# 1 Introduction

How would you describe

- The flickering of a flame?
- The texture of an oil painting?
- Highway traffic during a rush hour?
- Twinkling stars?
- Breaking glass?
- A bowling ball hitting pins?
- Melting ice?
- The flight of a paper airplane?
- The sound of a violin?

These questions do not have simple answers: many are active research areas. There cannot be a single recipe that covers this whole menu. There are many possible levels of description; choosing among them depends on your goals and on the available tools. This text is a tour through those spaces. For example, if you seek to make a mathematical model of a violin, you could use a numerical model based on a first-principles description. This lets you match your model parameters to measurements on a real instrument, and change parameters between a Stradivarius and a Guarneri. However, running it in real time will require a supercomputer, and the effort to find good parameters for the model is almost as much work as building a real violin. Alternatively, you could try to use an analytical (pencil-and-paper) solution to the governing equations; in return for some large approximations you may be able to find a useful explicit solution, but it might not sound very good. Finally, you could forget about the underlying governing equations entirely and experimentally try to find an effective description of how the player's actions are related to the sound made by the instrument (which is a reasonable thing to do because dissipation and symmetries in a system reduce the effective number of degrees of freedom [Temam, 1988]). These three approaches (analytical, numerical, and observational) comprise the three parts of this book.

To build a model there are many decisions that must be made, either explicitly or more often implicitly. Some of these are shown in Figure 1.1. Each of these is a continuum rather than a discrete choice. This list is not exhaustive, but it's important to keep

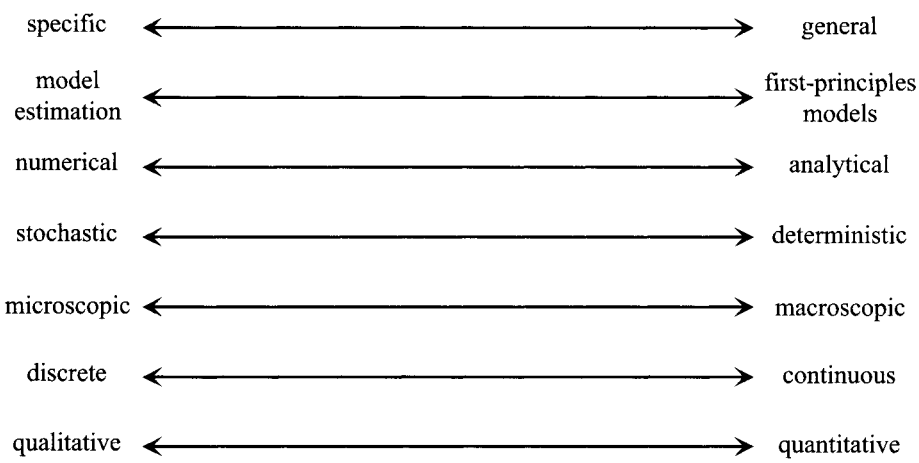


Figure 1.1. Some levels of description for mathematical model building.

returning to it: many efforts fail because of an unintentional attempt to decribe either too much or too little.

These are *meta-modeling* questions. There are no rigorous ways to make these choices, but once they've been decided there are rigorous ways to use them. There's no single definition of a "best" model, although quasi-religious wars are fought over the question. One good attempt is the *Minimum Description Length* principle [Rissanen, 1986], essentially Occam's Razor: the best model is the one that is the smallest (including the information to specify both the form of the model and the values of the parameters). Unfortunately, this has two serious problems: finding the minimum description length for a given problem is an uncomputable task, and it says nothing about the error metric that will be used to judge the model. A stock trader, civil engineer, cardiologist, and video game designer have very different standards for success. They differ in the prior information they have about their problem, and the posterior criteria that they will use to evaluate and update their model. Ultimately, the strongest useful statement is that the best model is the one that works best for you.

Surprisingly little ambition is needed to exceed the performance of almost any available computer, and conversely computer hardware speeds have been racing ahead of the development of software tools to use them effectively. Where computational speed is most important, the examples in this book will use efficient portable low-level tools (such as C and X Windows). On the other hand, where algorithm clarity is most important, high-level environments will be used (such as Matlab). The appendices provide brief introductions to these environments.

No single reference text covers the range of subjects in this book. To help access the literature, each chapter ends with a list of relevant general sources, and then cites the more specialized literature as needed throughout. Where important ideas are introduced without any references they are either so well known that they need no further citation, or are my own results that I have not published elsewhere (the context should make this distinction clear). And I've used *URLs* (*World Wide Web Uniform Resource Locators*) where possible to provide pointers to information on the Internet.

## 1.1 SELECTED REFERENCES

[Press *et al.*, 1992] Press, William H., Teukolsky, Saul A., Vetterling, William T., & Flannery, Brian P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edn. New York, NY: Cambridge University Press.

This is warmly recommended for almost any numerical problem. The numerical analysis literature is full of rigorous results that have little bearing on solving practical problems; *Numerical Recipes* gracefully merges theoretical insights with practical tricks for most useful algorithms. It's one of those rare books that's immediately useful by a beginner but that continues to hold new insights for an expert.

[Pearson, 1990] Pearson, Carl E. (1990). *Handbook of Applied Mathematics: Selected Results and Methods*. 2nd edn. New York: Van Nostrand Reinhold.

This is a good example of one of a number of such large reference volumes that survey applied mathematics.

# *Part One*
# Analytical Models

The first part of the book looks at *analytical models*. These are models that you can at least in theory write down with nothing more than a pencil and a piece of paper, hopefully arriving at an explicit *closed-form* solution. Analytical modeling is often, but not always, done with *analytic functions* [Saff & Snider, 1993], and so we will usually assume that the functions we encounter can be expanded in a power series. Analytical models have been, and continue to be, of great importance because of their power: where they are applicable, it can be possible to deduce everything there is to know about a system. The cost for this power is limited applicability – much of the world is simply too complicated to describe this way.

Analytical models are still important in approximate techniques that do require computers. This includes numerical methods, which can use pieces of analytical solutions to make the numerical steps much more effective, and symbolic methods that can quantitatively expand the effective size of your piece of paper with significant qualitative implications (such as the ability to do much higher-order perturbation theory).

The first chapter covers *ordinary differential equations*, where a collection of variables change as a function of one independent variable (such as time). The orbits of the planets are a classic example. The next chapter adds more independent degrees of freedom (such as space) to arrive at *partial differential equations* to describe, for example, the ripples on the surface of a lake. There is an intimate connection between local differential equations and the global properties of a system, introduced in the following chapter on *variational methods*. The last chapter looks at solutions for *stochastic systems*. While being exact about something random might appear to be paradoxical, there are many powerful techniques for exactly describing the *distribution* of a random variable without saying anything about the particular value of the variable.

# 2 Ordinary Differential and Difference Equations

## 2.1 LINEAR DIFFERENTIAL EQUATIONS

Change is the most interesting aspect of most systems, hence the central importance across disciplines of differential equations. An *ordinary differential equation* (*ODE*) is an equation (or system of equations) written in terms of an unknown function and its derivatives with respect to a single independent variable (such as time). Examples include the familiar equations of classical mechanics and electrical circuits. In the next chapter we will consider *partial differential equations* (*PDEs*), which have multiple independent variables (such as space, for example in fluid flow or electrodynamics). The subject of differential equations can appear to be quite tedious. In part it is: it is like learning spelling and grammar as a necessary prelude to the study of Shakespeare. And in part it isn't: there can be beautiful structure lurking behind what appear to be very simple differential equations. This chapter will concentrate on the canon of linear (or nearly linear) differential equations; after detouring through many other supporting topics the book will return to consider nonlinear differential equations in the closing chapter on time series.

The simplest differential equation can immediately be solved by integration

$$\frac{dy}{dt} = f(t) \Rightarrow dy = f(t)\,dt$$
$$\Rightarrow y(t_1) - y(t_0) = \int_{t_0}^{t_1} f(t)\,dt \tag{2.1}$$

(a point that is surprisingly often forgotten). The *order* of a differential equation is the highest derivative that occurs, and so the preceeding example is a first-order equation. If every term involves either the unknown function or its derivatives the equation is said to be *homogeneous*; if there is a term that depends on the independent variable alone (i.e., a forcing term) then the equation is *inhomogeneous*. If the unknown function does not appear within powers or more complicated functions, then the differential equation is linear, and can be written in terms of a linear operator $L_N(y)$ defined by

$$L_N(y) \equiv \frac{d^N y}{dt^N} + A_1(t)\frac{d^{N-1}y}{dt^{N-1}} + \cdots + A_{N-1}(t)\frac{dy}{dt} + A_N(t)y \quad . \tag{2.2}$$

There is no need for an $A_0$ coefficient because it can be eliminated by dividing all the other terms by it. $L_N(y) = f(t)$ is an inhomogeneous equation, and $L_N(y) = 0$ is the associated homogeneous equation.

Linear differential equations are particularly important, in part because they occur so

often (particularly in systems that are not strongly driven), and in part because general techniques exist for solving them (whether or not they really apply to what might be a nonlinear problem). Although this can be a bit like the proverbial drunk looking for lost change under a street lamp, it is sensible if it is the only illumination available.

The solution of an $N$th-order linear differential equation will contain $N$ unknown constants that are determined by *boundary conditions*. If it is an *initial-value problem*, the initial values of $N$ independent functions of the variable and its derivatives are given (usually, $y(0)$, $dy/dt(0)$, ..., $d^{N-1}y/dt^{N-1}(0)$). For a *boundary-value problem*, boundary conditions are given at both the beginning and the end of an interval.

An $N$-order homogeneous equation $L_N(y) = 0$ will have $N$ linearly independent solutions $u_1(t)$, $u_2(t)$, ..., $u_N(t)$. By superposition, an arbitrary linear combination of them will also be a solution:

$$y_g(t) = \sum_{n=1}^{N} C_n u_n(t) \quad . \tag{2.3}$$

This is the *general solution*; any solution of the homogeneous equation can be represented by an appropriate choice of the $C_n$'s. If a *particular solution* of the inhomogeneous problem can be found ($L_N(y_p) = f(t)$), then the *complete solution* is $y(t) = y_g(t) + y_p(t)$. The general solution represents the transient response of the system to the boundary conditions, and the particular solution is the result of the forcing of the system by the inhomogeneous term.

The simplest linear differential equation has constant coefficients:

$$\frac{d^N y}{dt^N} + A_1 \frac{d^{N-1} y}{dt^{N-1}} + \cdots + A_{N-1} \frac{dy}{dt} + A_N y = f(t) \quad . \tag{2.4}$$

An important technique for solving differential equations is to guess the functional form of a solution (called an *ansatz*, or trial answer), substitute it in, and then see if the free parameters can be adjusted to make the solution work. Because the solution of a differential equation is unique as long as the functions defining it are reasonably smooth and bounded [Coddington & Levinson, 1984], if you find *a* solution then that is *the* solution. If we try the guess $y = e^{rt}$ for the solution of the homogeneous part of equation (2.4), the result of substituting it in is the *characteristic equation*

$$r^N + A_1 r^{N-1} + \cdots + A_{N-1} r + A_N = 0 \quad . \tag{2.5}$$

This $N$th-order polynomial has $N$ roots. The real part of the roots represent exponentially growing or decaying solutions, and the complex part oscillatory behavior. If all of the roots are distinct:

$$r^N + A_1 r^{N-1} + \cdots + A_{N-1} r + A_N = (r - r_1)(r - r_2) \cdots (r - r_N) \tag{2.6}$$

then the general solution is

$$y_g = \sum_{n=1}^{N} C_n e^{r_n t} \quad . \tag{2.7}$$

This gives the $N$ linearly independent solutions required for a general solution. However, if a root has a higher multiplicity

$$r^N + A_1 r^{N-1} + \cdots + A_{N-1} r + A_N = (r - r_1)^M (r - r_{M+1}) \cdots (r - r_N) \tag{2.8}$$

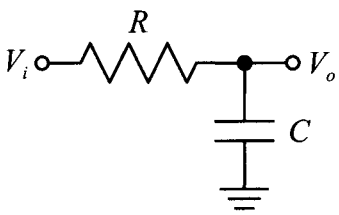


Figure 2.1. An RC circuit.

then this will provide fewer than $N$ solutions. The missing solutions are found by recognizing that if $L_N(e^{rt}) = 0$ then

$$\frac{\partial}{\partial r} L_N(e^{rt}) = L_N\left(\frac{\partial e^{rt}}{\partial r}\right) = L_N(te^{rt}) = 0 \quad ,$$

$$\frac{\partial^2}{\partial r^2} L_N(e^{rt}) = L_N\left(\frac{\partial^2}{\partial r^2} e^{rt}\right) = L_N(t^2 e^{rt}) = 0 \quad , \tag{2.9}$$

and so forth. Therefore, the $M$ functions

$$(C_1 + C_2 t + C_3 t^2 + \cdots + C_M t^{M-1}) e^{r_1 t} \tag{2.10}$$

are linearly independent solutions to $L_N(y) = 0$, and so these can be used as the $M$ solutions associated with the $M$-fold root. It might appear that this trick can be used to generate arbitrarily many solutions by continuing to differentiate, but this is not so: a derivative of an order higher than the multiplicity of equation (2.8) will give the useless equation $0 = 0$.

As a simple example of a linear constant-coefficient differential equation, consider a circuit consisting of a resistor and a capacitor (Figure 2.1). The current into the node from the resistor is $(V_i - V_o)/R$, and the current out of the node into the capacitor is $C\dot{V}_o$, and so the governing equation for this circuit is

$$C\dot{V}_o = \frac{V_i - V_o}{R} \tag{2.11}$$

or

$$RC\dot{V}_o + V_o = V_i \quad . \tag{2.12}$$

The characteristic equation gives

$$RC \cdot r + 1 = 0 \;\Rightarrow\; r = \frac{-1}{RC} \;\Rightarrow\; V_o = Ae^{-t/RC} \quad . \tag{2.13}$$

The undriven response of the circuit is to exponentially discharge the capacitor. Now, let's assume periodic forcing $V_i = \exp(i\omega t)$ and look for a particular solution at this frequency. The voltage in the circuit is of course a real number; by representing it as a complex number we can simultaneously keep track of both phase components (sin and cos). Plugging in the ansatz $V_o = A\exp(i\omega t)$ gives

$$RCAi\omega + A = 1 \;\Rightarrow\; A = \frac{1}{1 + i\omega RC} \quad . \tag{2.14}$$

At low frequencies the output is equal to the input; at high frequencies it rolls off as $1/\omega$ (it is a low-pass filter) and is out of phase by 90°. Problem 2.1 covers the important example of a damped, driven harmonic oscillator.

This completes (more-or-less) everything that there is to know about solving linear differential equations. The theory is simple and useful. The situation is very different for nonlinear differential equations, where amidst a sea of insoluble problems live special tricks for some tractable equations, approximation methods based on some nearby exactly soluble problems, and qualitative insights into the behavior of classes of solutions. Because of this, the study of nonlinear differential equations requires either a lot of specialized attention or else numerical methods.

Another extension of this basic theory is to coupled systems of equations. Once again, little general can be said about nonlinear systems, but for the case of linear couplings it is possible to find exact solutions. The next section looks at this for the important case of coupled oscillators.

## 2.2 SYSTEMS OF DIFFERENTIAL EQUATIONS AND NORMAL MODES

The $N$th-order linear differential equation (2.4) can be written as a first-order equation for an $N$-dimensional vector

$$\frac{d}{dt}\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-2} \\ y_{N-1} \end{pmatrix} = \tag{2.15}$$

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -A_N & -A_{N-1} & -A_{N-2} & \cdots & -A_1 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-2} \\ y_{N-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ f(t) \end{pmatrix} .$$

This transformation does not make the problem any simpler (it can be solved by diagonalizing the matrix, which requires solving exactly the same characteristic equation), but it can be convenient to simplify notation by using a vector first-order equation [Gershenfeld *et al.*, 1983].

This is a simple example of a system of differential equations. Such systems also arise whenever there are interactions; an important special case is an unforced, undamped system of masses with coordinates $(y_1, y_2, \ldots, y_N) \equiv \vec{y}$ that have a restoring force that is an arbitrary linear combination of their positions. The corresponding vector equation is

$$\frac{d^2\vec{y}}{dt^2} + \mathbf{A} \cdot \vec{y} = 0 \quad . \tag{2.16}$$

If the coupling matrix **A** is diagonal ($A_{ij} = 0$ for $i \neq j$) then the oscillators will be independent, but if it isn't then they won't. Let's look for a new set of variables $\vec{z} \equiv \mathbf{M}^{-1} \cdot \vec{y}$, defined by an unknown transformation **M**, for which these equations decouple:

$$\frac{d^2\vec{z}}{dt^2} + \mathbf{D} \cdot \vec{z} = 0 \quad , \tag{2.17}$$

where **D** is a diagonal matrix. The required transformation **M** can be found by changing variables:

$$\begin{aligned} \frac{d^2\vec{y}}{dt^2} + \mathbf{A} \cdot \vec{y} &= 0 \\ \mathbf{M} \cdot \frac{d^2\vec{z}}{dt^2} + \mathbf{A} \cdot \mathbf{M} \cdot \vec{z} &= 0 \\ \frac{d^2\vec{z}}{dt^2} + \mathbf{M}^{-1} \cdot \mathbf{A} \cdot \mathbf{M} \cdot \vec{z} &= 0 \\ \frac{d^2\vec{z}}{dt^2} + \mathbf{D} \cdot \vec{z} &= 0 \end{aligned} \tag{2.18}$$

and so $\mathbf{M}^{-1} \cdot \mathbf{A} \cdot \mathbf{M} = \mathbf{D}$ or $\mathbf{A} \cdot \mathbf{M} = \mathbf{M} \cdot \mathbf{D}$. This will be the case (remember that **D** is diagonal) if the columns of **M** are the eigenvectors of **A** (the diagonal elements of **D** will then be the eigenvalues of **M**). This procedure is called *diagonalizing*. The new variables here are called *normal modes* [Goldstein, 1980, Scheck, 1990] and behave exactly like independent oscillators. There will be as many normal modes as there are degrees of freedom, unless there are fewer distinct eigenvectors because of degenerate eigenvalues. Problem 2.2 finds the normal modes for a simple system.

## 2.3 LAPLACE TRANSFORMS

Using the characteristic equation to solve a differential equation requires separate steps to find the general solution, search for a particular solution, and solve for the coefficients to match the boundary conditions. *Laplace transforms* provide a convenient alternative, turning many differential equations into an algebraic problem and giving the complete solution in a single step.

The *one-sided Laplace transform* of a function $f(t)$ is defined by

$$\mathcal{L}\{f(t)\} \equiv F(s) = \int_0^\infty e^{-st} f(t)\, dt \quad . \tag{2.19}$$

If the integral extended from $-\infty$ to $\infty$ this would be the *two-sided* Laplace transform. The one-sided transform explicitly includes the initial conditions of the system at $t = 0$, and for this reason we will use it; the two-sided transform is used for steady-state problems for which the initial conditions do not matter.

The Laplace transform is a generalization of the Fourier transform to an arbitrary complex argument. Its usefulness for differential equations comes from recognizing that