Systems of Linear Equations

Greg Baker

October 9, 2003

Overview

To further our knowledge in the sciences and engineering, we try to formulate mathematial models that quantify behavior. We trust these models only when they can predict observed behavior. Normally, this step requires the solution of the unknown quantities in our model when written as a system of equations. So the first major question is "Do our equations have solutions?"

In general, the question of existence of solutions to mathematical equations is very challenging. Fortunately, the answer is known when the equations are linear. The simplest linear equation

$$ax = b$$

provides amazing insight into the question of existence even for systems of linear equations. I will use this simple equation as the starting point for developing understanding about existence of solutions for increasingly larger systems of equations.

Most who pose mathematical models expect there to be just a single solution. When there are no solutions they are faced with a dilemma; either their model is wrong, or the equations are describing an impossible situation. What is frequently overlooked is that the equations contain the information which may lead to a resolution of the difficulty. On the other hand, there may be many solutions (for systems of linear equations an infinite number of solutions), and the modeler faces a different challenge. Somehow not enough information has been provided. I will use simple examples to illustrate each of these possibilities and what they mean.

There is another very important reason why we want to know whether solutions exist. Computer codes used to compute solutions will often find solutions where none exist. How then can we trust our model if the computer provides false information? We need ways to detect when false solutions are generated by computer codes.

The theme in this material is to face squarely the question of existence of solutions to systems of linear equations. Specifically, we will learn how to determine when no solutions exist, many solutions exist or there is just one solution.

1 Introduction

We are constantly solving linear equations every day simply because they arise from calculating ratios. As an example, suppose we are in a grocery store and we want to buy some eggs. The standard offer is \$1.40 for a dozen, but there is a special offer of \$0.98 for a package of 8 eggs. Is this a better deal? There are several ways to compare these prices. Let us decide to compare the price per dozen. Let x be the price per dozen for the special offer. Since 8 eggs is $2/3$ of a dozen, we know that $2/3$ of the price for a dozen must be \$0.98, the advertized price. Rewriting this statement as an equation, we have

$$\frac{2}{3}x = 0.98\,. \tag{1.1}$$

The solution is easily found:

$$x = \frac{3}{2} \times 0.98 = 1.47$$

which is more expensive – so much for the special offer.

Besides the almost unconscious, everyday use of proportionality, its extension to the formulation of systems of linear equations lies at the heart of many disciplines, such as science, engineering, economics, medicine and so on. While all these applications are important, I'll take several examples from mechanics, in part because of their historical origins.

One of the first mechanical devices that has proved extremely useful is the spring. There are many types of springs, but the one most used to introduce the subject of mechanics in education is the spiral spring. We imagine the spring is attached so that it can hang freely. Then we attach a weight and observe that the spring is stretched. The additional length of the spring is called the extension. It is the unstretched length of the spring subtracted

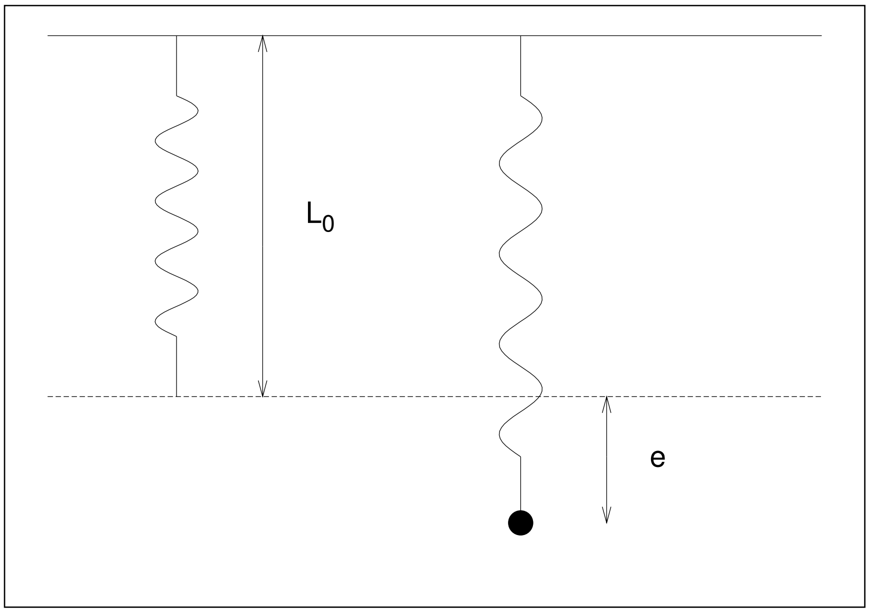


Figure 1: Schematic of a spring that is stretched by some weight

from its stretched length. Let us take what we have stated in English and state it in Mathematics. Let us denote the unstretched length of the spring as L_0. After we attach a weight W to the spring, its new length becomes L_s. We are at liberty to choose any symbol to represent length and it seems natural to pick the letter L. Since we have two lengths to measure, we may distinguish between them by including a subscript. The extension, called e, is the difference between these lengths (see Figure 1)

$$e = L_s - L_0 \,. \tag{1.2}$$

Hooke's law (the mid-seventeenth century) states that the weight is proportional to the extension. The constant of proportionality is called the spring constant which I denote as k. Thus

$$W = ke \,. \tag{1.3}$$

Notice that the spring constant must have the units of weight per length, for example lb/ft, which reflect the nature of the ratio.

The relationship (1.3) serves as a reminder about some important aspects of equations. In general, mathematical equations may be used to represent processes whereby some input produces output. Whenever we establish a mathematical model for processes, we must be mindful of what we consider known and what is to be determined. Let's use (1.3) to illustrate the possibilities. There are three symbols in the relationship and they can take on different meanings depending on the circumstances. At issue is the need to determine what is the input and what is to be determined (the output).

Case 1: First we imagine that we want to determine the weight of some object. We must know the spring constant k. We attach the object to the spring and record the extension e. Then we use (1.3) to determine W. Here, the extension is the input and the weight is the output. The spring constant plays the role of a parameter. It is fixed for a specific choice of spring. If it helps, just consider k as some number. Notice that the evaluation of W is direct. We simply substitute the values of k and e into (1.3).

Case 2: Now suppose we have a spring but we don't know what its spring constant is. We attach a known weight W to the spring and record its extension. The extension is the input and the spring constant is the output. The weight plays the role of a parameter. To calculate k we divide (1.3) by e to obtain $k = W/e$. Although trivial, we had to execute a solution strategy to determine the output.

Case 3: Finally, we wish to design a scale using a spring with a known constant. All we need to know is the extension for a certain weight. Then we can mark off multiples of the extension to create the scale. In this case, the weight is the input and the extension is the output, while the spring constant is just a parameter. Using (1.3) we may calculate the extension by dividing by k. Thus $e = W/k$.

What we learn from these three cases is that circumstances determine what we consider known and what we must calculate. The mathematician assumes that you know what is known and what is unknown. The job of the mathematician is to show you how to calculate the output, and these days that includes the use of numerical algorithms, the basis of computational science. Consequently, the mathematician assumes the unknown is called x and describes the solution strategy to calculate it. As we can see from the three cases above, (1.3) can be used in many ways to state the connection between what is known and unknown, and there are two cases, 2 and 3, that require a solution strategy to find the result. In case 2 the unknown is $x = k$, while in case 3 the unknown is $x = e$.

In general, a single linear equation has the form,

$$ax = b\,. \tag{1.4}$$

In English, this equation says the multiplication of a parameter a with an unknown quantity x must balance the input b. There is only one value of x for which this will be true. It is the balance of the left- and right-hand sides that determines x and so we call this an equation – different from an identity when the result is always true for any value of x). Mathematicians usually consider both quantities a and b as parameters. I draw a distintion because we usually solve the system many times with different choices for b – hence the term input. For example, we may be manufactoring scales with different units, such as pounds or kilograms, but using the same type of spring. The situation fits case 3 where $x = e$, $a = k$ and $b = W$ – see (1.3).

Why do we call the equation linear? Basically, because the unknown x appears just as itself. By that I mean it doesn't appear as x^2 or $\exp(x)$ or some other function of x. Because x appears by itself in (1.4), we can calculate the solution by simply dividing (1.4) by a. Thus

$$x = \frac{b}{a}\,. \tag{1.5}$$

On the other hand, suppose the equation is $ax^2 = b$. Then, $x^2 = b/a$, and $x = \sqrt{b/a}$. Unfortunately, we will not always have a real solution (b/a may be negative), and when we do, we have two solutions, $x = \pm\sqrt{b/a}$. An equation such as $ax^2 = b$ is called nonlinear to emphasize this important difference. Other examples of nonlinear equations include trigonometric equations such as $\cos(ax) = b$ or $a\sin(x) = b$, which will have many solutions and exponential equations such as $\exp(ax) = b$, which has only one solution. Notice that in these examples of nonlinear equations, when the input b is doubled the solution x is not doubled. In other words, there is no proportionality between input and output.

Since the solution to (1.4) is (1.5), we seem to know everything about a single linear equation. But wait! When we divided (1.4) by a we unconsciously assumed that $a \neq 0$. Instead of (1.5) we should have

$$x = \frac{b}{a}, \quad \text{provided } a \neq 0\,. \tag{1.6}$$

I call such a restriction a warning flag – it alerts me to potential problems. It is also a part of good computer programming. Before attempting division, taking square roots, etc., we should always test to see whether the arguments satisfy the necessary requirements.

If we regard (1.4) as the replacement for (1.3), then it would appear unlikely that the spring constant $k = 0$lb/ft. Or stated a different way, "What physical interpretation can we give to $k = 0$lb/ft?" To explore this question, let's imagine that k is very, very small, $k = 10^{-1000}$lb/ft for example. If we attach a weight $W = 1$lb, then the extension would be $e = 10^{1000}$ft, a very long distance. But if the weight is only $W = 10^{-1000}$lb, then $e = 1$ft, which is reasonable. After a moment's thought we may conclude that if $W \neq 0$, then $e \to \infty$ as $k \to 0$ – the solution doesn't exist! On the other, if $W \to 0$ as well, then e is undetermined – W/k takes the form $0/0$. Note that when $k = W = 0$ (see footnote[1]), the original equation (1.3) reads

$$0 \times e = 0\,, \tag{1.7}$$

which is obviously true no matter what value for e. All of this means that we must replace (1.6) by

$$x = \begin{cases} \dfrac{b}{a} & \text{provided } a \neq 0, \\ \text{doesn't exist} & \text{if } a = 0 \text{ but } b \neq 0, \\ \text{any value} & \text{if } a = b = 0. \end{cases} \tag{1.8}$$

Comments:

1. As we shall see, the result (1.8) carries all the possibilities that can occur for systems of equations. There are always three possibilities: solutions, no solutions, undetermined solutions. Of course, the restrictions are more complicated for systems of equations, but the essence of the ideas is the same.

[1] This is simply a shorthand way of expressing $k = 0$ and $W = 0$. Mathematicians strive continually to find ways to write ideas in ever more compact ways. If you don't understand the expression, ask the mathematician to explain. It is their responsibility to express ideas in a clear and understandable form.

2. Some aspects of (1.8) may seem rather silly. After all, who would write down equation (1.7). The reason it is important involves the nature of computer arithmetic. In short, a number such as 10^{-1000} cannot be represented on a computer. It simply stores the number as 0 which means that if we specify $k = W = 10^{-1000}$, the computer would store the variables as $k = W = 0$ and evaluate e according to (1.8) as undetermined. We must not only understand precisely the true mathematically result – in this case (1.8), but we must also understand how the computer performs arithmetic. The nature of computer arithmetic may lead to cases that we may not otherwise anticipate.

Surprising as it may seem, we have covered all the important aspects to solving systems of linear equations. What remains is to uncover the details.

2 Two Linear Equations

This section provides a stepping stone from the basic ideas expressed in the Introduction to the more general case of many linear equations. The goal is to find the appropriate form of (1.8) for two equations. Once done, the extension to more equations is relatively straightforward. Of course, mathematical models for mechanical devices or various processes typically involve many equations, but the way they arise can be understood from examples with just two equations, another advantage to taking this stepping stone.

Equations arise when we seek to balance different quantities. The simplest example is the conservation of a quantity such as mass. The flow of water from a tap into a hose pipe must be the same as the flow of water through the nozzle. The number of atoms must remain the same before and after a chemical reaction. In biology, the population of a species is controlled by the number of births and deaths. But Newton's Laws provide perhaps the most famous example of balances: all the forces on a body that is at rest,or moving uniformly, must cancel – sum up to zero. Let's use this principle to consider the balance of forces on a body suspended by two wires from two supports. Figure 2 provides a schematic. The objective is to adjust the length of the wires by winches so that the the body is in a certain position marked by the distances l and h. What needs to be checked is that the tension in the wires does not exceed a critical value T_c, otherwise the wires will snap.

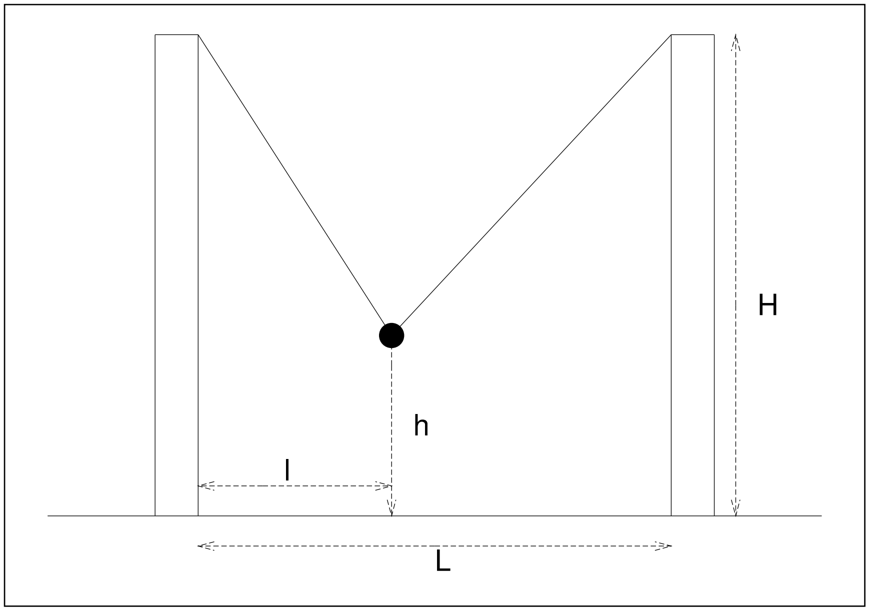


Figure 2: Schematic of a body suspended by two wires

The first part of the problem then is to determine the length of the wires. Replace Figure 2 with a simple geometric drawing Figure 3 that aids in determining the lengths L_1 and L_2. Introducing the angles θ_1 and θ_2 with θ_1 measured as positive in the clockwise direction and θ_2 measured positive in the anti-clockwise direction, geometrical considerations lead to the following relationships:

$$L_1 \cos(\theta_1) = l\,, \qquad L_1 \sin(\theta_1) = H - h\,, \tag{2.1a}$$

$$L_2 \cos(\theta_2) = L - l\,, \qquad L_2 \sin(\theta_2) = H - h\,. \tag{2.1b}$$

The parameters L and H are associated with the particular design of the device. They are considered fixed in value. As the body moves to different locations, l and h will have different values. They are the inputs. The system of four equations (2.1) determine L_1, L_2, θ_1 and θ_2. They are the unknowns. The four equations are nonlinear because two of the unknowns appear as arguments of trigonometric functions and, moreover, they appear as products with the other unknowns. Generally, nonlinear equations require special tricks to solve them. Fortunately, the equations in (2.1a) can be

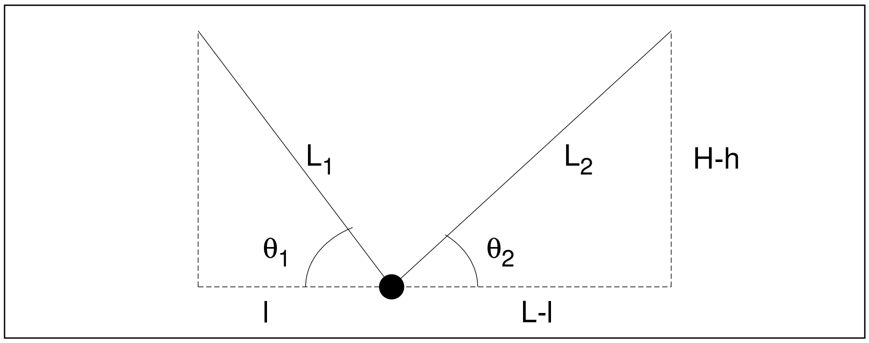


Figure 3: Geometrical Sketch

replaced with

$$L_1^2 = l^2 + (H-h)^2 \,, \qquad \tan(\theta_1) = \frac{H-h}{l} \,, \tag{2.1c}$$

and (2.1b) with

$$L_2^2 = (L-l)^2 + (H-h)^2 \,, \qquad \tan(\theta_2) = \frac{H-h}{L-l} \,. \tag{2.1d}$$

These new equations allow us to determine the unknowns very easily by simply taking square roots and inverse tangents.

The balance of forces will determine the tensions T_1 and T_2 in the two wires. The tension in the wires act along the wires as shown in the force diagram Figure 4. Obviously, the weight of the body acts downwards – we neglect the weight of the wires. Since the body is at rest, the net force on it must vanish. The horizontal and vertical components of the force lead to the two equations

$$T_1 \cos(\theta_1) - T_2 \cos(\theta_2) = 0 \,, \tag{2.2a}$$

$$T_1 \sin(\theta_1) + T_2 \sin(\theta_2) = W \,. \tag{2.2b}$$

Since θ_1 and θ_2 have been determined by(2.1), these two equation determine the unknowns T_1 and T_2 once the input W is specified. We follow standard practice by writing the terms with the unknowns on the left hand side of

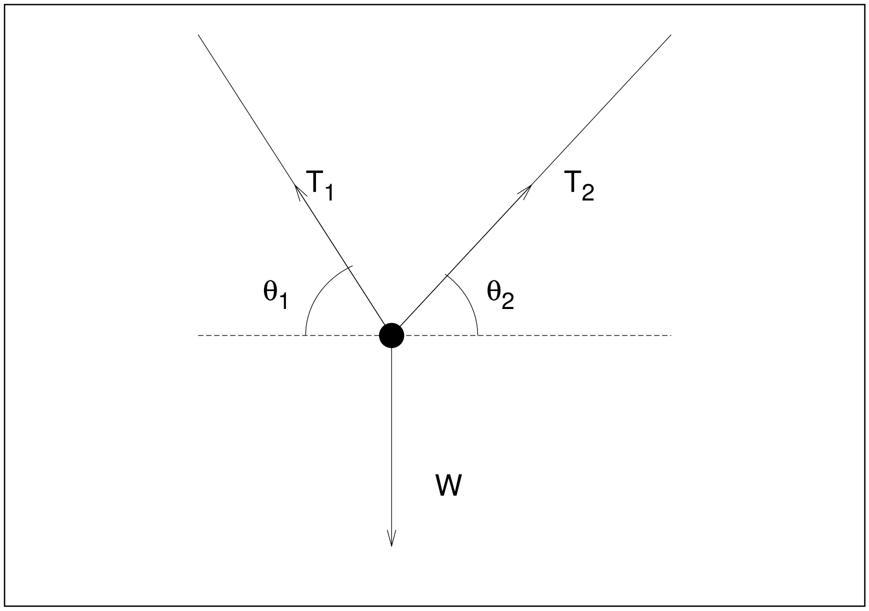


Figure 4: Force Diagram

the equations and the terms with the knowns (or inputs) on the right. Fortunately, we may use (2.1a) and (2.1b) to express the coefficents of (2.2) directly in terms of the physical parameters:

$$\frac{l}{L_1} T_1 - \frac{L-l}{L_2} T_2 = 0\,, \tag{2.3a}$$

$$\frac{H-h}{L_1} T_1 + \frac{H-h}{L_2} T_2 = W\,. \tag{2.3b}$$

We have a system of two equations for the two unknowns T_1 and T_2. We consider the location of the supports as fixed and the weight W is specified. Our interest, then, is knowing the tensions as we vary the location of the weight. That means h and l are the inputs. We must solve the system for any choice of these inputs.

There are several ways the equations (2.3) can be solved. The way adopted here is a forerunner of the computer algorithm that is currently in use. Multiply (2.3a) by $(H-h)/L_1$ and (2.3b) by l/L_1:

$$\frac{(H-h)l}{L_1^2}\,T_1 - \frac{L-l)(H-h)}{L_1L_2}\,T_2 = 0\,, \tag{2.4a}$$

$$\frac{H-h)l}{L_1^2}\,T_1 + \frac{(H-h)l}{L_1L_2}\,T_2 = \frac{l}{L_1}\,W\,. \tag{2.4b}$$

The point of these multiplications is to produce the same expression multiplying T_1 in both equations. Now when the equations are subtracted the result is

$$\left[\frac{(H-h)l}{L_1L_2} + \frac{(H-h)(L-l)}{L_1L_2}\right] T_2 = \frac{l}{L_1}\,W\,, \tag{2.5}$$

and the unknown T_1 is absent from the new equation (2.5). All that needs to be done is a division by the expression in brackets and T_2 will be known:

$$T_2 = \frac{L_2 l}{(H-h)L}\,W\,, \quad \text{provided } h \neq H\,. \tag{2.6a}$$

Notice the warning flag! The astute observer will immediately notice the similarity between (2.6a) and (1.6). The solution strategy eliminates one of the unknowns leaving a single equation in a single unknown. All the considerations in solving (1.4) now apply to the solution of (2.5). Before discussing the implications when $h = H$, let's complete the solution.

Now that T_2 is unknown – it can be directly evaluated when values are given for the parameters l and h and a value is given for W, what about T_1? We can return to (2.3) and follow the strategy that would eliminate T_1 instead of T_2. That means multiplying (2.3a) by $(H-h)/L_2$ and (2.3b) by $(L-l)/L_2$ and then adding. Complete the details yourself to see that the same result is obtained. Alternatively, T_2 can be substituted into either (2.3a) or (2.3b). Work is minimized if we use (2.3a), for then

$$\begin{aligned} T_1 &= \frac{(L-l)L_1}{L_2 l}\,T_2\,, \\ &= \frac{(L-l)L_1}{(H-h)L}\,W\,. \end{aligned} \tag{2.6b}$$

What happens when $h = H$? The weight lies somewhere on the horizontal line connecting the tops of the support. Imagine trying to pull the weight up towards height H. More and more tension will be needed to raise it. The solutions (2.6) show that T_1 and T_2 become larger and larger as $h \to H$. Ultimately we would need an infinite tension to raise the weight all the way to the top. Of course, the wires would snap before then. On the other hand, we can raise the weight higher if it is lighter. The lighter and lighter the weight, the higher we can raise it. If there is no weight ($W = 0$) we can pull the wires taut across the opening – we neglected the weight of the wires. Even so, the tensions must balance $T_1 = T_2$.

To summarize these results we need a simple way to represent the solution. The common choice is to write both unknowns in a column and place brackets around them,

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} . \tag{2.7}$$

To specify the solution then, we provide a column with the results:

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} \dfrac{(L-l)L_1}{(H-h)L} W \\ \dfrac{L_2 l}{(H-h)L} W \end{pmatrix} .$$

Unfortunately the right hand column is difficult to read, and ugly! But we do notice each entry has a common factor

$$\alpha = \frac{W}{(H-h)L} .$$

So introduce the convention that a common factor to all entries can be written outside the brackets. In other words,

$$\begin{pmatrix} \alpha(L-l)L_1 \\ \alpha L_2 l \end{pmatrix} = \alpha \begin{pmatrix} (L-l)L_1 \\ L_2 l \end{pmatrix} . \tag{2.8}$$

Using this convention we may write the solution as

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \frac{W}{(H-h)L} \begin{pmatrix} (L-l)L_1 \\ L_2 l \end{pmatrix} , \qquad \text{provided } h \neq H . \tag{2.9}$$

Let's be clear about what this notation means. If you want to evaluate T_1, then you take the corresponding column entry in the right hand side of (2.9) and multiply it by the factor in front. Thus we obtain (2.6a). T_2 can be evaluated similarly.

Comments:

1. You may recognize that (2.7) is a vector written as a column with its components in order.

2. You may recognize that (2.8) is a statement about multiplying a vector by a number. Each component of the vector must be multiplied by that number.

3. When two vectors are made to balance with an equal sign, they must have the same number of components (number of entries in the column), and each column entry must balance its counterpart.

4. The warning flag posted in (2.9) is similar to the warning flag posted in the first entry of (1.8).

Let's now consider what happens when $h = H$. There are two possibilities. Either $W \neq 0$ and the factor in (2.9) becomes infinite and

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \text{doesn't exist}\,, \tag{2.10}$$

or $W = 0$ and the solution is undetermined because the factor in (2.9) has the form $0/0$ and cannot be evaluated. Let's examine the latter case more carefully. Set $h = H$ which gives $L_1 = l$ and $L_2 = L - l$. Then (2.3) takes the form

$$T_1 - T_2 = 0 \tag{2.11a}$$

$$T_1 \times 0 + T_2 \times 0 = 0\,. \tag{2.11b}$$

The important point is that (2.11b) provides no information – it is just like (1.7) – and that is the cause for the indeterminacy. On the other hand, (2.11a) tells us that $T_1 = T_2$. Recall that if the body has no weight then we can pull the wires taut so that the body is at height H, but the tensions must still balance. Since we can pull on the wires as much as we like (without snapping the wires), we are free to specify how much tension we want. Let us designate this choice as C. In other words, $T_1 = T_2 = C$ and we have expressed the unknowns in terms of a new parameter C:

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} C \\ C \end{pmatrix} = C \begin{pmatrix} 1 \\ 1 \end{pmatrix}\,. \tag{2.12}$$

When the above results are stated more concisely the similarity with (1.8) is more obvious.

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{cases} (2.9) & \text{provided } h \neq H\,, \\ \text{doesn't exist} & \text{if } h = H \text{ but } W \neq 0\,, \\ (2.12) & \text{if } h = H \text{ and } W = 0\,. \end{cases} \tag{2.13}$$

The main difference occurs in the third case where the solution is not just any value but any multiple of a certain vector. In other words, the solution it is not uniquely determined. Instead, there are many possible solutions depending on the choice for C.

The specific example of a suspended weight has served to illustrate all the possibilities for solutions to two equations with two unknowns. It is time to repeat the solution strategy for the general case with a view to identifying a successful algorithm that can be coded for general use.

First, introduce a notation that covers the general case. Unknowns are usually represented by x, but we have two unknowns. Rather than introduce another symbol (we'll some run out of symbols if we continue this way), we introduce a subscript and consider the two unknowns to be x_1 and x_2. This is also done in our example where T_1 and T_2 are the unknowns. Now we must represent the coefficients of the unknowns in the two equations. While for two equations with two unknowns there are only four coefficients, it is best to pick a notation that can be extended to many equations with many unknowns. We choose a single symbol, a say, but with subscripts. How many subscripts? Well, we want to know to which equation the coefficients refer, and to which unknown they are attached. So two subscripts! Let the first subscript refer to the equation number and the second subscript refer to the unknown to which it is attached. So a_{ij} will refer to the coefficient multiplying the unknown x_j in the i-th equation. A system of two equations with two unknowns takes the general form

$$a_{11}x_1 + a_{12}x_2 = b_1\,, \tag{2.14a}$$

$$a_{21}x_1 + a_{22}x_2 = b_2\,. \tag{2.14b}$$

The quantities b_1 and b_2 are the inhomogeneous terms in the equations and are considered to be the components of a vector. For example, $b_1 = 0$ and $b_2 = W$ in the system (2.2). Mathematicians regard all the coefficients a_{ij} and b_i as parameters. By that we mean we consider them to be some fixed numbers. As the example of the suspended weight illustrates, the coefficients may be expressions and they may arise from parameters, such as L and H, or from inputs, such as l, h and W. We imagine that choices for all parameters and inputs of the problem have been made so that the coefficients a_{ij} and b_i are all known. We now face the challenge of executing a solution strategy to determine the unknowns.

The basic ideas to solve (2.14) have already been applied to solving (2.3), but a few modifications prove useful.

Solution Strategy

1. Check whether $a_{11} = 0$. If so, see comment 1 below. If $a_{11} \neq 0$, continue.

2. Divide the first equation (2.14a) by a_{11} – this is legitimate since we have just checked that $a_{11} \neq 0$. We now think of the first equation as

$$x_1 + \frac{a_{12}}{a_{11}} x_2 = \frac{b_1}{a_{11}} . \tag{2.15}$$

3. Check whether $a_{21} \neq 0$. If it is, jump to step (5).

4. Multiply (2.15) by a_{21} and subtract the result from the second equation (2.14b). Thus

$$\begin{array}{lcllcl} a_{21}x_1 & + & a_{22}x_2 & = & b_2 \\ a_{21}x_1 & + & a_{21}a_{12}/a_{11}x_2 & = & a_{21}b_1/a_{11} \\ \hline 0 & + & (a_{22} - a_{21}a_{12}/a_{11})\, x_2 & = & b_2 - a_{21}b_1/a_{11} \end{array}$$

Of course, the point of dividing the first equation by a_{11} and then multiplying it by a_{21} is to make sure the quantity multiplying x_1 is the same in both equations. When we subtract them we eliminate x_1 as the result shows. We may rewrite the result as

$$\left(a_{22} - \frac{a_{21}}{a_{11}} a_{12} \right) x_2 = b_2 - \frac{a_{21}}{a_{11}} b_1 ,$$

or

$$\left(\frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11}} \right) x_2 = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}} . \tag{2.16}$$

5. Now identify (2.16) with (1.4) where the following connections are made.

$$a = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11}}\,, \qquad b = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}}\,, \qquad x = x_2\,.$$

   This is another example where parameters such as a and b in (1.4) can represent expressions. Alternatively, it is often useful to replace expressions by parameters!

   In essence, the strategy has been to reduce the system of equations to a single equation in a single unknown. According to (1.8), the solution is

$$x_2 = \begin{cases} \dfrac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} & \text{provided } a_{11}a_{22} - a_{12}a_{21} \neq 0, \\ \text{doesn't exist} & \text{if } a_{11}a_{22} - a_{12}a_{21} = 0 \text{ but } a_{11}b_2 - a_{21}b_1 \neq 0, \\ \text{undetermined} & \text{if } a_{11}a_{22} - a_{12}a_{21} = a_{11}b_2 - a_{21}b_1 = 0. \end{cases} \tag{2.17}$$

6. Now that x_2 is known, we return to (2.15) to find x_1. Of course, we now assume $a_{11}a_{22} - a_{12}a_{21} \neq 0$ otherwise x_2 is not uniquely determined and we are unsure how to proceed.

$$\begin{aligned} x_1 &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}\,x_2 \\ &= \frac{1}{a_{11}}\left(b_1 - a_{12}\,\frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}}\right) \\ &= \frac{b_1(a_{11}a_{22} - a_{12}a_{21}) - a_{12}(a_{11}b_2 - a_{21}b_1)}{a_{11}(a_{11}a_{22} - a_{12}a_{21})} \\ &= \frac{b_1a_{11}a_{22} - a_{12}a_{11}b_2}{a_{11}(a_{11}a_{22} - a_{12}a_{21})} \\ &= \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}}\,. \end{aligned} \tag{2.18}$$

The procedure gives all the steps to construct the solution under general circumstances because the coefficents in the equations have all been represented by parameters. The disadvantage is that the algebra is somewhat complicated and disguises the simplicity of the ideas. To show how simple the procedure really is, let's consider a special case of (2.3): $\theta_1 = \theta_2 = \pi/4$ and $W = 1$lb. This choice arises, for example, when $l = L/2$ and $h = H - L/2$ which means the object hangs in the middle at just the right height. Then,

$$\frac{T_1}{\sqrt{2}} - \frac{T_2}{\sqrt{2}} = 0\,, \tag{2.19a}$$

$$\frac{T_1}{\sqrt{2}} + \frac{T_2}{\sqrt{2}} = 1\,, \tag{2.19b}$$

and this system is (2.14) with the following identification of the symbols:

$$a_{11} = \frac{1}{\sqrt{2}}\,, \qquad a_{12} = -\frac{1}{\sqrt{2}}\,, \qquad b_1 = 0\,,$$

$$a_{21} = \frac{1}{\sqrt{2}}\,, \qquad a_{22} = \frac{1}{\sqrt{2}}\,, \qquad b_2 = 1\,.$$

Now, let's execute the solution strategy.

1. Clearly, $a_{11} \neq 0$.

2. The first equation becomes

$$T_1 - T_2 = 0\,. \tag{2.19c}$$

3. After subtraction, (2.16) will be

$$\sqrt{2}\,T_2 = 1\,. \tag{2.19d}$$

4. The solution (2.17) will be

$$T_2 = \frac{1}{\sqrt{2}}\,. \tag{2.19e}$$

   Note that $a_{11}a_{22} - a_{12}a_{21} = 1$, so the first case in (2.17) applies.

5. Finally,

$$T_1 = T_2 = \frac{1}{\sqrt{2}}\,, \tag{2.19f}$$

   which is much simpler than all the messy algebra in (2.18). As a check, substitute our choices for θ_1 and θ_2 in (2.9).

Comments:

1. What do we do if $a_{11} = 0$? We check whether $a_{21} = 0$! Suppose $a_{21} \neq 0$, then we should interchange the first and second equation – the order of the equations is usually up to us to decide. Thus we should write

$$a_{21}x_1 + a_{22}x_2 = b_2\,, \tag{2.20a}$$
$$0x_1 + a_{12}x_2 = b_1\,. \tag{2.20b}$$

   At the same time we should redefine our coefficients with the following replacements:

$$a_{11} \leftarrow a_{21}\,, \qquad a_{12} \leftarrow a_{22}\,, \qquad b_1 \leftarrow b_2\,,$$
$$a_{21} = 0\,, \qquad a_{22} \leftarrow a_{12}\,, \qquad b_2 \leftarrow b_1\,.$$

   Now we can continue the procedure at step 2.

2. If $a_{11} = 0$ and $a_{21} = 0$, then we have the situation that x_1 doesn't appear in the equations and (2.14) takes the form

$$0x_1 + a_{12}x_2 = b_1\,, \tag{2.21a}$$
$$0x_1 + a_{22}x_2 = b_2\,. \tag{2.21b}$$

   Not only doesn't x_1 appear in the equations but there are two equations to determine only one unknown x_2. There are several possibilities:

   (a) $a_{12} \neq 0$ and $a_{22} \neq 0$, then we appear to have two different solutions

$$x_2 = \frac{b_1}{a_{12}} \quad \text{and} \quad x_2 = \frac{b_2}{a_{22}}\,, \tag{2.22}$$

   which means no solution exists – both cannot be true – unless we are lucky and $b_1/a_{12} = b_2/a_{22}$ and both results agree – the result is the solution.

   (b) if either $a_{12} = 0$ or $a_{22} = 0$, then there is no solution unless $b_1 = 0$ or $b_2 = 0$ respectively. To illustrate what happens, suppose $a_{12} = b_1 = 0$, but $a_{22} \neq 0$. Then the first equation is just $0 = 0$ and there never really was a first equation and all we really have is $a_{22}x_2 = b_2$ with the obvious solution

$$x_2 = \frac{b_2}{a_{22}}\,, \qquad \text{provided } a_{22} \neq 0\,. \tag{2.23}$$

Usually we can see what's happening by just looking at the equations. Unfortunately, computers cannot just look at equations. To design a safe algorithm, we must plot a systematic course through the possibilities. So here is the procedure:

(a) Check whether $a_{12} = 0$. If it is not, then divide the first equation in (2.21a) by a_{12} to obtain

$$0x_1 + x_2 = \frac{b_1}{a_{12}}\,, \tag{2.24a}$$

$$0x_1 + a_{22}x_2 = b_2\,. \tag{2.24b}$$

Although (2.24a) appears to give the solution for x_2, we must still deal with (2.24b).

(b) Check whether $a_{22} = 0$. If it is, then there is no solution when $b_2 \neq 0$. If $b_2 = 0$ also, then the last equation reads as $0 = 0$ and may be ignored. In other words, x_2 has the solution given in (2.24a).

(c) If $a_{22} \neq 0$, then multiply (2.24a) by a_{22} and subtract it from (2.24b) with the result

$$0x_1 + x_2 = \frac{b_1}{a_{12}}\,, \tag{2.25a}$$

$$0x_1 + 0x_2 = b_2 - \frac{a_{22}b_1}{a_{12}} = \frac{a_{12}b_2 - a_{22}b_1}{a_{12}}\,. \tag{2.25b}$$

The last equation (2.25b) makes no sense unless $a_{12}b_2 - a_{22}b_1 = 0$. In other words, there is no solution if $a_{12}b_2 - a_{22}b_1 \neq 0$. When $a_{12}b_2 - a_{22}b_1 = 0$ there are many solutions where x_2 is given by (2.25a) but x_1 can be any value and becomes a free parameter to be selected.

3. The expression $a_{11}a_{22} - a_{12}a_{21}$ plays a prominent role in the solution (2.17) and (2.18), and is called the determinant. Of course, we want a shorthand icon to represent this expression and the standard choice is $\det(A)$. The abbreviation "det" is obvious but what is A? Well, A stands for the four quantites a_{11}, a_{12}, a_{21} and a_{22}. we pick the capital letter that corresponds to the lowercase symbol used to denote each entry. In short,

$$\det(A) \equiv a_{11}a_{22} - a_{12}a_{21}\,. \tag{2.26}$$

Notice that instead of an equals sign we have used the equivalence sign. That's because (2.24) is *not* an equation, but a definition. When we write $\det(A)$, all we mean is the expression $a_{11}a_{22} - a_{12}a_{21}$. Using this notation, we may write the solution (2.17) and (2.18) as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} a_{22}b_1 - a_{12}b_2 \\ a_{11}b_2 - a_{21}b_1 \end{pmatrix}, \qquad \text{provided } \det(A) \neq 0\,. \quad (2.27)$$

4. The statement (2.27) is correct, but the information is incomplete. What it states is that when $\det(A) \neq 0$, there is a single solution given in (2.25). It does not say that when $\det(A) = 0$ there is no solution. That is just one possibility. As (2.17) shows, there is another possibility. When in addition $a_{11}b_2 - a_{21}b_1 = 0$, then x_2 is undetermined and can be any value. That means we are free to convert it to the status of a parameter. Suppose we reset $x_2 = C$. Then (2.15) must be used to determine x_1:

   $$x_1 = \frac{b_1 - a_{12}C}{a_{11}}\,.$$

   So the solution takes the form

   $$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{a_{11}} \begin{pmatrix} b_1 - a_{12}C \\ C \end{pmatrix}, \qquad \text{provided } a_{11} \neq 0\,. \qquad (2.28)$$

   The key observation now is that there are many solutions, depending on the value we prescribe to C. We say that the solution is not unique since it depends on the choice for C, whereas the solution (2.25) is unique because it is completely determined. An example of non-unique solutions is given in the example of a suspended weight, see (2.12).

At this stage, there seems to be a confusing list of possibilities: unique solutions, no solutions or many solutions. Unfortunately, this is so. Fortunately, there is a better way to organize the information so that these possibilities are easy to identify.

3 Matrix Version of Two Linear Equations

By looking at (2.14) and then at the solutions (2.17) and (2.18), it is clear that all manupilations involve only the coefficients a_{ij} and the "inhomogeneous terms" b_i. Arrange the coefficients in a table

$$A \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} . \tag{3.1}$$

Notice that the subscripts correspond to a certain pattern. The first subscript gives the row location ansd is the same as equation number. The second subscript identifies a column number and is the same as the unknown subscript. At the same time, the symbol A is introduced to represent this table. Since the solution and the vector of inhomogeneous terms have already been expressed as column vectors $\mathbf{x}$ and $\mathbf{b}$ respectively, it is now tempting to write the system of equations (2.14) in the form

$$A\mathbf{x} = \mathbf{b} , \tag{3.2a}$$

so that it has the appearance of (1.4). This shorthand notation corresponds to

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} . \tag{3.2b}$$

Since this arrangement of symbols must stand for the system of equations, they must obey the rule

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} . \tag{3.2c}$$

A table that satisfies this rule (3.2c) is called a matrix (plural matrices) to distinguish it from just any table. So (3.2c) tells us how to take a matrix and multiply it with a vector. The result is a vector. There is a simple pattern to this rule. To obtain the i-th component of the result, take the coefficients along the i-th row of the matrix, multiply them with the corresponding component of $\mathbf{x}$ and sum the products.

There are many useful and interesting properties of matrices which will be discussed in another module, but for now let's use a matrix representation to see how the solution strategy outlined after (2.14) is applied to solve (3.2). We start by writing down the matrix and the right-hand side vector

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} , \qquad \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} . \tag{3.3a}$$

Assume $a_{11} \neq 0$ which means we may proceed to the second step. After the second step, the first equation becomes (2.15) and the matrix and right-hand side vector for the new system becomes

$$\begin{bmatrix} 1 & \dfrac{a_{12}}{a_{11}} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \begin{pmatrix} \dfrac{b_1}{a_{11}} \\ b_2 \end{pmatrix}. \tag{3.3b}$$

Clearly what has happended is that the first row of both the matrix and the right-hand side vector have been divided by a_{11}. This step illustrates the principle that whatever we do to the equations, we do to the rows of the matrix and the inhomogeneous terms in the right-hand side vector. We may as well then combine the forcing vector and the matrix into one big table, called the augmented matrix

$$\begin{bmatrix} 1 & \dfrac{a_{12}}{a_{11}} & \vdots & \dfrac{b_1}{a_{11}} \\ a_{21} & a_{22} & \vdots & b_2 \end{bmatrix}. \tag{3.3c}$$

The vertical dots serve to remind us that the entries on the left are part of the coefficient matrix while those on the right are the components of the right-hand side vector. Suppose $a_{21} \neq 0$ and proceed to the fourth step. Effectively we multiply each entry in the first row by a_{21} and subtract the result from the entry below in the second row which produces

$$\begin{bmatrix} 1 & \dfrac{a_{12}}{a_{11}} & \vdots & \dfrac{b_1}{a_{11}} \\ 0 & a_{22} - \dfrac{a_{21}}{a_{11}}\, a_{12} & \vdots & b_2 - \dfrac{a_{21}}{a_{11}}\, b_1 \end{bmatrix}. \tag{3.3d}$$

This step is equivalent to replacing the second row of (3.3c) by the result in (2.16). The fifth step requires dividing the entries in the second row by the quantity

$$a_{22} - \frac{a_{21}}{a_{11}}\, a_{12}$$

with the result

$$\begin{bmatrix} 1 & \dfrac{a_{12}}{a_{11}} & \vdots & \dfrac{b_1}{a_{11}} \\ 0 & 1 & \vdots & \dfrac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \end{bmatrix}. \tag{3.3e}$$

To emphasize the use of the notation, let's unravel what the augmented matrix in (3.3e) really means.

$$\begin{bmatrix} 1 & \dfrac{a_{12}}{a_{11}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + \dfrac{a_{12}}{a_{11}} x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} \dfrac{b_1}{a_{11}} \\ \dfrac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \end{pmatrix} . \tag{3.3f}$$

Clearly the second row is giving us the solution for x_2! And notice that it agrees with the solution given in (2.17).

The final step, step 6, substitutes the solution for x_2 into the first equation and then solves for x_1 (2.18). Unfortunately, the procedure messes up the structure of the augmented matrix. Instead, multiply the second row by a_{12}/a_{11} and subtract it from the first row. The result is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \dfrac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} \\ \dfrac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \end{pmatrix} , \tag{3.3g}$$

and now both components of the solution are obvious. Returning to the augmented form

$$\left[\begin{array}{ccc} 1 & 0 & \vdots & \dfrac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} \\ 0 & 1 & \vdots & \dfrac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \end{array}\right] , \tag{3.3h}$$

we see that the components of the solution have replaced the components of the $\mathbf{b}$. The last column gives the solution.

Comments:

1. The augmented matrix and the row operations that produce the solution are easily coded on a computer.

2. The matrix

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.4}$$

is called the identity matrix. It has the important property that $I\mathbf{x} = \mathbf{x}$ for any vector $\mathbf{x}$.

3. We may use the rule for matrix multiplication with a vector to write the result (3.3g) in the form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \dfrac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} \\ \dfrac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \end{pmatrix} = \begin{bmatrix} \dfrac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}} & -\dfrac{a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \\ -\dfrac{a_{21}}{a_{11}a_{22} - a_{12}a_{21}} & \dfrac{a_{11}}{a_{11}a_{22} - a_{12}a_{21}} \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} . \tag{3.5a}$$

The easiest way to see the how this product is found is by multiplying the matrix with $\mathbf{b}$. Let's represent the matrix in (3.5a) by the symbol B. So

$$\mathbf{x} = B\mathbf{b}\,, \tag{3.5b}$$

where

$$B = \begin{bmatrix} \dfrac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}} & -\dfrac{a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \\ -\dfrac{a_{21}}{a_{11}a_{22} - a_{12}a_{21}} & \dfrac{a_{11}}{a_{11}a_{22} - a_{12}a_{21}} \end{bmatrix} \tag{3.5c}$$

As noticed before – see (2.25), $\det(A) = a_{11}a_{22} - a_{12}a_{21}$ is a common denominator. Inspired by the definition (2.8) of the multiplication of a vector by a number, we define the multiplication of a matrix by a number α as

$$\alpha \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \equiv \begin{bmatrix} \alpha a_{11} & \alpha a_{12} \\ \alpha a_{21} & \alpha a_{22} \end{bmatrix} . \tag{3.5d}$$

To multiply a matrix by a number, you must multiply every entry by that number. Using this convention,

$$B = \frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ a_{21} & a_{11} \end{bmatrix} . \tag{3.5e}$$

4. Note that there is a natural interpretation for B. Recall the solution to a single linear equation, $ax = b$, is $x = b/a$ ($a \neq 0$). It is tempting to write $B \equiv 1/A$, so that $\mathbf{x} = \mathbf{b}/A$. However, this way of writing the result is misleading because it suggests we take a vector and divide it somehow from the right, whereas what we do is multiply $\mathbf{b}$ from the

left with the matrix B. We could write the result rather as $\mathbf{x} = (1/A)\mathbf{b}$ with a clearer meaning. We go one step further and write $B = A^{-1}$ borrowing from the identity $1/a = a^{-1}$ valid for numbers. So $\mathbf{x} = A^{-1}\mathbf{b}$ is the standard way to write the solution in matrix form. If we substitute the solution into the equation $A\mathbf{x} = \mathbf{b}$, we find $A(A^{-1}\mathbf{b}) = \mathbf{b}$, which means AA^{-1} is the same as I, the identity matrix. We call the matrix A^{-1} the inverse of A.

Time for some examples. Consider the system of equations

$$2x_1 + 4x_2 = 6\,, \tag{3.6a}$$
$$4x_1 + 4x_2 = 8\,, \tag{3.6b}$$

which leads to the augmented matrix

$$\begin{bmatrix} 2 & 4 & \vdots & 6 \\ 4 & 4 & \vdots & 8 \end{bmatrix}. \tag{3.7a}$$

Since $a_{11} = 2 \neq 0$, we pass by step 1 of the solution strategy and proceed to step 2. Divide the first row by 2:

$$\begin{bmatrix} 1 & 2 & \vdots & 3 \\ 4 & 4 & \vdots & 8 \end{bmatrix}. \tag{3.7b}$$

Now multiply the first row by 4 and subtract it from the second row (step 4, since $a_{21} \neq 0$):

$$\begin{bmatrix} 1 & 2 & \vdots & 3 \\ 0 & -4 & \vdots & -4 \end{bmatrix}. \tag{3.7c}$$

Divide the second row by -4 (step 5):

$$\begin{bmatrix} 1 & 2 & \vdots & 3 \\ 0 & 1 & \vdots & 1 \end{bmatrix}. \tag{3.7d}$$

Finally, multiply the second row by 2 and subtract from the first row (replacement step 6 – see discussion before (3.3g)):

$$\begin{bmatrix} 1 & 0 & \vdots & 1 \\ 0 & 1 & \vdots & 1 \end{bmatrix}. \tag{3.7e}$$

The solution to the system (3.6) is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} . \tag{3.7f}$$

In this example, we did not run into any of the flags. Both conditions, $a_{11} \neq 0$ and $\det(A) \neq 0$, were satisfied. Now, through a series of further examples, let's see how to treat the augmented matrix when these conditions are not satisfied.

Case 1. Consider the augmented matrix

$$\begin{bmatrix} 0 & 4 & \vdots & 8 \\ 2 & 1 & \vdots & 4 \end{bmatrix} . \tag{3.8a}$$

Since $a_{11} = 0$, we simply interchange the rows – see comment with (2.20)

$$\begin{bmatrix} 2 & 1 & \vdots & 4 \\ 0 & 4 & \vdots & 8 \end{bmatrix} , \tag{3.8b}$$

and now the new a_{11} is not zero. Divide the first row by 2 (step 2):

$$\begin{bmatrix} 1 & \frac{1}{2} & \vdots & 2 \\ 0 & 4 & \vdots & 8 \end{bmatrix} . \tag{3.8c}$$

Since the new $a_{21} = 0$, jump to step 5 and divide the second row by 4:

$$\begin{bmatrix} 1 & \frac{1}{2} & \vdots & 2 \\ 0 & 1 & \vdots & 2 \end{bmatrix} . \tag{3.8d}$$

Finally, multiply the second row by a half and subtract from the first row (modified step 6):

$$\begin{bmatrix} 1 & 0 & \vdots & 1 \\ 0 & 1 & \vdots & 2 \end{bmatrix} , \tag{3.8e}$$

and clearly the solution is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} . \tag{3.8f}$$

Case 2. Consider the augmented system

$$\begin{bmatrix} 2 & 4 & \vdots & 6 \\ 3 & 6 & \vdots & 6 \end{bmatrix}. \tag{3.9a}$$

Divide the first equation by 2 (step 2, since $a_{11} = 2 \neq 0$):

$$\begin{bmatrix} 1 & 2 & \vdots & 3 \\ 3 & 6 & \vdots & 6 \end{bmatrix}. \tag{3.9b}$$

Since $a_{21} = 3 \neq 0$, proceed to step 4. Subtract three times the first row from the second row:

$$\begin{bmatrix} 1 & 2 & \vdots & 3 \\ 0 & 0 & \vdots & -3 \end{bmatrix}. \tag{3.9c}$$

Disaster! The equation corresponding to the second row is

$$0x_1 + 0x_2 = -3\,, \tag{3.9d}$$

and clearly there is no solution.

Case 3. If, by some lucky miracle, the right-hand vector is

$$\mathbf{b} = \begin{pmatrix} 6 \\ 9 \end{pmatrix}, \tag{3.10a}$$

the augmented system (3.9c) is replaced by

$$\begin{bmatrix} 1 & 2 & \vdots & 3 \\ 0 & 0 & \vdots & 0 \end{bmatrix}. \tag{3.10b}$$

The last row just gives an equation that reads $0 = 0$; there is no equation. How did this happen? If we write down the original system with the new right-hand side vector,

$$\begin{bmatrix} 2 & 4 & \vdots & 6 \\ 3 & 6 & \vdots & 9 \end{bmatrix}. \tag{3.10c}$$

we notice that the second equation is just a factor 3/2 of the first equation. The system contains just one equation, expressed twice. One equation can't determine two unknowns. This situation can only arise if the right-hand side vector is just right. If it is not, as in (3.9a), then there is no solution.

Given the system (3.10b), how do we proceed? One way is to view the lack of information in the second row as a statement that x_2 is undetermined and that it must be given rather than be viewed as an unknown. We shift the status of x_2 from an unknown to a parameter by setting $x_2 = C$. Then we use the first equation to express x_1 in terms of the parameter C

$$x_1 = 3 - 2C\,. \tag{3.10d}$$

While we can't completely determine the solution, we can express the solution in a specific form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 - 2C \\ C \end{pmatrix}\,. \tag{3.10e}$$

It is time to introduce the idea of adding (or subtracting) vectors. Suppose there are two vectors, $\mathbf{x}$ and $\mathbf{y}$. Their sum is just the addition of their components,

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \equiv \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}\,. \tag{3.11}$$

In another module there is a detailed discussion about vectors and their properties. For now we just use (3.11) to rewrite (3.10e) as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix} + C \begin{pmatrix} -2 \\ 1 \end{pmatrix}\,. \tag{3.12}$$

The solution is given as a distinct vector plus an arbitrary multiple of another vector. The idea it expresses is that there is a particular solution (the distinct vector) to which must be added any multiple of the homogeneous solution (the other vector). This idea will be fully exploited in another module.

Let's complete our study of two equations in two unknowns by reviewing all the possible outcomes of our elimination strategy. There are three different results:

$$\underset{\text{case 1}}{\begin{bmatrix} 1 & 0 & \vdots & X \\ 0 & 1 & \vdots & X \end{bmatrix}}\,, \quad \underset{\text{case 2}}{\begin{bmatrix} 1 & X & \vdots & X \\ 0 & 0 & \vdots & X \end{bmatrix}}\,, \quad \underset{\text{case 3}}{\begin{bmatrix} 1 & X & \vdots & X \\ 0 & 0 & \vdots & 0 \end{bmatrix}}\,. \tag{3.13}$$

where X is used to represent some number, not necessarily the same. Case 1 comes from (3.8e) and represents the successful determination of a unique solution. Notice that the procedure works for all possible right-hand side vectors $\mathbf{b}$, and this result reflects the fact that $\det(A) = -8$ which is not zero. Case 2 comes from (3.9c). Whenever we find a row of zeros, there is no solution unless there is a corresponding zero on the right-hand side. Finally, case 3 comes from (3.10b). Because the row of zeroes has a corresponding zero in the right-hand side, we have "lost" an equation. One consequence is the existence of many solutions.

All three cases constitute what is called the "reduced row echelon form" – abbreviated as rref. We have reduced the system to as simple a form as we can, and in this form we know everything about the system and its solution.

The results in (3.13) are the generalization of (1.8). The same three possibilities occur in general systems of linear equations, but they may arise in many more different ways.

4 General Systems

In science and engineering today, systems of equations are often large. Two parameters characterize such systems: m is the number of unknowns which are ordered as though they are the components x_i of a vector $\mathbf{x}$ and n is the number of equations which are taken to be linear. Our objective, then, is to construct solutions to these equations, and as the previous sections have demonstrated there may be unique solutions, no solutions, or many solutions. Our solution strategy must allow for all these possibilities. We already have in place all the ideas we need to achieve this objective. So the real purpose of this section is to apply these ideas to general systems of linear equations.

There are several ways we may represent systems of linear equations mathematically. The obvious way is to indicate a pattern

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \,, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \,, \\ &\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n \,. \end{aligned} \tag{4.1}$$

While the pattern is clear, it is rather detailed and probably unnecessarily so. By using the sum symbol, we can also write (4.1) as

$$\sum_{j=1}^{m} a_{ij}x_j = b_i\,, \qquad \text{for } i = 1, 2, \ldots n\,. \tag{4.2}$$

Finally, we can use the matrix notation we introduced in the previous section 3. Define

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_m \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}. \tag{4.3a}$$

Then,

$$A\mathbf{x} = \mathbf{b}\,, \tag{4.3b}$$

which is an extremely compact way to represent the system. It does have the disadvantage that unless (4.3a) is also given the parameters n and m are not revealed. Instead, we can specify n and m in a separate way. We call A a matrix and indicate its size by the symbol $R^{n\times m}$. What this symbol represents is the set of all matrices with entries that are real numbers – to distinguish from matrices with complex numbers – and with n rows and m columns. For (4.3b) to make sense, $\mathbf{x}$ must be a column vector with m entries, the number of columns of A, and $\mathbf{b}$ must be a column vector of n rows, the number of rows of A. For example, the system of equations

$$2x_1 - 4x_2 + 2x_3 - 6x_4 = 4\,, \tag{4.4a}$$
$$x_1 - x_3 + 5x_4 = -2\,, \tag{4.4b}$$
$$4x_1 - 4x_2 + 4x_4 = 1\,, \tag{4.4c}$$

may be written as

$$\begin{bmatrix} 2 & -4 & 2 & -6 \\ 1 & 0 & -1 & 5 \\ 4 & -4 & 0 & 4 \end{bmatrix} \begin{pmatrix} z_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix}. \tag{4.4d}$$

Just in case you didn't notice before, the coefficients with terms that are subtracted in the equations are written with negative signs. The reason is that the coefficients in (4.1) are all written as though we are adding the terms. Notice also that if an unknown doesn't appear in an equation, then its coefficient is zero – an example is the absense of a term with x_2 in the second equation (4.4b), leading to $a_{22} = 0$ (also $a_{33} = 0$).

The augmented matrix follows easily from the ideas in the previous section 3. For example, the augmented matrix for (4.4) is

$$\begin{bmatrix} 2 & -4 & 2 & -6 & \vdots & 4 \\ 1 & 0 & -1 & 5 & \vdots & -2 \\ 4 & -4 & 0 & 4 & \vdots & 1 \end{bmatrix} . \tag{4.4e}$$

Time now to describe how we solve these equations by extending the procedure given in (3.3). Rather than describing the steps in mathematical terms, I'll use common English and show the consequences on a simple example. Suppose the system of equations gives rise to the augmented matrix

$$\begin{bmatrix} 3 & -3 & 6 & \vdots & 3 \\ 2 & 0 & 10 & \vdots & 8 \\ 2 & 1 & 15 & \vdots & 13 \end{bmatrix} . \tag{4.5a}$$

1. The procedure works by starting with the first column, performing algebraic steps to make all entries zero below the top one. *First, divide the first equation by its first row entry* (making the top entry 1) – this is the same as the first step in the procedure described in (2.15) and (3.3c):

$$\begin{bmatrix} 1 & -1 & 2 & \vdots & 1 \\ 2 & 0 & 10 & \vdots & 8 \\ 2 & 1 & 15 & \vdots & 13 \end{bmatrix} . \tag{4.5b}$$

2. *Now subtract multiples of the first row from the subsequent rows so that their first row entries are zero.* This means we must subtract 2 times the first row from the second row, and 2 times the first row from the third row:

$$\begin{bmatrix} 1 & -1 & 2 & \vdots & 1 \\ 0 & 2 & 6 & \vdots & 6 \\ 0 & 3 & 11 & \vdots & 11 \end{bmatrix} . \tag{4.5c}$$

Notice what we have achieved. The only equation that contains the unknown x_1 is the first equation. The remaining equations form a system for the other unknowns. An obvious name for these remaining equations is a sub-system, and the matrix they form, a sub-matrix. In our example, the sub-matrix is

$$\begin{bmatrix} 2 & 6 & \vdots & 6 \\ 3 & 11 & \vdots & 11 \end{bmatrix} . \tag{4.5d}$$

Rather than solving (4.5d) as we do in (3.3), it is better to return to (4.5c) and continue with the strategy of elimination. The approach proves easier to handle, especially for much larger systems.

3. *Proceed to the next column and row and repeat the elimination strategy given in the first two steps above.* For our example, that means dividing the second row by 2

$$\begin{bmatrix} 1 & -1 & 2 & \vdots & 1 \\ 0 & 1 & 3 & \vdots & 3 \\ 0 & 3 & 11 & \vdots & 11 \end{bmatrix} , \tag{4.5e}$$

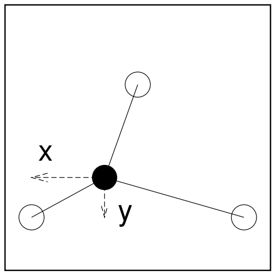
and then subtracting three times the second row from the third row, and adding the second row to the first row:

$$\begin{bmatrix} 1 & 0 & 5 & \vdots & 4 \\ 0 & 1 & 3 & \vdots & 3 \\ 0 & 0 & 2 & \vdots & 2 \end{bmatrix} . \tag{4.5f}$$

By including the eliminating of x_2 from the first row, we see the emergence of the identity matrix which allows us, when the steps are all completed, to read off the solution directly.

4. *Continue the process recursively until there are no more columns or rows.* In our example, we need to repeat the elimination process one more time; we proceed to row 3, column 3. Divide the third row by 2

$$\begin{bmatrix} 1 & 0 & 5 & \vdots & 4 \\ 0 & 1 & 3 & \vdots & 3 \\ 0 & 0 & 1 & \vdots & 1 \end{bmatrix} , \tag{4.5g}$$



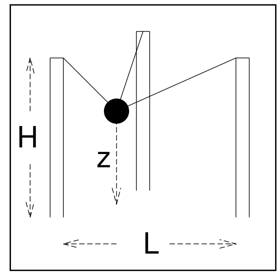


Figure 5: Top and Side Views

then subtract 3 times the third row from the second and 5 times the third row from the first to obtain

$$\begin{bmatrix} 1 & 0 & 0 & \vdots & -1 \\ 0 & 1 & 0 & \vdots & 0 \\ 0 & 0 & 1 & \vdots & 1 \end{bmatrix} . \tag{4.5h}$$

Each row now gives the solution to one of the components of $\mathbf{x}$.

The example used above was just made up of numbers to make the algebra simple, but it lacks the excitement of an appreciation for the meaning of the result. Nor does it convey the fact that we usually solve a system of equations derived from some application many times to appreciate the full range of consequences for the solutions. To illustrate these points, let us return to the example of a weight suspended between two posts as discussed in section 2. Suppose now that the weight is suspended from three posts of equal height H placed at the corners of a triangle with equal sides of length L. As before, there are winches at the top of the posts that can adjust the length of the wires that support the weight and thus move the weight through a certain volume instead of just an area when two posts are used. Figure 5 illustrates the situation.

Just as in Figures 3 and 4 the angles subtended by the wires can be used to determine the balance of forces. However, the geometry is more complicated in three-dimensions and it is easier to proceed via the use of vectors. First choose the origin, for convenience, at the foot of the leftmost post. Let the rightmost post lie on the x-axis and let the y-axis lie in the other horizontal direction perpendicular to the x-axis. The z-axis points upwards. In this choice of coordinate system we represent the location of the weight as a vector from the origin at the foot of the leftmost post to the location of the weight

$$\mathbf{x} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}\,, \tag{4.6a}$$

where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors that lie along the x- ,y- and z-axes respectively. Three additional vectors locate the tops of the posts:

$$\mathbf{x}_1 = H\mathbf{k}\,, \tag{4.6b}$$

$$\mathbf{x}_2 = L\mathbf{i} + H\mathbf{k}\,, \tag{4.6c}$$

$$\mathbf{x}_3 = \frac{1}{2}\,L\mathbf{i} + \frac{\sqrt{3}}{2}\,L\mathbf{j} + H\mathbf{k}\,. \tag{4.6d}$$

Notice that $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are vectors and not the components of a vector – we should always understand what we mean by our choice of symbols. In general, we can specify a solid structure by a list of vectors that identify all the parts of the structure.

Now turn to the balance of forces on the object. There will be three tensions directed along each of the wires, and the weight acts downwards on the object. The sum of all forces on the weight must vanish. in vector notation,

$$\mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 - W\mathbf{k} = 0 \tag{4.7}$$

where $\mathbf{T}_i$ is the tension along the wire to the top of post labelled i. The easiest way to proceed is to construct unit vectors $\mathbf{i}_i$ that lie along the wires from the body to the top of the posts labelled i. Then, $\mathbf{T}_i = T_i\mathbf{i}_i$ where T_i are just numbers giving the magnitude of the tensions. Fortunately, the construction of these unit vectors is easy. To illustrate, consider the vector $\mathbf{x}_1 - \mathbf{x}$. It lies along the wire going up to the top of the first post, so it gives the direction of $\mathbf{i}_1$. Since the length of a vector $\mathbf{a} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ is given by $\sqrt{a^2 + b^2 + c^2}$, the length d_1 of $\mathbf{x}_1 - \mathbf{x}$ is

$$d_1 = \sqrt{x^2 + y^2 + (H - z)^2}\,. \tag{4.8a}$$

Similarly, the vectors $\mathbf{x}_2 - \mathbf{x}$ and $\mathbf{x}_3 - \mathbf{x}$ lie along the wires from the weight to the top of the other posts and they have lengths d_2 and d_3 respectively, given by

$$d_2 = \sqrt{(L - x)^2 + y^2 + (H - z)^2}\,, \tag{4.8b}$$

$$d_3 = \sqrt{\left(\tfrac{1}{2}\,L - x\right)^2 + \left(\tfrac{\sqrt{3}}{2}\,L - y\right)^2 + (H - z)^2}\,. \tag{4.8c}$$

Once we divide a vector by its length, we create a new vector obviously with unit length. The desired unit vectors $\mathbf{i}_i$ lying along the wires are

$$\mathbf{i}_1 = \frac{1}{d_1}\big(-x\mathbf{i} - y\mathbf{j} + (H - z)\mathbf{k}\big)\,, \tag{4.9a}$$

$$\mathbf{i}_2 = \frac{1}{d_2}\big((L - x)\mathbf{i} - y\mathbf{j} + (H - z)\mathbf{k}\big)\,, \tag{4.9b}$$

$$\mathbf{i}_3 = \frac{1}{d_3}\big((\tfrac{1}{2}L - x)\mathbf{i} - (\tfrac{\sqrt{3}}{2}L - y)\mathbf{j} + (H - z)\mathbf{k}\big)\,. \tag{4.9c}$$

Now the balance of forces on the object may be written as

$$T_1\mathbf{i}_1 + T_2\mathbf{i}_2 + T_3\mathbf{i}_3 - W\mathbf{k} = 0\,. \tag{4.10}$$

By expressing this balance for each component, we obtain a system of linear equations for the tensions T_i

$$-\frac{x}{d_1}T_1 + \frac{L - x}{d_2}T_2 + \frac{L - 2x}{2d_3}T_3 = 0\,, \tag{4.11a}$$

$$-\frac{y}{d_1}T_1 - \frac{y}{d_2}T_2 + \frac{\sqrt{3}\,L - 2y}{2d_3}T_3 = 0\,, \tag{4.11b}$$

$$\frac{H - z}{d_1}T_1 + \frac{H - z}{d_2}T_2 + \frac{H - z}{d_3}T_3 = W\,. \tag{4.11c}$$

Comments:

1. The balance of forces applied at one location results in three equations. If the balance of forces on the posts are included, we would have three more vector equations for the three unknown forces on the posts, each force having three components. There would be a total of 12 unknowns and twelve equations. You can see how quickly the size of the system of equations grows, and all this for a very simple structure. Just imagine how big the system will be for a building or a bridge, etc.

2. The unknowns are T_i and not x_i. The point has been made before that unknowns are often not expressed in what is considered standard mathematical notation.

3. The coefficients of the system (4.11) are not just numbers, but are expressions that contain other parameters. This happens whenever we derive systems from applications in the sciences or engineering. Thus we may not know in advance whether the system will satisfy the conditions for a solution to exist, not exist, or have many solutions.

Just for fun, let's determine what the tensions will be if we place the object in the middle of the device. From geometric considerations the object will be at $x = L/2$, $y = L/(2\sqrt{3})$, $z = H/2$. Consequently,

$$d_1^2 = d_2^2 = d_3^2 = \frac{L^2}{3} + \frac{H^2}{4}\,.$$

In English, the distance from the object to the top of any of the posts is the same. Let $d = d_1$ (or $= d_2 = d_3$). The system (4.11) becomes

$$-\frac{L}{2d}T_1 + \frac{L}{2d}T_2 = 0\,, \tag{4.12a}$$

$$-\frac{L}{2\sqrt{3}\,d}T_1 - \frac{L}{2\sqrt{3}\,d}T_2 + \frac{L}{\sqrt{3}\,d}T_3 = 0\,, \tag{4.12b}$$

$$\frac{H}{2d}T_1 + \frac{H}{2d}T_2 + \frac{H}{2d}T_3 = W\,, \tag{4.12c}$$

and the augmented matrix is

$$\begin{bmatrix} -\frac{L}{2d} & \frac{L}{2d} & 0 & \vdots & 0 \\ -\frac{L}{2\sqrt{3}\,d} & -\frac{L}{2\sqrt{3}\,d} & \frac{L}{\sqrt{3}\,d} & \vdots & 0 \\ \frac{H}{2d} & \frac{H}{2d} & \frac{H}{2d} & \vdots & W \end{bmatrix}. \tag{4.13a}$$

Since the top entry in the first column is non-zero, we divide the first row by $-L/(2d)$:

$$\begin{bmatrix} 1 & -1 & 0 & \vdots & 0 \\ -\frac{L}{2\sqrt{3}\,d} & -\frac{L}{2\sqrt{3}\,d} & \frac{L}{\sqrt{3}\,d} & \vdots & 0 \\ \frac{H}{2d} & \frac{H}{2d} & \frac{H}{2d} & \vdots & W \end{bmatrix}. \tag{4.13b}$$

Next, we add $L/(2\sqrt{3}\,d)$ times the first row to the second row and subtract $H/(2d)$ times the first row from the third row:

$$\begin{bmatrix} 1 & -1 & 0 & \vdots & 0 \\ 0 & -\frac{L}{\sqrt{3}\,d} & \frac{L}{\sqrt{3}\,d} & \vdots & 0 \\ 0 & \frac{H}{d} & \frac{H}{2d} & \vdots & W \end{bmatrix}. \tag{4.13c}$$

Now we proceed to the second column. Divide the second row by $-L/\sqrt{3}$:

$$\begin{bmatrix} 1 & -1 & 0 & \vdots & 0 \\ 0 & 1 & -1 & \vdots & 0 \\ 0 & \dfrac{H}{d} & \dfrac{H}{2d} & \vdots & W \end{bmatrix}. \tag{4.13d}$$

Next, add the second row to the first row and subtract H/d times the second row from the third row:

$$\begin{bmatrix} 1 & 0 & -1 & \vdots & 0 \\ 0 & 1 & -1 & \vdots & 0 \\ 0 & 0 & \dfrac{3H}{2d} & \vdots & W \end{bmatrix}. \tag{4.13e}$$

Now proceed to the third column. Divide the third row by $3H/(2d)$:

$$\begin{bmatrix} 1 & 0 & -1 & \vdots & 0 \\ 0 & 1 & -1 & \vdots & 0 \\ 0 & 0 & 1 & \vdots & \dfrac{2dW}{3H} \end{bmatrix} \tag{4.13f}$$

Finally, add the third row to both the first and second rows:

$$\begin{bmatrix} 1 & 0 & 0 & \vdots & \dfrac{2dW}{3H} \\ 0 & 1 & 0 & \vdots & \dfrac{2dW}{3H} \\ 0 & 0 & 1 & \vdots & \dfrac{2dW}{3H} \end{bmatrix}. \tag{4.13g}$$

Equivalently,

$$T_1 = T_2 = T_3 = \frac{2dW}{3H}. \tag{4.13h}$$

The result states that all three tensions have the same value. Surely the tensions in all three wires would have to be the same to make the object lie in the middle, so the result makes perfect sense.

It is tempting to believe that when we derive "sensible" equations from an application in science or engineering, we will always find unique solutions,

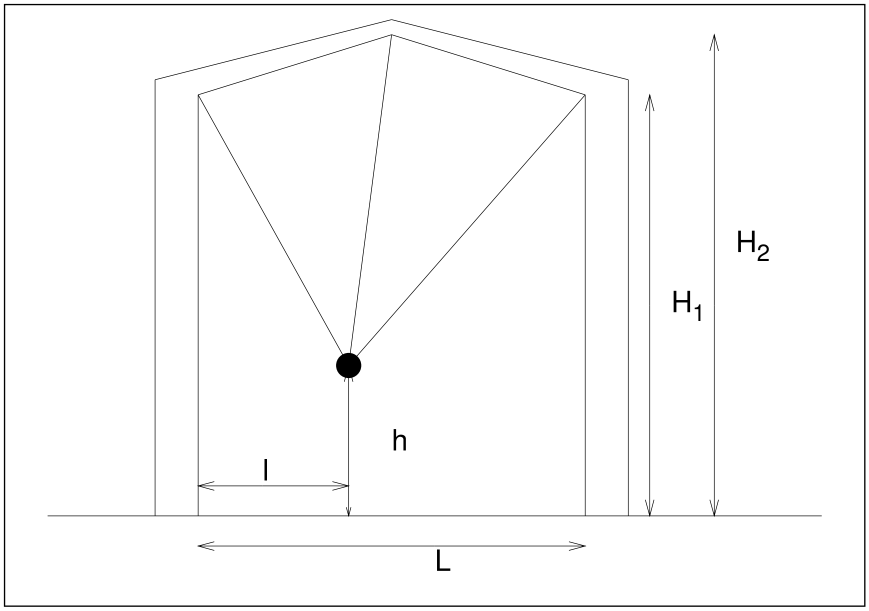


Figure 6: Schematic of a body suspended by three wires

just as we did in (4.13h). However, this is not so, and the following extension of the example in Figure 2 demonstrates why. Add a triangular arch across the supports and place another winch at the peak to pull a third wire attached to the weight. Figure 6 provides a schematic. Now we will be able to raise the object without excessive tensions in the other two wires; see (2.6b). The three winches are located at $(0, H_1)$, (L, H_1) and $(L/2, H_2)$. In the same way that we obtain (2.1) from geometrical considerations, we now obtain

$$L_1 \cos(\theta_1) = l\,, \qquad L_1 \sin(\theta_1) = H_1 - h\,, \tag{4.14a}$$

$$L_2 \cos(\theta_2) = L - l\,, \qquad L_2 \sin(\theta_2) = H_1 - h\,, \tag{4.14b}$$

$$L_3 \cos(\theta_3) = \frac{L}{2} - l\,, \qquad L_3 \sin(\theta_3) = H_2 - h\,, \tag{4.14c}$$

where the angle θ_3 is measured from the horizontal to the third wire in the same sense as θ_2. The inclusion of a third tension modifies the system of equations (2.3) and we have instead

$$T_1\cos(\theta_1) - T_2\cos(\theta_2) - T_3\cos(\theta_3) = 0\,, \tag{4.15a}$$

$$T_1\sin(\theta_1) + T_2\sin(\theta_2) + T_3\sin(\theta_3) = W\,. \tag{4.15b}$$

After substituting (4.14),

$$\frac{l}{L_1}T_1 - \frac{L-l}{L_2}T_2 - \frac{L-2l}{2L_3}T_3 = 0\,, \tag{4.16a}$$

$$\frac{H_1-h}{L_1}T_1 + \frac{H_1-h}{L_2}T_2 + \frac{H_2-h}{L_3}T_3 = W\,, \tag{4.16b}$$

and the augmented matrix is

$$\left[\begin{array}{ccc:c} \dfrac{l}{L_1} & -\dfrac{L-l}{L_2} & -\dfrac{L-2l}{2L_3} & 0 \\ \dfrac{H_1-h}{L_1} & \dfrac{H_1-h}{L_2} & \dfrac{H_2-h}{L_3} & W \end{array}\right]. \tag{4.17}$$

This time the coefficient matrix A for (4.16) is lobsided – it has three columns and two rows reflecting the fact that the system has only two equations for three unknowns. Let's proceed with the elimination and discover what the mathematical results tell us. But let's do so with the special choice $\theta_1 = \theta_2 = \pi/4$ and $\theta_3 = \pi/2$. Recall this choice was made when we considered the object hanging in the middle of the device – see (2.19). The only difference is the presence of the third wire pulling the object directly upwards. Consequently, (4.17) is replaced with

$$\left[\begin{array}{ccc:c} \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} & 0 & 0 \\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & 1 & W \end{array}\right]. \tag{4.18a}$$

Now let's execute the solution strategy. First, we divide the first row by $1/\sqrt{2}$ (or multiply by $\sqrt{2}$):

$$\left[\begin{array}{ccc:c} 1 & -1 & 0 & 0 \\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & 1 & W \end{array}\right]. \tag{4.18b}$$

Next, we multiply the first row by $1/\sqrt{2}$ and subtract it from the second row:

$$\begin{bmatrix} 1 & -1 & 0 & \vdots & 0 \\ 0 & \sqrt{2} & 1 & \vdots & W \end{bmatrix} . \tag{4.18c}$$

We have completed the actions required for the first column and we move to the second column, second row. Divide the second row by $1/\sqrt{2}$:

$$\begin{bmatrix} 1 & -1 & 0 & \vdots & 0 \\ 0 & 1 & \dfrac{1}{\sqrt{2}} & \vdots & \dfrac{W}{\sqrt{2}} \end{bmatrix} . \tag{4.18d}$$

Finally, add the second row to the first row:

$$\begin{bmatrix} 1 & 0 & \dfrac{1}{\sqrt{2}} & \vdots & \dfrac{W}{\sqrt{2}} \\ 0 & 1 & \dfrac{1}{\sqrt{2}} & \vdots & \dfrac{W}{\sqrt{2}} \end{bmatrix} . \tag{4.18e}$$

We have completed the steps for the second column. Note that when we complete the steps for the first two columns successfully, the results show a single entry (the number 1) in the column at the appropriate row.

At this stage, we must terminate the procedure. We can move over to another column (column 3) but we have no more rows. Obviously, if there were another equation, we would have another row and we could proceed to the third column just as we did in (4.5g) to obtain (4.5h). In the case (4.18e), we are unable to complete the process to determine T_3. Instead, we view T_3 as a parameter that must be given before the solution can be fully determined. Set $T_3 = C$ to indicate our choice for T_3. Then the second row of (4.18e) has the meaning

$$T_2 + \frac{1}{\sqrt{2}} C = \frac{W}{\sqrt{2}} ,$$

or

$$T_2 = \frac{W}{\sqrt{2}} - \frac{C}{\sqrt{2}} . \tag{4.19a}$$

Similarly,

$$T_1 = \frac{W}{\sqrt{2}} - \frac{C}{\sqrt{2}} . \tag{4.19b}$$

What does this result mean? From a physical view point, C is the amount of tension we decide to apply at the winch at the top of the structure. The effect is to reduce the weight of the object to $W - C$ which must then be supported by the remaining two wires. Of course, we require $C \leq W$ to make physical sense.

From a mathematical view, (4.19) expresses the presence of many solutions, dependent on the choice for C. We may bring out more clearly the impact of the choice for C by writing the result as the addition of two vectors

$$\begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} = W \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + C \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \end{pmatrix} . \tag{4.20}$$

The solution has two parts; a particular solution, the vector multiplied by W, and a homogeneous solution, the vector multiplied by C. To understand better the meaning of these two solutions, let's express the original system (4.16) in matrix form. Define

$$A = \begin{bmatrix} \dfrac{l}{L_1} & -\dfrac{L-l}{L_2} & -\dfrac{L-2l}{2L_3} \\ \dfrac{H_1 - h}{L_1} & \dfrac{H_1 - h}{L_2} & \dfrac{H_2 - h}{L_3} \end{bmatrix} , \quad \mathbf{x} = \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} , \quad \mathbf{b} = \begin{pmatrix} 0 \\ W \end{pmatrix} . \tag{4.21}$$

Then (4.17) is simply $A\mathbf{x} = \mathbf{b}$. Now let's write the two solutions as vectors

$$\mathbf{x}_p = W \begin{pmatrix} \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} \\ 0 \end{pmatrix} , \quad \mathbf{x}_h = \begin{pmatrix} \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} \\ 1 \end{pmatrix} .$$

Note that $A\mathbf{x}_p = \mathbf{b}$, but that $A\mathbf{x}_h = 0$. The latter result is a consequence of setting $W = 0$, and the terminology reflects the fact that $\mathbf{x}_h$ satisfies the homogeneous equations – there are no inhomogeneous terms and the right-hand side vector contains just zeros. The other solution $\mathbf{x}_p$ is just one particular choice of solutions (based on $C = 0$) for the inhomogeneous equations. This decomposition pervades all linear problems where the solutions are not unique.

Now that there are several examples of the procedure, it is time to consider what the general pattern in the results will be. For example, all the possible outcomes of the procedure for a system of three equations in three unknowns are listed below.

$$\underset{\text{case 1}}{\begin{bmatrix} 1 & 0 & 0 & \vdots & X \\ 0 & 1 & 0 & \vdots & X \\ 0 & 0 & 1 & \vdots & X \end{bmatrix}} \tag{4.22a}$$

is the most common case and the result is a unique solution. Compare with case 1 in (3.13). In the following three cases

$$\underset{\text{case 2a}}{\begin{bmatrix} 1 & 0 & X & \vdots & X \\ 0 & 1 & X & \vdots & X \\ 0 & 0 & 0 & \vdots & X \end{bmatrix}} \quad \underset{\text{case 2b}}{\begin{bmatrix} 1 & X & 0 & \vdots & X \\ 0 & 0 & 1 & \vdots & X \\ 0 & 0 & 0 & \vdots & X \end{bmatrix}} \quad \underset{\text{case 2c}}{\begin{bmatrix} 1 & X & X & \vdots & X \\ 0 & 0 & 0 & \vdots & X \\ 0 & 0 & 0 & \vdots & X \end{bmatrix}}, \tag{4.22b}$$

at least one row of zeros occurs in the matrix. Since the corresponding entry in the right-hand side (the rightmost column) is nonzero in general, there is no solution to the system. Compare with case 2 in (3.13). Because the system is larger, there are more ways for a zero row to occur. There are three corresponding cases

$$\underset{\text{case 3a}}{\begin{bmatrix} 1 & 0 & X & \vdots & X \\ 0 & 1 & X & \vdots & X \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix}} \quad \underset{\text{case 3b}}{\begin{bmatrix} 1 & X & 0 & \vdots & X \\ 0 & 0 & 1 & \vdots & X \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix}} \quad \underset{\text{case 3c}}{\begin{bmatrix} 1 & X & X & \vdots & X \\ 0 & 0 & 0 & \vdots & 0 \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix}} \tag{4.22c}$$

where there is a zero in the right-hand side of the row of zeros in the new coefficient matrix. Now there are many possible solutions. Before describing how to express these many solutions, it is worthwhile to make some observations about the above patterns.

Comments:

1. Typically, the entries along the rows in the reduced matrix start with zeros followed by 1. The first row may start with a 1, and some of the last rows may contain only zeros. The column number in which the first entry with 1 occurs is the unknown that the equation is determining. All other entries in that column are zero.

2. In some cases, such as case 2b, a column is skipped because there are only zero entries in all the remaining rows of that column. This means there are no equations to determine the known associated with that column, and we are forced to regard the unknown as a parameter. In case 2c (or case 3c), the second and third columns are skipped indicating that two unknowns are not determined.

3. The pattern has an upside-down staircase profile where the presence of a 1 following the first zeros along a row indicates the presence of a step. As (4.22) shows, the steps may be uneven in length. It should be relatively easy to imagine the continuation of this pattern for any size matrix, including those that are not square. This pattern in the matrix is referred to as the row reduced echelon form (rref).

Since Case 2b introduces a new entry in the patterns not seen in previous examples, let me show how it might arise. Consider the system

$$2x_1 + 4x_2 + 2x_3 = 6\,, \tag{4.23a}$$

$$x_1 + 2x_2 = 1\,, \tag{4.23b}$$

$$3x_1 + 6x_2 + 2x_3 = 7\,, \tag{4.23c}$$

which has the augmented matrix

$$\begin{bmatrix} 2 & 4 & 2 & \vdots & 6 \\ 1 & 2 & 0 & \vdots & 1 \\ 3 & 6 & 2 & \vdots & 7 \end{bmatrix} . \tag{4.24a}$$

The first step requires a division of the first row by 2:

$$\begin{bmatrix} 1 & 2 & 1 & \vdots & 3 \\ 1 & 2 & 0 & \vdots & 1 \\ 3 & 6 & 2 & \vdots & 7 \end{bmatrix} . \tag{4.24b}$$

Next, subtract the first row from the second row and subtract three times the first row from the third row:

$$\begin{bmatrix} 1 & 2 & 1 & \vdots & 3 \\ 0 & 0 & -1 & \vdots & -2 \\ 0 & 0 & -1 & \vdots & -2 \end{bmatrix} . \tag{4.24c}$$

Observe there is already something odd; the second and third rows are the same. The next step requires moving to the second column and second row. However, the entry is zero. Further, an interchange with the third row doesn't help. We are forced to move to the third column, second row. Divide the second row by -1:

$$\begin{bmatrix} 1 & 2 & 1 & \vdots & 3 \\ 0 & 0 & 1 & \vdots & 2 \\ 0 & 0 & -1 & \vdots & -2 \end{bmatrix} . \tag{4.24d}$$

Finally, add the second row to the third row and subtract the second row from the first row:

$$\begin{bmatrix} 1 & 2 & 0 & \vdots & 1 \\ 0 & 0 & 1 & \vdots & 2 \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix} . \tag{4.24e}$$

The result matches the pattern in case 2b.

How do we write down the solution given in (4.24e)? Note first that the last row gives no information and may be discarded. The second row determines $x_3 = 2$. The second column doesn't determine x_2; so set $x_2 = C$ where C becomes a parameter to be specified. Thus the first row becomes $x_1 = 1 - 2C$. Now write the solution as a vector

$$\mathbf{x} = \begin{pmatrix} 1-2C \\ C \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + C \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} . \tag{4.24f}$$

It is that simple!

The final example illustrates how to write down the solution in case 2c. Take the following specific example for the augmented matrix

$$\begin{bmatrix} 1 & 2 & 3 & \vdots & 4 \\ 0 & 0 & 0 & \vdots & 0 \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix} . \tag{4.25a}$$

Here there are two columns that don't determine x_2 and x_3. Both these unknowns must be shifted to the status of parameters. Set $x_2 = C_1$ and

$x_3 = C_2$. The first row – the only row containing useful information – gives $x_1 = 4 - 2C_1 - 3C_2$ and the solution vector is

$$\mathbf{x} = \begin{pmatrix} 4 - 2C_2 - 3C_2 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix} + C_1 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + C_2 \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} . \tag{4.25b}$$

There are two vectors that constitute the homogenous solution (the vectors multiplied by C_1 and C_2), and there are two parameters that influence how they are added to the particular solution (the first vector).

5 Summary

The algorithm described leads to a matrix that is said to be in reduced row echelon form (rref), and the pattern in this matrix tells us everything about the solution; whether it exists, doesn't exist, or allows many solutions, and even allows us to specify the nature of the multiple solutions. Further, the algorithm can be coded on a computer to handle large matrices where executing the algorithm would be laborious or impractical.

Unfortunately, this wonderful news is dampened by practical matters. The implementation requires precise arithmetic, which is not possible on computers. Computer algorithms have small errors that can have disastrous consequences. For example, suppose the execution of the algorithm produces small errors, represented by the symbol Y, in the three case given in (4.22c) so that they take the modified forms

$$\underset{\text{case 3a}}{\begin{bmatrix} 1 & 0 & X & \vdots & X \\ 0 & 1 & X & \vdots & X \\ 0 & 0 & Y & \vdots & Y \end{bmatrix}} \quad \underset{\text{case 3b}}{\begin{bmatrix} 1 & X & 0 & \vdots & X \\ 0 & Y & 1 & \vdots & X \\ 0 & 0 & Y & \vdots & Y \end{bmatrix}} \quad \underset{\text{case 3c}}{\begin{bmatrix} 1 & X & X & \vdots & X \\ 0 & Y & Y & \vdots & Y \\ 0 & 0 & Y & \vdots & Y \end{bmatrix}} . \tag{5.1}$$

The algorithm could proceed as normal to produce the form (4.22a) and the results would be completely wrong!

There are two directions to take to deal with this difficulty. One is to design an algorithm that would alert us to the possibility that errors have been made. This approach demands first an understanding of how errors arise and some method of evaluating their importance. The other is to find an alternate way to recognize when one of the cases in (4.22b) and (4.22c) must occur. This approach leads us to an even deeper understanding of the properties of linear systems.