1

Problem Solving in Engineering

In chemical and biological engineering, students find that the sequence of steps outlined in Figure 1.1 is an effective problem-solving approach for the vast majority of the problems they encounter.

In most courses, students practice all the steps outlined in Figure 1.1, but the focus is usually on the construction of the system diagram and developing the mathematical equations for every unique type of process that is described in a particular course. Only limited attention is usually given to solving the mathematical equations that arise in a particular course because the assumption is that the student should have learned how to do that in their mathematics courses or some other course. Many engineering curricula have a course that is focused on the use of computers to solve the many different types of equations that arise in a student's engineering courses. The focus of this textbook is just "using computers to solve the equation(s) that students typically encounter throughout the engineering curriculum."

The timing of a course on computational or numerical methods for solving engineering problems varies considerably from one curriculum to the next. One approach is to schedule the course near the end of the curriculum. As an upper level course, students are able to review most of the engineering principles and mathematics that they learned previously and develop a new set of tools (specifically, computational tools) for solving those same problems. Two disadvantages are associated with this approach. First, students do not have the computational tools when they first learn a new engineering principle, which limits the scope of problems they can solve to problems that can be largely solved without a computer (i.e., problems that can be solved with paper and pencil). The second disadvantage is that the third and fourth years of many engineering curricula are already filled with other required courses and it is difficult to find time for yet another course.

A second approach is to schedule the computational methods course early in the curriculum, before students have taken most of the engineering courses in which they learn to derive, construct, and identify the mathematical equations they need to solve and that sometime require a computational approach. There are also two problems with this approach. First, the students have

Chemical and Biomedical Engineering Calculations Using Python®, First Edition. Jeffrey J. Heys.

Companion Website: www.wiley.com/go/heys/engineeringcalculations_python

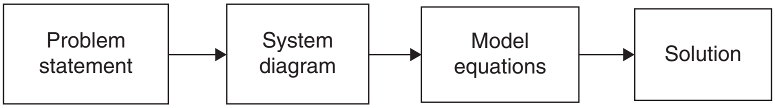

Figure 1.1 Engineering problem-solving process.

typically not taken all the required mathematics courses, and, as a result, it is difficult to teach a computational approach to solving a differential equation when a student is not yet familiar with differential equations or techniques for solving them. The second disadvantage is that the student has not taken courses on separations, kinetics, transport, and so on in which they learn to derive or identify the appropriate mathematical equation(s) for their particular problem. It is, of course, difficult to teach a computational approach to solving an equation when the importance or relevance of that equation is not known.

A third approach for addressing this dilemma is to simply not teach a stand-alone computational methods course and instead cover the relevant computational approaches as they are needed in each individual course. We will continue our listing of the "top two challenges" and identify two potential difficulties with this approach. First, instead of learning and becoming comfortable with two or three computational tools (i.e., mathematical software packages), students under this format often need to learn 4 or 5 computational tools because every one of their instructors prefers a different tool, and the students never really become proficient with any single tool. The second difficulty is that there are a few important concepts that play a role in many of the various computational methods, for example, rounding error, logical operators, and accuracy, that may never be taught if there is not a single course focused on computational methods.

This textbook, and the course that it was originally written to support, is focused on the second approach – a course that appears in the first year or early in the second year of an engineering curriculum. The main reason for adopting this approach is simply the belief that it is critical for students to understand both the potential power and flexibility of computational methods and also the important limitations of these methods before using them to solve problems in engineering. For a student to use a computational tool in a course and blindly trust that tool because they do not understand the algorithms behind the tool is probably more destructive than never learning the tool at all. Further, to limit a student to only problems that can be solved with paper and pencil for most of their undergraduate education is similarly unacceptable. Addressing the limitations associated with teaching computational methods before most of the fundamental engineering and some mathematics courses is difficult. The basic strategy employed by this book is to teach students to recognize the type of mathematical equation they need to solve, and, once they know the type of equation, they can take advantage of the appropriate computational approach that is presented here (or, more likely, refer back to this book for the appropriate algorithm for their particular equation).

There is a second, and possibly more important, reason for learning this material early in the engineering education process. It is related to the fact that one of the most difficult skills for many science, engineering, and mathematics students to master is the ability to combine a number of small, simple pieces together into a more complex framework. In most science, engineering, and mathematics courses in high school and early in college, students learn to find the right equation to solve the question they are asked to answer. Most problems can be completed in *one or two* steps. Problems in later courses, on the other hand, can often require *5–10* or more steps and can require multiple pages of equations and mathematics to solve. This transition from small problems that only require a few lines to large problems that require a few pages can be very challenging for many science, engineering, and mathematics students. I believe that programming in general, and numerical computations, in particular, can be a great way to develop the skills associated with solving larger problems. Programming requires one to combine a number of simple logical commands and variables together into a more complex framework. Programming develops the parts of our brains that allow us to synthesize a number of smaller pieces into a much larger whole. A good analogy is building something complex (e.g., the Death Star) with LEGO bricks. This process requires one to properly and carefully combine a number of simple pieces into a much larger structure. The entire process requires one to simultaneously think on both the large scale ("What is my design objective?") and the small scale ("Will these two pieces stay connected? Are they compatible?"). This skill is necessary for both programming and engineering. It is a skill that almost everyone is capable of developing, but it takes practice – so, we might as well start early!

This textbook advocates that students develop the following skills: (1) recognize the type of mathematical equation that needs to be solved – algebraic or differential? linear or nonlinear? interpolation or regression? ordinary or partial differential equation (PDE)?, and (2) select and implement the appropriate algorithm. If students are able to develop these two skills, they will be equipped with a set of tools that will serve them well in their later engineering courses. These tools can be used by a student to check their work, even when they are primarily using paper and pencil to solve a problem. It is not optimal that students learn how to approximately solve mathematical equations before they know why the equation is relevant, but every effort is made in this book to at least try and explain the relevance of equations when possible.

1.1 Equation Identification and Categorization

We identified two categories of skills that we wish to develop throughout this book: (1) recognizing the type of mathematical equation(s) and (2) selecting and implementing an appropriate computational method. The first skill will be covered in this chapter and then the remainder of the book is for developing the second set of skills.

1.1.1 Algebraic versus Differential Equations

The distinction between algebraic and differential equations is trivial – a differential equation is a relationship between the derivatives of a variable and some function. Differential equations described the rate of change of a variable; typically the rate of change with respect to space or time. Equations can have both independent and dependent variables. It is usually simplest to identify the dependent variables because their value depends on the value of another variable. For example, in both $v(t) = 2\pi + t^2$ and $\frac{dv}{dt} = 3 + v \cdot t$, v is the dependent variable because its value depends on the value of t and t is the independent variable. There can be multiple independent variables, for example, multiple spatial dimensions and time, and the value of dependent variable may depended on the value of all independent variables. The density of air, for example, varies with location: latitude, longitude, and elevation above sea level, as well as time. Therefore, if we have an equation that describes the density of air as a function of location and time, then, in that equation, density is the dependent variable and location and time are the independent variables. Similarly, the ideal gas law can be used to calculate the density of air: $\rho(P, T) = \frac{P}{R \cdot T}$. For this equation, ρ is a function of temperature and pressure, so ρ is the dependent variable and P and T are the independent variables. Alternatively, this equation could be seen such that pressure, P, is the dependent variable that depends on density, ρ, and temperature, T, that is, $P(\rho, T) = \rho \cdot R \cdot T$.

For differential equations, there are three different notation styles that are commonly used for derivatives.

Leibniz notation The derivative of the function, $f(x)$, with respect to x is written as

$$\frac{df}{dx}$$

and the second derivative is written:

$$\frac{d^2f}{dx^2}.$$

The *partial* derivative of $f(x, y)$ with respect to x is

$$\frac{\partial f}{\partial x}.$$

Lagrange notation The derivative of the function, $f(x)$, with respect to x is written:

$$f'(x)$$

and the second derivative is written:

$$f''(x).$$

The notation is not easily extended to partial derivatives and there is no universal standard, but one style that is used is to switch from the prime mark, $'$, to a subscript so that the partial derivative of $f(x, y)$ with respect to x is

$$f_x.$$

Euler notation The derivative of the function, $f(x)$, with respect to x is written:

$$D\ f$$

and the second derivative is written:

$$D^2\ f.$$

The *partial* derivative of $f(x, y)$ with respect to x is

$$D_x\ f.$$

In summary, differential equations have at least one derivative and algebraic equations do not. The presence of a derivative has a significant impact on the computational method used for solving the problem of interest.

1.1.2 Linear versus Nonlinear Equations

A linear function, $f(x)$, is one that satisfies both of the following properties:

additivity: $f(x + y) = f(x) + f(y)$.
homogeneity: $f(c \cdot x) = cf(x)$.

In practice, this means that the dependent variables cannot appear in polynomials of degree two or higher (i.e., $f(x) = x^2$ is nonlinear because $(x + y)^2 \neq x^2 + y^2$), in nonlinear arguments within the function (i.e., $f(x) = x + \sin(x)$ is nonlinear because $\sin(x + y) \neq \sin(x) + \sin(y)$), or as products of each other (i.e., $f(x, y) = x + xy$ is nonlinear).

For algebraic equations, it is typically straightforward to solve linear systems of equations, even very large systems consisting of millions of equations and millions of unknowns. Two different methods for solving linear systems of equations will be covered in Chapter 6. Nonlinear algebraic equations can sometimes be solved exactly using techniques learned in algebra or using symbolic mathematics algorithms, especially when there is only a single equation. However, if we have more than one nonlinear equation or even a single, particularly complex nonlinear algebraic equation (or if we are simply feeling a little lazy), we may need to take advantage of a computational technique to try and find an approximate solution. Algorithms for solving nonlinear algebraic equations are described in Chapter 8.

It is important to note that the distinction between linear and nonlinear equations can also be extended to differential equations and all of the same principles apply. For example, $\frac{dc}{dt} = 4c$ and $\frac{d^2c}{dt^2} = 2\sin(\pi t)$ are linear while $\frac{dc}{dt} = c^2$ is nonlinear. In some cases, the nonlinearity will not significantly increase the computational challenge, but, in other cases like the Navier–Stokes equations, the nonlinearity can significantly increase the difficulty in obtaining even an approximate solution.

Linear versus Nonlinear Examples

Linear:

- single linear equation: $5 \cdot x + \frac{1}{3} = x$
- linear system of equations:

$$3 \cdot x + \frac{y}{4} = 10$$
$$x = 6 \cdot y.$$

Nonlinear:

- single nonlinear equation: $5 \cdot x - \frac{1}{3} = \sqrt{x}$
- single nonlinear equation: $x^2 - 8 \cdot x - 9 = 0$
- nonlinear system of equations:

$$3 \cdot x \cdot y + \frac{y}{4} = 10$$
$$x - 6 \cdot y = 0$$

- nonlinear system of equations:

$$x + y = 4$$
$$\log(x) - 7 \cdot y = 0.$$

1.1.3 Ordinary versus Partial Differential Equations

An ordinary differential equation (ODE) has a single independent variable. For example, if a differential equation only has derivatives with respect to time, t, or a single spatial dimension, x, it is an ODE. A differential equation with two or more independent variables is a PDE. The following are examples of ODEs.

$$t \cdot \frac{dp}{dt} + \frac{d^2p}{dt^2} = \sin(t) \quad \text{(linear, second-order ODE).}$$

If you have not taken a differential equations course, this equation may look a little intimidating or confusing. To solve this equation, we need to find a function $p(t)$ where the first derivative of the function, multiplied by t, plus the second derivative of the function is equal to $\sin(t)$. If that sounds difficult, do not worry, by the end of this textbook, you will know how to get an approximate solution, that is, a numerical approximation of the function $p(t)$. It is also important to emphasize that multiplying the dependent variable p by the independent variable t did not make the equation nonlinear. A nonlinearity only arises if, for example, p is multiplied by itself.

$$\frac{dx}{dt} = x^2 + 3\cos(t) \quad \text{(nonlinear, first-order ODE)}.$$

Again, if you have not had a differential equations course, solving this equation requires finding a function $x(t)$ that has a derivative equal to $(x(t))^2$ plus $3\cos(t)$. Do not worry if that makes your head spin, we will also cover the solution of this class of problems.

Some examples of PDEs are included below.

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \quad \text{(linear, second-order PDE)}.$$

This is an equation that describes unsteady, conductive heat transport in one spatial dimension. You could use this equation to describe, for example, the warming of the ground when the sun comes up in the morning, among many other examples. Solving this equation requires finding a function $T(x, t)$ of both time t and space x where the first derivative with respect to time is equal to α times the second derivative with respect to space.

$$m\frac{\partial m}{\partial x} + \frac{\partial m}{\partial y} = 0 \quad \text{(nonlinear, first-order PDE)}.$$

By now it is probably obvious that the standard mathematical convention is to use ∂ for derivatives in a PDE while ODEs use d. The order of the equation is determined by the order of the highest derivative.

Solving a Differential Equation

Even though you may not have taken a differential equations course, you might be able to solve a simplified version of the first ODE example. Try to solve

$$\frac{d^2p}{dt^2} = \sin(t).$$

Notice that we have eliminated the difficult term with t multiplied by the first derivative. Let us start by integrating both sides of the equation with respect to t:

$$\int \frac{d}{dt}\left(\frac{dp}{dt}\right) dt = \int \sin(t) dt.$$

(Continued)

Solving a Differential Equation (Continued)

Recalling that an integral is just an antiderivative, we get

$$\frac{dp}{dt} + c_1 = -\cos(t) + c_2.$$

The two constants of integration can simply be combined into a single constant, c_0, which can be placed on the right-hand side giving:

$$\frac{dp}{dt} = -\cos(t) + c_0.$$

Now, let us integrate both sides once more with respect to t:

$$p(t) + c_3 = -\sin(t) + c_0 t + c_4,$$

which we can simplify once again by combining the two new constants of integration to a single constant c, to give

$$p(t) = -\sin(t) + c_0 t + c.$$

In order to fully determine our unknown function $p(t)$, we need two additional conditions to solve for the value of our two remaining unknown constants, c_0 and c. Typically, this additional information would be initial conditions, that is, the value of p when $t = 0$, and the value of $\frac{dp}{dt}$ at $t = 0$.

It is always a good idea to check the solution to your problem by substituting $p(t)$ back into the original differential equation and checking to make sure that the left side (i.e., the second derivate of $p(t)$) is equal to the right-hand side.

1.1.4 Interpolation versus Regression

Within engineering, it is often necessary to obtain an equation, usually a polynomial equation, that "fits" a given set of data. If we want an equation that exactly matches the data, then we must interpolate the data so that we obtain a function (e.g., a polynomial) that has the same value as the data for a given value of the independent variable (Figure 1.2). In order to determine an interpolant, the number of adjustable parameters that we determine in the equation must equal the number of data points. For example, if we want to interpolate three data points, we must use an equation that has three adjustable parameters, such as a quadratic polynomial, $ax^2 + bx + c$.

In practice, it is actually pretty rare that we want to exactly interpolate a given set of data because we hopefully have a large amount of data (and we do not want to use a very high-order polynomial) and that data contains some amount of error. In most cases, we want to approximately fit our data with an equation of some form (Figure 1.3). In order to do this, we must first decide

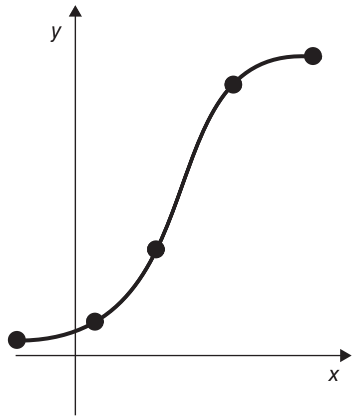

Figure 1.2 An example of interpolation for a set of data. The data is usually represented using points (circles) and the interpolant function is usually represented using a line.

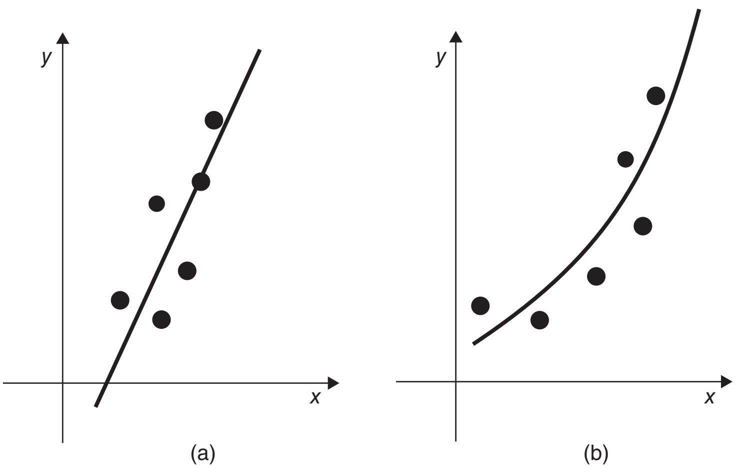

Figure 1.3 An example of linear (a) regression and nonlinear (b) regression for a set of data.

how we want to measure the "goodness" of a fit. Maybe we want to fit an equation so that the sum of the distances from the best fit equation to each and every point is minimized. Another option (the option that is almost always selected) is to minimize the sum of the *square* of the distance between every data point and the "best" fit approximation. This is the so- called least-squares regression approach. The function that gives us the best fit based on our chosen criteria is called the regression function and the process of determining the regression function is called regression analysis. The most popular type of regression, linear regression (Figure 1.3) using least-squares, and nonlinear polynomial regression are both covered in Chapter 7.

Problems

1.1 Determine the type (linear or nonlinear) of algebraic equation assuming x, y, and z are unknown variables:
a) $x^2 + y^2 = 1.0$
b) $x + y = \sqrt{2}$
c) $y = 2 \cdot \sin(x)$
d) $x + y + z^2 = 0$

1.2 Determine the type (linear or nonlinear; ordinary or PDE) of differential equation assuming that z, x, and t are independent variables and g, $\mathcal{D}$, and k are known parameters:
a) $\frac{d^2y}{dt^2} = -g$ (Newton's first law)
b) $\frac{\partial C_A}{\partial t} + v \cdot \frac{\partial C_A}{\partial z} + kC_A = \frac{\partial}{\partial z}\left(\mathcal{D}\frac{\partial C_A}{\partial z}\right)$
c) $f'(x) = \sin(x) + 4$

1.3 If you want to determine the polynomial that interpolates 6 data points, what is the minimum order polynomial that is required? Write the polynomial with x as the independent variable and $a, b, c, \ldots$ as the unknown coefficients.

1.4 You are asked to use regression to determine the best linear polynomial fit for a given set of data. A colleague encourages you to determine the best fit by minimizing the sum of the distance between each point and the line instead of minimizing the sum of the *square* of the distance, which is the standard practice. The colleague claims that this will reduce the influence of a few outlying data points. Is the colleague correct?

1.5 You have been hired to produce an exact replacement part for a classic Porsche because the part is no longer available. Another engineer collects precise measurements of the location of a number of points on the surface of the part. You need to produce a new part with corresponding points at the same locations. Before machining the new part, you need to develop a continuous function that fits the measurement points because the continuous function will provide a representation of the surface connecting the points. Should you develop the continuous function using regression or interpolation between the precisely measured locations on the surface of the part? Why?

1.6 While studying a particular system, you collect some data on a measurable variable (y) versus an adjustable variable (x). Your next task is to use regression to approximately fit the data with a continuous mathematical function. Most engineers would start by trying to fit the data with a polynomial. You are not like most engineers because you wisely start by plotting the data. While examining the plot, you notice that the data has a pattern that is repeated as the adjustable variable is continuously changed. The measured variable increases and decreases regularly as the adjustable variable is increased. Should you fit this data with a polynomial? If so, what order polynomial? If not, what function(s) would you use instead?

Additional Resources

An understanding of how to solve differential equation problems is not required for understanding the material in this book. However, an ability to classify or recognize the type of equation that one is trying to solve is required. Most differential equation textbooks include a comprehensive set of definitions that enable the classification of mathematical equations. Some popular differential equation textbooks for engineers are:

- Differential Equations for Engineers and Scientists by Çengel and Palm [1]
- Advanced Engineering Mathematics by Zill and Cullen [2]
- Advanced Engineering Mathematics by Kreyszig [3]

and a helpful resource for data plotting and regression using Microsoft Excel is:

- Engineering with Excel by Larsen [4].

References

1 Çengel, Y. and Palm, W. III (2013) *Differential Equations for Engineers and Scientists*, McGraw-Hill, New York, NY, 1st edn.
2 Zill, D. and Cullen, M. (2006) *Advanced Engineering Mathematics*, Jones and Barlett, Sudbury, MA, 3rd edn.
3 Kreyszig, E. (2011) *Advanced Engineering Mathematics*, John Wiley and Sons, Inc., Hoboken, NJ, 10th edn.
4 Larsen, R.W. (2009) *Engineering with Excel*, Pearson Prentice Hall, Upper Saddle River, NJ, 3rd edn.