the nonlinearity is such that symmetry is broken in u, as evident from panel 2 of Fig. 4.1. However, motion is still periodic via the combination of a number of frequencies with a common factor (the fundamental frequency). For the right panel of Fig. 4.1, the motion may be bounded or unbounded, depending on the value of H_0, as seen in

the right panel of Figure 4.1.

When motion is periodic, the period can be estimated from the energy, since

$$\frac{du}{dt} = \pm\sqrt{\frac{2}{m}}\sqrt{H_0 - \phi}. \tag{4.14}$$

The plus or minus sign depends on which side of the phase portrait the particle is (i.e. whether it is found in the $y \geq 0$ or $y < 0$ half-plane). Considering the plus sign and inverting, one has

$$\mathrm{d}t = \sqrt{\frac{m}{2}}\frac{\mathrm{d}u}{\sqrt{H_0 - \phi}}. \tag{4.15}$$

Thus, half the period is obtained by integrating between u_1 and u_2, which are the points in the phase plane where $du/dt = 0$. Denoting the period τ, one has

$$\tau = \sqrt{2m}\int_{u_1}^{u_2}\frac{\mathrm{d}u}{\sqrt{H_0 - \phi}}. \tag{4.16}$$

In general, this equation does not have a closed-form solution. Some cases are exceptional, including the case of simple harmonic motion. For that, $u_1 = -u_2$, and $H_0 = H(t) = \frac{Ku_2^2}{2}$, where the last equality holds since $du/dt|_{u=u_2} = 0$. Hence, the period is obtained as the integral over one quadrant (one-quarter of a loop):

$$\tau = 2\sqrt{2m}\int_0^{u_2}\frac{\mathrm{d}u}{\sqrt{\frac{Ku_2^2}{2} - \frac{Ku^2}{2}}} = 4\sqrt{\frac{m}{K}}\arctan\left(\frac{u}{\sqrt{u_2^2 - u^2}}\right)\Big|_0^{u_2} = 2\pi\sqrt{\frac{m}{K}}. \tag{4.17}$$

It is convenient to introduce the *radian frequency* $\omega_0 = \sqrt{K/m} = 2\pi f_0$, where f_0 is a linear frequency (in Hz). Hence, one has

$$\tau = \frac{2\pi}{\omega_0} = \frac{1}{f_0}. \tag{4.18}$$

4.3 Frequency domain analysis

Before proceeding, it is worth introducing the definitions of the Laplace and Fourier transforms in continuous time. These powerful tools find application in analysing *linear and time-invariant* (LTI) systems, a restricted but important subclass of the above examples. Through these transforms, the model problem defined in the time domain is described in the frequency domain, where certain properties may be ascertained more easily.

4.4 The Laplace and Fourier transforms

For the continuous function $u(t)$, the Laplace transform is obtained as

$$\hat{u}(s) = \int_{\{-\infty,0\}}^{\infty} u(t)e^{-st}\,\mathrm{d}t := \mathcal{L}\{u\}(s), \tag{4.19}$$

where $s = j\omega + \sigma \in \mathbb{C}$ is a complex variable. The term *transform* indicates both the integral operation of turning u into $\hat{u}$, as well as the transformed variable $\hat{u}$. We will use this terminology interchangeably. The lower bound in the integral means that the transform can be defined as two-sided (starting from $-\infty$), or one-sided (starting from 0), allowing initial conditions to be incorporated. In the analysis of linear, time-invariant (LTI) systems, both continuous and discrete, one usually computes s from the given model problem via direct substitution of the transforms. Then, the stability of the underlying system may be inferred by direct inspection of s, as will be stated below. Applying the Laplace transform to the time derivative of $u(t)$ gives:

$$\mathcal{L}\left\{\frac{du}{dt}\right\} = \int_{-\infty}^{\infty} \frac{du(t)}{dt} e^{-st}\,\mathrm{d}t = s\int_{-\infty}^{\infty} u(t)e^{-st}\,\mathrm{d}t = s\mathcal{L}\{u\}, \tag{4.20}$$

where integration by parts was used, and where it was assumed that $u(t)$ dies out fast enough as $t \to \pm\infty$. Using analogous arguments, one can show that higher-order derivatives transform as:

$$\mathcal{L}\left\{\frac{d^p u}{dt^p}\right\} = s^p \mathcal{L}\{u\}. \tag{4.21}$$

In practice, it is often useful to substitute simpler expressions than the transforms via an *ansatz*. So, one may employ the test solutions

$$u(t) = \hat{u}e^{st}, \tag{4.22}$$

for an appropriate constant complex amplitude $\hat{u}$. It is immediate to verify that the derivatives of these test solutions transform in the same way as the derivatives of the Laplace in (4.21). For this reason, for LTI systems, the use of the test solutions ultimately yields the same qualitative analysis as the substitution of the full transforms, and we shall use such solutions accordingly. When one transforms a model problem according to Laplace, the transformed variable $\hat{u}$ is usually expressed as a *rational function*:

$$\hat{u} = A(s)B^{-1}(s). \tag{4.23}$$

For a rational Laplace transform, the order of the polynomial A or B is the power of the highest power of s. The roots of the numerator polynomial are called the *zeros* of the Laplace transform, and the roots of the denominator polynomial are called the *poles*. The pole locations border the region of convergence of a rational Laplace transform. Hence, a linear, time-invariant system is stable if and only if all the poles of its transfer function lie in the left half of the complex plane. All the poles must have *negative real parts*. This is one of the most important and useful results in Laplace transform theory.

A second powerful property, already seen from (4.21), is the transformation of derivatives as multiplications. In theory, once the problem is solved in the frequency domain, one can return to the time domain by computing the inverse transforms. This approach, however, presents some drawbacks. Most notably, inverse transforms are difficult to compute, and closed-form solutions are available only in a few cases. Second, these techniques do not generalise to nonlinear, time-variant systems (though some exceptions exist, such as Volterra kernels). Nonetheless, frequency domain techniques remain a popular analysis tool

since most nonlinear systems reduce to linear under suitable conditions, and one may infer (at least qualitatively) some useful properties of the model systems under study.

Laplace transforms (and inverses) may be difficult to compute, and usually, one resorts to table look-ups. One useful transform pair, used later on, is:

$$\mathcal{L}\left\{e^{-\sigma(t-t')}\sin(a(t-t'))\Theta(t,t')\right\}(s) = \frac{ae^{-st'}}{(s+\sigma)^2+a^2}, \tag{4.24}$$

The function $\Theta(t,t')$ is the *step function* (i.e. $\Theta(t<t')=0, \Theta(t\geq t')=1$)

When one considers $\sigma = 0$ in (4.19), in the two-sided form, the continuous *Fourier transform* is obtained. This is useful to compute the solutions' magnitude and phase (i.e., the spectrum) and will also be used throughout. The definition of this transform is as:

$$\hat{u}(\omega) = \int_{-\infty}^{\infty} u(t)e^{-j\omega t}\,\mathrm{d}t := \mathcal{F}\{x\}(\omega). \tag{4.25}$$

A couple of useful transform pairs are given here as

$$\mathcal{F}\left\{u(t)\sin(at)\right\}(\omega) = \frac{\hat{u}(\omega-a)-\hat{u}(\omega+a)}{2j}, \tag{4.26a}$$

$$\mathcal{F}\left\{e^{-\sigma t}\Theta(t,t')|\sigma\geq 0\right\}(\omega) = \frac{e^{-j\omega t'}}{\sqrt{2\pi}(\sigma+j\omega)}. \tag{4.26b}$$

In the second transform, $\Theta(t,t')$ is again the step function as in (4.24).

Chapter 5

Harmonic Motion in Continuous Time

The discussion in the previous chapter suggests studying the simplest kind of oscillation, obtained when the potential function in (4.1) is a parabola that, is $\phi(u) = Ku^2/2$. In this case:

$$m\frac{d^2u}{dt^2} = -Ku. \tag{5.1}$$

This is a model for the *simple harmonic oscillator*, where K has the interpretation of a stiffness constant, or a rigidity, measured in $\mathrm{N{\cdot}m^{-1}}$. The same mathematical model describes various other physical systems, such as the series LC circuit, a mass hanging from the tip of a cantilever beam, the pendulum, the Helmholtz resonator, and many others. The constants in the model change, and so does the physical interpretation of the state variable u, but the mathematical model and its properties remain unchanged, so we will keep studying form (5.1) without loss of generality, knowing that the results derived below apply equally to all the other systems, see also Figure 5.1.

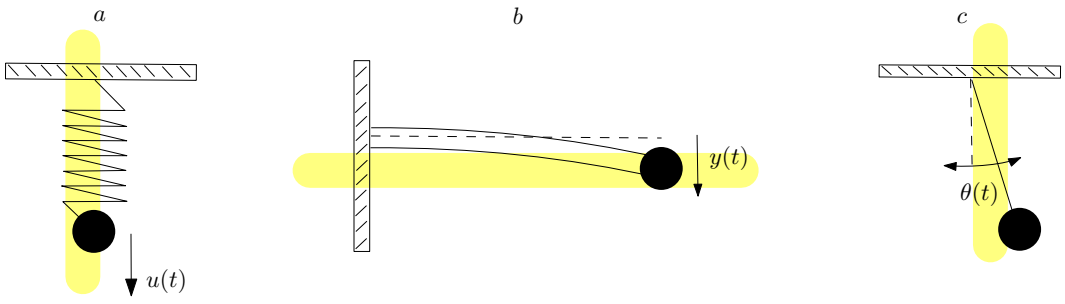


Figure 5.1: Examples of harmonic oscillators. The mass-spring system (a), the cantilever beam with a point mass (b), and the pendulum (c) may display harmonic motion in the case of small amplitude vibration.

5.1 The undamped oscillator

Scaling the equation by mass, and in the absence of friction and sources, a model for simple harmonic motion is thus obtained as:

$$\frac{d^2u}{dt^2} = -\omega_0^2 u, \tag{5.2}$$

where ω_0 is as per (4.18). This simple equation is the basic building block of vibration analysis and simulation, and its properties are worth studying in some detail. First, the equation admits analytic solutions taking a variety of forms:

$$u(t) = \hat{u}_+ e^{j\omega_0 t} + \hat{u}_- e^{-j\omega_0 t}, \tag{5.3a}$$
$$u(t) = A\cos(\omega_0 t + \zeta), \tag{5.3b}$$
$$u(t) = B\sin(\omega_0 t + \theta), \tag{5.3c}$$
$$u(t) = p\cos(\omega_0 t) + q\sin(\omega_0 t). \tag{5.3d}$$

Here, $(\hat{u}_-, \hat{u}_+) \in \mathbb{C}$, $(A, B, \zeta, \theta, p, q) \in \mathbb{R}$. It is immediate to see that all these forms solve equation (5.2). However, it is unclear which form one should work with. This apparent confusion is resolved by completing the equation of motion with the *initial conditions*, as per (4.4). Once the initial displacement u_0 and the initial velocity v_0 are set, all the forms above yield the same solution, as one can verify immediately. For instance, solution (5.3d) becomes:

$$u(t) = u_0\cos(\omega_0 t) + \frac{v_0}{\omega_0}\sin(\omega_0 t), \tag{5.4}$$

whilst solution (5.3b) becomes:

$$u(t) = \sqrt{u_0^2 + \frac{v_0^2}{\omega_0^2}}\cos\left(\omega_0 t + \arctan\frac{v_0}{\omega_0 u_0}\right). \tag{5.5}$$

Form (5.5) can be converted immediately to (5.4) by using the trigonometric angle sum identities and by using $\sin(\arctan(x)) = x(\sqrt{1+x^2})^{-1}$, $\cos(\arctan(x)) = (\sqrt{1+x^2})^{-1}$. Proving the equivalence of the other solutions in (5.3) is left as an exercise for the reader. In particular, note that solution (5.3a) is, in fact, *real*, even though it is expressed in complex form when the initial conditions are themselves real.

One remarkable property of harmonic motion, already encapsulated in (4.18), is the independence of the period of oscillation from the amplitude. Regardless of the initial conditions, the period of the oscillations depends exclusively on ω_0. This can be seen immediately from (5.4) and (5.5) since the trigonometric functions are periodic with period 2π. This conclusion is valid mathematically, but it is not always true in physical systems: when the amplitude of the vibration becomes large enough, various kinds of amplitude-dependent phenomena ensue, and the system becomes *nonlinear*. This is the case of the pendulum, for instance, for which the period of the oscillation *decreases* as the initial amplitude gets larger. For now, however, the interest lies in understanding the mathematical properties of the simple harmonic oscillator, which will be extremely useful when analysing more complex systems.

5.1.1 Energy analysis

The energy analysis was carried out in the general case in section 4.1. Here, we want to derive more specific identities valid for the simple harmonic oscillator. From (4.9), one has:

$$H = \frac{m}{2}\left(\frac{du}{dt}\right)^2 + \frac{Ku^2}{2}, \tag{5.6}$$

and using form (5.3b):

$$H = \frac{m}{2}\left(-\omega_0 A \sin(\omega_0 t + \zeta)\right)^2 + \frac{K}{2}\left(A\cos(\omega_0 t + \zeta)\right)^2. \tag{5.7}$$

The sine and the cosine in the above expression lag each other by a phase difference of $\pi/2$, meaning that, as the sine takes on a zero value, the cosine is either plus or minus one. Conversely, as the cosine takes on a zero value, the sine is either plus or minus one. Since H is conserved, one has:

$$H = \frac{KU^2}{2} = \frac{mV^2}{2}, \tag{5.8}$$

where $U := A$ is the displacement amplitude of the oscillator i.e., its largest displacement; $V := \omega_0 A$ is the velocity amplitude of the oscillator, i.e. the largest velocity during a cycle. The formula above reveals that the oscillator's energy can be expressed in a "displacement-only" or a "velocity-only" form, though its value always remains conserved.

5.2 The damped oscillator

Of course, no such thing as a perfectly energy-conserving oscillation can occur. Physical systems are subjected to all sorts of damping through mechanical friction in their constituent parts or other physical effects. One common type of damping is *viscous damping*, proportional to the velocity. Viscosity models various important cases, such as the friction of a sphere dropping into a dense liquid and the loss effects experienced by elastic waves in air. Thus, the oscillator equation (5.1) is modified to take into account the effects of viscosity as follows:

$$m\frac{d^2u}{dt^2} = -Ku - R\frac{du}{dt}, \tag{5.9}$$

where $R := 2\sigma m \geq 0$ is the *mechanical resistance* and σ, measured in s^{-1}, is a loss factor. When scaled by mass, the equation is:

$$\frac{d^2u}{dt^2} = -\omega_0^2 u - 2\sigma\frac{du}{dt}. \tag{5.10}$$

Finding a solution in this case may be accomplished by various methods. A particularly useful one is through the following *ansatz*. Assume $u := \hat{u}e^{st}$, where $(\hat{u}, s) \in \mathbb{C}$. Inserting this test solution in the above equation, one obtains:

$$(s^2 + 2\sigma s + \omega_0^2)\hat{u} = 0, \tag{5.11}$$

where, conveniently, the equation was scaled by mass. The equation may be solved by requiring the following:

$$s^2 + 2\sigma s + \omega_0^2 = 0 \implies s_\pm = -\sigma \pm \sqrt{\sigma^2 - \omega_0^2}. \tag{5.12}$$

The solutions $s_\pm$ look like real numbers, but, in fact, they are more conveniently expressed as:

$$s_\pm = -\sigma \pm j\sqrt{\omega_0^2 - \sigma^2}. \tag{5.13}$$

This expression is convenient because the loss factor σ is usually smaller than the natural frequency ω_0, meaning that, in most cases, $s_\pm$ are complex numbers with real and imaginary parts. This scenario defines the *underdamped oscillator*. Thus, the general solution will be:

$$u(t) = \hat{u}_+ e^{(-\sigma + j\sqrt{\omega_0^2 - \sigma^2})t} + \hat{u}_- e^{(-\sigma - j\sqrt{\omega_0^2 - \sigma^2})t} = e^{-\sigma t}\left(\hat{u}_+ e^{j\sqrt{\omega_0^2 - \sigma^2}t} + \hat{u}_- e^{-j\sqrt{\omega_0^2 - \sigma^2}t}\right).$$

Defining $\omega_d := \sqrt{\omega_0^2 - \sigma^2}$, one has:

$$u(t) = e^{-\sigma t}\left(\hat{u}_+ e^{j\omega_d t} + \hat{u}_- e^{-j\omega_d t}\right). \tag{5.14}$$

This form is reminiscent of the form (5.3a) for the undamped case. However, two notable differences exist: the presence of the decay envelope $e^{-\sigma t}$, and a different vibration frequency $\omega_d \leq \omega_0$. When the initial conditions are substituted in, the solution takes the form:

$$u(t) = e^{-\sigma t}\left(u_0 \cos(\omega_d t) + \frac{v_0 + \sigma u_0}{\omega_d}\sin(\omega_d t)\right), \tag{5.15}$$

generalising (5.4) to the damped case.

The decay envelope is an exponential function of time scaled by the loss factor σ. In acoustics, it is preferable to characterise decays via other constants. One such constant, denoted here τ_{60}, denotes the time the signal takes to decay by 60 *decibels* (dB) compared to the amplitude at the time $t = 0$. Using the definition of decibels, deriving τ_{60} is accomplished by:

$$20\log_{10} e^{\sigma\tau_{60}} = 60. \tag{5.16}$$

The logarithm base ten is turned into a natural logarithm to give:

$$20\log e^{\sigma\tau_{60}} = 60\log(10), \tag{5.17}$$

which, ultimately, yields:

$$\tau_{60} := 3\sigma^{-1}\log(10). \tag{5.18}$$

Different constants are sometimes used. A widely used characterisation of damped oscillations is through the "quality" factor (*Q-factor*), defined as

$$Q := (2\sigma)^{-1}\omega_0. \tag{5.19}$$

This is the number of cycles required for the amplitude of motion to reduce to $e^{-\pi}$ of its original value. Larger values of Q imply more oscillations occur before the mass halts.

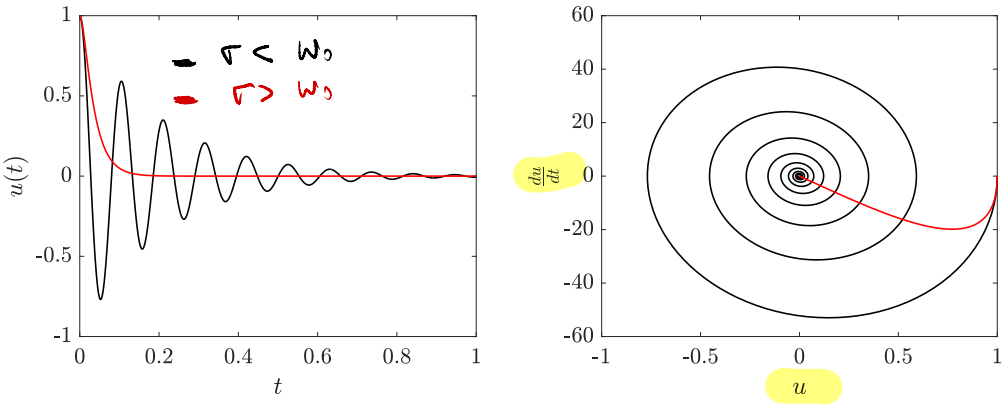


Figure 5.2: Time domain plots and phase portraits of an underdamped oscillator (black) and an overdamped oscillator (red).

5.2.1 Behaviour of the damped oscillator for large σ

When $\sigma \geq \omega_0$, the roots in (5.12) become real numbers. The case $\sigma = \omega_0$ defines the *critically damped* oscillator, whereas when $\sigma > \omega_0$, one recovers the *overdamped oscillator*. Note that, in all such cases, $s_\pm$ remains strictly negative, and, thus, the solution u is strictly decaying: no oscillations take place. Examples of the underdamped and overdamped behaviours are given in Figure 5.2.

5.2.2 Energy analysis

Just like the conservative case, useful energy relations will be derived for the underdamped oscillator. Solution (5.14) can be written equivalently as:

$$u(t) = A(t)\cos(\omega_d t + \zeta), \tag{5.20}$$

generalising (5.3b) to the underdamped case. Here, $A(t) := A_0 e^{-\sigma t}$, for some amplitude A_0 set by the initial data. The instantaneous energy is again of the form (5.6):

$$H(t) = \frac{m}{2}\left(\frac{du}{dt}\right)^2 + \frac{Ku^2}{2}. \tag{5.21}$$

Substituting expression (5.20) yields:

$$H(t) = \frac{m\omega_d^2 A^2}{2} - m\omega_d A\frac{dA}{dt}\sin(\omega_d t + \zeta)\cos(\omega_d t + \zeta) + \frac{m}{2}\left(\frac{dA}{dt}\right)^2\cos^2(\omega_d t + \zeta). \tag{5.22}$$

Furthermore, using the definition of $A(t)$ above, and remembering that $K = m\omega_d^2 + m\sigma^2$, the expression for the energy becomes:

$$H(t) = \frac{m\omega_d^2 A^2}{2} + m\sigma^2 A^2\cos^2(\omega_d t + \zeta) - m\omega_d A\frac{dA}{dt}\sin(\omega_d t + \zeta)\cos(\omega_d t + \zeta). \tag{5.23}$$

This expression can be averaged over one cycle (of length $\tau := 2\pi\omega_d^{-1}$, see (4.18)) to give the average available energy:

$$\lceil H \rfloor_{t_0}^{t_0+\tau} := \tau^{-1} \int_{t_0}^{t_0+\tau} H(t)\ \mathrm{d}t. \tag{5.24}$$

Note that the last term in (5.23) integrates to zero. Then, one has:

$$\lceil H \rfloor_{t_0}^{t_0+\tau} = \tau^{-1} m A^2(t_0) \left(\frac{{\omega_d}^4 + 2{\omega_d}^2\sigma^2 + 2\sigma^4}{4\sigma\left({\omega_d}^2 + \sigma^2\right)} - \frac{\mathrm{e}^{-\frac{4\pi\sigma}{\omega_d}} \left({\omega_d}^4 + 2{\omega_d}^2\sigma^2 + 2\sigma^4\right)}{4\sigma\left({\omega_d}^2 + \sigma^2\right)} \right). \tag{5.25}$$

Since σ is small, one may expand the resulting expression in a Taylor series to get:

$$\lceil H \rfloor_{t_0}^{t_0+\tau} \approx \tau^{-1} m A^2(t_0) \left(\pi\omega_d - 2\pi^2\sigma\right) = \frac{m A^2(t_0) \omega_d^2 (1 - 2\pi\sigma\omega_d^{-1})}{2} \tag{5.26}$$

Here, $A(t_0)$ represents the amplitude of the oscillation at the beginning of the cycle. This expression is reminiscent of the expression for the energy of the conserved system, (5.8), except for the presence of a correction term due to the loss. In other words, one has:

$$\lceil H \rfloor_{t_0}^{t_0+\tau} \approx \frac{K U^2(t_0)(1 - 2\pi\sigma\omega_d^{-1})}{2} = \frac{m V^2(t_0)(1 - 2\pi\sigma\omega_d^{-1})}{2}, \tag{5.27}$$

where U and V are the displacement and velocity amplitudes. This expression represents the available energy at the time t_0, the energy that has not yet been converted to heat and that the oscillator can use to sustain its motion.

A second important energy relation involves the rate of change of the total energy. To that end, derive (5.21) with respect to time:

$$\frac{dH}{dt} = \frac{d}{dt}\left(\frac{m}{2}\left(\frac{du}{dt}\right)^2 + \frac{Ku^2}{2}\right) = \frac{du}{dt}\left(m\frac{d^2u}{dt^2} + Ku\right) = -R\left(\frac{du}{dt}\right)^2 \leq 0. \tag{5.28}$$

This expresses the *instantaneous energy balance* of the oscillator: energy decreases over time.

5.3 The forced oscillator

The final case under consideration pertains to the forced oscillator. Frequently, a system enters into oscillation due to its connection with another oscillating system, referred to here as the *driving system*. The driven system absorbs energy from the driving system and oscillates accordingly. In such instances, the driven system typically does not return significant energy to the driving system. This lack of feedback may occur because the linkage between the two systems is weak or because the driving system possesses ample reserve energy, rendering the feedback negligible. Thus, assume the following equation of motion:

$$m\frac{d^2u}{dt^2} = -Ku - R\frac{du}{dt} + F(t), \tag{5.29}$$

where $F(t)$ represents the external driving force.

5.3.1 Harmonic forcing

As a first example, assume the forcing to be periodic and characters by some radian frequency ω. Its form may be expressed via a complex exponential, $F(t) := f_0 e^{j\omega t}$. It is convenient to derive a particular solution and to study its behaviour as the input driving frequency ω varies. Its form is obtained immediately by assuming that the oscillator vibrates at the same frequency as the driver:

$$u = \hat{u} e^{j\omega t}, \tag{5.30}$$

for some complex amplitude $\hat{u}$. Inserting this expression in (5.29) results in:

$$j\omega \left(j\omega m - jK\omega^{-1} + R \right) \hat{u} = f_0. \tag{5.31}$$

Defining the complex velocity amplitude as $\hat{v} := j\omega\hat{u}$, one obtains two equivalent definitions of the *mechanical impedance* Z_m, as:

$$Z_m := f_0 (j\omega\hat{u})^{-1}, \quad Z_m := f_0 \hat{v}^{-1}. \tag{5.32}$$

The mechanical impedance is a complex number whose real part corresponds to the mechanical resistance R and whose imaginary part determines the phase difference between the input force and the output displacement or velocity. The complex expression for the impedance is, thus:

$$Z_m = R + jm\omega^{-1}(\omega^2 - \omega_0^2). \tag{5.33}$$

In complex polar form, the impedance is expressed as:

$$Z_m = r\, e^{j\theta}, \; r := \sqrt{R^2 + m^2\omega^{-2}(\omega^2 - \omega_0^2)^2}, \; \tan\theta := m(\omega R)^{-1}(\omega^2 - \omega_0^2). \tag{5.34}$$

Magnitude and phase plots for the impedance under two different choices of the mechanical resistance R are given in Figure 5.3. One gets:

$$\hat{u} = f_0 (\omega\, r)^{-1}\, e^{-j(\theta + \frac{\pi}{2})}, \quad \hat{v} = f_0 r^{-1}\, e^{-j\theta}. \tag{5.35}$$

A quick study of the displacement and velocity magnitude responses reveals the presence of stationary points (maxima) at the frequencies:

$$\omega^{\hat{u}}_{\max} = m^{-1} \sqrt{m^2\, \omega_0{}^2 - (R/2)^2} \approx \omega_0 - (4\omega_0 m)^{-1} R^2, \quad \omega^{\hat{v}}_{\max} = \omega_0. \tag{5.36}$$

Since R is small, the two maxima are approximately the same. At their respective maxima, the displacement and velocity amplitudes become:

$$|\hat{u}(\omega^{\hat{u}}_{\max})| \approx (\omega_0 R)^{-1} f_0, \quad |\hat{v}(\omega^{\hat{v}}_{\max})| = R^{-1} F_0 \tag{5.37}$$

Examples of the displacement and velocity magnitude and phase responses are visible in Figure 5.4. An estimate for the mechanical resistance R can be obtained via the "half-power point" method. Since power is proportional to the square of $|\hat{u}|$, half of the maximum "power", from (5.37), is obtained as $2^{-1/2}(\omega_0 R)^{-1} f_0$. Thus, the half-power frequencies are obtained by solving the following according to ω:

$$|\hat{u}| - 2^{-1/2}(\omega_0 R)^{-1} f_0 = 0 \implies (\omega r)^{-1} - 2^{-1/2}(\omega_0 R)^{-1} = 0. \tag{5.38}$$

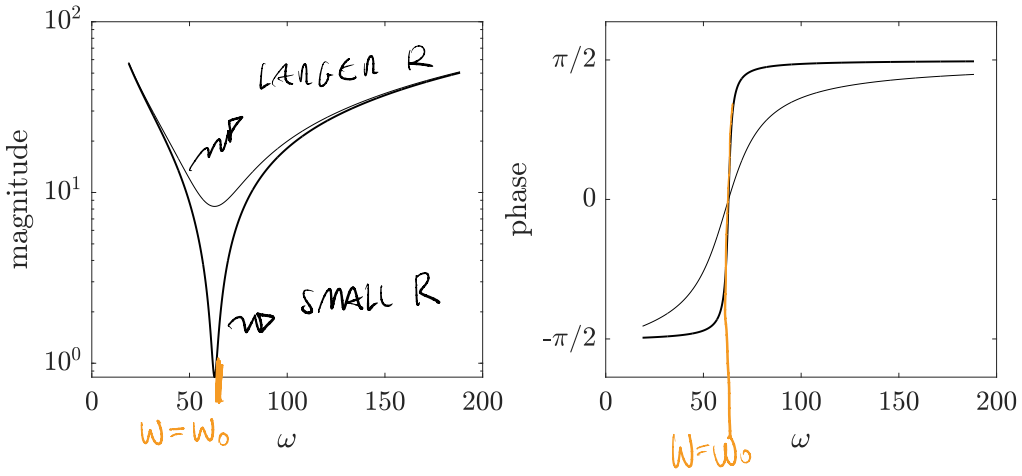


Figure 5.3: Magnitude and phase response of the impedance Z_m under two different values of the mechanical resistance R. Here, $\omega_0 = 63.8$ rad·s^{-1}, $f_0 = 1$ N, $m = 0.3$ kg, and $\tau_{60} = 5$ s (thick line) and 0.5 s (thin line).

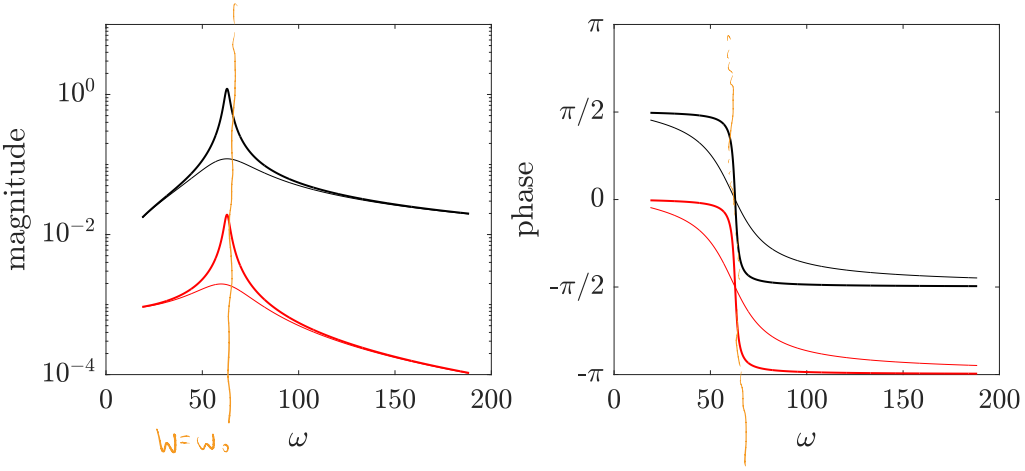
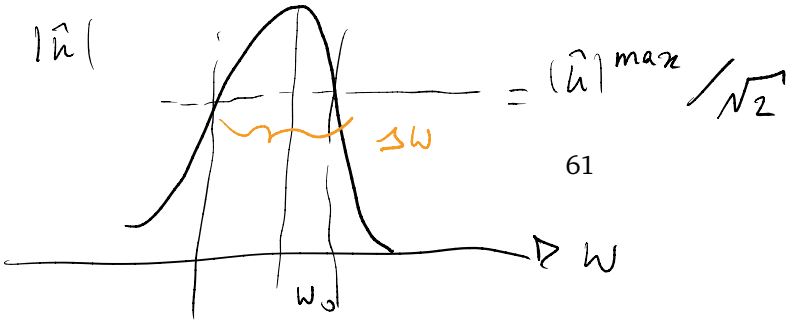


Figure 5.4: Magnitude and phase response of the displacement amplitude $\hat{u}$ (red) and velocity amplitude $\hat{v}$ (black) under two different values of the mechanical resistance R. Here, $\omega_0 = 63.8$ rad·s^{-1}, $f_0 = 1$ N, $m = 0.3$ kg, and $\tau_{60} = 5$ s (thick) and 0.5 s (thin).



This defines a quadratic equation in ω which one can solve analytically. Computing a Taylor expansion about the small parameter R, the approximate half-power points take the following form:

$$\omega^{\hat{u}}_{\text{half}\pm} :\approx \omega_0 \pm (2m)^{-1}R. \tag{5.39}$$

An estimate for R is obtained by computing the *bandwidth* of the response:

$$\Delta\omega^{\hat{u}}_{\text{half}} := \omega^{\hat{u}}_{\text{half}+} - \omega^{\hat{u}}_{\text{half}-} \approx m^{-1}R. \tag{5.40}$$

From here, the decay time τ_{60} is obtained immediately, as $\tau_{60} = 6(\Delta\omega^{\hat{u}}_{\text{half}})^{-1}\log(10)$. Similar results may be obtained using the velocity amplitude $\hat{v}$. Their derivation is left as an exercise for the reader.

Turning now to the phase, from (5.35), it is clear that the displacement lags the velocity by 90 degrees, or a quarter of a cycle. Furthermore, by inspection of (5.34), it is clear that for small values of $\theta \approx -\pi/2$ for small ω, and $\theta \approx \pi/2$ for large ω. Thus, the limiting phase angles for $\hat{u}$ and $\hat{v}$ are easily derived. Examples of magnitude and phase responses for the displacement and the velocity amplitudes are given in Figure 5.4.

Before proceeding, note that the complete solution to (5.29) is given by the superposition of the particular solution just found and the solution to the homogeneous equation, (5.20). Thus, taking the real part of $\hat{u}$ in (5.35), one obtains:

$$u(t) = A_0 e^{-\sigma t}\cos(\omega_d t + \zeta) + f_0(\omega r)^{-1}\sin(\omega t - \theta), \quad t \geq 0. \tag{5.41}$$

As before, the initial conditions fix A_0 and ζ. Initially, the motion comprises two frequencies and may be completely aperiodic. Once the contribution from the initial condition has died out, the system vibrates at the sole frequency of the driver.

The expressions for the impedance in (5.32) may be simplified when either m, R or K are large. Motion is said to be:

- *stiffness controlled* when $K\omega^{-1}$ is large, in which case $Z_m \approx -jK\omega^{-1}$, $\hat{u} \approx f_0 K^{-1}$,
- *resistance controlled* when R is large, in which case $Z_m \approx R$, $d\hat{u}/dt \approx f_0 R^{-1}$,
- *mass controlled* when $m\omega$ is large, in which case $Z_m \approx j\omega m$, $d^2\hat{u}/dt^2 \approx f_0 m^{-1}$.

It is remarked that every driven oscillator is mass-controlled in the frequency range well above ω_0, is resistance controlled around ω_0, and is stiffness controlled at lower frequencies.

5.3.2 Impulse response and Green's function

The previous section detailed the case of a steady driving force. Of course, the oscillator may be acted upon by various forces: stationary, non-stationary, periodic, aperiodic, wideband, narrowband, and so on. Studying (5.29) in the general case can get quickly out of hand. There is, however, a general result allowing to obtain the response of the oscillator under a general force $F(t)$. Sometimes, the oscillator is activated impulsively, quickly imparting energy into the system. When the impulsive loading is modelled via a *Dirac delta*, the oscillator's behaviour is known as its *impulse response*. Thus, assume $F(t) := \delta(t - t')$. The Dirac delta is a distribution for which two important properties hold:

$$\delta(t - t') = 0 \;\text{ if } t \neq t', \quad \int_{-\infty}^{\infty} \delta(t - t')f(t)\,\mathrm{d}t = f(t'), \tag{5.42}$$

for any function $f(t)$ continuous in a neighbourhood of t'. The solution to (5.29) under this choice of $F(t)$ is known as *Green's function* $G(t,t')$. The Green's function is useful as it allows computing the solution to any force via the *convolution integral*. The mathematical definition of the Green's function is as follows:

$$\left(m\frac{d^2}{dt^2}+K+R\frac{d}{dt}\right)G(t,t')=\delta(t-t'), \tag{5.43}$$

where it is assumed that $(t,t')\geq 0$. Now, assume to be wanting to compute $u(t)$ from (5.29) under a generic force $F(t)$. It is immediate to see that:

$$u(t)=\int_{-\infty}^{\infty}G(t,t')F(t')\,\mathrm{d}t', \tag{5.44}$$

where the integral above is called *convolution integral*. To prove this, substitute (5.44) into the left-hand side of (5.29):

$$\left(m\frac{d^2}{dt^2}+K+R\frac{d}{dt}\right)u(t)=\int_{-\infty}^{\infty}\left(m\frac{d^2}{dt^2}+K+R\frac{d}{dt}\right)G(t,t')F(t')\,\mathrm{d}t'=F(t), \tag{5.45}$$

which is the right-hand side of (5.29).

To compute the Green's function, one first transforms (5.43) in the frequency domain via a two-sided Laplace transform, as defined in (4.19):

$$\int_{-\infty}^{\infty}e^{-st}\left(m\frac{d^2}{dt^2}+K+R\frac{d}{dt}\right)G(t,t')\,\mathrm{d}t=\int_{-\infty}^{\infty}e^{-st}\delta(t-t')\,\mathrm{d}t \tag{5.46}$$

Defining the transformed Green's function as:

$$\hat{G}(s,t'):=\int_{-\infty}^{\infty}e^{-st}G(t,t')\,\mathrm{d}t, \tag{5.47}$$

and remembering that derivatives under the Laplace transform become multiplications of the transformed variable, one gets:

$$\hat{G}(s,t')=m^{-1}e^{-st'}(s^2+2\sigma s+\omega_0^2)^{-1}. \tag{5.48}$$

The second-order polynomial is rewritten as follows:

$$s^2+2\sigma s+\omega_0^2=(s+\sigma)^2+\omega_d^2, \tag{5.49}$$

where ω_d^2 is the oscillation frequency of the damped oscillator, as defined in (5.14) (assume $\sigma<\omega_0$ here). Using the first Laplace transform pair in (4.24), one can invert $\hat{G}$ and compute the time-domain Green's function:

$$G(t,t')=e^{-\sigma(t-t')}(m\omega_d)^{-1}\sin\left(\omega_d(t-t')\right), \tag{5.50}$$

valid for $t\geq t'$, and $G(t,t')=0$ otherwise.

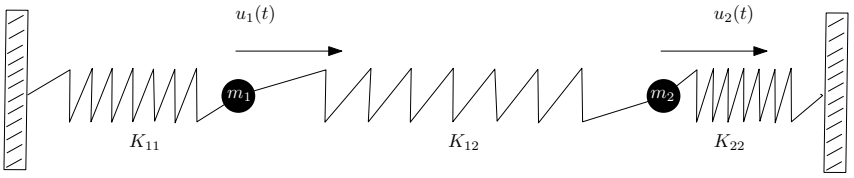


Figure 5.5: A two-mass oscillator.

5.4 Multiple degrees of freedom

The case of the single oscillator can be readily extended to systems comprising more degrees of freedom. Physically, multiple degrees of freedom are found in systems such as the mass-spring chain, where masses move under the action of the forces due to the extension and compression of the interconnecting springs. Like the mass-spring oscillator, this is a rather abstract system, seldom encountered in practice. As we will see in forthcoming chapters, however, the motion of a taut string can be thought of as a particular case of the mass-spring chain and, hence, it is useful to study this system in some detail here. Just like the simple harmonic oscillator, it is natural to describe motion by tracking the displacement of the masses as a function of time. As opposed to the simple harmonic oscillator, however, the overall motion may not be harmonic because it may not repeat itself at exact time intervals. Employing a special set of coordinates called the *modal coordinates*, the overall motion may still be described in terms of periodic oscillations of the *modes* of the system. Orthogonality is a distinctive feature of the modes in that the overall motion is a *linear superposition* of these elementary periodic motions. A generalisation of the frequency domain techniques is possible, yielding a description of the system in terms of *eigenvalues* and *eigenvectors*. The eigenvalues have a direct interpretation in terms of the resonant frequencies of the system and the eigenvectors as their mode shapes.

5.4.1 Two masses, three springs

The device sketched in Figure 5.5 represents a simple example of a system comprising two degrees of freedom. This is a useful test case for developing mathematical techniques for analysing larger systems. Describing the system is most naturally accomplished by tracking the evolution over time of the displacements $u_1(t)$, $u_2(t)$ of the two masses m_1, m_2 from their rest position. The equations of motion for this system are as follows:

$$m_1 \frac{d^2 u_1}{dt^2} = -K_{11}u_1 - K_{12}(u_1 - u_2), \tag{5.51a}$$

$$m_2 \frac{d^2 u_2}{dt^2} = -K_{22}u_2 + K_{12}(u_1 - u_2). \tag{5.51b}$$

The matrix-vector formalism allows writing (5.51) in a more compact way. To that end, consider the following:

$$\mathbf{M}\,\frac{d^2\mathbf{u}}{dt^2} = -\mathbf{K}\,\mathbf{u}, \tag{5.52}$$

where:

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \ \mathbf{K} = \begin{bmatrix} K_{11} + K_{12} & -K_{12} \\ -K_{12} & K_{22} + K_{12} \end{bmatrix}, \ \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \tag{5.53}$$

In the above, $\mathbf{M}$ is called the *mass matrix* and $\mathbf{K}$ is called the *stiffness matrix*. Clearly, (5.52) is a much more compact version of the corresponding "unrolled" version (5.51). Form (5.52) allows applying a direct extension of the frequency domain techniques from Section 4.3, regardless of whichever form $\mathbf{M}$ and $\mathbf{K}$ have. To that end, one can conveniently define a test solution of the form:

$$\mathbf{u} = \hat{\mathbf{u}} e^{st}, \tag{5.54}$$

which is a generalisation of ansatz (4.22) to the vector case. Like the scalar case, $\hat{\mathbf{u}}$ is a constant complex amplitude, and s is the Laplace complex variable. Since in this problem, damping is neglected, we can simplify the discussion by requiring s to be purely imaginary, $s = j\omega$, so that, after substituting the test solution in (5.52), one gets:

$$\left(\mathbf{K} - \omega^2 \mathbf{M}\right) \hat{\mathbf{u}} = 0. \tag{5.55}$$

This defines an eigenvalue problem in which the resonant frequencies ω are obtained as the square root of the eigenvalues. Nontrivial solutions are obtained when the determinant of $\mathbf{K} - \omega^2 \mathbf{M}$ is zero. This is a general result since we made no assumptions about $\mathbf{M}$, $\mathbf{K}$ or $\mathbf{u}$. Specialising this result to the two-mass case of Figure 5.5, one has:

$$(K_{11} + K_{12} - \omega^2 m_1)(K_{22} + K_{12} - \omega^2 m_2) - K_{12}^2 = 0, \tag{5.56}$$

or

$$a\omega^4 + b\omega^2 + c = 0, \tag{5.57}$$

with:

$$a = m_1 m_2, \ b = -K_{11} m_2 - K_{12}(m_1 + m_2) - K_{22} m_1, \ c = K_{11} K_{12} + K_{11} K_{22} + K_{12} K_{22}.$$

Solving for ω^2, one gets

$$\omega_{\pm}^2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{5.58}$$

It is easy to show that both $\omega_{\pm}$ are *real*: one only needs to show the positivity of the quantity under the square root. To that end, note that:

$$b^2 - 4ac = (K_{11} m_2 - K_{22} m_1 + K_{12}(m_2 - m_1))^2 + 4K_{12} m_1 m_2 \geq 0. \tag{5.59}$$

However, there is no guarantee that $\omega_{\pm}^2$ are also *positive*. A negative ω^2 would result in exponentially growing behaviour of the test solution. Checking the positivity of the roots can be quite laborious in the general case. To make things a little easier, we make the assumption $m_1 = m_2 = m$, for which one has

$$2m\omega_{\pm}^2 = K_{11} + K_{22} + 2K_{12} \pm \sqrt{(K_{11} - K_{22})^2 + 4K_{12}^2}. \tag{5.60}$$

The positivity of *both* solutions is enforced when

$$K_{11} + K_{22} + 2K_{12} \geq 0, \quad \text{and} \quad (K_{11} + K_{22} + 2K_{12})^2 \geq (K_{11} - K_{22})^2 + 4K_{12}^2. \tag{5.61}$$

Solving for K_{12}, one has

$$K_{12} \geq -\frac{K_{11}+K_{22}}{2} \quad \text{and} \quad \begin{cases} K_{12} \geq -\dfrac{K_{11}K_{22}}{K_{11}+K_{22}} & \text{if } K_{11}+K_{22} > 0, \qquad (5.62a) \\ K_{12} \leq -\dfrac{K_{11}K_{22}}{K_{11}+K_{22}} & \text{if } K_{11}+K_{22} < 0. \qquad (5.62b) \end{cases}$$

If instead $K_{11}+K_{22}=0$, at least one eigenvalue is negative, as seen immediately from (5.61). For example, consider the case $K_{11}=K_{22}=1$. The conditions above give $K_{12} \geq -1/2$: this is an interesting case since one may allow the coupling to have negative stiffness whilst guaranteeing oscillating solutions overall. As a second example, consider $K_{11}=K_{22}=-1$: here, there is no range allowable for K_{12}.

Once the eigenvalues are computed, one may compute the eigenvector $\hat{\mathbf{u}}$ from (5.55). Since the determinant is null, the two equations in system (5.55) define the same equation. Using the first row, one obtains:

$$\hat{u}_2 = \frac{K_{11}+K_{12}-\omega_{\pm}^2 m_1}{K_{12}}\hat{u}_1, \tag{5.63}$$

where either $\hat{u}_1$ or $\hat{u}_2$ is arbitrary. When $\hat{u}_1 = (\sqrt{2})^{-1}$, the general solution is given by:

$$\mathbf{u} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \frac{K_{11}+K_{12}-\omega_+^2 m_1}{K_{12}} \end{bmatrix}\left(a\, e^{j\omega_+ t} + b\, e^{-j\omega_+ t}\right) + \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \frac{K_{11}+K_{12}-\omega_-^2 m_1}{K_{12}} \end{bmatrix}\left(c\, e^{j\omega_- t} + d\, e^{-j\omega_- t}\right),$$

where a, b, c, d, are four complex constants depending on the initial conditions $\mathbf{u}(t=0)$, $\frac{d\mathbf{u}(t=0)}{dt}$. As per the case of the harmonic oscillator, the complex exponential form can be recast in various other forms employing trigonometric functions, such that an equivalent form of the above is obtained as:

$$\mathbf{u} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \frac{K_{11}+K_{12}-\omega_+^2 m_1}{K_{12}} \end{bmatrix} A_+ \cos(\omega_+ t + \zeta_+) + \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ \frac{K_{11}+K_{12}-\omega_-^2 m_1}{K_{12}} \end{bmatrix} A_- \cos(\omega_- t + \zeta_-),$$

where the constants A_+, A_-, ζ_+, ζ_- are set by the initial conditions. The formula above reveals that the system's global motion can be written as the sum of two basic harmonic motions, independent of each other, one with frequency ω_+, the other with frequency ω_-. As an example, consider the case $K_{11}, K_{22}, K_{12}, m_1, m_2 = 1$. In this case, one has $\omega_+ = \sqrt{3}$, $\omega_- = 1$, and the solution is:

$$\mathbf{u} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} A_+ \cos(\sqrt{3}t + \zeta_+) + \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} A_- \cos(t + \zeta_-). \tag{5.64}$$

Remarkably, the motion of the two masses is, generally speaking, not periodic since the ratio of the two fundamental frequencies is an irrational number. In some cases, If one chooses $\mathbf{u}(t=0) = [1, -1]^\intercal$, $d\mathbf{u}(t=0)/dt = [0, 0]^\intercal$, the solution is:

$$\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \cos\left(\sqrt{3}t\right). \tag{5.65}$$

This shows that a two-mode system may collapse to a single mode when the system is started in that mode. There is no trace of the other mode of vibration! One may, of course, start the system in the other mode and observe it oscillating in that mode only. Using $\mathbf{u}(t=0) = [1,1]^\intercal$, $d\mathbf{u}(t=0)/dt = [0,0]^\intercal$, one gets:

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cos(t)\,. \tag{5.66}$$

Figure 5.6 shows the motion of the masses when the system is started in either of the two modes.

5.4.2 Eigenvalue decomposition

It is easy to generalise the study above to a case comprising M degrees of freedom. The discussion above suggests that the motion of system (5.52) may be decomposed onto linearly independent blocks, called the *modes*. A *generalised eigenvalue problem* can be formulated as:

$$\mathbf{K}\hat{\mathbf{U}} = -\mathbf{M}\,\hat{\mathbf{U}}\mathbf{S}^2\,, \tag{5.67}$$

where $\hat{\mathbf{U}}$ is a matrix comprising the column *eigenvectors*, and where $\mathbf{S}$ is a diagonal matrix containing the square roots of the *eigenvalues*. In practice:

$$\hat{\mathbf{U}} := \left[\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_M\right], \quad \mathbf{S} := \begin{bmatrix} s_1 & 0 & 0 & ... & 0 \\ 0 & s_2 & 0... & 0 & \\ 0 & 0 & \ddots & ... & 0 \\ 0 & 0 & 0 & ... & s_M \end{bmatrix} \tag{5.68}$$

The eigenvectors have the interpretation of a *modal shape*. Under oscillating conditions, the entries of $\mathbf{S}$ are purely imaginary, such that $\mathbf{S} := j\boldsymbol{\Omega}$, with:

$$\boldsymbol{\Omega} := \begin{bmatrix} \omega_1 & 0 & 0 & ... & 0 \\ 0 & \omega_2 & 0 & ... & 0 \\ 0 & 0 & \ddots & ... & 0 \\ 0 & 0 & 0 & ... & \omega_M \end{bmatrix}, \tag{5.69}$$

where ω_i has the interpretation of a *modal frequency*, or an *eigenfrequency*. Note that this generalises the approach adopted to derive (5.55). When the eigenfrequencies are all distinct, the columns of $\hat{\mathbf{U}}$ are linearly independent and, hence, form a basis in $\mathbb{R}^M$. If some of the eigenvalues are repeated (that is, ω_i has an *algebraic multiplicity* greater than one for some i), there is no guarantee that the eigenvectors are all linearly independent, and $\hat{\mathbf{U}}$ may be *rank-deficient*. We will, for now, treat $\hat{\mathbf{U}}$ as full-rank. That is, even in the case of repeated eigenvalues, we will assume that $\hat{\mathbf{U}}$ is invertible. A special case is obtained when $\mathbf{M}^{-1}\mathbf{K}$ is *symmetric*, in which case $\hat{\mathbf{U}}$ is an *orthogonal matrix*. An orthogonal matrix $\hat{\mathbf{U}}$ is full-rank, and furthermore:

$$\hat{\mathbf{U}}^{-1} = \hat{\mathbf{U}}^\intercal, \tag{5.70}$$

that is, the transpose of $\hat{\mathbf{U}}$ and its inverse are the same matrix.

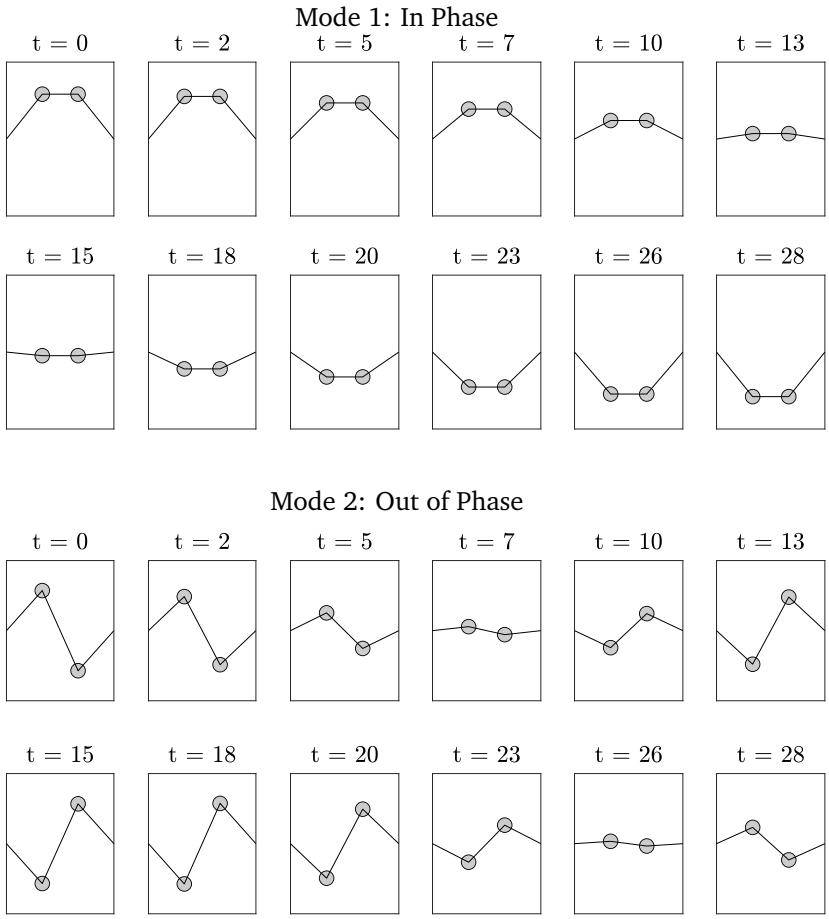


Figure 5.6: Time evolution of the eigenvectors (modal shapes) of model (5.51). The out-of-phase mode is faster. The time stamps have arbitrary units.

Defining $\mathbf{w} := \hat{\mathbf{U}}^{-1}\mathbf{u}$, and using the eigendecomposition (5.67) in (5.52), one gets:

$$\frac{d^2\mathbf{w}}{dt^2} = -\boldsymbol{\Omega}^2\,\mathbf{w}, \tag{5.71}$$

Since $\boldsymbol{\Omega}$ is a diagonal matrix, this system is completely uncoupled. Each line defines the equation of a harmonic oscillator for the modal coordinate w_i and resonant frequency ω_i. One may decide to analyse and simulate the system in its diagonal form (5.71), and to switch back to the "physical" coordinates $\mathbf{u}$, using $\mathbf{u} = \hat{\mathbf{U}}\,\mathbf{w}$.

Coming back again to the example (5.64), here one has

$$\hat{\mathbf{U}} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix},\ \boldsymbol{\Omega} = \begin{bmatrix}1 & 0\\ 0 & \sqrt{3}\end{bmatrix}. \tag{5.72}$$

Here, it is immediate to verify that $\hat{\mathbf{U}}\,\hat{\mathbf{U}}^\intercal = \hat{\mathbf{U}}^\intercal\,\hat{\mathbf{U}} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. It is also immediate to check that (5.67) is verified.

5.4.3 Energy analysis

It is remarked that the allowable ranges for the stiffness constants K_{ij} are such that the potential energy of the system is *non-negative*, so that one may bound the growth of the solutions in some manner. The total energy of the system can be found by multiplying (5.51a) by $\frac{du_1}{dt}$, and (5.51b) by $\frac{du_2}{dt}$, and summing. Using the usual identities, this gives:

$$\frac{d}{dt}\left(\frac{m_1}{2}\left(\frac{du_1}{dt}\right)^2 + \frac{m_2}{2}\left(\frac{du_2}{dt}\right)^2 + \frac{K_{11}u_1^2}{2} + \frac{K_{22}u_2^2}{2} + \frac{K_{12}(u_1-u_2)^2}{2}\right) = 0, \tag{5.73}$$

implying energy conservation:

$$\frac{dH}{dt} = 0 \implies H(t) = H(0) := H_0. \tag{5.74}$$

The form of the energy comprises the energy of the two harmonic oscillators in isolation plus the coupling energy, proportional to K_{12}. The coupling is a function of the relative distance between the masses so that whenever $u_1 = u_2$, the coupling energy is null.

The energy can be written more compactly via vector norms. Consider the non-negative, symmetric matrix $\mathbf{A}$. The inner product and related norm are defined as:

$$\langle \mathbf{u}, \mathbf{v}\rangle_{\mathbf{A}} := \mathbf{u}^\intercal\mathbf{A}\mathbf{v} = \mathbf{v}^\intercal\mathbf{A}\mathbf{u}, \quad \|\mathbf{u}\|_{\mathbf{A}}^2 := \mathbf{u}^\intercal\,\mathbf{A}\,\mathbf{u}, \tag{5.75}$$

and note that $\|\mathbf{u}\|_{\mathbf{A}}^2 \geq 0$ following the non-negativity of the matrix $\mathbf{A}$. The norm can be conveniently bounded in terms of the smallest and largest eigenvalue of the matrix $\mathbf{A}$:

$$0 \leq \min(\lambda_{\mathbf{A}})\,\|\mathbf{u}\|^2 \leq \|\mathbf{u}\|_{\mathbf{A}}^2 \leq \max(\lambda_{\mathbf{A}})\,\|\mathbf{u}\|^2, \tag{5.76}$$

where $\lambda_{\mathbf{A}}$ is an eigenvalue of the matrix, and min and max are its smallest and largest elements, respectively. Proving the bounds is left as an exercise for the reader. In the inequality above, the following notation is implied:

$$\|\mathbf{u}\| := \|\mathbf{u}\|_{\mathbf{I}}, \tag{5.77}$$

where $\mathbf{I}$ is the identity matrix. Using the norms, the conserved energy takes the expression:

$$H = \frac{1}{2}\left\|\frac{d\mathbf{u}}{dt}\right\|_{\mathbf{M}}^2 + \frac{1}{2}\|\mathbf{u}\|_{\mathbf{K}}^2 . \tag{5.78}$$

From here, bounds on the norms are obtained as:

$$\left\|\frac{d\mathbf{u}}{dt}\right\| \leq \sqrt{2H_0/\min(\lambda_{\mathbf{M}})}, \quad \|\mathbf{u}\| \leq \sqrt{2H_0/\min(\lambda_{\mathbf{K}})}, \tag{5.79}$$

generalising the bounds given in Section 4.2 for the scalar case.

5.4.4 Loss and Forcing

System (5.51) may be generalised to include losses and external forcing. The system reads

$$m_1\frac{d^2u_1}{dt^2} = -K_{11}u_1 - K_{12}(u_1-u_2) - R_1\frac{du_1}{dt} + F_1(t), \tag{5.80a}$$

$$m_2\frac{d^2u_2}{dt^2} = -K_{22}u_1 + K_{12}(u_1-u_2) - R_2\frac{du_2}{dt} + F_2(t). \tag{5.80b}$$

Finding a general solution to this case can be hard in the general case but, as seen for the single-mass case, some cases are more easily treated. For now, assume that the external forcing is sinusoidal with frequency ω. Then, a particular solution is obtained by assuming that both u_1, u_2 vibrate at the same frequency as the forcing. Thus:

$$F_1(t) = \hat{f}_1 e^{j\omega t}, \quad F_2(t) = \hat{f}_2 e^{j\omega t}, \quad u_1(t) = \hat{u}_1 e^{j\omega t}, \quad u_2(t) = \hat{u}_2 e^{j\omega t}. \tag{5.81}$$

Substituting these expressions in (5.80) yields the system:

$$j\omega\mathbf{Z}\hat{\mathbf{u}} = \hat{\mathbf{f}}, \quad \text{or} \quad \mathbf{Z}\hat{\mathbf{v}} = \hat{\mathbf{f}} \tag{5.82}$$

defining the *impedance matrix* $\mathbf{Z}(j\omega)$ in terms of the complex displacement and velocity amplitude vectors $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$. Note that this generalises the definition of the mechanical impedance given for the single-mass case in (5.32). For system (5.80), the impedance matrix is of the form:

$$\mathbf{Z} := \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \tag{5.83}$$

with

$$Z_{11} = j\omega m_1 - j\omega^{-1}(K_{11}+K_{12}) + R_1 := r_1 e^{j\theta_1}, \tag{5.84a}$$

$$Z_{22} = j\omega m_2 - j\omega^{-1}(K_{22}+K_{12}) + R_2 := r_2 e^{j\theta_2}, \tag{5.84b}$$

$$Z_{12} = Z_{21} = j\omega^{-1}K_{12} := \omega^{-1}K_{12}e^{j\frac{\pi}{2}}. \tag{5.84c}$$

Note that the matrix is symmetric, following the symmetry of the stiffness matrix. Deriving the expressions of $r_1, r_2, \theta_1\theta_2$ is left as an exercise for the reader. It is useful to derive some analytic results in special cases. Consider again the case the case $K_{11}, K_{22}, K_{12}, m_1, m_2 = 1$

leading to the solution (5.64) in the undamped case. When R_1, R_2 are small, the eigenvalues of the impedance matrix are:

$$\lambda_{\mathbf{Z},+} \approx \frac{R_1+R_2}{2} + j\omega^{-1}(\omega^2-\omega_+^2), \quad \lambda_{\mathbf{Z},-} \approx \frac{R_1+R_2}{2} + j\omega^{-1}(\omega^2-\omega_-^2), \tag{5.85}$$

where $\omega_\pm$ are as per the undamped case (5.64). These expressions are analogous to the complex expression of the impedance in the single-mass case, as per (5.33). The eigenvectors are:

$$\hat{\mathbf{z}}_+ \approx \begin{bmatrix} 1 \\ -1 \end{bmatrix} + j\omega \begin{bmatrix} (R_1-R_2)/2 \\ 0 \end{bmatrix}, \quad \hat{\mathbf{z}}_- \approx \begin{bmatrix} 1 \\ 1 \end{bmatrix} + j\omega \begin{bmatrix} (R_2-R_1)/2 \\ 0 \end{bmatrix}. \tag{5.86}$$

These results turn out to be useful when deriving the expressions for $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ by inverting $\mathbf{Z}$. It is interesting to plot the magnitude and phase angle of, say, $\hat{\mathbf{v}}$ under various choices of the forcing amplitude vector $\hat{\mathbf{f}}$, as per Figure 5.7. In most cases, the velocity presents two peaks at the two modal frequencies of the unforced system. However, when $\hat{\mathbf{f}}$ is itself chosen in the neighbourhood of an eigenvector of $\mathbf{Z}$, the resulting velocity amplitude presents a single peak at the modal frequencies corresponding to the same eigenmode. The other resonance is cancelled out! As per the one-mass case, solutions (5.82) represent only the steady-state. The complete solution is given by adding the steady-state solution to the solution of the homogeneous problem, incorporating the initial conditions.

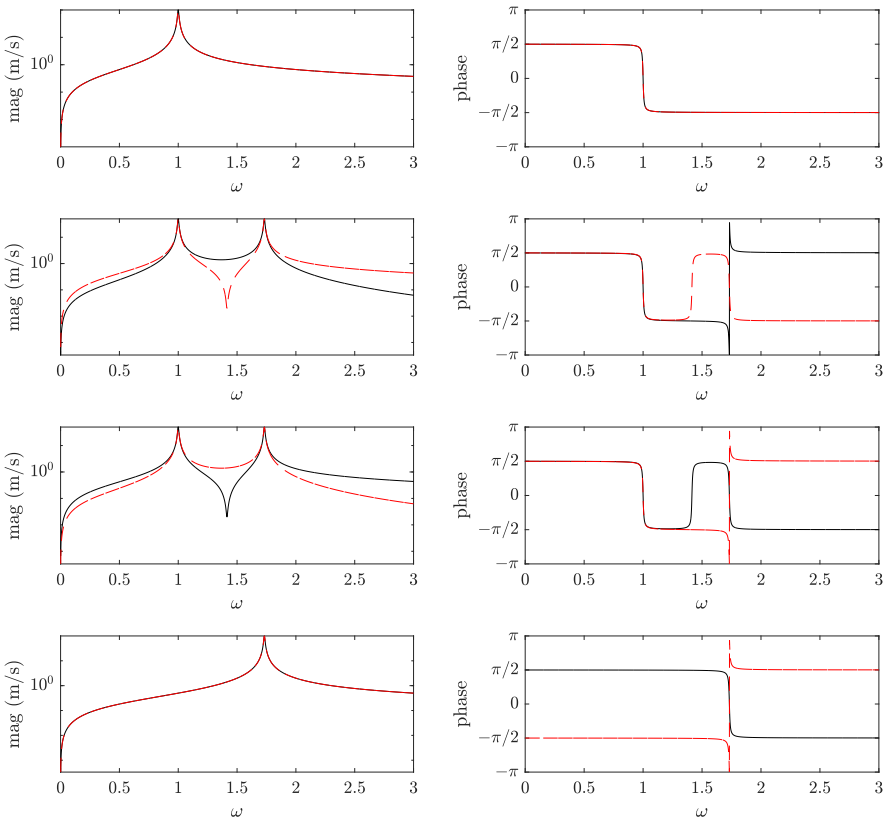


Figure 5.7: Amplitude and phase of the velocity amplitude $\hat{\mathbf{v}}$, computed via (5.82). In this example, $K_{11}, K_{22}, K_{12}, m_1, m_2 = 1$, $R_1 = R_2 = 0.01$. The four cases (top to bottom) correspond to $\hat{\mathbf{f}} = [1,1]^\intercal, [0,1]^\intercal, [1,0]^\intercal, [1,-1]^\intercal$. The solid black line is $\hat{v}_1$, whereas the dashed red line is $\hat{v}_2$.

Chapter 6

Time Difference Operators

This chapter introduces the notation and the principles of difference calculus, upon which the method of *finite differences* is constructed. Though finite differences are used for algorithmic solutions of differential equations, this method has surprisingly long roots, reaching as far as the work of Courant, Friedrichs, and Lewy in 1928 (before digital computers were even available). The underlying principle of finite differences is straightforward and involves the discrete approximation of differential operators. In digital applications, whether, e.g. performing a measurement or in computer-aided simulation, time is most often discretised by means of a *sample rate*, f_s. In practice, one usually knows (or is interested in knowing) the state of a system at discrete time intervals of length k seconds, the time step. The relationship between sample rate and time step is simple:

$$kf_s = 1, \tag{6.1}$$

Finite differences aim to compute a *time series* u^n approximating the "true" solution $u(t)$ of a given model problem. The index $n \in \mathbb{N}_0$ in u^n is shorthand for $t = t_n := nk$, i.e. n is the time index, approximating the true solution $u(t)$ at the time $t_n = nk$. The fundamental relationship between the approximate time series u^n and the true solution $u(t)$ is as follows:

$$\varepsilon := u(t_n) - u^n, \tag{6.2}$$

where ε defines the *error* of the approximation. In general, $\varepsilon \neq 0$, and, generally, it will not be possible to obtain an exact expression for ε since $u(t)$ is generally unknown! However, some cases are exceptional, such as the case of the harmonic oscillator presented in Chapter 5. However, the study of finite differences is almost entirely devoted to the design of schemes for which the error remains provably *bounded* by a given power of k, and we shall, of course, spend considerable effort studying such schemes.

6.1 Shift, difference and averaging operators

Given the time series u^n, the identity, forward and backward shift operators are given as

$$1u^n := u^n, \quad e_{t+}u^n := u^{n+1}, \quad e_{t-}u^n := u^{n-1}. \tag{6.3}$$

From these, one may define the time difference operators, all approximating the first time derivative, as

$$\delta_{t+} := \frac{e_{t+} - 1}{k} \approx \frac{d}{dt}, \quad \delta_{t-} := \frac{1 - e_{t-}}{k} \approx \frac{d}{dt}, \quad \delta_{t\cdot} := \frac{e_{t+} - e_{t-}}{2k} \approx \frac{d}{dt}. \tag{6.4}$$

When applied to the time series u^n, the difference operators above take the explicit form:

$$\delta_{t+}u^n = \frac{u^{n+1} - u^n}{k}, \quad \delta_{t-}u^n = \frac{u^n - u^{n-1}}{k}, \quad \delta_{t\cdot}u^n = \frac{u^{n+1} - u^{n-1}}{2k}. \tag{6.5}$$

An approximation to the second time derivative is constructed from the above as

$$\delta_{tt} := \delta_{t+}\delta_{t-} = \delta_{t-}\delta_{t+} \approx \frac{d^2}{dt^2}, \tag{6.6}$$

which takes the explicit form:

$$\delta_{tt}u^n = \frac{u^{n+1} - 2u^n + u^{n-1}}{k^2}. \tag{6.7}$$

Averaging operators (all approximating the identity) are also used throughout the text and are:

$$\mu_{t+} := \frac{e_{t+} + 1}{2} \approx 1, \quad \mu_{t-} := \frac{1 + e_{t-}}{2} \approx 1, \quad \mu_{t\cdot} := \frac{e_{t+} + e_{t-}}{2} \approx 1. \tag{6.8}$$

When applied to the time series u^n, the averaging operators above take the explicit form:

$$\mu_{t+}u^n = \frac{u^{n+1} + u^n}{2}, \quad \mu_{t-}u^n = \frac{u^n + u^{n-1}}{2}, \quad \mu_{t\cdot}u^n = \frac{u^{n+1} + u^{n-1}}{2}. \tag{6.9}$$

Whilst these expressions look at least reasonable, there is not a clear indication as to what we mean by the "approximate" equalities in (6.4), (6.6), (6.8). These are not well-defined so long as they are not applied to a smooth function. As an example, apply the forward difference operator to the smooth function $u(t)$, and compute its Taylor series:

$$\begin{aligned}\delta_{t+}u(t_n) = \frac{u(t_{n+1}) - u(t_n)}{k} &\approx \frac{u(t_n) + k\frac{du(t_v)}{dt} + \frac{k^2}{2}\frac{d^2u(t_n)}{dt^2} - u(t_n)}{k} \\ &= \frac{du(t_n)}{dt} + \frac{k}{2}\frac{d^2u(t_n)}{dt^2}.\end{aligned}$$

Since $d^2u(t_n)/dt^2$ is a value independent of k, in the limit of high sample rate, the expression above reduces to $du(t_n)/dt$. The rate at which such approximation is satisfied is linear in k, so that applying definition (6.2) one gets:

$$\frac{du(t_n)}{dt} - \delta_{t+}u(t_n) := \varepsilon_{\delta_{t+}} = \mathcal{O}(k). \tag{6.10}$$

The "big-Oh" notation $\mathcal{O}(k^p)$ means that the rate of the approximation goes as k^p. For the current case, $p = 1$ and δ_{t+} is said to be *first-order accurate*. Using similar arguments and a

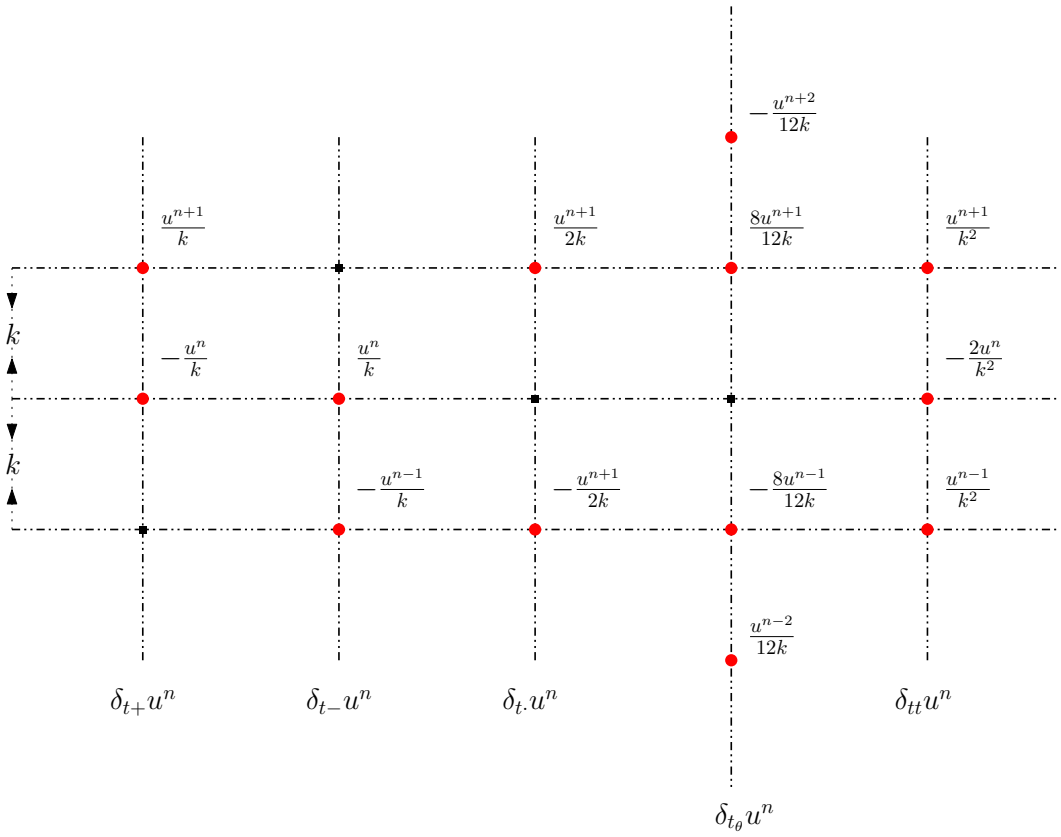


Figure 6.1: Stencil width and coefficients of various difference operators approximating the first and second time derivatives.

similar notation to denote the truncation error of the difference operators, it is easy to show that:

$$\varepsilon_{\delta_{t-}} = \mathcal{O}(k), \quad \varepsilon_{\delta_{t\cdot}} = \mathcal{O}(k^2), \quad \varepsilon_{\delta_{tt}} = \mathcal{O}(k^2), \tag{6.11}$$

and, hence, δ_{t-} is first-order accurate, while $\delta_{t\cdot}$ and δ_{tt} are second-order accurate. Similar relationships can be derived for the averaging operators, and the derivation of the errors is left as an exercise for the reader.

All such identities define the *truncation errors* of the finite difference approximations. Clearly, some operators have a higher accuracy than others. In particular, $\delta_{t\cdot}$ and δ_{tt} have higher accuracy, as is typical of *centred operators*. Obtaining even higher-accurate approximations is possible. For example, consider the following approximation to the first time derivative:

$$\delta_{t_\theta}\, u(t_n) := \frac{u(t_{n-2}) - 8u(t_{n-1}) + 8u(t_{n+1}) - u(t_{n+2})}{12k} = \frac{du(t_n)}{dt} + \mathcal{O}(k^4). \tag{6.12}$$

This defines a fourth-order accurate operator. When applied to the time series u^n, this becomes:

$$\delta_{t_\theta} u^n = \frac{u^{n-2} - 8u^{n-1} + 8u^{n+1} - u^{n+2}}{12k}. \tag{6.13}$$

This operator has one fundamental difference compared to the operators defined in (6.4): the *stencil* (or footprint) is larger and "reaches out" to further points in the time series, defined at $n \pm 2$. The *stencil width* of δ_{t_θ} is, thus, four. The operators δ_{t-}, δ_{t+}, μ_{t-}, μ_{t+} all have a stencil of width one. $\delta_{t\cdot}$, $\mu_{t\cdot}$, δ_{tt} have a stencil of width two. See also Figure 6.1. At this point of the discussion, one may be tempted to think that constructing higher-order schemes to solve a model problem amounts to merely employing difference operators with an appropriately large stencil. Things are, of course, more complicated than this: primarily, operators with wide stencils in time tend to yield *unstable* simulations, as will be seen in forthcoming examples. Other problems exist: for instance, it is unclear how to initialise an operator with a stencil of width M, when the model problem is of order $N < M$. A similar problem arises when discretising differential operators in space for boundary-values problems, where one must set values for the "ghost points" (i.e. points located outside the boundary of the grid). Constructing higher-order schemes is, of course, possible and sometimes desirable, and these words of caution suffice for the moment as a cursory understanding of the underlying difficulties.

Before proceeding, it is worth commenting on the definition of the error ε, as defined in (6.2). When applied to infer the order of approximation of a given discrete-time operator, such as in (6.10), ε defines the *local truncation error* (LTE). This is, essentially, the error made in truncating the Taylor series to a finite order. However, when solving a differential problem, it is customary to define the *rate of convergence* of the scheme r, which may or may not be equal to the truncation error p of the constitutive difference operators. As mentioned above, using large-stencil operators increases p but does not guarantee convergence, so r may be undefined. In general, one should hope that the order of convergence is at least as good as the order of accuracy so that $r \geq p$. For the finite difference method to be used in numerical analysis and simulation, the conditions under which such a relation holds must be found, and we shall study the problems accordingly in the forthcoming chapters.

6.1.1 Interleaved time operators

A time series such as u^n is defined at integer locations $n = 0, 1, 2, \ldots$. In some cases, it is useful to define time series on an interleaved time grid, defined at half-integer steps $n + \frac{1}{2}$. For example, the action of the operator δ_{t+} on u^n can be defined in terms of the interleaved grid function $v^{n+\frac{1}{2}}$ as:

$$v^{n+\frac{1}{2}} := \delta_{t+} u^n. \tag{6.14}$$

This definition may look arbitrary, as one may force v to be defined on the integer time grid. However, note how the action of δ_{t+}, δ_{t-} on the interleaved grid function results in a centred operation:

$$\delta_{t-} v^{n+\frac{1}{2}} := \frac{v^{n+\frac{1}{2}} - v^{n-\frac{1}{2}}}{k} := \delta_{t+} v^{n-\frac{1}{2}}. \tag{6.15}$$

Previously, the Taylor series of the centred operators yielded higher-accurate difference and averaging operators. One can expect to observe the same behaviour for the interleaved time series:

$$\begin{aligned}
\delta_{t+} v(t_{n-\frac{1}{2}}) &= \frac{v(t_{n+\frac{1}{2}}) - v(t_{n-\frac{1}{2}})}{k} \\
&\approx \frac{v(t_n) + \frac{k}{2}\frac{dv(t_n)}{dt} + \frac{k^2}{8}\frac{d^2v(t_n)}{dt^2} + \frac{k^3}{48}\frac{d^3v(t_n)}{dt^3} - v(t_n) + \frac{k}{2}\frac{dv(t_n)}{dt} - \frac{k^2}{8}\frac{d^2v(t_n)}{dt^2} + \frac{k^3}{48}\frac{d^3v(t_n)}{dt^3}}{k} \\
&= \frac{dv(t_n)}{dt} + \frac{k^2}{24}\frac{d^3v(t_n)}{dt^3},
\end{aligned}$$

and second-order accuracy is recovered. The formal definition of the averaging operators is as follows:

$$\mu_{t-} v^{n+\frac{1}{2}} := \frac{v^{n+\frac{1}{2}} + v^{n-\frac{1}{2}}}{2} := \mu_{t+} v^{n-\frac{1}{2}}, \tag{6.16}$$

and are again second-order accurate when expanded about $t = t_n$. Using these definitions, one has:

$$\delta_{t\cdot} u^n = \mu_{t+} v^{n-\frac{1}{2}} = \mu_{t-} v^{n+\frac{1}{2}}, \quad \delta_{tt} u^n = \delta_{t+} v^{n-\frac{1}{2}} = \delta_{t-} v^{n+\frac{1}{2}}, \tag{6.17}$$

see also Figure 6.2.

6.2 Frequency domain analysis

Before proceeding, it is worth introducing the frequency-domain equivalents of the Laplace and Fourier transforms defined in Section 4.3. As per their continuous counterparts, these techniques are really only useful when the model problem is linear and time-invariant. In such cases, however, they are extremely powerful and insightful, and we will use such techniques accordingly.

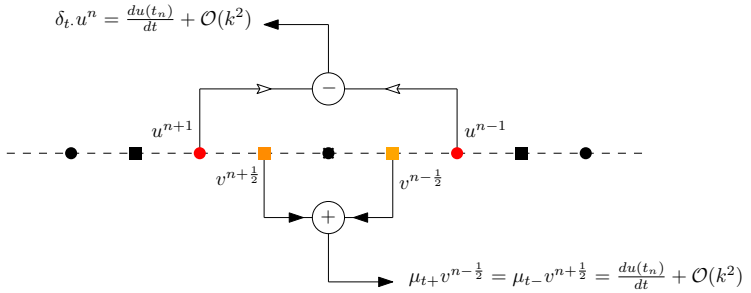


Figure 6.2: Approximation of the first time derivative via a centred difference on the integer grid function u^n, or a centred average of the interleaved grid function $v^{n+\frac{1}{2}}$. When $v^{n+\frac{1}{2}}$ is defined as per (6.14), the two expressions yield the same result. In the figure, disks denote integer grid locations, and squares denote half-integer locations.

6.2.1 z and discrete time Fourier transforms

A discrete counterpart to the Laplace transform defined in (4.19) is the closely related z transform:

$$\hat{u}(z) = \sum_{n=\{-\infty,0\}}^{\infty} u^n z^{-n} := \mathcal{Z}\{u\}(z), \tag{6.18}$$

for a number $z \in \mathbb{C}$. This definition may look confusing at first, but the properties of the z transform will become clearer when treating difference equations in applied cases. The meaning of the index n in the definition (6.18) is often a source of confusion: when applied to u, n defines a *time index*; when applied to z, it is an *exponent*. So, the z transforms transforms the original time series u^n into a power series of the complex variable z. This relation is important because discrete counterparts of (4.21), holding for time derivatives, can be derived here for time shifts:

$$\mathcal{Z}\left\{e_{t\pm}u^n\right\} = \sum_{n=-\infty}^{\infty} u^{n\pm1}z^{-n} = z^{\pm1}\sum_{n=-\infty}^{\infty} u^{n\pm1}z^{-(n\pm1)} = z^{\pm1}\mathcal{Z}\{u^n\}. \tag{6.19}$$

Applying these serially, higher-order shifts transform as:

$$\mathcal{Z}\left\{e^p_{t\pm}u^n\right\} = z^{\pm p}\mathcal{Z}\left\{u^n\right\}. \tag{6.20}$$

In practice, substituting a simpler expression than the actual transform is often useful via an *ansatz*. So, one may employ the test solution:

$$u^n = \hat{u}z^n, \tag{6.21}$$

for an appropriate constant complex amplitude $\hat{u}$. It is immediate to verify that shifts of this test solution transform in the same way as the z transform in (6.20); hence, we will use the test solution accordingly. When substituted into a difference equation discretising a differential problem, the complex amplitude $\hat{u}$ is expressed analogously to (4.23), as:

$$\hat{u} = A(z)B^{-1}(z), \tag{6.22}$$

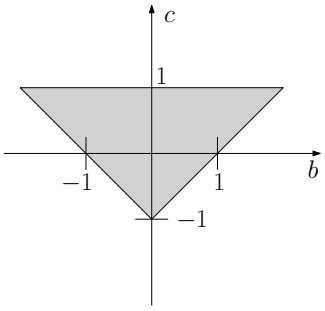


Figure 6.3: The Schur-Cohn stability region, i.e. the region of the $(b, c)-$plane for which the roots of (6.23) have magnitude less than unity.

that is, a rational function. As for the Laplace transform, also for a rational z transform the order of the polynomial A or B is the power of the highest power of z. The roots of the numerator and denominator polynomials are called again the *zeros* and the *poles*, respectively. Stability is guaranteed for linear, time-invariant systems when the complex poles *lie within the unit circle*. This is the most important and useful z transform theory result. Often, for the problems of interest here, the order of the denominator B is two and, thus, the poles are recovered by solving:

$$z^2 + bz + c = 0, \tag{6.23}$$

for given model-dependent coefficients $(b, c) \in \mathbb{R}$. The solutions are given by:

$$z_\pm = \frac{-b \pm \sqrt{b^2 - 4c}}{2}. \tag{6.24}$$

Useful bounds on the coefficients b, c can be derived when one considers $|z_\pm| < 1$ (i.e. when the poles are within the unit circle). These are

$$|c| = |z_+ z_-| = |z_+||z_-| < 1, \quad |b| < 1 + |c|. \tag{6.25}$$

These have the interpretation of *necessary and sufficient* conditions to enforce $|z_\pm| < 1$, and are known as *Schur-Cohn stability test*, see also Figure 6.3.

When z is forced to lie exactly on the unit circle, i.e. $z = e^{j\omega k}$ in (6.18), the *discrete-time Fourier transform* (DTFT) is obtained. Thus:

$$\hat{u}(\omega) = \sum_{n=-\infty}^{\infty} u^n e^{-j\omega kn} := \mathcal{F}\{u^n\}(\omega). \tag{6.26}$$

We remark, however, that the DTFT is a *continuous function* of the radian frequency ω and is thus not to be confused with the *discrete Fourier transform* (DFT), which is evaluated at discrete frequency bins ω_m. When the simpler ansatz (6.21) is used instead, the DTFT takes the following form:

$$u^n = \hat{u} e^{j\omega kn}. \tag{6.27}$$

It is convenient to compute the action of the z and the discrete-time Fourier transforms on the difference operators defined in Section 6.1. Consider, for example, the second difference operator δ_{tt}. This is

$$\mathcal{Z}\left\{\delta_{tt}u^n\right\} = \frac{z-2+z^{-1}}{k^2}\mathcal{Z}\left\{u^n\right\} = \frac{z-2+z^{-1}}{k^2}\hat{u}(z). \tag{6.28}$$

When $z = e^{j\omega k}$, the DTFT is recovered and can be written conveniently as:

$$\mathcal{F}\left\{\delta_{tt}u^n\right\} = -\frac{4}{k^2}\sin^2\left(\frac{\omega k}{2}\right)\mathcal{F}\left\{u^n\right\}. \tag{6.29}$$

Analogously, one has:

$$\mathcal{F}\left\{\delta_{t\cdot}u^n\right\} = \frac{j}{k}\sin(\omega k)\mathcal{F}\left\{u^n\right\}. \tag{6.30}$$

Obtaining the expressions above using simple trigonometric identities is left as an exercise for the reader. It is useful to recast the expressions above using the sinc function $\operatorname{sinc} x := x^{-1}\sin x$. Thus:

$$\mathcal{F}\left\{\delta_{tt}u^n\right\} = -\omega^2\operatorname{sinc}^2\left(\frac{\omega k}{2}\right)\mathcal{F}\left\{u^n\right\}, \quad \mathcal{F}\left\{\delta_{t\cdot}u^n\right\} = j\omega\operatorname{sinc}(\omega k)\mathcal{F}\left\{u^n\right\}. \tag{6.31}$$

These expressions highlight the warping effect due to the discretisation, compared to the Fourier transforms of the continuous operators $d^2/dt^2 \to -\omega^2$, $d/dt \to j\omega$. See also Figure 6.4.

A useful z transform, which will be used later, is:

$$\mathcal{Z}\{a^n\sin(\mathfrak{w}k)\Theta^{n,n'}\} = \frac{az\sin(\mathfrak{w}k)z^{-n'}}{z^2 - 2a\cos(\mathfrak{w}k) + a^2}, \tag{6.32}$$

where $\mathfrak{w}$ has the interpretation of a radian frequency, and where $\Theta^{n,n'}$ is the discrete-time step function (i.e. equal to one when $n \geq n'$, and zero otherwise). This is a discrete-time version of (4.24).

6.3 Discrete-time energy identities

Energy analysis is central to the study of oscillations. In continuous time, closed orbits appear in phase space whenever the potential is non-negative, given the non-negativity of the kinetic energy. While the frequency-domain techniques are extremely powerful in analysing LTI systems, their use in the context of nonlinear systems is limited. In many cases, unstable numerical behaviour results when the discrete energy associated with a given model problem becomes negative. Thus, identifying a discrete counterpart of the system's energy may be beneficial to analysing any difference system of equations, whether linear or not. Keeping track of the energy error is also a useful debugging strategy.

In continuous time, the energy balance of the oscillator was expressed as the difference between the instantaneous power released by the energy-storing components of the oscillator (mass and stiffness) and the dissipated power due to the mechanical resistance. As

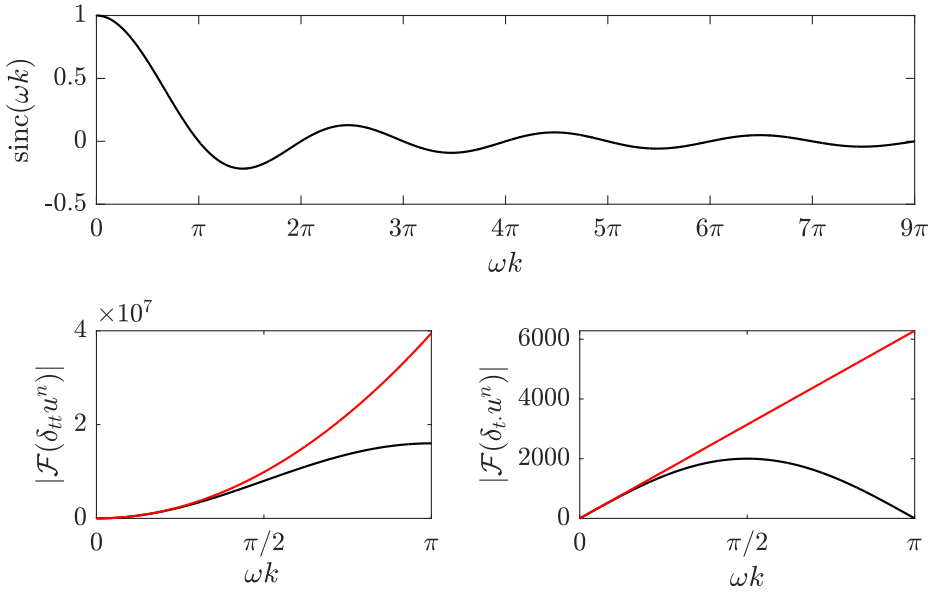


Figure 6.4: Sinc function, and absolute values of the DTFT of the δ_{tt} and $\delta_{t\cdot}$. The red lines are the absolute values of the Fourier transforms of the continuous differential operators.

encapsulated in e.g. (5.28), the energy balance can be obtained after multiplying the equation of motion by the velocity (yielding the power or the force times velocity). Obtaining analogous expressions in discrete time is not as straightforward, as many approximations to the time derivatives exist. Hence, it is worth spending some time working out useful identities to be used in the forthcoming chapters.

Suppose one wants to find a discrete-time equivalent of the instantaneous power yielded by the inertial term of a vibrating object. In continuous time:

$$\frac{du}{dt}\frac{d^2u}{dt^2} = \frac{1}{2}\frac{d}{dt}\left(\frac{du}{dt}\right)^2, \tag{6.33}$$

from which an expression of the kinetic energy may be derived. In discrete time, analogous identities are obtained as:

$$\delta_{t\cdot}u^n\,\delta_{tt}u^n = \frac{1}{2}\delta_{t+}\left(\delta_{t-}u^n\right)^2 = \frac{1}{2}\delta_{t-}\left(\delta_{t+}u^n\right)^2. \tag{6.34}$$

These identities are proven by simple algebra:

$$\begin{aligned}\frac{1}{2}\delta_{t+}\left(\delta_{t-}u^n\right)^2 &= \frac{1}{2}\delta_{t+}\left(\frac{(u^n-u^{n-1})^2}{k^2}\right) = \frac{1}{2k}\frac{(u^{n+1}-u^n)^2-(u^n-u^{n-1})^2}{k^2}\\ &= \frac{1}{2k}\frac{(u^{n+1}-u^{n-1})(u^{n+1}-2u^n+u^{n-1})}{k^2} = \delta_{t\cdot}u^n\,\delta_{tt}u^n.\end{aligned}$$

A list of useful identities is presented here. The proof is left as an exercise for the reader.

$$\delta_{t\cdot}u^n\,\delta_{tt}u^n = \frac{1}{2}\delta_{t+}\left(\delta_{t-}u^n\right)^2 = \frac{1}{2}\delta_{t-}\left(\delta_{t+}u^n\right)^2 \tag{6.35a}$$

$$u^n\,\delta_{t\cdot}u^n = \frac{1}{2}\delta_{t+}\left(u^n e_{t-}u^n\right) = \frac{1}{2}\delta_{t-}\left(u^n e_{t+}u^n\right) \tag{6.35b}$$

$$u^n e_{t-}u^n = \left(\mu_{t-}u^n\right)^2 - \frac{k^2}{4}\left(\delta_{t-}u^n\right)^2 \tag{6.35c}$$

$$u^n e_{t+}u^n = \left(\mu_{t+}u^n\right)^2 - \frac{k^2}{4}\left(\delta_{t+}u^n\right)^2 \tag{6.35d}$$

$$\delta_{t\cdot}u^n\,\mu_{t\cdot}u^n = \frac{1}{2}\delta_{t\cdot}\left(u^n\right)^2 = \frac{1}{2}\delta_{t+}\left(\mu_{t-}\left(u^n\right)^2\right) = \frac{1}{2}\delta_{t-}\left(\mu_{t+}\left(u^n\right)^2\right) \tag{6.35e}$$

$$\delta_{t+}v^{n-\frac{1}{2}}\,\mu_{t+}v^{n-\frac{1}{2}} = \frac{1}{2}\delta_{t+}\left(v^{n-\frac{1}{2}}\right)^2 \tag{6.35f}$$

Chapter 7

Harmonic Motion in Discrete Time

This chapter introduces the finite difference method to treat the case of the harmonic oscillator in some detail. Like the Laplace and Fourier transforms allow finding a solution to the forced and unforced oscillator in continuous time, the difference equations in this chapter can be analysed using the frequency domain techniques developed in Section 6.2.1. Many possible discretisations exist for the differential equations shown in Chapter 5, differing in terms of their stability and accuracy properties: analysis via the z and Fourier transforms will uncover all such properties. Stability conditions through energy analysis will be carried out in parallel. These techniques will translate directly to the nonlinear case presented in later chapters and allow further insight in the properties of the scheme and their relation to the continuous model problems.

7.1 The undamped oscillator

Consider the time series u^n, approximating the true solution $u(t)$ of (5.2). As a first example of a working finite difference scheme, consider:

$$\delta_{tt} u^n = -\omega_0^2 u^n. \tag{7.1}$$

Expanding out the operator, one gets

$$u^{n+1} = u^n(2 - \omega_0^2 k^2) - u^{n-1}, \tag{7.2}$$

Hence, the update requires one multiply and one sum. In (7.2), u^n, u^{n-1} are known values. The starting values u^0 and u^1 are fixed by the *numerical initial conditions*, which will be described later. Furthermore, note that $2 - \omega_0^2 k^2$ is a constant, so its value can be stored offline.

Though scheme (7.1) looks reasonable, there is no guarantee that the computed solutions are an approximate form of the true solution $u(t)$. In some cases, as will be seen shortly, the time series computed by (7.1) diverges; in some other cases, it remains bounded. The next few sections will explain the idea of *convergence* and the closely linked idea of *stability*.

7.1.1 Stability via frequency domain analysis

As anticipated, frequency-domain techniques may be employed to analyse the stability of linear, time-invariant discrete systems, such as (7.1). For that, *ansatz* (6.21) is substituted, yielding:

$$\left(z - (2 - \omega_0^2 k^2) + z^{-1}\right) \hat{u} z^n = 0, \quad \rightarrow \quad z_\pm = \frac{2 - \omega_0^2 k^2 \pm \omega_0^2 k^2 \sqrt{1 - 4(\omega_0 k)^{-2}}}{2}. \tag{7.3}$$

Thus, the solution to the difference equation (7.1) is obtained as a linear superposition of $z_\pm$:

$$u^n = A_+ z_+^n + A_- z_-^n, \tag{7.4}$$

for complex constants $A_\pm$. We assume the scheme is started using two starting values u^0, u^1 (obtained from u_0, v_0 of the continuous problem, as will be seen shortly). Then:

$$u^0 = A_+ + A_-, \quad u^1 = A_+ z_+ + A_- z_-. \tag{7.5}$$

From these, the complex constants are obtained as

$$A_+ = \frac{u^0 z_- - u^1}{z_- - z_+}, \;\; A_- = \frac{u^1 - u^0 z_+}{z_- - z_+}. \tag{7.6}$$

If the square root in $z_\pm$ is a real number, then z_- has a magnitude larger than unity. The solution u^n grows exponentially over time: this is an instance of *instability* since the absolute value of u^n cannot be bounded by a constant independent of k. On the other hand, when the square root is imaginary, then $z_\pm$ are complex conjugates and u^n oscillates. This condition is obtained whenever:

$$k < 2\omega_0^{-1}, \tag{7.7}$$

which is an upper bound on the time step once the natural frequency of the oscillator is set. (7.7) is an important relationship, as it sets the bound between oscillating (i.e., *stable*) solutions and divergent (i.e., *unstable*) solutions. It is therefore referred to as the *stability condition* for scheme (7.1). When such a condition is respected, the roots $z_\pm$ can be expressed in complex polar form:

$$z_\pm = a e^{\pm j \mathfrak{w}_0 k}, \tag{7.8}$$

with:

$$a = 1, \quad \tan \mathfrak{w}_0 k = \left(\omega_0^2 k^2 \sqrt{4(\omega_0 k)^{-2} - 1}\right) / (2 - \omega_0^2 k^2). \tag{7.9}$$

Thus, the expression for the output time series becomes:

$$u^n = A_+ e^{j \mathfrak{w}_0 k n} + A_- e^{-j \mathfrak{w}_0 k n}. \tag{7.10}$$

To check stability, one may bound the absolute value of u^n directly:

$$|u^n| = |A_+ e^{j \mathfrak{w}_0 k n} + A_- e^{-j \mathfrak{w}_0 k n}| \leq |A_+| + |A_-| \leq (|u^0| + |u^1|) |\sin \mathfrak{w}_0 k|^{-1}, \tag{7.11}$$

and, thus, the absolute value of the solution at the time $n > 1$ is bounded in terms of the values at $n = 0, 1$. The first inequality in (7.11) was obtained via the triangle inequality.

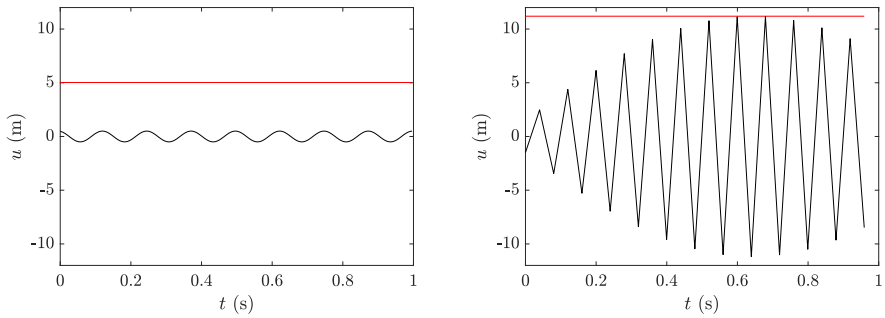


Figure 7.1: Simple harmonic oscillator, numerical check of the stability bound via z domain analysis, (7.11). Here, $\omega_0 = 50$, $u^0 = u^1 = 0.5$. The red horizontal lines are the bounds on $|u^n|$, as per (7.11). The time step is chosen as $0.2\omega_0^{-1}$ (left) and $1.9\omega_0^{-1}$ (right).

Then, the fact that $|e^{\pm j\mathfrak{w}_0 k}| = 1$ was used. Finally, the values from (7.6) were substituted using $z_- - z_+ = -2j \sin \mathfrak{w}_0 k$. A numerical check of the current bound is given in Figure 7.1.

When the initial conditions are substituted in (7.10), the following solution is obtained:

$$u^n = u^0 \cos(\mathfrak{w}_0 kn) + \frac{u^1 - u^0 \cos(\mathfrak{w}_0 k)}{\mathfrak{w}_0 k \operatorname{sinc}(\mathfrak{w}_0 k)} \sin(\mathfrak{w}_0 kn), \tag{7.12}$$

discretising the exact solution in continuous time (5.4).

7.1.2 Stability via energy analysis

The discussion in Section 4.2 suggests that if the model problem can be shown to have conserved total energy with non-negative kinetic and potential terms, then the solution can be bounded in terms of the energy. Whether this property translates to the discrete-time setting depends critically on the choice of the design parameters. It will now be shown that when (7.7) is respected, one may identify a form of conserved energy with non-negative kinetic and potential components. To that end, (7.1) is multiplied by $m\,\delta_{t\cdot}u^n$, to get:

$$m\,\delta_{t\cdot}u^n(\delta_{tt}u^n + \omega_0^2 u^n) = 0. \tag{7.13}$$

Using identities (6.35b) and (6.35b) in (7.13), one obtains a discrete energy balance of the kind:

$$\delta_{t+}\mathfrak{h}^{n-\frac{1}{2}} = 0, \tag{7.14}$$

where the discrete energy is:

$$\mathfrak{h}^{n-\frac{1}{2}} := \frac{m}{2}(\delta_{t-}u^n)^2 + \frac{K}{2}u^n e_{t-}u^n, \tag{7.15}$$

discretising (5.6). Note that multiplication by m in (7.13) was here used to yield units of energy in the expression within the brackets, though it is unnecessary in the following

derivation. The *interleaved* energy time series $\mathfrak{h}^{n-1/2}$ is conserved in discrete time, so that:

$$\mathfrak{h}^{n-1/2} = \mathfrak{h}^{1/2}. \tag{7.16}$$

The problem here is that $\mathfrak{h}^{n-1/2}$ may *not* be positive since the potential energy is of indefinite sign. Instances leading to negative potential energy overall manifest instability and must be avoided. It may be useful, then, to bound the potential term in the energy expression. Using (6.35c) and (6.35d), the total energy becomes:

$$\mathfrak{h}^{n-1/2} = \left(1 - \frac{\omega_0^2 k^2}{4}\right) \frac{m\left(\delta_{t-} u^n\right)^2}{2} + \frac{K(\mu_{t+} u^n)^2}{2}. \tag{7.17}$$

Compare this expression with (7.15): the two expressions are *equivalent*, but presenting two different forms for the discrete kinetic and potential energies. This is a fundamental distinction compared to the continuous case: not only do many difference equations approximating a given continuous model problem exist, but the energy expression for a given discretisation also has many forms! Finding a form that suits the analysis is key to successfully implementing the finite difference method. In (7.17), the form of the potential energy is clearly non-negative, and the kinetic energy is non-negative when (7.7) is satisfied. Figure 7.2 shows the energy components as per (7.17) and the *energy error* for scheme (7.1). The error is defined as:

$$\Delta H := 1 - \mathfrak{h}^{n-\frac{1}{2}} / \mathfrak{h}^{\frac{1}{2}}, \tag{7.18}$$

and is of the order of *machine accuracy*. In double precision, this is of the order of 10^{-15} since 52 out of 64 bits are used to represent the *significand*, and $2^{52} \approx 4 \cdot 10^{15}$.

When the stability condition is respected, one obtains bounds on the velocity and displacement from the equivalent forms (7.15) (7.17):

$$|\delta_{t-} u^n| \leq \sqrt{2m^{-1}\mathfrak{h}^{\frac{1}{2}}}, \quad |\mu_{t-} u^n| \leq \sqrt{2K^{-1}\mathfrak{h}^{\frac{1}{2}}}, \tag{7.19}$$

mimicking the bounds obtained in continuous time in Section 4.2.

7.1.3 Consistency, accuracy and convergence

When stability condition (7.7) is respected, u^n oscillates and is bounded. It is not yet clear, however, how the scheme performs with respect to the "true" solution. The case of the harmonic oscillator is special since an analytic solution does exist, and an assessment of the performance of the scheme may be constructed by direct comparison against the exact solution. In many cases, however, an analytic solution is not readily available, and the scheme's performance cannot be obtained similarly. Whilst this seems limiting, the numerical performance of a given scheme can nonetheless be characterised in terms of the behaviour appropriately defined error measures.

The first such measure is the *local truncation error* (LTE), denoted here ε. Applying the finite difference scheme to the true solution $u(t)$ yields a definition of the LTE as:

$$\delta_{tt} u(t) + \omega_0^2 u(t) = \varepsilon. \tag{7.20}$$

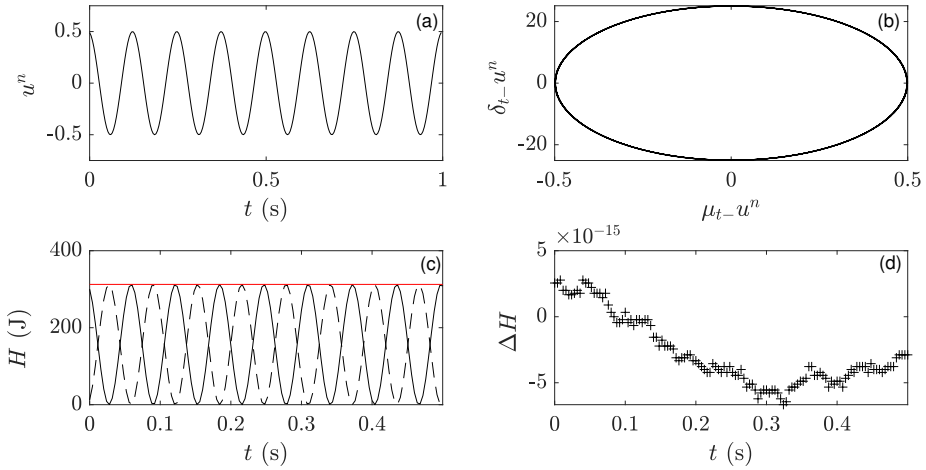


Figure 7.2: Simulation of the simple harmonic oscillator. (a): time domain plot of the displacement u^n; (b): phase portrait highlighting bounds (7.19); (c): potential (continuous black), kinetic (dashed black) and total (red) discrete energy; (d): total energy error, defined as $\Delta H := 1 - \mathfrak{h}^{n-\frac{1}{2}}/\mathfrak{h}^{\frac{1}{2}}$. Here, $\omega_0 = 50$, $m = 1$, $k = 0.004$, and the scheme is initialised with $u^0 = u^1 = 0.5$.

Using Taylor series arguments, one gets:

$$\frac{d^2u(t)}{dt^2} + \omega_0^2 u(t) + \mathcal{O}(k^2) = \varepsilon, \tag{7.21}$$

and since $u(t)$ is the true solution, (5.2) holds, and one recovers $\varepsilon = O(k^2)$. The behaviour of the LTE as a function of k describes the idea of *consistency*: a scheme is said to be consistent if:

$$\lim_{k\to 0} \varepsilon = 0. \tag{7.22}$$

In practice, consistent schemes are such that the local error becomes small as k is decreased. In general:

$$\varepsilon = \mathcal{O}(k^p), \tag{7.23}$$

and the scheme is said to be p^{th}*-order accurate*. This is only partly useful since the question of accuracy is tightly bound to the ideas of *stability and convergence*: schemes employing higher-accurate discrete operators may *never* converge for a given model problem. The idea of accuracy will only be meaningful when a scheme is provably stable. As an example, consider a fourth-order accurate difference operator discretising the second time derivative:

$$\left(\frac{-e_{t+}^2 + 16e_{t+} - 30 + 16e_{t+} - e_{t+}^2}{12k^2}\right) u(t) = \frac{d^2u}{dt^2} + \mathcal{O}(k^4). \tag{7.24}$$

Though technically "higher" accurate, this approximation is always unstable, even for the simple problem of a free particle ($\phi = 0$). Using the test solution $u^n = \hat{u}z^n$ for this test case,

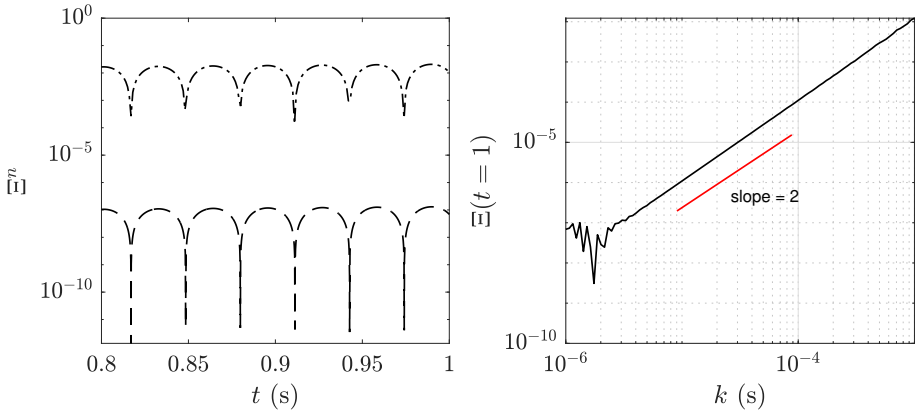


Figure 7.3: Global error of scheme (7.1), for $\omega_0 = 100$. Initial conditions are given as $u_0 = 0.5$, $v_0 = 1$, giving $u(t) = 0.5\cos(\omega_0 t) + \omega_0^{-1}\cos(\omega_0 t)$. The numerical schemes are initialised exactly using $u^0 = u_0$, $u^1 = u(k)$. Left: global error Ξ^n as a function of time for $k = 10^{-3}$ (dashed-dotted) and $k = 10^{-6}$ (dashed). Right: global error as a function of k, computed at $nk = 1$.

one has

$$\left(-z^2 + 16z - 30 + 16z^{-1} - z^{-2}\right)\hat{u}z^n = 0, \tag{7.25}$$

and it is easy to verify that there exists one (real) root $z \approx 13.9282$ for which clearly $|z| > 1$. The scheme is unstable, and the higher accuracy of this difference operator is useless. In turn, it is the evolution of some more global measures of the scheme's performance that we seek. Consider the following definition of the *global error*:

$$\Xi^n := u(t_n) - u^n. \tag{7.26}$$

This defines a *time series* as a difference between the "true" solution $u(t_n)$ and the scheme output at the corresponding time step u^n. Convergence is assured when:

$$\lim_{k\to 0} \Xi^n = 0, \; \forall n. \tag{7.27}$$

For stable, consistent schemes, the global error Ξ^n can be expected to maintain the same trend as the local truncation error ε, as formalised in the *Lax equivalence theorem*. In Figure 7.3, the output of scheme (7.1) is compared against the exact solution given in (5.4). The behaviour of the error Ξ^n is assessed as a function of time and as a function of the time step. In this latter case, the expected slope of two is recovered, as visible in the double log plot.

7.1.4 Initialisation

For the test in Figure 7.3, the scheme was initialised exactly using knowledge coming from the exact solution $u(t)$ as per (5.4). Of course, generally, an exact solution is unavailable, and schemes must be initialised in some other manner. The two initial steps in the time

series u^0 and u^1 are obtained as a function of the initial data in continuos time u_0, v_0. Obviously, one may set:

$$u^0 = u_0. \tag{7.28}$$

Applying a forward difference to discretise the time derivative, u^1 can be extracted from:

$$u^1 = u^0 + kv_0. \tag{7.29}$$

This approximation to the initial conditions is only *first-order accurate*, as this is the truncation error of the operator δ_{t+}, see (6.10). The operator $\delta_{t\cdot}$ returns a higher-accurate discretisation of the first time derivative, and one may be tempted to use its definition directly to obtain u^1. This, however, requires knowledge of u^{-1}, which is undefined. However, consider the following expression for the centred time difference:

$$\delta_{t\cdot} = \delta_{t+} - \frac{k}{2}\delta_{tt}. \tag{7.30}$$

Applying this to discretise the continuous velocity initial condition yields:

$$\delta_{t+}u^0 + \frac{k\omega_0^2}{2}u_0 = v_0, \tag{7.31}$$

where the identity $\delta_{tt} = -\omega_0^2$ was used. This identity is obtained directly from scheme (7.1). Thus, a second-order accurate initial condition can be given as:

$$u^1 = \left(1 - \frac{k^2}{2}\omega_0^2\right)u^0 + kv_0. \tag{7.32}$$

Higher-order accurate approximations are possible:

$$\delta_{t+}x^0 = v_0 \quad \text{first order} \tag{7.33}$$

$$\left(\delta_{t+} - \frac{k}{2}\delta_{tt}\right)u^0 = v_0 \quad \text{second order} \tag{7.34}$$

$$\left(\delta_{t+} - \frac{k}{2}\delta_{tt} - \frac{k^2}{6}\delta_{t+}\delta_{tt}\right)u^0 = v_0 \quad \text{third order} \tag{7.35}$$

$$\left(\delta_{t+} - \frac{k}{2}\delta_{tt} - \frac{k^2}{6}\delta_{t+}\delta_{tt} - \frac{k^3}{24}\delta_{tt}^2\right)u^0 = v_0 \quad \text{fourth order} \tag{7.36}$$

In the expressions above, substituting $(\delta_{tt})^p = (-\omega_0)^p$ gives a way to compute u^1, knowing u^0 and v_0.

7.1.5 Frequency warping and modified equation techniques

Order accuracy is a useful measure quantifying how fast the error of a finite difference decreases as the time step is reduced. The error is measured as a residual on the displacement $u(t_n)$ and, hence, does not convey information regarding the frequency domain behaviour

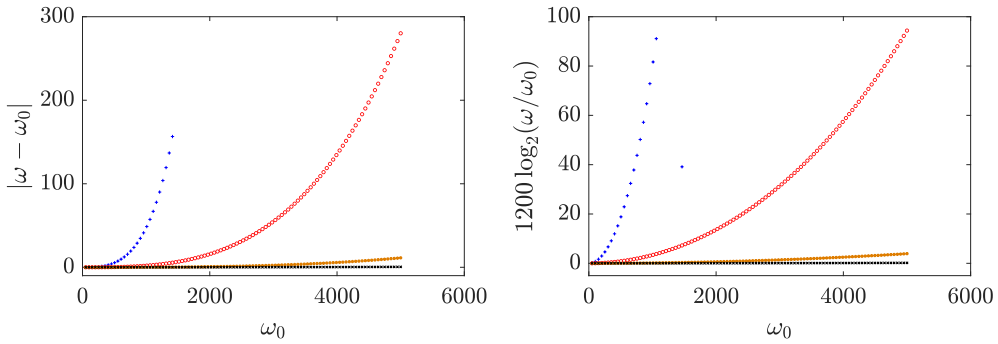


Figure 7.4: Frequency warping (left) and cent deviation (right) of scheme (7.1), under four different choices of k: 10^{-3} (blue); $2.1 \cdot 10^{-4}$ (red); $4.6 \cdot 10^{-5}$ (orange); 10^{-6} (black). Note that the case $k = 10^{-3}$ goes unstable for $\omega_0 > 2000$, as per (7.7).

of the solution. Sound, however, is often analysed in the frequency domain, and thus, a suitable measure of the frequency warping due to the numerical approximations is desirable. Solution (7.4) suggests that the time series u^n oscillates with natural frequency:

$$\mathfrak{w}_0 = \frac{1}{k} \arctan \frac{\omega_0^2 k^2 \sqrt{4(\omega_0 k)^{-2} - 1}}{(2 - \omega_0^2 k^2)} = \omega_0 \left(1 + \frac{\omega_0^2 k^2}{24} + \frac{3\omega_0^4 k^4}{640} + \mathcal{O}(\omega_0^6 k^6)\right), \tag{7.37}$$

and hence, the natural frequency computed by scheme (7.1) is second-order accurate compared to the natural frequency of the continuous system. See also Figure 7.4. The frequency error can be quite audible as ω_0 approaches the limit of stability $\omega_{max} = 2/k$. In practice, the cent deviation from (7.37) is given by

$$1200 \log_2 \frac{\mathfrak{w}_0}{\omega_0} = 1200 \log_2 \left(1 + \frac{\omega_0^2 k^2}{24} + \frac{3\omega_0^4 k^4}{640} + \mathcal{O}(\omega_0^6 k^6)\right), \tag{7.38}$$

and this is shown in the right panel of Figure 7.4. The frequency warping effects are quite evident. Constructing higher-order schemes cannot be done by employing difference operators with wider stencils, as the simple example in Section 7.1.3 shows: unconditionally unstable discretisations arise! Taylor series arguments, however, can be employed again to "correct" the behaviour of the numerical scheme without affecting its stability properties. This is known as the *modified equation method*. Consider the Taylor series expansion of the second difference operator:

$$\delta_{tt} = \frac{d^2}{dt^2} + \frac{k^2}{12}\frac{d^4}{dt^4} + \frac{k^4}{360}\frac{d^6}{dt^6} + \ldots = \sum_{l=1}^{\infty} \frac{2k^{2(l-1)}}{(2l)!}\frac{d^{2l}}{dt^{2l}}. \tag{7.39}$$

Considering for the moment the expansion up to the term $l = 2$, it is natural to add a term $-\frac{k^2}{12}\delta_{tt}\delta_{tt}u^n$ on the left-hand side of (7.1), in order to cancel the $\mathcal{O}(k^2)$ error. Remembering the identity $\delta_{tt} = -\omega_0^2$, this gives:

$$\delta_{tt} u^n = \frac{1}{k^2}\left(-\omega_0^2 k^2 + \frac{\omega_0^4 k^4}{12}\right) u^n, \tag{7.40}$$

and, of course, the local truncation error ε is now $\mathcal{O}(k^4)$. Considering now the next term in the series, proportional to k^4, and using $\delta_{tt} = -\omega_0^2$ three times, one gets

$$\delta_{tt}u^n = \frac{1}{k^2}\left(-\omega_0^2 k^2 + \frac{\omega_0^4 k^4}{12} - \frac{\omega_0^6 k^6}{320}\right)u^n, \tag{7.41}$$

and this approximation is $\mathcal{O}(k^6)$. This operation can be repeated indefinitely, ultimately giving:

$$\frac{1}{k^2}\left(-\omega_0^2 k^2 + \frac{\omega_0^4 k^4}{12} - \frac{\omega_0^6 k^6}{320} + ...\right) = \frac{2}{k^2}\left(-1 + \underbrace{1 - \frac{\omega_0^2 k^2}{2} + \frac{\omega_0^4 k^4}{4!} - \frac{\omega_0^6 k^6}{6!} + ...}_{\cos(\omega_0 k)}\right), \tag{7.42}$$

and thus, when $\omega_0 k < 1$, the series converges to:

$$\delta_{tt}u^n = \frac{2}{k^2}\left(-1 + \cos(\omega_0 k)\right)u^n. \tag{7.43}$$

Since the series expansion converges to a known function in this case, we can say that scheme (7.43) solves (5.2) *exactly*. Of course, this is somewhat too strong a statement: following the discussion in Section 7.1.4, we know that the scheme will only be as accurate as its initial conditions in this case. However, scheme (7.43) does not warp the frequency axis, and u^n oscillates exactly at ω_0. The scheme is also *unconditionally stable*, as one may prove immediately via energy analysis in a way analogous to (7.13). Such a proof is left as an exercise for the reader.

7.2 The damped oscillator

Following along, and after scaling by mass, a basic discretisation for (5.10) is obtained as:

$$\delta_{tt}u^n = -\omega_0^2 u^n - 2\sigma\delta_{t\cdot}u^n, \tag{7.44}$$

yielding an update equation:

$$u^{n+1} = (\sigma k + 1)^{-1}\left((2 - \omega_0^2 k^2)u^n + (\sigma k - 1)u^{n-1}\right). \tag{7.45}$$

Since the coefficients are constants, the values $(\sigma k + 1)^{-1}(2-\omega_0^2 k^2)$, and $(\sigma k + 1)^{-1}(\sigma k - 1)$ can be stored offline, reducing the number of operations per time step to one sum and two products, almost the same as the undamped case. The local truncation error for this scheme, as defined in the lossless case in (7.20), is $\varepsilon = O(k^2)$, $\forall n$, showing that the LTE is second order. Stability may be inferred using either frequency-domain analysis. Applying the test solution (6.21), one obtains:

$$\left((1 + \sigma k)z - (2 - \omega_0^2 k^2) - (\sigma k - 1)z^{-1}\right)\hat{u}z^n = 0, \tag{7.46}$$

whose roots are given by:

$$z_\pm = \frac{2 - \omega_0^2 k^2 \pm \sqrt{\omega_0^4 k^4 - 4\omega_0^2 k^2 + 4\sigma^2 k^2}}{2(1 + \sigma k)}. \tag{7.47}$$

Concerning stability, an application of the Schur-Cohn condition stability test (6.25) allows to obtain $|z_\pm| < 1$ when:

$$k < 2{\omega_0}^{-1}, \tag{7.48}$$

that is the same as (7.7). The same condition may be arrived at via energy analysis. To that end, multiply (7.44) by $m\,\delta_{t\cdot}u^n$,

$$m\,\delta_{t\cdot}u^n\,(\delta_{tt}u^n + \omega_0^2 u^n) = -R(\delta_{t\cdot}u^n)^2. \tag{7.49}$$

Using identities (6.35), one gets:

$$\delta_{t+}\left(\frac{m}{2}(\delta_{t-}u^n)^2 + \frac{K}{2}u^n e_{t-}u^n\right) = -R(\delta_{t\cdot}u^n)^2 \leq 0, \tag{7.50}$$

which is a discrete counterpart of (5.28). Thus, the discrete energy is non-increasing, and when the total energy is itself non-negative, the boundedness of the solution results. Condition (7.7) is thus necessary and sufficient for stability. Of course, bounds (7.19) hold in this case too.

An analysis of the roots $z_\pm$ is revealing. As before, it is convenient to analyse the factor under the square root first. By studying the behaviour of the quadratic equation in $\omega_0^2 k^2$, it is easy to show that for

$$2 - 2\sqrt{1-\sigma^2k^2} \leq \omega_0^2k^2 \leq 2 + 2\sqrt{1-\sigma^2k^2}, \text{ and } \sigma k < 1 \tag{7.51}$$

the square root in (7.47) is a purely imaginary number, and the time series u^n oscillates. Thus:

$$z_\pm = ae^{\pm j\mathfrak{w}_d k}, \tag{7.52}$$

and:

$$u^n = A_+a^n e^{j\mathfrak{w}_d kn} + A_-a^n e^{-j\mathfrak{w}_d kn}, \tag{7.53}$$

The amplitude and frequency have the following expressions:

$$a := \sqrt{\frac{1-\sigma k}{1+\sigma k}} < 1, \quad \tan\mathfrak{w}_d k := \frac{\sqrt{4\omega_0^2k^2 - 4\sigma^2k^2 - \omega_0^4k^4}}{2-\omega_0^2k^2}. \tag{7.54}$$

From here, one sees that frequency of vibration $\mathfrak{w}_d$ of the discrete-time solution u^n is obtained as an approximation to ω_d, as defined in (5.14), i.e. $\mathfrak{w}_d = \omega_d + \mathcal{O}(k^2)$. Furthermore, the amplitude of the time series decreases over time. When the initial conditions are substituted in (7.53), one gets:

$$u^n = a^n\left(u^0\cos(\mathfrak{w}_d kn) + \frac{u^1 - au^0\cos(\mathfrak{w}_d k)}{a\mathfrak{w}_d k\,\mathrm{sinc}(\mathfrak{w}_d k)}\sin(\mathfrak{w}_d kn)\right), \tag{7.55}$$

which approximates the exact solution in continuous time (5.15).

7.2.1 Higher-order schemes

Higher-order accurate schemes may be obtained in this case as well. To that end, consider the definition of the LTE, as:

$$(\delta_{tt} + 2\sigma\delta_{t\cdot})\, u(t) = -\omega_0^2 u(t) + \varepsilon, \tag{7.56}$$

where $u(t)$ is the true solution. Expanding in a Taylor series, one has:

$$\left(\frac{d^2}{dt^2} + 2\sigma\frac{d}{dt}\right)\left(1 + \frac{k^2}{6}\frac{d^2}{dt^2}\right) u(t) - \frac{k^2}{12}\frac{d^4}{dt^4}u(t) + \mathcal{O}(k^4) = -\omega_0^2 u(t) + \varepsilon. \tag{7.57}$$

This suggests the use of the following modified scheme to cancel the terms proportional to k^2:

$$(\delta_{tt} + 2\sigma\delta_{t\cdot})\left(1 - \frac{k^2}{6}\delta_{tt}\right) u^n + \frac{k^2}{12}\delta_{tt}\delta_{tt}u^n = -\omega_0^2 u^n. \tag{7.58}$$

Since $\delta_{tt} + 2\sigma\delta_{t\cdot} = -\omega_0^2 + \mathcal{O}(k^2)$, the scheme above can be written as (to the order $\mathcal{O}(k^4)$):

$$(\delta_{tt} + 2\sigma\delta_{t\cdot})\, u^n - \frac{k^2}{6}(-\omega_0^2)\delta_{tt}u^n + \frac{k^2}{12}\delta_{tt}\delta_{tt}u^n = -\omega_0^2 u^n. \tag{7.59}$$

A suitable approximation to $\delta_{tt}\delta_{tt}$ involving, at most, a stencil of width two is needed. This can be accomplished in the following way:

$$\delta_{tt}\delta_{tt} \approx \left(-\omega_0^2 e_{t-} - 2\sigma\delta_{t-}\right)\left(-\omega_0^2 e_{t+} - 2\sigma\delta_{t+}\right) = \omega_0^4 + 4\sigma\omega_0^2\delta_{t\cdot} + 4\sigma^2\delta_{tt}. \tag{7.60}$$

This expression can now be substituted in (7.59) to give:

$$\left(1 + \frac{k^2}{6}(\omega_0^2 + 2\sigma^2)\right)\delta_{tt}x^n = -\omega_0^2\left(1 + \frac{\omega_0^2 k^2}{12}\right)u^n - 2\sigma\left(1 + \frac{\omega_0^2 k^2}{6}\right)\delta_{t\cdot}u^n. \tag{7.61}$$

which is a fourth-order accurate approximation to (5.10). Higher-order accurate schemes may be obtained this way, i.e. finding approximations to δ_{tt}^p, involving only operators of width two. A sketch of the idea is given briefly here. From (7.60), one may construct $\delta_{tt}\delta_{tt}\delta_{tt}$ in the following way:

$$\begin{aligned}\delta_{tt}\delta_{tt}\delta_{tt} \approx \left(\omega_0^4 + 4\sigma\omega_0^2\delta_{t\cdot} + 4\sigma^2\delta_{tt}\right)\delta_{tt} \approx \left(\omega_0^4 e_{t-} + 4\sigma\omega_0^2\delta_{t-} + 4\sigma^2(-\omega_0^2 e_{t-} - 2\sigma\delta_{t-})\right)\delta_{tt} \approx \\ \left(\omega_0^4 e_{t-} + 4\sigma\omega_0^2\delta_{t-} + 4\sigma^2(-\omega_0^2 e_{t-} - 2\sigma\delta_{t-})\right)\left(-\omega_0^2 e_{t+} - 2\sigma\delta_{t+}\right) = \\ \left(-\omega_0^6 + 4\sigma^2\omega_0^4\right) + \left(-6\sigma\omega_0^4 + 16\sigma^3\omega_0^2\right)\delta_{t\cdot} + \left(-8\sigma^2\omega_0^2 + 16\sigma^4\right)\delta_{tt},\end{aligned}$$

showing that $\delta_{tt}\delta_{tt}\delta_{tt}$ can be approximated using a stencil of width two. One may use the modified equation technique described above to any desired order. Luckily, the oscillator with loss also possesses an exact solution, where "exact" is intended in the same way as for (7.43) (i.e. exact up to the accuracy order of the initial conditions). Under the following transformation:

$$U(t) = e^{\sigma t}u(t), \tag{7.62}$$

the continuous equation (5.10) becomes:

$$\frac{d^2U}{dt^2} + \omega_d^2\, U = 0. \tag{7.63}$$

Thus, the exact scheme (7.43) for the undamped oscillator can be applied to the transformed variable U. When transformed back to u, this gives

$$\delta_{tt} u^n = \left(-\frac{2}{k^2} \left(1 - \cos\left(\omega_d\, k\right)\right) - \frac{e_{t+}(e^{\sigma k} - 1) + e_{t-}(e^{-\sigma k} - 1)}{k^2} \right) u^n. \tag{7.64}$$

This scheme solves (7.43) exactly; in particular, the frequency of oscillation and the numerical decay time are exact.

7.3 The forced oscillator

The analysis of the forced oscillator in continuous time extends directly to the discrete case. As a basic finite difference scheme discretising (5.29) is:

$$m\, \delta_{tt} u^n = -K u^n - R \delta_{t\cdot} u^n + F^n, \tag{7.65}$$

where the input time series is defined as a sampled version of the continuous input, i.e. $F^n := F(t_n)$. The update equation for this scheme is

$$u^{n+1} = (\sigma k + 1)^{-1} \left((2 - \omega_0^2 k^2) u^n + (\sigma k - 1) u^{n-1} + k^2 m^{-1} F^n \right), \tag{7.66}$$

and, like previously, one may store the constant coefficients offline. The update can be computed with a total number of two sums and three products per iteration. The cases of a steady harmonic input forcing and a Dirac forcing will now be analysed. Most of the properties seen for the continuous case will translate directly to the discrete case, though warping effects introduced by the scheme's finite bandwidth emerge.

7.3.1 Harmonic forcing

When the input forcing is pure harmonic, $F^n = F_0 e^{j\omega k n}$ for some input forcing frequency ω. A particular solution is obtained by assuming that u^n oscillates at the same frequency as the input, so that $u^n = \hat{u} e^{j\omega k n}$ i.e., the DTFT of u^n as per (6.27). Thus, the difference operators in (7.65) transform as shown in (6.31). The transformed equation is:

$$j\omega\, \mathrm{sinc}(\omega k) \left(\frac{j\omega\, \mathrm{sinc}(\omega k/2) m}{\cos(\omega k/2)} - \frac{jK}{\omega\, \mathrm{sinc}(\omega k/2) \cos(\omega k/2)} + R \right) \hat{u} = F_0, \tag{7.67}$$

from which the discrete-time mechanical impedance $\mathfrak{Z}_m$ takes the two equivalent definitions:

$$\mathfrak{Z}_m := (j\omega\, \mathrm{sinc}(\omega k) \hat{u})^{-1} F_0, \quad \mathfrak{Z}_m := \hat{v}^{-1} F_0, \tag{7.68}$$

where, conveniently, $\hat{v} := j\omega\, \mathrm{sinc}(\omega k) \hat{u}$, as one obtains by transforming $v^n = \delta_{t\cdot} u^n$. Clearly, (7.67) approximates the continuous-time impedance (5.31). At small frequencies, the discrete-time impedance is approximately equal to the continuous impedance (remember that $\mathrm{sinc}\, x \approx 1$ for small x, see also Figure 6.4). In other words:

$$\lim_{\omega k \to 0} \mathfrak{Z}_m(\omega) = Z_m(\omega). \tag{7.69}$$

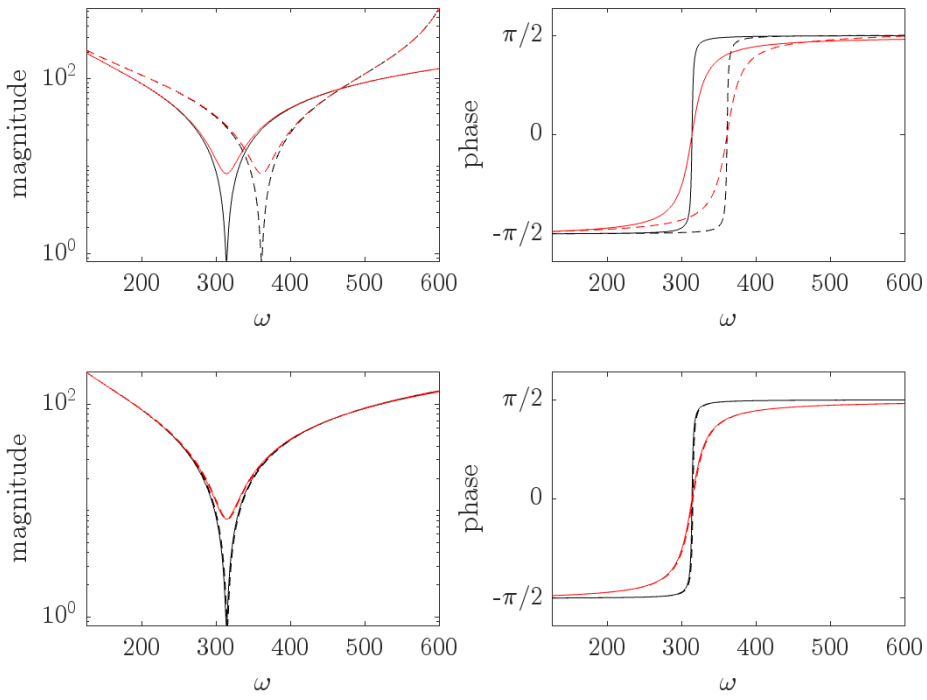


Figure 7.5: Magnitude and phase plots of the discrete (dashed) and continuous (solid) mechanical impedance, under two different choices of the decay time τ_{60}: 5 s (black) and 0.5 s (red). The oscillator's mass is $m = 0.3$, and the natural frequency is $\omega_0 = 100\pi$. Top: $k = 5 \cdot 10^{-3}$. Bottom: $k = 10^{-3}$.

On the other hand, as ωk approaches π (the Nyquist limit), the warping effects become dominant, and the imaginary part of $\mathfrak{Z}_m$ departs considerably from its continuous expression. Figure 7.5 reports the plots of the discrete and continuous mechanical impedance under various choices of the decay time $\tau_6 0$ and time step k. In the top plots, both the magnitude and the phase plots show considerable discrepancies. In this case, $\omega_0 k \approx 1.5$ and, hence, the stability condition (7.7) is respected, the scheme is stable, but the warping effects are evident. In the bottom plot, $\omega_0 k \approx 0.3$ and the behaviour of the discrete-time impedance $\mathfrak{Z}_m$ near resonance is a much better approximation of the continuous-time impedance Z_m. When the discrete-time impedance is written in complex polar form:

$$\mathfrak{Z}_m = re^{j\theta}, \tag{7.70}$$

the displacement and velocity amplitudes become:

$$\hat{u} = F_0(r\omega\,\mathrm{sinc}(\omega k))^{-1}e^{-j(\theta+\frac{\pi}{2})}, \quad \hat{v} = F_0 r^{-1}e^{-j\theta}. \tag{7.71}$$

The derivation of the expressions for r and θ is left as an exercise for the reader.

Like the continuous case, the global solution is obtained as the sum of the solution to the homogeneous difference equation (7.55) plus the particular solution just given. Hence:

$$u^n = a^n\left(u^0\cos(\mathfrak{w}_d kn) + \frac{u^1 - au^0\cos(\mathfrak{w}_d k)}{a\sin(\mathfrak{w}_d k)}\sin(\mathfrak{w}_d kn)\right) + \frac{F_0\sin(\omega kn-\theta)}{r\omega\,\mathrm{sinc}(\omega k)}, \tag{7.72}$$

where, for convenience, the real part of (7.71) was retained. The solution above is assumed valid for $n \geq 0$. Examples of the oscillator's behaviour under sinusoidal input forcings are given in Figure 7.6.

7.3.2 Impulse response and discrete-time Green's function

In continuous time, the solution of the differential equation under Dirac delta input forcing allows computing the impulse response, or the Green's function. An extension of Green's function theory to the discrete case is relatively straightforward. Consider the following definition for the *discrete-time* Green's function:

$$(m\delta_{tt} + K + R\delta_{t\cdot})\mathfrak{G}^{n,n'} = \delta^{n,n'}, \tag{7.73}$$

where δ is here Kroenecker's delta, i.e. zero when $n \neq n'$ and one when $n = n'$. It is easy to show that, if $\mathfrak{G}^{n,n'}$ is known, a particular solution to (7.65) is obtained via the *convolution sum*:

$$u^n = \sum_{n'=-\infty}^{\infty}\mathfrak{G}^{n,n'}F^{n'}. \tag{7.74}$$

This is proven by substituting this expression for u^n in the left-hand side of (7.65):

$$(m\delta_{tt} + K + R\delta_{t\cdot})\sum_{n'=-\infty}^{\infty}\mathfrak{G}^{n,n'}F^{n'} = \sum_{n'=-\infty}^{\infty}\underbrace{(m\delta_{tt} + K + R\delta_{t\cdot})\mathfrak{G}^{n,n'}}_{\delta^{n,n'}}F^{n'} = F^n, \tag{7.75}$$

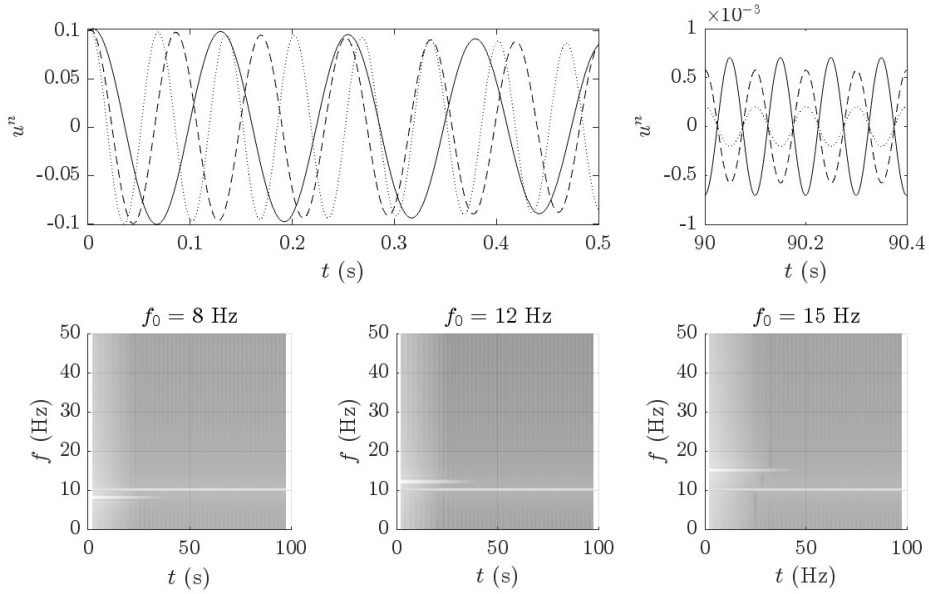


Figure 7.6: Simple harmonic oscillator under sinusoidal input forcing. Here, three oscillators with different natural frequencies are simulated, under a steady sinusoidal input forcing with forcing frequency $f = 10$ Hz. The natural frequencies of the oscillators are as indicated. In the spectrograms, note that the contribution of the initial conditions dies out after sufficient time has elapsed. For the simulations, the initial displacement is set at $u_0 = 0.1$, and the initial velocity is $v_0 = 1$. The loss factor is $\sigma = 0.3$, yielding a decay time $\tau_{60} \approx 23$ s. For the simulations, the sample rate is selected as $f_s = 2000$.

from which the claim follows. Computing the discrete-time Green's function is immediate in the z domain. To that end, compute the z transform of (7.73):

$$\sum_{n=-\infty}^{\infty} (m\delta_{tt} + K + R\delta_{t\cdot})\mathfrak{G}^{n,n'} z^{-n} = z^{-n'}, \tag{7.76}$$

and, hence:

$$\sum_{n=-\infty}^{\infty} \mathfrak{G}^{n,n'} z^{-n} = \frac{k^2}{m} \frac{z^{-n'}}{(1+\sigma k)z - (2-\omega_0^2 k^2) - (\sigma k - 1)z^{-1}}. \tag{7.77}$$

Rearranging the right-hand side, and using expressions (7.52) to factor the denominator, leads to:

$$\sum_{n=-\infty}^{\infty} \mathfrak{G}^{n,n'} z^{-(n-n')} = \frac{k^2}{m\sqrt{1-\sigma^2 k^2}\sin(\mathfrak{w}_d k)} \frac{az\sin(\mathfrak{w}_d k)}{z^2 - 2a\cos(\mathfrak{w}_d k)z + a^2}, \tag{7.78}$$

which may be inverted using the tabulated inverse given in (6.32), yielding

$$G^{n,n'} = \frac{ka^n \sin(\mathfrak{w}_d k n)\Theta^{n,n'}}{\sqrt{1-\sigma^2 k^2}\,\mathrm{sinc}(\mathfrak{w}_d k) m \mathfrak{w}_d}. \tag{7.79}$$

7.4 Multiple Degrees of Freedom

In this section, the case of coupled mass-spring systems is considered. Some results can be derived as direct extensions of the single-mass case, but we must pay special attention to understanding the differences and subtleties involved here. Like the continuous-time case, it is convenient to start from the analysis of system (5.52). A basic finite difference scheme can be derived by simply discretising the second-time derivative:

$$m_1\, \delta_{tt} u_1^n = -K_{11} u_1^n - K_{12}(u_1^n - u_2^n), \tag{7.80a}$$
$$m_2\, \delta_{tt} u_2^n = -K_{22} u_2^n + K_{12}(u_1^n - u_2^n). \tag{7.80b}$$

Like previously, u_i^n denotes a *time series* approximating the continuous function $u_i(t)$ at the time $t = t_n := kn$. The system above can be written compactly using the matrix-vector notation:

$$\mathbf{M}\, \delta_{tt} \mathbf{u}^n = -\mathbf{K}\, \mathbf{u}^n, \tag{7.81}$$

where the form of the matrices is as per (5.53). Expanding out the difference operators, one gets

$$\mathbf{M}\, \mathbf{u}^{n+1} = \left(2\mathbf{M} - k^2\mathbf{K}\right) \mathbf{u}^n - \mathbf{M}\mathbf{u}^{n-1}. \tag{7.82}$$

Since $\mathbf{M}$ is fully diagonal, this scheme is *explicit*, and the update may be computed by merely multiplying both sides by the diagonal matrix $\mathbf{M}^{-1}$, and by performing the trivial matrix-vector operations on the right-hand side. In fact, the inverse of a diagonal matrix reduces to elementwise division, so there is no need to store the full form of the inverse matrix.

Scheme (7.81) is consistent and second-order accurate, as one may show immediately after substituting the continuous vector $\mathbf{u}(t)$ in the scheme, and Taylor-expanding the operator δ_{tt} around $t = t_n$. If the scheme is stable, then it is also convergent following an application of the *Lax equivalence theorem*, as noted in Section 7.1.3. Stability analysis may be performed via frequency domain techniques or energy analysis. Let us begin with the former. Multiply (7.80a) by $\delta_{t\cdot}u_1^n$, and (7.80b) by $\delta_{t\cdot}u_2^n$, and sum. Using the same identities as per the scalar case, listed in (6.35), one obtains the following energy balance:

$$\delta_{t+}\mathfrak{h}^{n-\frac{1}{2}} = 0, \tag{7.83}$$

where the form of the energy is:

$$\mathfrak{h}^{n-\frac{1}{2}} = \frac{m_1(\delta_{t-}u_1^n)^2}{2} + \frac{m_2(\delta_{t-}u_2^n)^2}{2} + \frac{K_{11}u_1^n u_1^{n-1}}{2} + \frac{K_{22}u_2^n u_2^{n-1}}{2} + \frac{K_{12}(u_1^n - u_2^n)(u_1^{n-1} - u_2^{n-1})}{2}.$$

Clearly, (7.83) discretises (5.73) to second-order accuracy, however, there is no guarantee that the discrete energy is, in fact, positive. The energy can be written more compactly using the definitions of norm and inner product introduced in (5.75). One has:

$$\mathfrak{h}^{n-1/2} = \frac{1}{2}\left\|\delta_{t-}\mathbf{u}^n\right\|_{\mathbf{M}}^2 + \frac{1}{2}\left\langle e_{t-}\mathbf{u}^n, \mathbf{u}^n\right\rangle_{\mathbf{K}}, \tag{7.84}$$

and note that, clearly, this approximates the continuous energy expression (5.78). Like the case of the harmonic oscillator, the kinetic energy is non-negative by definition, but the potential energy has an indefinite sign. For the symmetric, non-negative matrix $\mathbf{K}$, a vector equivalent of the discrete-time identity (6.35c) can be given as:

$$\left\langle e_{t-}\mathbf{u}^n, \mathbf{u}^n\right\rangle_{\mathbf{K}} = \left\|\mu_{t-}\mathbf{u}^n\right\|_{\mathbf{K}}^2 - \frac{k^2}{4}\left\|\delta_{t-}\mathbf{u}^n\right\|_{\mathbf{K}}^2. \tag{7.85}$$

Substituting this identity in (7.84), one gets:

$$\mathfrak{h}^{n-1/2} = \frac{1}{2}\left(\left\|\delta_{t-}\mathbf{u}^n\right\|_{\mathbf{M}}^2 - \frac{k^2}{4}\left\|\delta_{t-}\mathbf{u}^n\right\|_{\mathbf{K}}^2\right) + \frac{1}{2}\left\|\mu_{t-}\mathbf{u}^n\right\|_{\mathbf{K}}^2. \tag{7.86}$$

Thus, non-negativity of the energy overall is guaranteed whenever

$$\left\|\delta_{t-}\mathbf{u}^n\right\|_{\mathbf{M}}^2 - \frac{k^2}{4}\left\|\delta_{t-}\mathbf{u}^n\right\|_{\mathbf{K}}^2 \geq 0. \tag{7.87}$$

Applying the definition of the norm from (5.75), this translates to:

$$\mathrm{eig}(\mathbf{A}) \geq 0, \quad \mathbf{A} := \left(\mathbf{M} - \frac{k^2}{4}\mathbf{K}\right). \tag{7.88}$$

For the case $K_{11}, K_{22}, K_{12}, m_1, m_2 = 1$, the eigenvalue equation for eigenvalue $\lambda_{\mathbf{A}}$ is obtained as

$$\left(\frac{k^2}{2} + \lambda_{\mathbf{A}} - 1\right)^2 - \frac{k^2}{16} \geq 0. \tag{7.89}$$

The quadratic has two solution, $\lambda_{\mathbf{A}}^+ = 1 - \frac{k^2}{4}$, $\lambda_{\mathbf{A}}^- = 1 - \frac{3k^2}{4}$, and they are both positive if and only if $\lambda_{\mathbf{A}}^+ > 0$, i.e.

$$k \leq \frac{2}{\sqrt{3}}, \tag{7.90}$$

which may be interpreted as a stability condition for scheme (7.80).

7.4.1 Modal Decomposition

The stability condition (7.90) has a direct interpretation in terms of the stability condition of the simple harmonic oscillator, (7.7). From Section 5.4.1, we know that the "fastest" frequency in the system is precisely $\sqrt{3}$. It is easy to understand how this condition arises, using the eigenvalue decomposition (5.67). As before, defining $\mathbf{w} = \hat{\mathbf{U}}^{-1}\mathbf{u}$, the modal equations are obtained as:

$$\delta_{tt}\mathbf{w}^n = -\mathbf{\Omega}^2\,\mathbf{w}^n, \tag{7.91}$$

showing that motion can be completely uncoupled in the two modes of the system. Stability analysis here is immediate since the energy is now expressed as the sum of *independent* harmonic oscillators. Hence, the stability condition of the simple harmonic oscillator translates here directly, as:

$$k < \frac{2}{\max(\omega_i)} \quad i = 1, ..., M. \tag{7.92}$$

For the test case $K_{11}, K_{22}, K_{12}, m_1, m_2 = 1$, one recovers of course (7.90).

Loss and Forcing

Including losses and external forcing is immediate from the template above. A discrete-time version is obtained as

$$m_1\delta_{tt}u_1^n = -K_{11}u_1^n - K_{12}(u_1^n - u_2^n) - R_1\delta_{t\cdot}u_1^n + F_1^n, \tag{7.93a}$$
$$m_2\delta_{tt}u_2^n = -K_{22}u_2^n - K_{12}(u_2^n - u_1^n) - R_2\delta_{t\cdot}u_2^n + F_2^n. \tag{7.93b}$$

The system can be written much more compactly using the matrix-vector notation, as:

$$\mathbf{M}\delta_{tt}\mathbf{u}^n = -\mathbf{K}\mathbf{u}^n - \mathbf{R}\delta_{t\cdot}\mathbf{u}^n + \mathbf{F}^n. \tag{7.94}$$

The energy balance is obtained after taking an inner product of the equation above with $\delta_{t\cdot}u^n$. The result is:

$$\delta_{t+}\mathfrak{h}^{n-1/2} = -\left\|\delta_{t\cdot}\mathbf{u}^n\right\|_{\mathbf{R}}^2 + \left\langle\delta_{t\cdot}\mathbf{u}^n, \mathbf{F}^n\right\rangle. \tag{7.95}$$

Here, the discrete energy has the same expression as per (7.84). In the zero-input case ($\mathbf{F} = \mathbf{0}$), or after the force has decreased to a null value, the system is strictly dissipative, and the energy overall decreases. Hence, the non-negativity of the energy corresponds yields again a stability condition.

Bibliography

[1] P. Morse and K.U. Ingard. *Theoretical Acoustics*. Princeton Univ Pr, 1987.

[2] J. l.R. d'Alembert. Recherches sur la courbe que forme une corde tendue mise en vibration. *Mem Acad Sci Berlin*, 3:214–219, 1747.

[3] N. H. Fletcher and T. D Rossing. *The physics of musical instruments*. Springer Science & Business Media, 2012.

[4] J.O. Smith. *Music Applications of Digital Waveguides*. PhD thesis, Stanford University, 1987.

[5] J. O. Smith. *Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects*. W3K Publishing, Palo Alto, CA, 2010. Available online at https://ccrma.stanford.edu/ jos/pasp/.

[6] T. J. R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1987.

[7] K-J. Bathe. *Finite Element Procedures*. Prentice Hall, Upper Saddle River, NJ, 1996.

[8] R.J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations. Steady State and Time Dependent Problems*. SIAM, Philadelphia, USA, 2007.

[9] J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2nd edition, 2004.

[10] B. Hamilton and S. Bilbao. Optimised 25-point finite difference schemes for the three-dimensional wave equation. In *Proc. Meet. Acoust.*, volume 28, page 015022, 2016.

[11] S. Bilbao, B. Hamilton, J. Botts, and L. Savioja. Finite volume time domain room acoustics simulation under general impedance boundary conditions. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 24(1):161–173, 2016.

[12] K. Kowalczyk and M. van Walstijn. Room acoustics simulation using 3-d compact explicit fdtd schemes. *IEEE Trans. Audio Speech Lang. Process.*, 19(1):34–46, 2011.

[13] J. Botts and L. Savioja. Effects of sources on time-domain finite difference models. *J. Acoust. Soc. Am.*, 136(1):242–247, 2014.

[14] M. Ducceschi and S. Bilbao. Simulation of the geometrically exact nonlinear string via energy quadratisation. *J. Sound Vib.*, 534:117021, 2022.

[15] M. Ducceschi and S. Bilbao. Non-iterative, conservative schemes for geometrically exact nonlinear string vibration. In *Proc. Int. Congr. Acoust. (ICA)*, Aachen, Germany, 2019.

[16] M. Ducceschi, C. Touzé, and S. Bilbao. Dynamics of the wave turbulence spectrum in vibrating plates: A numerical investigation using a conservative finite difference scheme. *Physica D*, 280–281:73–85, 2014.

[17] S. Bilbao, O. Thomas, C. Touzé, and M. Ducceschi. Conservative numerical methods for the full von kármán plate equations. *Num. Meth. Part. Diff. Eq.*, 31(6):1948–1970, 2015.

[18] S. Bilbao. Immersed boundary methods in wave-based virtual acoustics. *J. Acoust. Soc. Am.*, 151(3):1627–1638, 2022.

[19] S. Bilbao and B. Hamilton. Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation. *IEEE Trans. Audio Speech Lang. Process.*, 21(7):1524–1533, 2013.

[20] R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *Math. Ann.*, 100:32–74, 1928.

[21] G. C. Goodwin, S. F. Graebe, and M. Salgado. *Control system design*, volume 240. Prentice Hall, Upper Saddle River, NJ, 2001.

[22] T. Ha-Duong and P. Joly. On the stability analysis of boundary conditions for the wave equation by energy methods. part i: The homogeneous case. *Math. Comput.*, 62(206):539–563, 1994.

[23] P. Joly. Exact boundary conditions for the finite element solution of time-dependent problems. *ESAIM: Math. Model. Numer. Anal.*, 23(3):329–346, 1989.

[24] S. Bilbao. *Numerical Sound Synthesis*. John Wiley & Sons, Ltd, Chichester, UK, 2009.

[25] M. Shashkov. *Conservative Finite-Difference Methods on General Grids*. CRC Press, Boca Raton, FL, 1996.

[26] L. Meirovitch. *Methods of Analytical Dynamics*. McGraw-Hill, New York, 1970.

[27] T. Okuzono and T. Yoshida. High potential of small-room acoustic modeling with 3d time-domain finite element method. *Front. Built Environ.*, 8:1006365, 2022.

[28] R. Tournemenne and J. Chabassier. A comparison of a one-dimensional finite element method and the transfer matrix method for the computation of wind music instrument impedance. *Acta Acust. united Acust.*, 105(5):838–849, 2019.

[29] A. Thibault and J. Chabassier. Dissipative time-domain one-dimensional model for viscothermal acoustic propagation in wind instruments. *J. Acoust. Soc. Am.*, 150(2):1165–1175, 2021.

[30] J. Chabassier and P. Joly. Time domain simulation of a piano. part 2: Numerical aspects. *ESAIM Math. Model. Numer. Anal.*, 49(5):1331–1372, 2015.

[31] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse. On the oscillations of musical instruments. *J. Acoust. Soc. Am.*, 74(5):1325–1345, 1983.

[32] A. Chaigne. On the use of finite differences for the synthesis of musical transients. application to plucked stringed instruments. *J. Phys. IV*, 2(C5):187–196, 1992.

[33] A. Chaigne and A. Askenfelt. Numerical simulations of piano strings. i. a physical model for a struck string using finite difference methods. *J. Acoust. Soc. Am.*, 95(2):1112–1118, 1994.

[34] V. Doutaut, D. Matignon, and A. Chaigne. Numerical simulations of xylophones. ii: Time-domain modeling of the resonator and of the radiated sound pressure. *J. Acoust. Soc. Am.*, 104(3):1633–1647, 1998.

[35] T. J. R. Hughes and H. S. Tzou. The finite element method in plate bending analysis. *Comput. Struct.*, 7:311–317, 1977.

[36] P. Joly. Finite element methods with continuous displacement. In Kurt Friedrichs and Hilary Mackenzie, editors, *Effective Computational Methods for Wave Propagation*, pages 247–266. Chapman & Hall/CRC, 2008.

[37] M. Duruflé, P. Grob, and P. Joly. Influence of gauss and gauss-lobatto quadrature rules on the accuracy of a quadrilateral finite element method in the time domain. *Numer. Methods Partial Differ. Equ.*, 25(3):526–551, 2009.

[38] G. Cohen, P. Joly, J. E. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.*, 32(4):1449–1467, 1995.

[39] J. D. Morrison and J.-M. Adrien. Mosaic: A framework for modal synthesis. *Comput. Music J.*, 17(1):45–56, 1993.

[40] J.-M. Adrien. The missing link: Modal synthesis. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representations of Musical Signals*, pages 269–298. MIT Press, 1991.

[41] J. Woodhouse. On the synthesis of guitar plucks. *Acta Acust. united Acust.*, 90(6):928–944, 2004.

[42] B. Bank and L. Sujbert. Generation of longitudinal vibrations in piano strings: From physics to sound synthesis. *J. Acoust. Soc. Am.*, 117(4):2268–2278, 2005.

[43] B. Bank. A modal-based real-time piano synthesizer. *IEEE Trans. Audio Speech Lang. Process.*, 18(4):809–821, 2010.

[44] M. van Walstijn, J. Bridges, and S. Mehes. A real-time synthesis oriented tanpura model. In *Proc. Int. Conf. Digital Audio Effects (DAFx-16)*, pages 175–182, 2016.

[45] M van Walstijn, V. Chatziioannou, and A. Bhanuprakash. Implicit and explicit schemes for energy-stable simulation of string vibrations with collisions: Refinement, analysis, and comparison. *J Sound Vib*, 569:117968, 2024.

[46] E. Maestre, G. Scavone, and J. O. Smith. Efficient rendering of saxophone sound by modal synthesis and wave scattering. *J. Acoust. Soc. Am.*, 144(3):1752, 2018.

[47] E. Maestre, G. P. Scavone, and J. O. Smith. Joint modeling of bridge admittance and body radiativity for efficient synthesis of string instrument sound by digital waveguides. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 25(5):1128–1139, 2017.

[48] P. J. Davis. *Interpolation and Approximation*. Dover Publications, Mineola, NY, 1975.

[49] R. W. Hamming. *Numerical Methods for Scientists and Engineers*. McGraw-Hill, New York, 2nd edition, 1973.

[50] P Langer, M Maeder, C Guist, M Krause, and S Marburg. More than six elements per wavelength: The practical use of structural finite element models and their accuracy in comparison with experimental results. *J Comput Acoust*, 25(04):1750025, 2017.

[51] S. Marburg. Six boundary elements per wavelength: Is that enough? *J Comp Acoust*, 10(01):25–51, 2002.

[52] F. Soares, J. Antunes, and V. Debut. Multi-modal tuning of vibrating bars with simplified undercuts using an evolutionary optimization algorithm. *Appl. Acoust.*, 173:107704, 2021.