Chapter 1
Strategies for solving problems

Physics involves a great deal of problem solving. Whether you are doing cutting-edge research or reading a book on a well-known subject, you are going to need to solve some problems. In the latter case (the presently relevant one, given what is in your hand right now), it is fairly safe to say that the true test of understanding something is the ability to solve problems on it. Reading about a topic is often a necessary step in the learning process, but it is by no means a sufficient one. The more important step is spending as much time as possible solving problems (which is inevitably an active task) beyond the time you spend reading (which is generally a more passive task). I have therefore included a very large number of problems/exercises in this book.

However, if I'm going to throw all these problems at you, I should at least give you some general strategies for solving them. These strategies are the subject of the present chapter. They are things you should always keep in the back of your mind when tackling a problem. Of course, they are generally not sufficient by themselves; you won't get too far without understanding the physical concepts behind the subject at hand. But when you add these strategies to your physical understanding, they can make your life a lot easier.

1.1 General strategies

There are a number of general strategies you should invoke without hesitation when solving a problem. They are:

1. **Draw a diagram, if appropriate.**

   In the diagram, be sure to label clearly all the relevant quantities (forces, lengths, masses, etc.). Diagrams are absolutely critical in certain types of problems. For example, in problems involving "free-body" diagrams (discussed in Chapter 3) or relativistic kinematics (discussed in Chapter 11), drawing a diagram can change a hopelessly complicated problem into a near-trivial one. And even in cases where diagrams aren't this crucial, they're invariably very helpful. A picture is definitely worth a thousand words (and even a few more, if you label things!).

2. **Write down what you know, and what you are trying to find.**

   In a simple problem, you may just do this in your head without realizing it. But in more difficult problems, it is very useful to explicitly write things out. For example, if there are three unknowns that you're trying to find, but you've written down only two facts, then you know there must be another fact you're missing (assuming that the problem is in fact solvable), so you can go searching for it. It might be a conservation law, or an $F = ma$ equation, etc.

3. **Solve things symbolically.**

   If you are solving a problem where the given quantities are specified numerically, you should immediately change the numbers to letters and solve the problem in terms of the letters. After you obtain an answer in terms of the letters, you can plug in the actual numerical values to obtain a numerical answer. There are many advantages to using letters:

   - IT'S QUICKER. It's much easier to multiply a g by an ℓ by writing them down on a piece of paper next to each other, than it is to multiply them together on a calculator. And with the latter strategy, you'd undoubtedly have to pick up your calculator at least a few times during the course of a problem.
   - YOU'RE LESS LIKELY TO MAKE A MISTAKE. It's very easy to mistype an 8 for a 9 in a calculator, but you're probably not going to miswrite a q for a g on a piece of paper. But if you do, you'll quickly realize that it should be a g. You certainly won't just give up on the problem and deem it unsolvable because no one gave you the value of q!
   - YOU CAN DO THE PROBLEM ONCE AND FOR ALL. If someone comes along and says, oops, the value of ℓ is actually 2.4 m instead of 2.3 m, then you won't have to do the whole problem again. You can simply plug the new value of ℓ into your final symbolic answer.
   - YOU CAN SEE THE GENERAL DEPENDENCE OF YOUR ANSWER ON THE VARIOUS GIVEN QUANTITIES. For example, you can see that it grows with quantities a and b, decreases with c, and doesn't depend on d. There is much, much more information contained in a symbolic answer than in a numerical one. And besides, symbolic answers nearly always look nice and pretty.
   - YOU CAN CHECK UNITS AND SPECIAL CASES. These checks go hand-in-hand with the previous "general dependence" advantage. But since they're so important, we'll postpone their discussion and devote Sections 1.2 and 1.3 to them.

   Having said all this, it should be noted that there are occasionally times when things get a bit messy when working with letters. For example, solving a system of three equations in three unknowns might be rather cumbersome unless you plug in the actual numbers. But in the vast majority of problems, it is highly advantageous to work entirely with letters.

4. **Consider units/dimensions.**

   This is extremely important. See Section 1.2 for a detailed discussion.

5. **Check limiting/special cases.**

   This is also extremely important. See Section 1.3 for a detailed discussion.

6. **Check order of magnitude if you end up getting a numerical answer.**

   If you end up with an actual numerical answer to a problem, be sure to do a sanity check to see if the number is reasonable. If you've calculated the distance along the ground that a car skids before it comes to rest, and if you've gotten an answer of a kilometer or a millimeter, then you know you've probably done something wrong. Errors of this sort often come from forgetting some powers of 10 (say, when converting kilometers to meters) or from multiplying something instead of dividing (although you should be able to catch this by checking your units, too).

You will inevitably encounter problems, physics ones or otherwise, where you don't end up obtaining a rigorous answer, either because the calculation is intractable, or because you just don't feel like doing it. But in these cases it's usually still possible to make an educated guess, to the nearest power of 10. For example, if you walk past a building and happen to wonder how many bricks are in it, or what the labor cost was in constructing it, then you can probably give a reasonable answer without doing any severe computations. The physicist Enrico Fermi was known for his ability to estimate things quickly and produce order-of-magnitude guesses with only minimal calculation. Hence, a problem where the goal is to simply obtain the nearest power-of-10 estimate is known as a "Fermi problem." Of course, sometimes in life you need to know things to better accuracy than the nearest power of 10 . . .

> How Fermi could estimate things!
> Like the well-known Olympic ten rings,
> And the one hundred states,
> And weeks with ten dates,
> And birds that all fly with one . . . wings.

In the following two sections, we'll discuss the very important strategies of checking units and special cases. Then in Section 1.4 we'll discuss the technique of solving problems numerically, which is what you need to do when you end up with a set of equations you can't figure out how to solve. Section 1.4 isn't quite analogous to Sections 1.2 and 1.3, in that these first two are relevant to basically any problem you'll ever do, whereas solving equations numerically is something you'll do only for occasional problems. But it's nevertheless something that every physics student should know.

In all three of these sections, we'll invoke various results derived later in the book. For the present purposes, the derivations of these results are completely irrelevant, so don't worry at all about the physics behind them – there will be plenty of opportunity for that later on! The main point here is to learn what to do with the result of a problem once you've obtained it.

1.2 Units, dimensional analysis

The units, or dimensions, of a quantity are the powers of mass, length, and time associated with it. For example, the units of a speed are length per time. The consideration of units offers two main benefits. First, looking at units before you start a problem can tell you roughly what the answer has to look like, up to numerical factors. Second, checking units at the end of a calculation (which is something you should *always* do) can tell you if your answer has a chance at being correct. It won't tell you that your answer is definitely correct, but it might tell you that your answer is definitely incorrect. For example, if your goal in a problem is to find a length, and if you end up with a mass, then you know it's time to look back over your work.

> "Your units are wrong!" cried the teacher.
> "Your church weighs six joules – what a feature!
> And the people inside
> Are four hours wide,
> And eight gauss away from the preacher!"

In practice, the second of the above two benefits is what you will generally make use of. But let's do a few examples relating to the first benefit, because these can be a little more exciting. To solve the three examples below exactly, we would need to invoke results derived in later chapters. But let's just see how far we can get by using only dimensional analysis. We'll use the "[ ]" notation for units, and we'll let M stand for mass, L for length, and T for time. For example, we'll write a speed as $[v] = L/T$ and the gravitational constant as $[G] = L^3/(MT^2)$ (you can figure this out by noting that Gm_1m_2/r^2 has the dimensions of force, which in turn has dimensions ML/T^2, from $F = ma$). Alternatively, you can just use the mks units, kg, m, s, instead of M, L, T, respectively.[1]

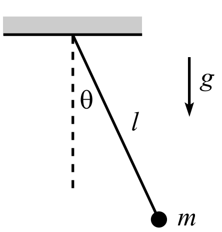

Fig. 1.1

Example (Pendulum): A mass m hangs from a massless string of length ℓ (see Fig. 1.1) and swings back and forth in the plane of the paper. The acceleration due to gravity is g. What can we say about the frequency of oscillations?

Solution: The only dimensionful quantities given in the problem are $[m] = M$, $[\ell] = L$, and $[g] = L/T^2$. But there is one more quantity, the maximum angle θ_0, which is dimensionless (and easy to forget). Our goal is to find the frequency, which

[1] When you check units at the end of a calculation, you will invariably be working with the kg,m,s notation. So that notation will inevitably get used more. But I'll use the M, L, T notation here, because I think it's a little more instructive. At any rate, just remember that the letter m (or M) stands for "meter" in one case, and "mass" in the other.

has units of $1/T$. The only combination of our given dimensionful quantities that has units of $1/T$ is $\sqrt{g/\ell}$. But we can't rule out any θ_0 dependence, so the most general possible form of the frequency is[2]

$$\omega = f(\theta_0)\sqrt{\frac{g}{\ell}}, \tag{1.1}$$

where f is a dimensionless function of the dimensionless variable θ_0.

REMARKS:

1. It just so happens that for small oscillations, $f(\theta_0)$ is essentially equal to 1, so the frequency is essentially equal to $\sqrt{g/\ell}$. But there is no way to show this by using only dimensional analysis; you actually have to solve the problem for real. For larger values of θ_0, the higher-order terms in the expansion of f become important. Exercise 4.23 deals with the leading correction, and the answer turns out to be $f(\theta_0) = 1 - \theta_0^2/16 + \cdots$.
2. Since there is only one mass in the problem, there is no way that the frequency (with units of $1/T$) can depend on $[m] = M$. If it did, there would be nothing to cancel the units of mass and produce a pure inverse-time.
3. We claimed above that the only combination of our given dimensionful quantities that has units of $1/T$ is $\sqrt{g/\ell}$. This is easy to see here, but in more complicated problems where the correct combination isn't so obvious, the following method will always work. Write down a general product of the given dimensionful quantities raised to arbitrary powers ($m^a \ell^b g^c$ in this problem), and then write out the units of this product in terms of a, b, and c. If we want to obtain units of $1/T$ here, then we need

$$M^a L^b \left(\frac{L}{T^2}\right)^c = \frac{1}{T}. \tag{1.2}$$

Matching up the powers of the three kinds of units on each side of this equation gives

$$M:\ a = 0, \qquad L:\ b + c = 0, \qquad T:\ -2c = -1. \tag{1.3}$$

The solution to this system of equations is $a = 0$, $b = -1/2$, and $c = 1/2$, so we have reproduced the $\sqrt{g/\ell}$ result. ♣

What can we say about the total energy of the pendulum (with the potential energy measured relative to the lowest point)? We'll talk about energy in Chapter 5, but the only thing we need to know here is that energy has units of ML^2/T^2. The only combination of the given dimensionful constants of this form is $mg\ell$. But again, we can't rule out any θ_0 dependence, so the energy must take the form $f(\theta_0)mg\ell$, where f is some function. That's as far as we can go with dimensional analysis. However, if we actually invoke a little physics, we can say that the total energy equals the potential energy at the highest point, which is $mg\ell(1 - \cos\theta_0)$. Using the Taylor expansion for $\cos\theta$ (see Appendix A for a discussion of Taylor series), we see that $f(\theta_0) = \theta_0^2/2 - \theta_0^4/24 + \cdots$. So in contrast with the frequency result above, the maximum angle θ_0 plays a critical role in the energy.

[2] We'll measure frequency here in radians per second, denoted by ω. So we're actually talking about the "angular frequency." Just divide by 2π (which doesn't affect the units) to obtain the "regular" frequency in cycles per second (hertz), usually denoted by ν. We'll talk at great length about oscillations in Chapter 4.

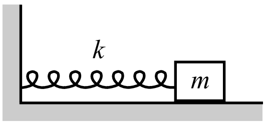

Fig. 1.2

Example (Spring): A spring with spring constant k has a mass m on its end (see Fig. 1.2). The spring force is $F(x) = -kx$, where x is the displacement from the equilibrium position. What can we say about the frequency of oscillations?

Solution: The only dimensionful quantities in this problem are $[m] = M$, $[k] = M/T^2$ (obtained by noting that kx has the dimensions of force), and the maximum displacement from the equilibrium, $[x_0] = L$. (There is also the equilibrium length, but the force doesn't depend on this, so there is no way it can come into the answer.) Our goal is to find the frequency, which has units of $1/T$. The only combination of our given dimensionful quantities with these units is

$$\omega = C\sqrt{\frac{k}{m}}, \tag{1.4}$$

where C is a dimensionless number. It just so happens that C is equal to 1 (assuming that we're measuring ω in radians per second), but there is no way to show this by using only dimensional analysis. Note that, in contrast with the pendulum above, the frequency cannot have any dependence on the maximum displacement.

What can we say about the total energy of the spring? Energy has units of ML^2/T^2, and the only combination of the given dimensionful constants of this form is Bkx_0^2, where B is a dimensionless number. It turns out that $B = 1/2$, so the total energy equals $kx_0^2/2$.

REMARK: A real spring doesn't have a perfectly parabolic potential (that is, a perfectly linear force), so the force actually looks something like $F(x) = -kx + bx^2 + \cdots$. If we truncate the series at the second term, then we have one more dimensionful quantity to work with, $[b] = M/LT^2$. To form a quantity with the dimensions of frequency, $1/T$, we need x_0 and b to appear in the combination $x_0 b$, because this is the only way to get rid of the L. You can then see (by using the strategy of writing out a general product of the variables, discussed in the third remark in the pendulum example above) that the frequency must be of the form $f(x_0 b/k)\sqrt{k/m}$, where f is some function. We can therefore have x_0 dependence in this case. This answer must reduce to $C\sqrt{k/m}$ for $b = 0$. Hence, f must be of the form $f(y) = C + c_1 y + c_2 y^2 + \cdots$. ♣

Example (Low-orbit satellite): A satellite of mass m travels in a circular orbit just above the earth's surface. What can we say about its speed?

Solution: The only dimensionful quantities in the problem are $[m] = M$, $[g] = L/T^2$, and the radius of the earth $[R] = L$.[3] Our goal is to find the speed, which has units of L/T. The only combination of our dimensionful quantities with these units is

$$v = C\sqrt{gR}. \tag{1.5}$$

It turns out that $C = 1$.

[3] You might argue that the mass of the earth, M_E, and Newton's gravitational constant, G, should be also included here, because Newton's gravitational force law for a particle on the surface of the earth is $F = GM_E m/R^2$. But since this force can be written as $m(GM_E/R^2) \equiv mg$, we can absorb the effects of M_E and G into g.

1.3 Approximations, limiting cases

As with units, the consideration of limiting cases (or perhaps we should say special cases) offers two main benefits. First, it can help you get started on a problem. If you're having trouble figuring out how a given system behaves, then you can imagine making, for example, a certain length become very large or very small, and then you can see what happens to the behavior. Having convinced yourself that the length actually affects the system in extreme cases (or perhaps you will discover that the length doesn't affect things at all), it will then be easier to understand how it affects the system in general, which will then make it easier to write down the relevant quantitative equations (conservation laws, $F = ma$ equations, etc.), which will allow you to fully solve the problem. In short, modifying the various parameters and observing the effects on the system can lead to an enormous amount of information.

Second, as with checking units, checking limiting cases (or special cases) is something you should *always* do at the end of a calculation. But as with checking units, it won't tell you that your answer is definitely correct, but it might tell you that your answer is definitely incorrect. It is generally true that your intuition about limiting cases is much better than your intuition about generic values of the parameters. You should use this fact to your advantage.

Let's do a few examples relating to the second benefit. The initial expressions given in each example below are taken from various examples throughout the book, so just accept them for now. For the most part, I'll repeat here what I'll say later on when we work through the problems for real. A tool that comes up often in checking limiting cases is the Taylor series approximations; the series for many functions are given in Appendix A.

Example (Dropped ball): A beach ball is dropped from rest at height h. Assume that the drag force from the air takes the form $F_{\rm d} = -m\alpha v$. We'll find in Section 3.3 that the ball's velocity and position are given by

$$v(t) = -\frac{g}{\alpha}\left(1 - e^{-\alpha t}\right), \quad \text{and} \quad y(t) = h - \frac{g}{\alpha}\left(t - \frac{1}{\alpha}\left(1 - e^{-\alpha t}\right)\right). \tag{1.6}$$

These expressions are a bit complicated, so for all you know, I could have made a typo in writing them down. Or worse, I could have completely botched the solution. So let's look at some limiting cases. If these limiting cases yield expected results, then we can feel a little more confident that the answers are actually correct.

If t is very small (more precisely, if $\alpha t \ll 1$; see the discussion following this example), then we can use the Taylor series, $e^{-x} \approx 1 - x + x^2/2$, to make approximations to leading order in αt. The $v(t)$ in Eq. (1.6) becomes

$$\begin{aligned} v(t) &= -\frac{g}{\alpha}\left(1-\left(1-\alpha t+\frac{(\alpha t)^2}{2}-\cdots\right)\right) \\ &\approx -gt, \end{aligned} \tag{1.7}$$

plus terms of higher order in αt. This answer is expected, because the drag force is negligible at the start, so we essentially have a freely falling body with acceleration g downward. For small t, Eq. (1.6) also gives

$$\begin{aligned} y(t) &= h-\frac{g}{\alpha}\left[t-\frac{1}{\alpha}\left(1-\left(1-\alpha t+\frac{(\alpha t)^2}{2}-\cdots\right)\right)\right] \\ &\approx h-\frac{gt^2}{2}, \end{aligned} \tag{1.8}$$

plus terms of higher order in αt. Again, this answer is expected, because we essentially have a freely falling body at the start, so the distance fallen is the standard $gt^2/2$.

We can also look at large t (or rather, large αt). In this case, $e^{-\alpha t}$ is essentially zero, so the $v(t)$ in Eq. (1.6) becomes (there's no need for a Taylor series in this case)

$$v(t) \approx -\frac{g}{\alpha}. \tag{1.9}$$

This is the "terminal velocity." Its value makes sense, because it is the velocity for which the total force, $-mg - m\alpha v$, vanishes. For large t, Eq. (1.6) also gives

$$y(t) \approx h - \frac{gt}{\alpha} + \frac{g}{\alpha^2}. \tag{1.10}$$

Apparently for large t, g/α^2 is the distance (and this does indeed have units of length, because α has units of T^{-1}, because $m\alpha v$ has units of force) that our ball lags behind another ball that started out already at the terminal velocity, $-g/\alpha$.

Whenever you derive approximate answers as we just did, you gain something and you lose something. You lose some truth, of course, because your new answer is technically not correct. But you gain some aesthetics. Your new answer is invariably much cleaner (sometimes involving only one term), and this makes it a lot easier to see what's going on.

In the above example, it actually makes no sense to look at the limit where t is small or large, because t has dimensions. Is a year a large or small time? How about a hundredth of a second? There is no way to answer this without knowing what problem you're dealing with. A year is short on the time scale of galactic evolution, but a hundredth of a second is long on the time scale of a nuclear process. It makes sense only to look at the limit of a small (or large) *dimensionless* quantity. In the above example, this quantity is αt. The given constant α has units of T^{-1}, so $1/\alpha$ sets a typical time scale for the system. It therefore makes sense to look at the limit where $t \ll 1/\alpha$ (that is, $\alpha t \ll 1$), or where $t \gg 1/\alpha$ (that is, $\alpha t \gg 1$). In the limit of a small dimensionless quantity, a Taylor series can be used to expand an answer in powers of the small quantity, as we did above. We sometimes get sloppy and say things like, "In the limit of small t." But you know that we really mean, "In the limit of some small dimensionless quantity that has a t in the numerator," or, "In the limit where t is much smaller that a certain quantity that has the dimensions of time."

REMARK: As mentioned above, checking special cases tells you that either (1) your answer is consistent with your intuition, or (2) it's wrong. It never tells you that it's definitely correct. This is the same as what happens with the scientific method. In the real world, everything comes down to experiment. If you have a theory that you think is correct, then you need to check that its predictions are consistent with experiments. The specific experiments you do are the analog of the special cases you check after solving a problem; these two things represent what you know is true. If the results of the experiments are inconsistent with your theory, then you need to go back and fix your theory, just as you would need to go back and fix your answer. If, on the other hand, the results are consistent, then although this is good, the only thing it really tells you is that your theory *might* be correct. And considering the way things usually turn out, the odds are that it's not actually correct, but rather the limiting case of a more correct theory (just as Newtonian physics is a limiting case of relativistic physics, which is a limiting case of quantum field theory, etc.). That's how physics works. You can't prove anything, so you learn to settle for the things you can't disprove.

Consider, when seeking gestalts,
The theories that physics exalts.
It's not that they're known
To be written in stone.
It's just that we can't say they're false. ♣

When making approximations, how do you know how many terms in the Taylor series to keep? In the example above, we used $e^{-x} \approx 1 - x + x^2/2$. But why did we stop at the x^2 term? The honest (but slightly facetious) answer is, "Because I had already done this problem before writing it up, so I knew how many terms to keep." But the more informative (although perhaps no more helpful) answer is that before you do the calculation, there's really no way of knowing how many terms to keep. So you should just keep a few and see what happens. If everything ends up canceling out, then this tells you that you need to repeat the calculation with another term in the series. For example, in Eq. (1.8), if we had stopped the Taylor series at $e^{-x} \approx 1 - x$, then we would have obtained $y(t) = h - 0$, which isn't very useful, since the general goal is to get the leading-order behavior in the parameter we're looking at (which is t here). So in this case we'd know we'd have to go back and include the $x^2/2$ term in the series. If we were doing a problem in which there was still no t (or whatever variable) dependence at that order, then we'd have to go back and include the $-x^3/6$ term in the series. Of course, you could just play it safe and keep terms up to, say, fifth order. But that's invariably a poor strategy, because you'll probably never in your life have to go out that far in a series. So just start with one or two terms and see what it gives you. Note that in Eq. (1.7), we actually didn't need the second-order term, so we in fact could have gotten by with only $e^{-x} \approx 1 - x$. But having the extra term here didn't end up causing much heartache.

After you make an approximation, how do you know if it's a "good" one? Well, just as it makes no sense to ask if a dimensionful quantity is large or small without comparing it to another quantity, it makes no sense to ask if an approximation is "good" or "bad" without stating the accuracy you want. In the above example, if you're looking at a t value for which $\alpha t \approx 1/100$, then the term we ignored in Eq. (1.7) is smaller than gt by a factor $\alpha t/2 \approx 1/200$. So the error is on the order of 1%. If this is enough accuracy for whatever purpose you have in mind, then the approximation is a good one. If not, it's a bad one, and you should add more terms in the series until you get your desired accuracy.

The results of checking limits generally fall into two categories. Most of the time you know what the result should be, so this provides a double-check on your answer. But sometimes an interesting limit pops up that you might not expect. Such is the case in the following examples.

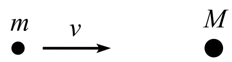

Fig. 1.3

Example (Two masses in 1-D): A mass m with speed v approaches a stationary mass M (see Fig. 1.3). The masses bounce off each other elastically. Assume that all motion takes place in one dimension. We'll find in Section 5.6.1 that the final velocities of the particles are

$$v_m = \frac{(m-M)v}{m+M}, \quad \text{and} \quad v_M = \frac{2mv}{m+M}. \tag{1.11}$$

There are three special cases that beg to be checked:

- If $m = M$, then Eq. (1.11) tells us that m stops, and M picks up a speed v. This is fairly believable (and even more so for pool players). And it becomes quite clear once you realize that these final speeds certainly satisfy conservation of energy and momentum with the initial conditions.
- If $M \gg m$, then m bounces backward with speed $\approx v$, and M hardly moves. This makes sense, because M is basically a brick wall.
- If $m \gg M$, then m keeps plowing along at speed $\approx v$, and M picks up a speed of $\approx 2v$. This $2v$ is an unexpected and interesting result (it's easier to see if you consider what's happening in the reference frame of the heavy mass m), and it leads to some neat effects, as in Problem 5.23.

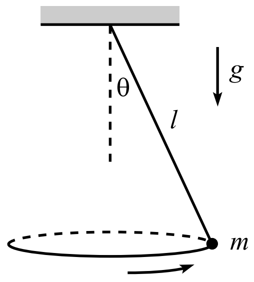

Fig. 1.4

Example (Circular pendulum): A mass hangs from a massless string of length ℓ. Conditions have been set up so that the mass swings around in a horizontal circle, with the string making a constant angle θ with the vertical (see Fig. 1.4). We'll find in Section 3.5 that the angular frequency, ω, of this motion is

$$\omega = \sqrt{\frac{g}{\ell \cos\theta}}. \tag{1.12}$$