Easy Algebra & Calculus

by *Jess H. Brewer*

October 28, 2020

Contents

Why Am I Doing This?

Once upon a time I wrote a book entitled "The Skeptic's Guide to Physics" to go with Physics 340, a course for Arts students at the University of British Columbia. After several experiments with existing textbooks, I decided to start my own, based on the usual collection of handwritten lecture notes. My reasons did not include any conviction that I could do a better job than anyone else; rather that I hadn't found any text that set out to do quite the same thing that I wanted to do, and I was too stubborn to revise my intentions to fit the literature. I have gotten worse with age.

What *did* I want to do? The impossible. Namely, to take my Arts students on a whirlwind tour of Physics from classical mechanics through modern elementary particle physics, without any patronizing appeals to faith in the experts. I especially wanted to avoid any hint of phrases like, "scientific tests prove…" that are employed with such poisonous efficiency by media manipulators. I wanted to treat them like savvy graduate students auditing a course outside their specialty, not like woodenheaded ignoramuses who had no intellect to appeal to. In particular, I believed that smart Arts people are as smart as (maybe smarter than!) smart Science people, and a good deal more eclectic on average.

Now I am retired from UBC, but I miss teaching; so I occasionally teach courses in the Elder College of Vancouver Island University in Nanaimo, BC. I found that folks over 50 who never took Physics before are a little reluctant to dive into the subject later in life, no matter how I promised to Keep It Simple.

So I decided to start a little less ambitiously: one of the reasons people are reluctant to tackle Physics is that they find the necessary *Mathematics* challenging. Now, a lot of Physicists offer courses entitled "Physics Without Mathematics" — that's ridiculous! Physics without Mathematics is like Poetry without Words.[1] Even babies acquire language skills, and small children are quite capable of appreciating poetry, though some of the more erudite nuances may escape them at first.

Hence this Elder College course. As it says in the course description, every discipline has easy parts and hard parts; the Easy parts of Algebra and Calculus are really all we need to make a satisfactory introduction to Physics (my ultimate goal) and the concomitant Mathematical understanding offers untold opportunities for enlightenment in myriad realms — Just wait and see!

So I will be addressing you as if you were in the Humanities, though you may just as well be a Nobel laureate or a short-order cook at a fast food restaurant. What do I care what you do for a living? I do want you to see Algebra and Calculus the way I see them, not some edited-for-television version. A tall order? You bet. I'm asking a lot? That's what I'm here for.

[1] Yes, yes, music and visual arts can be thought of metaphorically as forms of poetry, but I'm talking about *literal* poetry.

Chapter 1

Symbols

All communication relies upon abstract symbolism of one form or another. We can *feel* without symbols, but we can't *talk*. Before two people can communicate they must reach a *consensus* about the symbolic *representation* of reality they will employ in their conversation. This is so obvious that we usually take it for granted, but few experiences are so unsettling as to meet someone whose personal symbolic representation differs drastically from consensual reality.

How was this consensus reached? How arbitrary are symbolic conventions? Do they continue to evolve? They never represent quite the same things for different people; how do we know if there is a reality "out there" to be represented? These are questions that have perplexed philosophers for thousands of years; we are not going to find final answers to them here. But within the intrinsically abstract context of Mathematics we may find some instructive lessons in the interactions between tradition, convention, consensus and analytical logic. This is the focus of the present Chapter.

Each word in a dictionary plays the same role in writing or speech (or in "verbal" thought itself) as the hieroglyphic-looking symbols play in the equations of Algebra and Calculus. The big difference is *compactness* and in the degree to which ambiguity depends upon context. Obviously an algebraic symbol like t is rather compact relative to a word composed of several letters, like *time*. This allows storage of more information in less space, which is pragmatic but not always pleasing.

As for ambiguity in context, words are designed to have a great deal of ambiguity until they are placed in sentences, where the context partially dictates which meaning is intended. *But never entirely*. Part of the magic of poetry is its ambiguity; a good poet is offended by the question, "What exactly did you mean by that?" because *all* the possible meanings are intended. Great poetry does not highlight one meaning above all, but rather manipulates the interactions between the several possible interpretations so that each enriches the others and all unite to form a whole greater than the sum of its parts. As a result, no one ever knows for certain what another person is talking about; we merely learn to make good guesses.[1]

In Mathematics, every symbol must in principle be defined exhaustively and explicitly prior to its use. A meticulous Mathematician will *try* to provide an unambiguous definition of every *unusual* symbol introduced, but there are many symbols that are used so often in Mathematics to mean a certain thing that they have a well-known "default" meaning as long as they are used in a familiar *context*. In practice, some

[1] This seems to be holding up progress in Artificial Intelligence (AI) research, where people trying to teach computers to understand "natural language" (human speech) are stymied by the impossibility of reaching a unique logical interpretation of a typical sentence. Methinks they are trying too hard.

symbols (like ∂) are universally recognized as having a specific meaning that effectively *defines the context*.

The point is, algebraic notation follows a set of conventions, just like the grammar and syntax of verbal language, that defines the context in which each symbol is to be interpreted and thus provides a large fraction of the meaning of a given expression.

1.1 Number Systems

The first thing we learn about Mathematics as children is usually one of the most abstract concepts of all: the idea of *numbers* — zero, one, two, three, ... that have meaning independent of the concrete things they are the numbers *of*, like apples, oranges or universes. Numbers must also be assigned a conventional *representation* before they can be used to describe any practically useful examples (*e.g.* "arithmetic"). Such representations are *arbitrary*, based on rather simpleminded models of what is significant in a practical sense.

The *decimal* number system, based as it is upon a number whose only virtue is that most people have that number of fingers and thumbs, is a typical example. We all learned this convention as children and use it in our daily lives, but it is *neither unique nor optimal*. Let's examine the various ways we can count to ten:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

is "the decimal way". We can use our fingers to count to ten, but is it an *efficient* use of our fingers? If we had only thought to distinguish between fingers and thumbs, using thumbs for "carrying," and distinguishing between left and right hands, we could easily count to one hundred with our hands. If we were even cleverer, we could give each finger and thumb its own special significance and count to thirty-one on one hand and up to one thousand and twenty-three on both hands! That's as good as it can get, though; it requires a *binary* number system in which there are only *two* digits: 0 and 1. Counting to ten in binary looks like this:

$$1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010$$

This is how integers are stored in a digital computer. We could also count in *octal*:

$$1, 2, 3, 4, 5, 6, 7, 10, 11, 12$$

or *hexadecimal*:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, A$$

or in *trinary*:

$$1, 2, 10, 11, 12, 20, 21, 22, 100, 101$$

or in any other base we choose. I recommend learning to count in binary as well as decimal.

Is mathematics then arbitrary? Of course not. We can easily understand the distinction between the *representation* (which is arbitrary) and the *content* (which is not). Ten is still ten, regardless of which number system we use to write it "mathematically". Much more sophisticated notions can also be expressed in many ways; in fact it may be that we can only achieve a deep understanding of any concept by learning to express it in many alternate "languages."

The same is arguably true of Physics. Since I will often be using Physics examples to demonstrate the utility of Algebra and Calculus, it behooves me to briefly discuss some of the conventions and representations used in Physics.

1.2 Units & Dimensions

1.2.1 Time & Distance

Two of the most important concepts in Physics are "length" and "time." As is often the case with the most important concepts, neither can be defined except by example — *e.g.* "a meter is this long. . . ." or, "a second lasts from now

... to now." Both of these "definitions" completely beg the question, if you consider carefully what we are after; they merely define the *units* in which we propose to *measure* distance and time. Except for analogic reinforcements they do nothing at all to explain the "meaning" of the concepts "space" and "time."

Modern science has replaced the standard platinum-iridium reference **meter** (m) stick with the indirect prescription, "...the distance travelled by light in empty space during a time of 1/299,792,458 of a second," where a **second** (s) is now defined as the time it takes a certain frequency of the light emitted by cesium atoms to oscillate 9,192,631,770 times.[2] This represents a significant improvement inasmuch as we no longer have to resort to carrying our meter stick to the International Bureau of Weights and Measures in Sèvres, France (or to the U.S. National Bureau of Standards in Boulder, Colorado) to make sure it is the same length as the Standard Meter. We can just build an apparatus to count oscillations of cesium light and mark off how far light goes in 30.663318988 or so oscillations [well, it's easy if you have the right tools....] and make our own meter stick independently, confident that it will come out the same as the ones in France and Colorado, because our atoms are guaranteed to be just like theirs. We can even send signals to neighbors on Tau Ceti IV to tell them what size to make screwdrivers or crescent wrenches for export to Earth, since there is overwhelming evidence that their atoms also behave exactly like ours. This is quite remarkable, and unprecedented before the discovery of quantum physics; but unfortunately it does not make much difference to the dilemma we face when we try to define "distance." Nature has kindly provided us with an unlimited supply of accurate meter sticks, but it is still just a name we give to something.

[2] This is only the latest in a long sequence of redefinitions of the meter. Today's version reflects our recognition of the speed of light as a universal constant. (Here is a trick question for you: if the speed of light *were* different in one time and place from another, how could we tell?)

To learn the properties of that "something" which we call "distance" requires first that we believe that there is truly a physical entity, with intrinsic properties independent of our perceptions, to which we have given this name. This is extremely difficult to prove. Maybe not impossible, but I'll leave that to the philosophers. For the physicist it is really a matter of æsthetics to enter into conversations with Nature as if there were really a partner in such conversations. In other words, I cannot tell you what "distance" is, but if you will allow me to assume that the word refers to something "real," I can tell you a great deal about its properties, until at some point you feel the partial satisfaction of intimate familiarity where perfect comprehension is denied.

How do we begin to talk about time and space? The concepts are so fundamental to our language that all the words we might use to describe them have them built in! So for the moment we will have to give up and say, "Everyone knows pretty much what we mean by time and distance." This is always where we have to begin. Physics is just like poetry in this respect: you start by accepting a "basis set" of images, without discussion; then you work those images together to build new images, and after a period of refinement you find one day, miraculously, that the new images you have created can be applied to the ideas you began with, giving a new insight into their meaning. This "bootstrap" principle is what makes thinking profitable.

Later on, then, when we have learned to manipulate time and space more critically, we will acquire the means to break down the concepts and take a closer look.

1.2.2 Choice of Units

All choices of units are completely arbitrary and are made strictly for the sake of convenience. If you were a surveyor in 18th-Century England, you would consider the **chain** (66 feet by our standards) an extremely natural unit of length, and the **meter** would seem a completely artificial and useless unit, because people were shorter then and the **yard** (1 yard = 3600/3937 of a meter) was a better approximation to an average person's stride. **Feet** and **hands** were even better length units in those days; and if you hadn't noticed, an **inch** is just about the length of the middle bone in a small person's index finger.

If you couldn't get your hands on a timepiece with a second hand, the utility of **seconds** would seem limited to the (non-coincidental) fact that they are about the same as a resting heartbeat period. **Years** and **days** might seem less arbitrary to us, but we would have trouble convincing our friends on Tau Ceti IV.[3] Remember, our perspective in Physics is universal, and in that perspective all units are arbitrary.

We choose all our measurement conventions for convenience, often with monumental shortsightedness. The decimal number system is a typical example. At least when we realize this we can feel more forgiving of the clumsiness of many established systems of measurement. After all, a totally arbitrary decision is always wrong. (Or always right.)

Physicists are fond of devising "natural units" of measurement; but as always, what is considered "natural" depends upon what is being measured. Atomic physicists are understandably fond of the **Angstrom** (Å), which equals 10^{-10} m, which "just happens" to be roughly the diameter of a hydrogen atom. Astronomers measure distances in **light years**, the distance light travels in a year ($365 \times 24 \times 60 \times 60 \times 2.99 \times 10^8 = 9.43 \times 10^{15}$ m), **astronomical units** (a.u.), which I think have something to do with the Earth's orbit about the sun, or **parsecs**, which I seem to recall are related to seconds of arc at some distance. [I am not biased or anything. . . .]

Astrophysicists and particle physicists tend to use units in which the velocity of light (a fundamental constant) is dimensionless and has magnitude 1; then times and lengths are both measured in the same units. People who live near New York City have the same habit, oddly enough: if you ask them how far it is from Hartford to Boston, they will usually say, "Oh, about three hours." This is perfectly sensible insofar as the velocity of turnpike travel in New England is nearly a fundamental constant. In my own work at TRIUMF, I habitually measure distances in **nanoseconds** (billionths of seconds: 1 ns = 10^{-9} s), referring to the distance (about 29.97 cm) covered in that time by a particle moving at essentially the velocity of light.[4]

In general, physicists like to make *all* fundamental constants dimensionless; this is indeed economical, as it reduces the number of units one must use, but it results in some oddities from the practical point of view. A nuclear physicist is content to measure distances in *inverse pion masses*, but this is not apt to make a tailor very happy.

[3] This is a recurring problem in science fiction novels: will our descendents on other planets use a "local" definition of years, [months,] days, hours and minutes or try to stick with an Earth calendar despite the fact that it would mean the local sun would come up at a different time every day? Worse yet, how will a far-flung Galactic Empire reckon *dates*, especially considering the conditions imposed by Relativity? [The *Star Trek* solution is, of course, to ignore the laws of physics entirely.]

[4] Inasmuch as a *ns* is a roughly "person-sized" distance unit, it could actually be used rather effectively in place of feet and meters, which would get rid of at least *one* arbitrary unit. Oh well.

1.2.3 Perception Through Models

The upshot of all this is that you can't trust any units to carry lasting significance; all is vanity. Each and every choice of units represents essentially *a model of what is significant*. What is vitally relevant to one observer may be trivial and ridiculous to another. Lest this seem a depressing appraisal, consider that the same is true of all our means of perception, even including the physical sensing apparatus of our own bodies: our eyes are sensitive to an incredibly tiny fraction of the spectrum of electromagnetic radiation; what we miss is inconceivably vast compared to what we detect. And yet we see a lot, especially under the light of Sol, which at the Earth's surface happens to peak in just the region of our eyes' sensitivity. Our eyes are simply a model of what is important locally, and well adapted for the job.

The only understanding you can develop that is independent of units has to do with how dimensions can be combined, juxtaposed, *etc.* — their *relationships* with each other. The notion of a velocity as a ratio of distance to time is a concept which will endure all vagaries of fashion in measurment. This is the sort of concept that we try to pick out of the confusion. This is the sort of understanding for which the physicist searches.

1.3 Symbolic Conventions

In Physics we like to use a very compact notation for things we talk about a lot; this is æsthetically mandated by our commitment to making complicated things look [and maybe even *be*] simpler. Ideally we would like to have a single character to represent each paradigmatic "thing" in our lexicon, but in practice we don't have enough characters[5] and we have to re-use some of them in different contexts — just like in English!

In principle, any symbol can be used to represent any quantity, or even a non-quantity (like an "*operator*"), as long as it is explicitly and carefully defined. In practice, life is easier with some "default" *conventions* for what various symbols should be assumed to mean *unless otherwise specified*. On the next pages are some that I will be using a lot.[6]

[5] The wider availability of nice typesetting languages like LaTeX, in which this manuscript is being prepared, offers us the opportunity to add new symbols like $\aleph$, ϖ and $\heartsuit$, but this won't change the qualitative situation.

[6] (You may want to refer to these occasionally when trying to guess what I am trying to say with formulae. Don't worry if some are incomprehensible initially; for completeness, the list includes lots of "advanced" stuff.)

Table 1.1 Roman symbols commonly used in Physics

ROMAN LETTERS:

A = an *area*; Ampere(s).	a = *acceleration*; a general constant.
B = magnetic field.	b = a general constant.
C = heat capacity; Coulomb(s).	c = *speed of light*; a gen. constant.
D = a form of the electric field.	d = *differential* operator; *diameter*.
E = *energy*; electric field.	e = 2.71828...; electron's charge.
F = *force*; a general *function*.	f = a fraction; a *function* as in $f(x)$.
G = grav. constant; prefix *Giga*-.	g = *accel. of gravity* at Earth's surface.
H = magnetic field; Hamiltonian op.	h = Planck's constant; a height.
I = electric current.	$i = \sqrt{-1}$; an index (subscript).
J = *Joules*; spin; angular momentum.	j = a common integer index.
K = degrees Kelvin.	k = an integer index; a gen. constant; *kilo*-.
L = *angular momentum*; length.	l = an integer index; a *length*.
M = magnetization; mass; *Mega*-.	m = *metre(s)*; *mass*; an integer index.
N = *Newton(s)*; a large number.	n = a small number; prefix *nano*-.
$\mathcal{O}$ = "order of" symbol as in $\mathcal{O}(\alpha)$.	o = rarely used (looks like a 0).
P = probability; pressure; power.	p = *momentum*; prefix *pico*-.
Q = *electric charge*.	q = *elec. charge*; "canonical coordinate".
R = radius; electrical *resistance*.	r = *radius*.
S = *entropy*; surface area.	s = *second(s)*; distance.
T = *temperature*.	t = *time*.
U = potential energy; internal energy.	u = an abstract variable; a velocity.
V = *Volts*; *volume*; *potential energy*.	v = *velocity*.
W = *work*; weight.	w = a small weight; a width.
X = an abstract function, as $X(x)$.	x = *distance*; any *independent variable*.
Y = an abstract function, as $Y(y)$.	y = an abstract *dependent variable*.
Z = atomic number; $Z(z)$.	z = an abstract *dependent variable*.

Table 1.2 Greek symbols commonly used in Physics

GREEK LETTERS: (Capital Greek letters that look the same as Roman are omitted.)

	α = fine structure constant; an angle.
	$\beta = v/c$; an angle.
Γ = *torque*; a rate.	$\gamma = E/mc^2$; an angle.
Δ = "change in...", as in Δx.	δ = an infinitesimal; same as Δ.
	ϵ = an infinitesimal quantity.
$\mathcal{E}$ = "electromotive force".	ε = an energy.
	ζ = a general parameter.
	η = index of refraction.
Θ = an angle.	θ = an *angle* (most common symbol).
	ι = rarely used (looks like an i).
	κ = arcane version of k.
Λ = a rate; a type of baryon.	λ = *wavelength*; a rate.
	μ = reduced mass; muon; prefix *micro*-.
	ν = *frequency* in cycles/s (Hz); neutrino.
Ξ = a type of baryon.	ξ = a general parameter.
Π = *product* operator.	$\pi = 3.14159\ldots$; pion (a meson).
	ρ = *density* per unit volume; resistivity.
Σ = *summation* operator.	σ = cross section; area density; conductivity.
	τ = a *mean lifetime*; tau lepton.
Υ = an elementary particle.	υ = rarely used (looks like v).
Φ = a *wave function*; an angle.	ϕ = an angle; a wave function.
	χ = susceptibility.
Ψ = a *wave function*.	ψ = a *wave function*.
	ω = *angular frequency* (radians/s).

Table 1.3 Mathematical symbols commonly used in Physics

OPERATORS:

$\rightarrow$ = "...approaches in the limit..." (as in $\Delta t \rightarrow 0$).

∂ = *partial derivative* operator (as in $\frac{\partial F}{\partial x}$).

∇ = *gradient* operator (as in $\nabla\phi = \hat{x}\frac{\partial\phi}{\partial x} + \hat{y}\frac{\partial\phi}{\partial y} + \hat{z}\frac{\partial\phi}{\partial z}$).

$\int$ = *integral* operator as in $\int y(x)dx$

LOGICAL SYMBOLS: (Handy shorthand that I use a lot!)

$\therefore$ = "Therefore..." $\Rightarrow$ = "...implies..." $\equiv$ = "...is *defined* to be..."

$\exists$ = "there exists..." $\ni$ = "...such that..."

/ [a *slash* through any logical symbol] = *negation*; *e.g.* $\not\Rightarrow$ = "...does *not* imply..."

1.4 Functions

Mathematics is often said to be the language of Physics. This is not the whole truth, but it is part of the truth; one ubiquitous characteristic of Physics (the human activity), if not physics (the supposed methodology of nature), is the expression of relationships between measurable quantities in terms of mathematical formulae. The advantages of such notation are that it is concise, precise and "elegant," and that it allows one to calculate quantitative predictions which can be compared with measured experimental results to test the validity of the description.

The nearly-universal image used in such mathematical descriptions of nature is the FUNCTION, an abstract concept conventionally symbolized in the form $y(x)$ [read "y of x"] which formally represents *mathematical shorthand* for a *recipé* whereby a value of the "dependent variable" y can be calculated for any given value of the "independent variable" x.

The *explicit* expression of such a *recipé* is always in the form of an *equation*. For instance, the answer to the question, "What is $y(x)$?" may be "$y = 2 + 5x^2 - 3x^3$." This tells us how to get a numerical value of y to "go with" any value of x we might pick. For this reason, in Mathematics (the human activity) it is often formally convenient to think of a function as a *mapping* — *i.e.* a collection of *pairs* of numbers (x, y) with a concise prescription to tell us how to find the y which goes with each x. In this sense it is also easier to picture the "inverse function" $x(y)$ which tells us how to find a value of x corresponding to a given y. [There is not always a unique answer. Consider $y = x^2$.] On the other hand, whenever we go to *use* an explicit formula for $y(x)$, it is essential to think of it as a *recipé* — *e.g.* for the example described above, "Take the quantity inside the parentheses (whatever it is) and do the following arithmetic on it: first cube whatever-it-is and multiply by 3; save that result and subtract it from the result you get when you multiply 5 by the square of whatever-it-is; finally add 2 to the difference and *voila!* you have the value of y that goes with x = whatever-it-is."

This is most easily understood by working through a few examples, which we will do shortly.

1.4.1 Formulae *vs* Graphs

In Physics we often prefer the image of the GRAPH, because the easiest way to compare *data* with a theoretical function in a holistic manner is to plot both on a common graph. (The right hemisphere is best at holistic perception, so we go right in through the visual cortex.) Fortunately, the issue of whether a graph or an equation is "better" is entirely subjective, because *for every function there is a graph* — although sometimes the interesting features are only obvious when small regions are blown up, or when one or the other variable is plotted on a logarithmic scale, or suchlike.

Nevertheless, this process of translating between left and right hemispheres has far-reaching significance to the practice of Physics. When we draw a *graph*, we cathect the *pattern recognition* skills of our visual cortex, a large region of the brain devoted mainly to forming conceptual models of the "meaning" of visual stimuli arriving through the optic nerve. This is the part that learned to tell the difference between a leaf fluttering in the breeze and the tip of a leopard's tail flicking in anticipation; it performs such pattern recognition without our conscious intervention, and thus falls into the "intuitive" realm of mental functions. It is fantastically powerful, yet not entirely reliable (recall the many sorts of "optical illusions" you have seen).

The mere fact that many (not all) physicists like to display their results in graphical form offers a hint of our preferred procedure for hy-

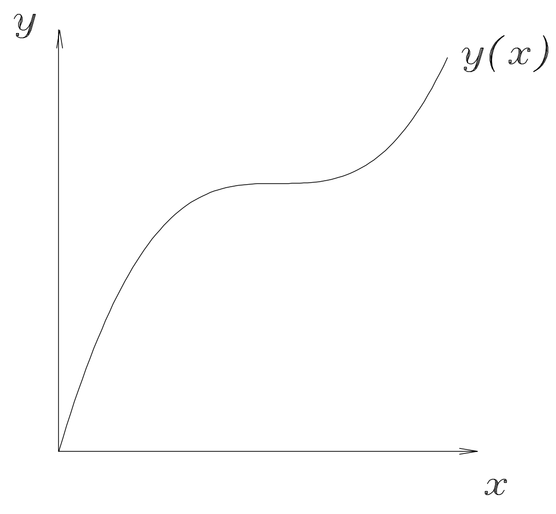


Figure 1.1 A typical graph of $y(x)$ [read "y as a function of x"].

pothesis formation (Karl Popper's *conjectures*). Namely, the data are "massaged" [not the same as "fudged" — massaging is strictly legitimate and all the steps are required to be explained clearly] until they can be plotted on a graph in a form that "speaks for itself" — *i.e.* that excites the strongest pattern-recognition circuit in the part of our visual cortex that we use on science — namely, the straight line. Then the author/speaker can enlist the collaboration of the audience in forming the hypothesis that there is a linear relationship between the two "massaged" variables.

For a simple example, imagine that a force F actually varies inversely with the square of distance r: $F(r) = k/r^2$ with k some appropriate constant. A graph of measured values of F *vs.* r will not be very informative to the eye except to show that, yes, F sure gets smaller fast as r increases. But if the ingenious experimenter discovers by hook or by crook that a plot of F *vs.* $1/r^2$ (or $1/F$ *vs.* r^2 or $\sqrt{F}$ *vs.* $1/r$ or…) comes out looking like a straight line, you can be sure that the data will be presented in that form in the ensuing talk or paper. The rigourous validity of this technique may be questionable, but it works great.

You may have perceived an alarmingly liberal use of *algebra* (or at least algebraic notation) in this last section. I have "pulled no punches" here, showing the "proper" Physics notation for functions and derivatives right at the beginning, for several reasons. First is simple intellectual honesty: this is the mathematical notation used in Physics; why pretend otherwise? Eventually you want to be able to translate this notation into your own favourite representation (words, graphs, whatever) so why not start getting used to it as soon as possible? Second, this is a sort of "implosion therapy" whereby I treat any math phobias by saturating the fear response: once you know it can't get any worse, it starts getting better. Be advised that we will spend the next few chapters (off and on) getting used to algebraic representations and their graphical counterparts.

Chapter 2

Basic Math Topics

2.1 Arithmetic

We have already dwelt upon the formalism of Number Systems in the previous Chapter, where we reminded ourselves that just counting to ten on paper involves a rather sophisticated and elaborate representational scheme that we all learned as children and which is now *tacit* in our thought processes until we go to the trouble of dismantling it and considering possible alternatives.

Arithmetic is the basic algebra of Numbers and builds upon our tacit understanding of their conventional representation. However, it would be emphatically wrong to claim that, "Arithmetic is made up of Numbers, so there is nothing to Arithmetic but Numbers." Obviously Arithmetic treats a new *level* of understanding of the properties of (and the relationships between) Numbers — something like the Frank Lloyd Wright house that was not there *in* the bricks and mortar of which it is built.[1]

We learn Arithmetic at two levels: the *actual* level ("If I have two apples and I get three more apples, then I have five apples, as long as nothing happens to the first two in the meantime.") and the *symbolic* level ("$2 + 3 = 5$"). The former level is of course both *concrete* (as in all the *examples*) and profoundly *abstract* in the sense that one learns to understand that two of anything added to three of the same sort of thing will make five of them, independent of words or numerical symbols. The latter level is more for *communication* (remember, we have to adopt and adapt to a notational convention in order to express our ideas to each other) and for *technology* — *i.e.* for developing *manipulative tricks* to use on Numbers.

Skipping over the simple Arithmetic I assume we all know tacitly, I will use *long division* as an example of the conventional technology of Arithmetic.[2] We all know (today) how to do long division. But can we *explain how it works*? Suppose you were Cultural Attaché to Alpha Centauri IV, where the local intelligent life forms were interested in Earth Math and had just mastered our ridiculous decimal notation. They understand addition, subtraction, multiplication and division perfectly and have developed the necessary skills in Earth-style

[1] One can argue that in fact the conceptual framework of Number Systems implicitly contains intimations of Arithmetic, but this is like arguing that the properties of atoms are implicit in the behaviour of electrons; let's leave that debate for later.

[2] No doubt the useful lifetime of this example is rapidly dwindling, since many students now learn to divide by punching the right buttons on a hand calculator, much to the dismay of their aged instructors. I am not so upset by this — one arithmetic manipulation technology is merely supplanting another — except that "long division" is *in principle* completely understood by its user, whereas few people have any idea what actually goes on inside an electronic calculator. This dependence on mysterious and unfamiliar technology may have unpleasant long-term psychological impact, perhaps making us all more willing to accept the judgements of authority figures without question.... But in Mathematics, as long as you have *once* satisfied yourself completely that some technology is indeed trustworthy and reliable, of course you should make use of it! (Do you *know* that your calculator *always* gives the right answers...?)

gimmicks (carrying, *etc.*) for the first three, but they have no idea how we actually go about dividing one multi-digit number by another. Try to imagine how you would explain the long division trick. Probably by example, right? That's how most of us learn it. Our teacher works out *beaucoup* examples on the blackboard and then gives us *beaucoup* homework problems to work out ourselves, hopefully arrayed in a sequence that sort of leads us through the process of *induction* (not a part of Logic, according to Karl Popper, but an important part of human thinking nonetheless) to a bootstrap grasp of the method. Nowhere, in most cases, does anyone give us a full rigourous derivation of the method, yet we all have a deep confidence in its universality and reliability — which, I hasten to add, I'm sure *can* be rigourously derived if we take the trouble. Still, we are awfully trusting. . . .

The point is, as Michael Polanyi has said, *"We know more than we can tell."* The *tacit* knowledge of Arithmetic that you possess represents an enormous store of

- sophisticated abstract understanding
- arbitrary conventions of representational notation
- manipulative technology

that have already coloured your thought processes in ways that neither you nor anyone else will ever be able to fathom. We are all brainwashed by our Grammar school teachers![3] This little book, if it is of any use whatsoever, will have the same sort of effect: it will "warp" your thinking forever in ways that cannot be anticipated. So if you are worried about being "contaminated" by Scientism (or whatever you choose to label the paradigms of the scientific community) then stop reading immediately before it's too late! (While you're at it, there are a few other activities you will also have to give up. . . .)

[3] It occurs to me that Grammar school is called Grammar school because it is where we learn *grammar* — *i.e.* the *conventional representations* for things, ideas and the relationships between them, whether in verbal language, written language, mathematics, politics, science or social behaviour. These are usually called "rules" or even (when a particularly heavy-handed emphasis is desired) "laws" of notation or manipulation or behaviour. We also pick up a little *technology*, which in this context begins to look pretty innocuous!

2.2 Geometry

In Grammar school we also learn to recognize (and learn the grammar of) geometrical shapes. Thus the "Right Hemisphere"[4] also gets early training. Later on, in High School, we get a bit more insight into the *intrinsic* properties of Euclidean space (*i.e.* the "flat" kind we normally *seem* to be occupying).

2.2.1 Areas of Plane Figures

- The area A of a *square* is the *square* of the length ℓ of any one of its 4 sides: $A = \ell^2$. In fact the question of which word "square" is named after which is a sort of chicken *vs.* egg problem for which there is no logical resolution (even though there may be an historically correct etymological answer).
- The area A of a *rectangle* (a bit more general) is the product of the length b of a long side ("base") and the length h of a short side ("height"): $A = bh$.
- The area A of a *triangle* with base b and height h (measured from the opposite vertex down perpendicular to the base) is $A = \frac{1}{2}bh$. (This is easy to see for a *right* triangle, which is obviously half a rectangle, sliced down the diagonal. You may

[4] The idea that the brain's Left Hemisphere does rational analysis and its Right Hemisphere does intuitive leaps is now considered an incorrect generalization, but I'll use it anyway, just for fun.

want to convice yourself that it is also true for "any old triangle.")

- The area A of a *circle* of radius r is given by $A = \pi r^2$ where π is a number, approximately 3.1415978 [it takes an infinite number of decimal digits to get it exactly; this is because π is an *irrational number*[5] — *i.e.* one which cannot be expressed as a ratio of integers], defined in turn to be the ratio of the *circumference* ℓ of a circle to its *diameter* d: $\pi = \ell/d$ or $\ell = \pi d$.

Were you able to visualize all these simple plane (2-dimensional) shapes "in your head" without resort to actual drawings? If so, you may have a "knack" for geometry, if not Geometry. If it was confusing without the pictures, they are provided in Fig. 2.1 with the appropriate labels.

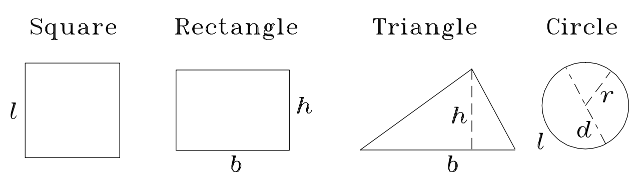


Figure 2.1 A few plane geometrical shapes, with labels.

2.2.2 The Pythagorean Theorem:

> The square of the length of the hypotenuse of a right triangle is equal to the sum of the squares of the lengths of the two shorter sides.

I.e. for the Left Hemisphere we have

$$c^2 = a^2 + b^2 \tag{1}$$

[5] I do not know the proof that π is an irrational number, but I have been told by Mathematicians that it is, and I have never had any cause to question them. In principle, this is reprehensible (shame on me!) but I am not aware of any practical consequences one way or the other; if anyone knows one, please set me straight!

where a, b and c are defined by the labelled picture of a right triangle, shown in Fig. 2.2, which cathects the Right Hemisphere and gets the two working together.

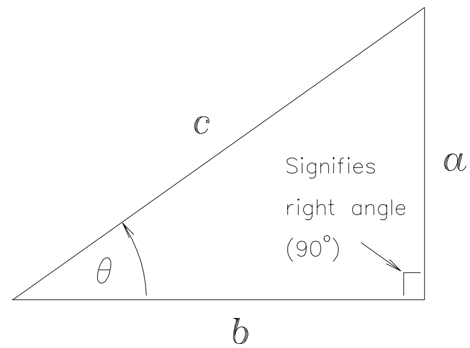


Figure 2.2 A right triangle with hypotenuse c and short sides a and b. The right angle is indicated and the angle θ is defined as shown. Note that a is always the (length of the) side "across from" the *vertex* forming the angle θ. This convention is essential in the *trigonometric* definitions to follow.

2.2.3 Solid Geometry

Most of us learned how to calculate the *volumes* of various solid or 3-dimensional objects even before we were told that the name for the system of conventions and "laws" governing such topics was "Solid Geometry." For instance, there is the *cube*, whose volume V is the *cube* (same chicken/egg problem again) of the length ℓ of one of its 8 *edges*: $V = \ell^3$. Similarly, a *cylinder* has a volume V equal to the product of its cross-sectional area A and its height h perpendicular to the base: $V = Ah$. Note that this works just as well for *any shape* of the cross-section — square, rectangle, triangle, circle or even some irregular oddball shape.

If you were fairly advanced in High School math, you probably learned a bit more abstract or general stuff about solids. But the really deep understanding that (I hope) you brought away with you was an awareness of the *qualitative* difference between 1-dimensional *lengths*, 2-dimensional *areas* and 3-dimensional *volumes*. This awareness can be amazingly powerful even without any "hairy Math details" if you consider what it implies about how these things change with *scale*.[6]

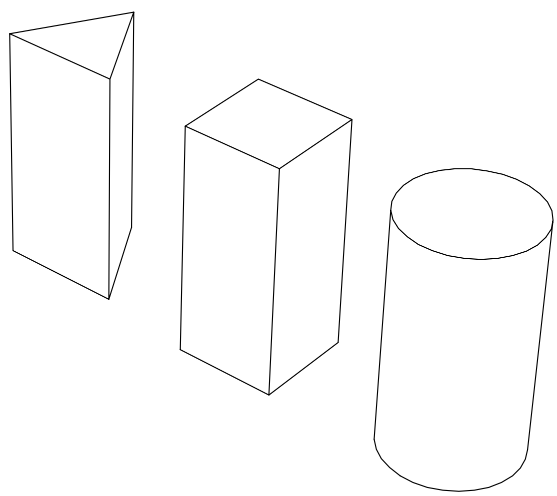


Figure 2.3 Triangular, square and circular right cylinders.

2.3 Trigonometry

Trigonometry is a specialized branch of Geometry in which we pay excruciatingly close attention to the properties of *triangles*, in particular *right* triangles. Referring to Fig. 2.2 again, we define the *sine* of the angle θ (abbreviated $\sin\theta$) to be the ratio of the "far side" a to the hypotenuse c and the *cosine* of θ (abbreviated $\cos\theta$) to be the ratio of the "near side" b to the hypoteneuse c:

$$\sin\theta \equiv \frac{a}{c} \qquad \cos\theta \equiv \frac{b}{c} \tag{2}$$

[6]For instance, it explains easily why the largest animals on Earth have to live in the sea, why insects can lift so many times their own weight, why birds have an easier time flying than airliners, why bubbles form in beer and how the American nuclear power industry got off to a bad start. All in due time. . . .

The other trigonometric functions can easily be defined in terms of the sin and cos:

tangent: $\tan\theta \equiv \frac{a}{b} = \frac{\sin\theta}{\cos\theta}$

cotangent: $\cot\theta \equiv \frac{b}{a} = \frac{\sin\theta}{\cos\theta} = \frac{1}{\tan\theta}$

secant: $\sec\theta \equiv \frac{c}{b} = \frac{1}{\cos\theta}$

cosecant: $\csc\theta \equiv \frac{c}{a} = \frac{1}{\sin\theta}$

For the life of me, I can't imagine why anyone invented the *cotangent*, the *secant* and the *cosecant* — as far as I can tell, they are totally superfluous baggage that just slows you down in any actual calculations. Forget them. [Ahhhh. I have always wanted to say that! Of course you are wise enough to take my advice with a grain of salt, especially is you want to appear clever to Mathematicians. . . .]

The *sine* and *cosine* of θ are our trigonometric workhorses. In no time at all, I will be wanting to think of them as *functions* — *i.e.* when you see "$\cos\theta$" I will want you to say, "cosine of theta" and think of it as $\cos(\theta)$ the same way you think of $y(x)$. Whether as simple ratios or as functions, they have several delightful properties, the most important of which is obvious from the Pythagorean Theorem:[7]

$$\cos^2\theta + \sin^2\theta = 1 \tag{3}$$

where the notation $\sin^2\theta$ means the *square* of $\sin\theta$ — *i.e.* $\sin^2\theta \equiv (\sin\theta)\times(\sin\theta)$ — and similarly for $\cos\theta$. This convention is adopted to avoid confusion, believe it or not. If we wrote "$\sin\theta^{\,2}$" it would be impossible to know for sure whether we meant $\sin(\theta^2)$ or $(\sin\theta)^2$; we could always put parentheses in the right places to

[7]Surely you aren't going to take my word for this! *Convince yourself* that this formula is really true!

remove the ambiguity, but in this case there is a convention instead. (People always have conventions when they are tired of thinking!)

I will need other trigonometric identities later on, but they can wait — why introduce math until we need it? [I have made an obvious exception in this Chapter as a whole only to "jump start" your Mathematical language (re)training.]

Chapter 3

Easy Algebra

Okay, now that you've read the first chapter on Symbols and Abstractions, we're ready to dive into Algebra. I'm going to pretend you are a bright 10-year-old who has never heard of Algebra, since I used exactly this approach on my son when he was 8.[1]

3.1 Adding and Subtracting

I'm thinking of a number, and its name is "x". If $x + 1 = 3$ then what is x? (That is, if we add 1 to x we get 3. What number is one less than three?) Write your answer like this: $\boxed{x = 2}$ (That is, x is **two**!)

OK, now I'm thinking of another number, and its name is "y". If $y + 2 = 4$, what is y?

$\boxed{y = \quad\quad\quad}$

Now I'm thinking of a number whose name is "A". If $A - 1 = 2$, what is A? (That is, if we *subtract* 1 from A we get 2. What number is one *more* than two?) $\boxed{A = \quad\quad\quad}$

I'll stop saying, "I'm thinking of a number, and its name is..." now. Any symbol I use like x or y or A or z or anything that is obviously not a word in a sentence will be meant as a "symbol" for an unknown number that you are supposed to figure out.

If $Z - 3 = 2$, what is Z? $\boxed{\quad\quad\quad}$

[1] Don't be insulted! He now has a Mathematics degree and is working as a Data Scientist, so I reckon something like this works.

If $x + 4 = 9$, what is x? $\boxed{\quad\quad\quad}$

You can do lots of these by yourself. It gets easy fast, if it's not already.

The next step (when I am doing this *in person*) is to ask, "*How do you know*?" The first answer is usually, "I don't know. It's *obvious*." I say, "Yes it is, but suppose you were trying to explain it to someone who didn't believe you: how would you *prove* you were right?" For this we need to agree on some *rules*.

The "algebra rule" for doing this kind of problem is that you can **add** or **subtract** the same thing from both sides of an equation and the equation is still true. (An equation is one of these things with an "equal sign" [=] in the middle, like $x + 4 = 9$.) Go back and check the problems you have done so far and notice that you can get the answer just by adding or subtracting the right number from both sides. For instance, in the first problem ($x + 1 = 3$) we subtract 1 from both sides ($x + 1 - 1 = 3 - 1$) and the $+1$ and -1 on the left side cancel each other out so that both disappear and we have ($x = 3 - 1$). But $3 - 1 = 2$ on the right side, so we have ($x = 2$) which is the answer! You probably can solve these easily "by inspection" without worring about the algebra rule, but it is interesting to know another way.

This works just as well for *other symbols* as it does for numbers: If $x + y = y - 1$, I can subtract y from both sides to get $\boxed{x = -1}$. Simple, eh?

Here's an example of how you would *use* algebra to solve a *practical problem*: if you have ten video games after just buying three, how many did you have before you bought the new ones? Let the answer be x. Then the equation that describes what happened is $x + 3 = 10$, and the solution is $\boxed{x = 7}$.

3.2 Multiplying and Dividing

Now let's do some multiplying and dividing. The algebra convention for multiplication is that any two symbols placed side by side with nothing in between are deemed to be *multiplied together*.[2] Another convention is that *numbers go to the left of algebraic symbols*. (Thus $x2$ is bad notation, while $2x$ is two *times* x.)

Suppose $2x = 4$. What is x? (That is, two *times* x is four. What number do we multiply by 2 to get 4?)

If $3y = 9$, what is y?

If $5A = 20$, what is A?

The trick here (other than just figuring it out by thinking, "Hmm, what number do I have to multiply by 5 to get 20?") is the algebra rule that you can **divide** both sides of an equation by the same thing and the resulting equation is still true. So if I have $2x = 4$, I can divide both sides by 2 which cancels the 2 multiplying x on the left side (leaving just x) and turns the 4 into a 2 on the right side: $x = 2$, the answer!

This also works fine for symbols as well as numbers: Suppose $ax + b = b + a$. Then first I subtract b from both sides, to get $ax = a$. (We always try to use *single characters* for symbols, so that "ax" means a *times* x or the *product* of a and x.) From $ax = a$ we get the answer by dividing both sides by a, which cancels the a multiplying x on the left side and turns the a on the right side into a 1: $\boxed{x = 1}$ is the answer!

Note that *anything*[3] divided by *itself* is 1, as in $a/a = 1$. (The way we write "B divided by C" is B/C, just as for numbers, like $4/2 = 2$.)

Try one for yourself: If $2XY+1 = 2Y+1$, what is X? (This is a little harder because X is multiplied by *both* the number 2 *and* the symbol Y, but you can do it the same way, either by dividing by 2 and then by Y or by dividing by $2Y$ all at once.)

Note that we *don't have to know what Y is*! It "drops out" of this equation! (The answer is $X = 1$ again. Did you get it right?) There are other cases where it *won't* drop out and you get an answer for X "in *terms* of Y". An example would be $xY = 1$, for which $x = 1/Y$ is the answer for x in terms of Y. (We could also "solve" the equation for Y in terms of x: $Y = 1/x$.)

The same thing works backwards, for equations like $X/2 = 3$. There the rule is that you can **multiply** both sides of an equation by the same thing and the resultant equation is still true. In this case we multiply both sides by 2. This cancels the $/2$ on the left side and turns the 3 on the right side into a 6. The answer is $\boxed{X = 6}$, which you could probably get just as easily by inspection ("Half of *what* is 3?"), but again it is nice to know the "rigorous" way of doing it, especially if you want to solve a more difficult equation like

$$\frac{(X-1)}{3} = 1$$

where the parentheses "$(\ldots)$" indicate that what is inside them $(X - 1)$ is *all* divided by 3.

Can you do this one? First multiply both sides by 3, then add 1 to both sides. What is your answer?

[2] Note that this is *not* the convention in computer programming, partly because many variables have multi-character names, but also because the compiler needs to see some kind of explicit instruction for an *operation*.

[3] (other than zero)

How about this? If $(aX + b)/2 = a$, what is X in terms of b?

Here's how this might be *applied* to another simple *practical* problem: Suppose your friend put some money in the bank at 10% *interest*. This means that for every dollar he gives the bank to keep for one year, at the end of the year the bank gives him back the dollar *plus ten cents* "interest" on his dollar. Let's suppose that, at the end of the year, the bank gives him \$220. *How much did he "deposit" [put in the bank] originally?* Let's let D be the number of dollars he deposited a year ago. For each dollar he put in, he gets back one dollar plus 1/10 of another dollar. But there are D dollars to start with, so we must have $D(1+1/10) = 220$. Now we can play a little trick with $1 + 1/10$: Since $1 = 10/10$ [any number divided by itself is 1], $1 + 1/10 = 10/10 + 1/10 = 11/10$ and so our equation reads $11D/10 = 220$. Now we multiply both sides by 10 to get $11D = 2200$ and then divide both sides by 11 to get $\boxed{D = 200}$, which is the answer. Your friend put \$200 in the bank a year ago!

3.3 Square Roots

The *square* of a number means *the number multiplied by itself*. It is written with a little ^{2} just above and to the right, like this: $3^2 = 3\times3 = 9$. (Three threes are nine, right?) The reason it is called the *square* is because that is how we calculate the *area* [size] of a square, flat surface from the length of a side. The area A is equal to the *square* of the length ℓ of a side:

$$A = \ell \times \ell = \ell^2$$

You can see this easily by looking at a picture of four *square pizzas* of different sizes.[4]

Suppose we want to order one of these square pizzas big enough to feed five kids who each want 20 bites. How long should its sides be, if one square inch makes a bite? Well, if each side is ℓ inches long then the total area is ℓ^2 square inches, or ℓ^2 bites. We can make up an equation: $\ell^2 = 5 \times 20 = 100$. What number multiplied by itself is 100? The answer is $\boxed{\ell = 10}$ — the pizza should be ten inches on a side.

[4] Round pizzas obey a similar rule, but it is easier to do the arithmetic for square ones. This is probably why the people who sell pizzas like to sell round ones, to make it harder to figure out how much you are really getting!

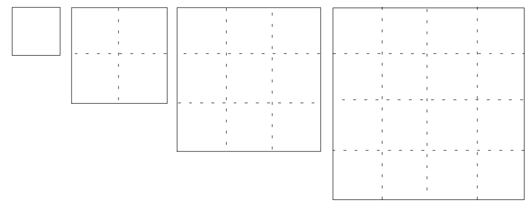


Figure 3.1 Four *square pizzas*. The first, on the left, is one inch on a side, making one small bite. The second is two inches on a side, making four small bites. How many bites in the third, which is three inches on a side?
How about the last one, on the right, which is four inches on a side?

3.4 Positive and Negative Roots

Suppose we have the equation $x^2 = 4$ and we want to know what x is. The equation tells us that $x \times x = 4$. What number multiplied by itself makes 4? Well, $2 \times 2 = 4$, so $x = 2$ is one answer. But there is another one! Suppose you multiply $(-2) \times (-2)$. the rules about multiplying *positive* numbers (like +2) and *negative* numbers (like −2) are as follows:

- A positive times a positive is a positive. $[+ \times + = +]$
- A positive times a negative is a negative. $[+ \times - = -]$
- A negative times a positive is a negative. $[- \times + = -]$

- A negative times a negative is a positive. $[- \times - = +]$

Thus $(-2) \times (-2) = (+4)$ [the square of (-2) is 4] so there are *two* correct solutions to the equation $x^2 = 4$, $x = 2$ and $x = -2$. A short way of writing this is to put the "+" and "−" signs together into a "±" ["plus-or-minus"] sign: $\boxed{x = \pm 2}$ which reads, "x is plus or minus two."

Now you try one: if $Y^2 = 9$, what is Y? $\boxed{\quad\quad}$

How about this: if $A^2 = 16$, what is A? $\boxed{\quad\quad}$

Now let's put together all we have learned so far. If $x^2 - 1 = 15$, what is x? $\boxed{\quad\quad}$

What if $2x^2 = 8$? [Note that $2x^2$ means $2 \times (x^2)$, *not* $(2 \times x)^2$.] $\boxed{\quad\quad}$

There is a special name for "the number which gives a when you multiply it by itself:" it is called the *square root* of a, and it has a special symbol too: $\sqrt{a}$. Thus $\sqrt{a} \times \sqrt{a} = a$ and of course $\sqrt{x^2} = x$. So what we are doing in these problems is using the algebra rule that says you can take the **square root** of both sides of an equation and the resultant equation is still true.

This will also work on a number whose square root is not an integer. [*Integers* are whole numbers like $1, 2, 3, 4, \ldots$, as opposed to fractions or "irrational" numbers, which we will see later.] For instance, $\sqrt{25} = 5$, but $\sqrt{2}$ is not an integer. This does not mean it is not a real number, but it is a very complicated one. In "decimal" notation it is $\sqrt{2} = 1.414213562373\ldots$, where the " ... " represents *an infinite number of significant decimal places*. This is what we call an *irrational* number, but it is still a perfectly good number. If you run across it, just leave it in the form $\sqrt{2}$. That is as good an answer as any. Same for $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$ and so on.

Thus if $x^2 = 3$, then $\boxed{x = \pm\sqrt{3}}$. [Note that we still have to remember those $\pm$ signs!]

You try one: if $x^2 = b$, what is x in terms of b? $\boxed{\quad\quad}$

Now try this: if $a\,x^2 - b = 0$, what is x in terms of a and b? $\boxed{\quad\quad}$

3.5 Quadratic Equations

A *quadratic* equation "in x" means one with x^2 and maybe x in it, but no "higher powers" of x like $x \times x \times x = x^3$. For instance, $x^2 - 1 = 0$ is a quadratic equation in x but $3x - 1 = 8$ is *not*. A quadratic equation in x *must* contain x^2 (possibly multiplied or divided by some number) but it may or may not include x or some constant. Let's look at some of the different kinds of quadratic equations in x.

The first kind we have already seen. These are quadratic equations that have x^2 and some constant (like 3 or b) but no term in x. All the examples so far are of this form. A trivial but important case is $x^2 = 0$, for which the solution is $\boxed{x = 0}$. [Note that we don't have to say ± 0 because $+0$ and -0 are the same thing!]

The second kind has a term proportional to x, as in $3x^2 - 6x = 0$. First we can divide both sides by 3 to get $x^2 - 2x = 0$. [Zero multiplied or divided by anything is still zero!] This "simplifies" the equation a bit. Then we ask ourselves the question, "How can this equation be true?" [That is, *for what values of* x is this equation correct?] Well, there are two ways: first, x can be zero. Then both of the terms on the left side are zero, so they add up to zero and the equation reads $0 = 0$, which is certainly true. This is often called the "trivial solution," $\boxed{x = 0}$, because it is not very interesting; but it is a "true" solution anyway. To see the other solution, suppose we add $2x$ to both sides of the latest equation, to get $x^2 = 2x$. We can then *divide both sides by* x to get $\boxed{x = 2}$, which is the second solution. $[x^2/x = (x \times x)/x = x.]$ Note that this time there is no $\pm$ sign because we are never taking a square root! But there are two answers, just as before.

We call the solutions to a quadratic equation the "roots" of the equation, to suggest the similarity to *square* roots. You try a few:

What are the roots of the equation $x^2 - x = 0$?

$x =$ and $x =$

What are the roots of the equation $5y^2 = 25y$?

$y =$ and $y =$

How about the equation $ax^2 - bx = 0$?

and

The most complicated type of quadratic equation in x is one that has a term proportional to x^2, another term proportional to x and a third *nonzero constant* term. An example would be $x^2 - 2x + 1 = 0$. [When they get complicated, we like to put all the nonzero terms on the left side of the equation and keep a zero on the right.] The trick to solving an equation like this is to see if it can be written as the product of two *pairs* of terms. In this case,

$$\begin{aligned}(x-1) \times (x-1) &= x \times (x-1) - 1 \times (x-1) \\ &= x^2 - x - x + 1 = x^2 - 2x + 1\end{aligned}$$

which means that our original equation is the same as $(x-1)(x-1) = 0$. [We can leave out the $\times$ ("times") symbol when we multiply together two things in parentheses: $(\dots)(\dots) = (\dots) \times (\dots)$.] So the equation will be true only if $(x - 1) = 0$ or if $\boxed{x = 1}$. Note that $x = -1$ is *not* a right answer in this case!

A tougher example would be this:

$$x^2 - 3x + 2 = 0$$

This has two answers; can you find them?

$x =$ and $x =$

A *really* tough example would be

$$2y^2 - 5y + 2 = 0$$

This also has two answers; can you find them?

and

Unlike the other kinds of algebra problems we have done before, these kind rely on "seeing the answer" or making a good guess. There is nothing wrong with this — seeing the answer is always the quickest and easiest way to solve a problem, as long as your guess is *right* ! Sometimes the best approach to a problem is to make a guess and then *check it* to see if it works; you can do this a few times before it gets to be a waste of time. But if after a while your guesses aren't working, you need a more rigourous approach to the problem. Fortunately, this kind of problem has a *general* solution, which is a little complicated, but if you memorize it you can always solve *any* quadratic equation just by "plugging in the parameters" that make it special. We will derive this below, but for now let's think about some more pieces to the puzzle.

3.6 Imaginary Numbers

OK, now you have solved problems where the answer was an *integer* (like 1 or 0 or -7) and problems where the answer was the *ratio* of two integers [the *ratio* of two things means one divided by the other] — which we call a "rational number" because it seems so reasonable — and you have even solved problems where the answer was an "irrational number" [one that *cannot* be expressed as the ratio of any two integers], like $\sqrt{2}$. All these different kinds of numbers are *real* numbers, even if they seem pretty unusual, because you can do arithmetic with them in the usual way. You have also seen that *symbols* like A or b can be used in answers just as if they were numbers, which lets you solve a whole bunch of different problems "formally" at the same time. Now it's time to tackle something really weird.

A long time ago, someone was telling her friends about this algebra stuff, showing that both $x = +1$ and $x = -1$ were solutions to the equation $x^2 = 1$, and someone said, "What about $x^2 = -1$? What would be the solution to that equation?" Well, all the other mathematicians laughed and said, "Boy, are you dumb! There *isn't any real number* that gives -1 when you multiply it by itself!" The first mathematician thought about it for a moment and said, "That is true, but wouldn't it be nice if there *were* such a number? You could solve *any* quadratic equation then!"

Her friends all laughed again and said, "Boy, do you have a great imagination!" And as they walked off laughing, she thought to herself, "OK, maybe I do; why not *call* the square root of -1 an *imaginary* number? I can even give it a special name "i" to signify that it is $i$maginary!" And she started figuring out how i would behave.

[This isn't really how i was invented, but it makes a good story!]

Anyway, if we have a solution to the equation $x^2 = -1$, then we can indeed solve *any* quadratic equation. For instance, because the square root of a product is the product of the square roots, like

$$\sqrt{36} = \sqrt{4 \times 9} = \sqrt{4} \times \sqrt{9} = \pm 2 \times \pm 3 = \pm 6$$

we can *separate out* i like this:

$$\sqrt{-4} = \sqrt{4 \times (-1)} = \sqrt{4} \times \sqrt{-1} = \pm 2\,i$$

where $2i$ means $2 \times i$ just as for any other symbol like $2x \equiv 2 \times x$. [The symbol "$\equiv$" is used as shorthand for "...means the same thing as...."]

Let's try a few. If $x^2 = -9$, what is x?

$x = \pm$

If $y^2 = -16$, what is y? | $y =$ |

What if $z^2 - 1 = -37$? | |

Great! Now (with a lot of work) we can actually write down the answer to *all possible quadratic equations* in *one formula*!

3.7 The Quadratic Theorem

The *most general possible quadratic equation* (after we have put all the terms on the lefthand side) looks like this:

$$\boxed{a\,x^2 + b\,x + c = 0}$$

where a, b and c are symbols representing some numbers that go into the equation. For instance, $4x^2 + 3x + 1 = 0$ would be a case where $a = 4$, $b = 3$ and $c = 1$. For another example, $x^2 - 1 = 0$, would be a case with $a = 1$, $b = 0$ and $c = -1$. So if we can "solve" this equation for x *in terms of* a, b and c, then we will be able to quickly convert the result into the specific solution for whatever quadratic equation we want to solve, just by substituting a, b and c into the result! Let's try.

First let's divide through by a to get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

Next we subtract $\frac{c}{a}$ from both sides to get

$$x^2 + \frac{b}{a}x = -\frac{c}{a}$$

Now we play an important trick known as "completing the square:" consider the equation $(x + d)^2 = x^2 + 2dx + d^2$. If we subtract d^2 from both sides we get $(x + d)^2 - d^2 = x^2 + 2dx$, which looks a lot like the left side of our previous equation, if only $2d$ were the same thing as $\frac{b}{a}$ — that is, if only $d = \frac{b}{2a}$. Well, let's put that in!

$$x^2 + \frac{b}{a}x = \left(x + \frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2$$

But the right side of this equation must be equal to the right side of the old equation:

$$\left(x + \frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 = -\frac{c}{a}$$

and if we add $\left(\frac{b}{2a}\right)^2$ to both sides we get

$$\left(x + \frac{b}{2a}\right)^2 = \left(\frac{b}{2a}\right)^2 - \frac{c}{a}.$$

Now we can take the square root of both sides [remembering our $\pm$ sign] and get

$$x + \frac{b}{2a} = \pm\sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}}$$

— from both sides of which we subtract $\frac{b}{2a}$ to get

$$x = -\frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}},$$

which is the answer, but still sort of messy. We can simplify a little by noting that

$$\frac{c}{a} = \frac{4ac}{4a^2} \qquad \text{and} \qquad \left(\frac{b}{2a}\right)^2 = \frac{b^2}{4a^2},$$

$$\text{so} \qquad \left(\frac{b}{2a}\right)^2 - \frac{c}{a} = \frac{b^2 - 4ac}{4a^2}$$

and therefore

$$\sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}} = \frac{\sqrt{b^2 - 4ac}}{\sqrt{4a^2}} = \frac{\sqrt{b^2 - 4ac}}{2a}$$

so that our answer now reads

$$x = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$$

or just

$$\boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

which is known as THE QUADRATIC THEOREM.

This formula, messy as it may look, is worth remembering, because you can use it to solve *any* quadratic equation! First you put your equation into the form $ax^2 + bx + c = 0$ and figure out what the values of a, b and c are. then you plug those values of a, b and c into the QUADRATIC THEOREM and out pops the answer!

Let's do an example. Suppose we want to solve $2y^2 - 5y + 2 = 0$ for y. This is in the standard form with $a = 2$, $b = -5$ and $c = 2$. The answer is thus

$$\begin{aligned} y &= \frac{+5 \pm \sqrt{25 - 4 \times 2 \times 2}}{2 \times 2} \\ &= \frac{5 \pm \sqrt{25 - 16}}{4} = \frac{5 \pm \sqrt{9}}{4} \\ &= \frac{5 \pm 3}{4} \end{aligned}$$

so there are two answers, $\boxed{y = \frac{8}{4} = 2}$ and $\boxed{y = \frac{2}{4} = \frac{1}{2}}$.

You can go back over all the examples mentioned so far and show that each one is a case of the QUADRATIC THEOREM for some values of a, b and c.

Go ahead and think up lots of cases on your own! As long as you put them in the standard form $ax^2 + bx + c = 0$, you can always get the answer!

Note that if $4ac$ is larger than b^2 you will have imaginary numbers in the answer — that makes it a "COMPLEX NUMBER".

Chapter 4

"Formal" Algebra

Now that you've seen a highly *informal* introduction to most of Algebra, let's cover exactly the same material the way you might see it from someone with a more pedantic respect for the subject....

In Algebra we learn to "solve" equations. What does that *mean*? Usually it means that we are to take a (relatively) complicated equation that has the "unknown" (often, but not always, called "x") scattered all over the place and turn it into a (relatively) simple equation with x on the left-hand side *by itself* and a bunch of other symbols (*not* including x) on the right-hand side of the "=" sign. Obviously this particular *format* is "just" a convention. But the *idea* is independent of the representation: we want to "solve" for the "unknown" quantity, in this case x, in terms of whatever else is in the equation: *numbers* like $1, 2, 3 \ldots$ or named *constants* like $a, b, c \ldots$.

4.1 Operations and Notation

Most algebra involves only a few simple operations:

- **Equality:** If we write $a = b$ we are saying that a and b are the same *kind* of thing and are *exactly the same size*.
- **Equivalence:** If we write $a \equiv b$, we are saying that a and b are *the same thing*. This may sound like the same thing as *equality*, but it's actually much stronger. Below it will be applied to equivalent *notations*.[1]
- **Addition:** If a and b are entities of the same type (usually just numbers), we can **add** them together as $a + b$ to get a new number or entity of the same type, called their **sum**. Example: $1 + 2 = 3$.
- **Subtraction:** By the same token, we can **subtract** b from a to get their **difference**, $a - b$. Example: $3 - 1 = 2$.
- **Multiplication:** The **product** of a and b is written as either $a \times b$ or $a \cdot b$ or just ab, with the understanding that each entity is represented by a single character. Example: $2 \times 3 = 6$.
- **Division:** Just as *subtraction* is sort of the opposite of *addition*, *division* (written $\frac{a}{b}$ or a/b or $a \div b$) is sort of the opposite of *multiplication*. Example: $\frac{6}{2} \equiv 6/2 \equiv 6 \div 2 = 3$.
- **Powers:** We can multiply a by *itself* n times (also called "raising a to the power

[1] The above operations are described using formal Mathematical notation and symbols that are easy to write longhand or typeset using LaTeX, but are notoriously difficult to express neatly in `HTML` or in the restricted `ASCII` character set used by computers. Most computer programming languages have `ASCII` conventions equivalent to the Mathematical symbols and operations. For instance, *equivalence* ($a \equiv b$) is expressed `A == B` in most programming languages; *multiplication* ($a \times b$) is written `A*B`; *division* is simply `A/B`; *powers* are almost universally expressed as `A**N` and *roots* as `B**(1/N)`. The Algebra tutorial is (of necessity) expressed in these forms.

n) to get a^n. Example: $2^5 = 16$.

- **Roots:** In something like the opposite of raising a to the power n, we can find the n^{th} **root** of b, written $\sqrt[n]{b} \equiv b^{1/n}$. Example: $16^{1/5} \equiv \sqrt[5]{16} = 2$.

4.2 LAWS

There are a few basic rules we use to "solve" problems in Algebra; these are called "laws" by Mathematicians who want to emphasize that you are not to question their content or representation.

- **Definition of Zero:**

$$a - a = 0 \tag{1}$$

- **Negative Values:** Along with the definition of zero, the *subtraction* operation allows us to assign a *negative value* to the expression $-a$ taken as a separate entity. That means we can think of $a - a$ as $a + (-a)$; *i.e.* if we *add* $-a$ to a we get zero again. Note that $-(-a) = a$: in mathematics, a *double negative* really **is** positive. Thus if a *itself* has a negative value, $-a$ is *positive*, and $a - a = 0$ still holds.[2]

- **Definition of Unity:**

$$\frac{a}{a} = 1 \tag{2}$$

- **Commutative Laws:**[3]

$$a + b = b + a \tag{3}$$

$$\text{and} \quad ab = ba \tag{4}$$

[2] Does this seem a bit circular? Right you are! It is!

[3] Note that *division* is *not* commutative: $a/b \neq b/a$! Neither is *subtraction*, for that matter: $a-b \neq b-a$. The Commutative Law for *multiplication*, $ab = ba$, holds for ordinary numbers (real and imaginary) but it does *not* necessarily hold for all the mathematical "things" for which some form of "multiplication" is defined! For instance, the *group* of *rotation operators* in 3-dimensional space is *not* commutative — think about making two successive rotations of a rigid object about perpendicular axes in different order and you will see that the final result is different! This seemingly obscure property turns out to have fundamental significance.

- **Distributive Law:**

$$a(b + c) = ab + ac \tag{5}$$

- **Sum or Difference of Two Equations:** Adding (or subtracting) the same thing from both sides of an equation gives a new equation that is still OK.

$$\begin{array}{rrcl} & x \;\; -a = & b & \\ +\,(& a = & a &) \\ & x \quad\;\; = & b + a & \end{array} \tag{6}$$

$$\begin{array}{rrcl} & x \;\; +c = & d & \\ -\,(& c = & c &) \\ & x \quad\;\; = & d - c & \end{array} \tag{7}$$

- **Product or Ratio of Two Equations:** Multiplying (or dividing) both sides of an equation by the same thing also gives a new equation that is still OK.

$$\begin{array}{rrcl} & x/a = & b & \\ \times\,(& a = & a &) \\ & x = & ab & \end{array} \tag{8}$$

$$\begin{array}{rrcl} & cx = & d & \\ \div\,(& c = & c &) \\ & x = & d/c & \end{array} \tag{9}$$

- **Imaginary and Complex Numbers:** So far we have limited ourselves to the *real numbers*. In that domain, $\sqrt{-1}$ is undefined: there is no real number that will yield -1 when squared. One imagines a particularly persistent student insisting, "But what if there *were* such a number?" The teacher would grumble, "You certainly have an active imagination!" And the student would say, 'Fine. Let's call it an *imaginary* number, and call it "i" for short!' The inclusion of multiples of i more than doubles the domain of algebra, since it means we can also have *combinations* of real and imaginary numbers, $z = a + ib$. These are called *complex* numbers.

These "laws" may seem pretty trivial (especially the first two) but they define the rules of Algebra whereby we learn to manipulate the form of equations and "solve" Algebra "problems." We quickly learn equivalent *shortcuts* like "moving a factor from the bottom of the left-hand-side [often abbreviated LHS] to the top of the right-hand side [RHS]:"

$$\frac{x-a}{b} = c + d \quad \Rightarrow \quad x - a = b(c+d) \tag{10}$$

and so on; but each of these is just a well-justified concatenation of several of the fundamental steps.

You may ask, "Why go to so much trouble to express the obvious in such formal terms?" Well, as usual the obvious is not necessarily the truth. While the real, imaginary and complex numbers may all obey these simple rules, there are perfectly legitimate and useful fields of "things" (usually some sort of *operators*) that do *not* obey all these rules, as we may see later. It is generally a good idea to be aware of your own assumptions; we haven't the time to keep reexamining them constantly, so we try to state them as plainly as we can and keep them around for reference "just in case. . . ."

4.3 The Quadratic Theorem

"I'm thinking of a number, and its name is 'x' . . ." So if

$$ax^2 + bx + c = 0, \tag{11}$$

what is x? Well, we can only say, "It depends." Namely, it depends on the values of a, b and c, whatever they are. Let's suppose the *dimensions* of all these "parameters" are mutually consistent[4] so that the equation makes sense. Then "it can be shown" (a classic phrase if there ever was one!) that the "answer" is *generally*[5]

$$\boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}} \tag{12}$$

This formula (and the preceding equation that defines what we mean by a, b and c) is known as the QUADRATIC THEOREM, so called because it offers "the answer" to *any* quadratic equation (*i.e.* one containing powers of x up to and including x^2). The power of such a *general* solution is prodigious. (Work out a few examples! There is a limited version implemented in the Algebra tutorial.)

It also introduces an interesting new way of looking at the relationship between x and the *parameters* a, b and c that determine its value(s). Having x all by itself on one side of the equation and no x's anywhere on the other side is what we call a "solution" in Algebra. Let's make a compact version of this sort of equation:

"I'm thinking of a number, and its name is 'y' . . ." So if $y = f(x)$, what is y? The answer is again, "It depends!" [In this case, upon the value of x and the detailed form of the *function* $f(x)$] . . . and that leads us into a new subject: CALCULUS!

[4] In Mathematics we never worry about such things; all our symbols represent *pure numbers*; but in Physics we *usually* have to express the value of some physical quantity in units which make sense and are consistent with the units of other physical quantities symbolized in the same equation!

[5] The ± symbol means that *both* signs (+ and −) should represent legitimate answers.

Chapter 5

Easy Calculus

In a *stylistic* sense, Algebra starts to become Calculus when we write the preceding example, $y = x^2$, in the form

$$y(x) = x^2$$

which we read as "y of x equals x squared." This is how we signal that we mean to think of y as a *function* of x, and right away we are leading into the terminology of Calculus. Recall the final sections of the preceding Chapter.

However, Calculus *really* begins when we start talking about the *rate of change* of y as x varies.

5.1 Rates of Change

One thing that is easy to "read off a graph" of $y(x)$ is the *slope* of the curve at any given point x. Now, if $y(x)$ is quite "curved" at the point of interest, it may seem contradictory to speak of its "slope," a property of a *straight* line. However, it is easy to see that as long as the curve is *smooth* it will always *look like a straight line* under sufficiently high *magnification*. This is illustrated in Fig. 5.1 for a typical $y(x)$ by a process of successive magnifications.

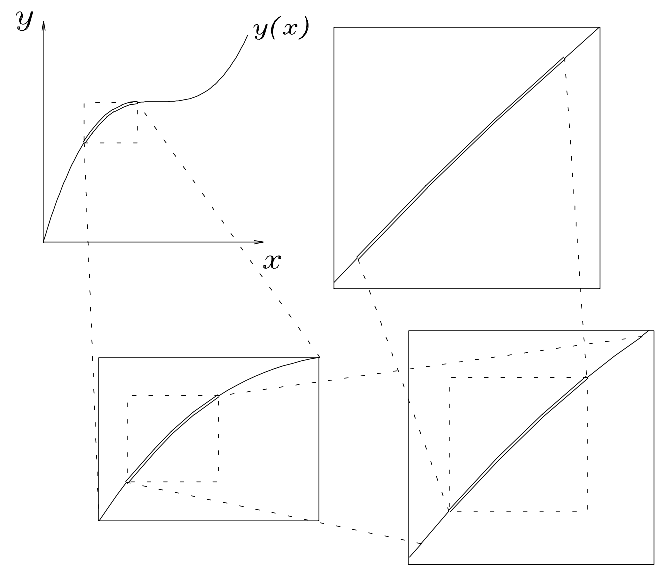


Figure 5.1 A series of "zooms" on a segment of the curve $y(x)$ showing how the *curved* line begins to look more and more like a *straight* line under higher and higher magnification.

We can also prescribe an algebraic method for *calculating* the slope, as illustrated in Fig. 5.2: the *definition* of the "slope" is the ratio of the increase in y to the increase in x on a vanishingly small interval. That is, when x goes from its initial value x_0 to a slightly larger value $x_0+\Delta x$, the curve carries y from its initial value $y_0 = y(x_0)$ to a new value $y_0+\Delta y = y(x_0+\Delta x)$, and the slope of the curve at $x = x_0$ is given by $\Delta y/\Delta x$ for a vanishingly small Δx. When a small change like Δx gets *really* small (*i.e.* small enough that the curve looks like a straight line on that interval, or "small enough to satisfy whatever criterion you want," then we write it differently, as dx, a "*differential*" (vanishingly small) change in x. Then the exact definition of the SLOPE of y with respect to x at some particular value of x, written in conventional

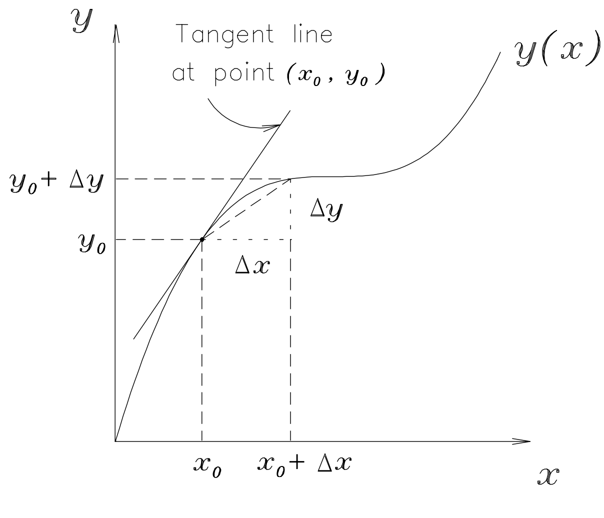


Figure 5.2 A graph of the function $y(x)$ showing how the *average* slope $\Delta y/\Delta x$ is obtained on a *finite* interval of the curve. By taking smaller and smaller intervals, one can eventually obtain the slope at a *point*, dy/dx.

Physics[1] language, is

$$\frac{dy}{dx} \equiv \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \equiv \lim_{\Delta x \to 0} \frac{y(x+\Delta x) - y(x)}{\Delta x} \quad (1)$$

This is best understood by an example: consider the simple function $y(x) = x^2$. Then

$$y(x+\Delta x) = (x+\Delta x)^2 = x^2 + 2x\Delta x + (\Delta x)^2$$

$$\text{and} \quad y(x+\Delta x) - y(x) = 2x\Delta x + (\Delta x)^2.$$

Divide this by Δx and we have

$$\frac{\Delta y}{\Delta x} = 2x + \Delta x.$$

Now let Δx shrink to zero, and all that remains is

$$\frac{\Delta y}{\Delta x} \underset{\Delta x \to 0}{\longrightarrow} \frac{dy}{dx} = 2x.$$

Thus the slope [or *derivative*, as mathematicians are wont to call it] of $y(x) = x^2$ is $dy/dx = 2x$. That is, the slope increases linearly with x. The slope *of* the slope — which we call[2] the *curvature*, for obvious reasons — is then trivially $d(dy/dx)/dx \equiv d^2y/dx^2 = 2$, a constant. Make sure you can work this part out for yourself.

We have defined all these algebraic solutions to the geometrical problem of finding the slope of a curve on a graph in completely abstract terms — "x" and "y" indeed! What *are* x and y? Well, the whole idea is that they can be anything you want! The most common examples in Physics are when x is the *elapsed time*, usually written t, and y is the *distance travelled*, usually (alas) written x. Thus in an elementary Physics context the function you are apt to see used most often is $x(t)$, the position of some object as a function of time. This particular function has some very well-known derivatives, namely $dx/dt = v$, the *speed* or (as long as the motion is in a straight line!) *velocity* of the object; and $dv/dt \equiv d^2x/dt^2 = a$, the *acceleration* of the object. Note that both v and a are themselves (in general) functions of time: $v(t)$ and $a(t)$. This example so beautifully illustrates the "meaning" of the slope and curvature of a curve as first and second derivatives that many introductory Calculus courses and virtually all introductory Physics courses use it as *the* example to explain these Mathematical conventions. I just had to be different and start with something a little more formal, because I think you will find that the idea of one thing being a *function* of another thing, and the associated ideas of graphs and slopes and curvatures, are handy notions worth putting to work far from their traditional realm of classical kinematics.

[1] Real Mathematicians prefer the "primed" notation, $dy/dx \equiv y'(x)$, for several reasons: first, it reminds us that dy/dx is also a *function of* x; the second reason will be obvious a little later....

[2] This differs from the conventional mathematical definition of *curvature*, $\kappa \equiv d\phi/ds$, where ϕ is the tangential angle and s is the arc length, but I like mine better, because it's simple, intuitive and useful. (OK, I'm a Philistine. So shoot me. ;-) Thanks to Mitchell Timin for pointing this out.

5.2 Second Derivatives

How about the rate of change **of** the rate of change? I slipped this in surreptitiously above when I defined the *curvature*,

$$\frac{d}{dx}\frac{dy}{dx} \equiv \frac{d^2y}{dx^2}$$

where the left hand side now explicitly displays the OPERATOR d/dx which means, "take the derivative with respect to x of whatever appears immediately to the right." (We will encounter other operators later on, so it's important to get used to this idea.)

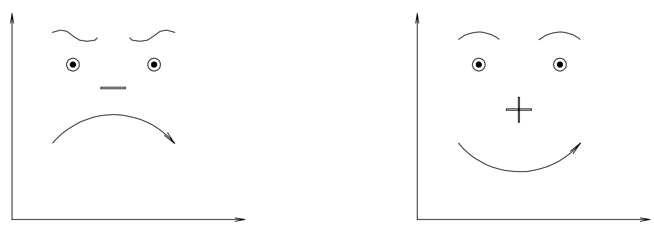


Figure 5.3 A graph of two functions, $y_-(x)$ [left] and $y_+(x)$ [right], having negative and positive curvature d^2y/dx^2, respectively. The frivolous cartoon format is an easy way to remember that a *negative* second derivative "curves downward" to make a *convex* "frowney face" whereas a *positive* second derivative "curves upward" to make a *concave* "smiley face".

In the prime Physics example where the vertical axis is *distance* and the horizontal axis is *time*, the *concave* graph corresponds to *acceleration* (speeding up of the speed) and the *convex* graph corresponds to *deceleration* (slowing down).

5.3 Higher Derivatives

One can, of course, take the derivative of the derivative of the derivative,

$$\frac{d}{dt}\frac{d}{dt}\frac{dx}{dt} \equiv \frac{d^3x}{dt^3},$$

a.k.a. (in Physics) as the "*jerk*". (No, I'm not kidding.) In Physics we rarely go this far, because Newton's Second Law relates the *second* time derivative of distance (the *acceleration*) to the *mass* of a body and the *force* applied to it. But Mathematicians know no such restraint. They will happily refer to the n^{th} derivative,

$$\frac{d^ny}{dx^n}$$

which has the d/dx operator applied n times to $y(x)$. Later on we will encounter a function for which the n^{th} derivative of $y(x)$ is both nonzero and as simple as $y(x)$ itself — in fact, for which

$$\frac{d^ny}{dx^n} = y(x).$$

Stay tuned. . . .[3]

5.4 Integrals

Suppose that y is the number of new COVID-19 cases per day and x is time in units of days. We have all seen many curves like this in 2020. Then the total number of COVID-19 cases between day x_0 and day x_1 is given by

$$\int_{x_0}^{x_1} y(x)dx$$

(read "the integral of $y(x)$ with respect to x from x_0 to x_1"), whose rigorous, formal meaning is simply the **area under the curve** of $y(x)$ from x_0 to x_1.

The usual approach to evaluating this quantity is to break the area up into a large number of very skinny vertical rectangles of very narrow width Δx and height $y(x)$ and then let $\Delta x \to 0$ as the number of tall skinny rectangles

[3] This would be a good time to remind you that Real Mathematicians prefer the notation $y'(x)$ instead of dy/dx. What do they use for the second derivative, d^2y/dx^2? Not surprisingly, they use $y''(x)$. For higher derivatives, I think the Physics notation d^5y/dx^5 is clearly preferable to the Mathematics notation $y'''''(x)$.

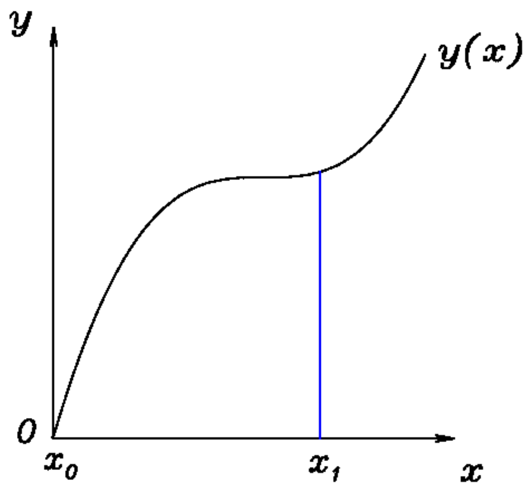


Figure 5.4 What is the **area under the curve** of $y(x)$?

becomes infinite. Although this formulation is easy to evaluate numerically on a computer, it does not lend itself to fun handwaving explanations that yield simple algebraic answers, so I'll be using the idea of *antiderivatives* — generally disdained by Real Mathematicians — to make it easier. Stay tuned.

Chapter 6

Derivatives

6.1 Definition

Recall the definition of the derivative: the rate of change [slope] of a function at a point is the limiting value of its *average* slope over an interval including that point, as the width of the interval shrinks to zero:

$$\boxed{\frac{dy}{dx} \equiv \lim_{\Delta x \to 0} \frac{y(x+\Delta x) - y(x)}{\Delta x}}$$

All the remaining Laws and Rules can be proven by algebraic manipulation of this definition.

6.2 Operator Notation

The symbol $\frac{d}{dx}$ (read "derivative with respect to x") can be thought of as a mathematical "verb" (called an *operator*) which "operates on" whatever we place to its *right*. Thus[1]

$$\frac{d}{dx}[y] \equiv \frac{dy}{dx}$$

[1] I should take this opportunity to emphasize the difference between the "=" *equals* sign (meaning the thing on the left is the same *size* as the thing on the right) and the "≡" *equivalence* sign (meaning the thing on the left is by definition the *same thing* as that on the right). The latter is like the "==" *definition* operator in many programming languages.

6.3 Product Rule

The derivative of the *product* of two functions is *not* the product of their derivatives! Instead,

$$\boxed{\frac{d}{dx}[f(x)\cdot g(x)] = \frac{df}{dx}\cdot g(x) + f(x)\cdot\frac{dg}{dx}}$$

Proof (Physicist's notation):
If $y(x) = f(x)\cdot g(x)$ then

$$\begin{aligned} y(x+\Delta x) &= f(x+\Delta x)\cdot g(x+\Delta x) \\ &= \left[f(x) + \frac{df}{dx}\Delta x\right]\left[g(x) + \frac{dg}{dx}\Delta x\right] \\ &= f(x)\cdot g(x) + \left[\frac{df}{dx}\cdot g(x) + f(x)\cdot\frac{dg}{dx}\right]\Delta x \\ &\quad +[\Delta x]^2\frac{df}{dx}\cdot\frac{dg}{dx} \end{aligned}$$

Divide this through by Δx and we have

$$\frac{y(x+\Delta x)}{\Delta x} = \frac{y(x)}{\Delta x} + \frac{df}{dx}\cdot g(x) + f(x)\cdot\frac{dg}{dx} + \Delta x\cdot\frac{df}{dx}\cdot\frac{dg}{dx}$$

Note that $y(x+\Delta x) - y(x) = \Delta y$ and let Δx shrink to zero, and all that remains is

$$\frac{\Delta y}{\Delta x} \underset{\Delta x\to 0}{\longrightarrow} \frac{dy}{dx} = \frac{df}{dx}\cdot g(x) + f(x)\cdot\frac{dg}{dx} \; . \; \mathcal{QED}$$

Now, you may find the expression for the change in $f(x)$,

$$\Delta f = \frac{df}{dx}\cdot\Delta x,$$

a little confusing: there's a Δx in the numerator and a dx in the denominator — which is which, and if we're going to make $\Delta x \to 0$ later, why not do it now and just cancel the two? We can't do that, and rather than try to explain why, I'll switch to Mathematician's notation, for the same reason *they* do!

Proof (Mathematician's notation):

If $y(x) = f(x) \cdot g(x)$ then

$$y(x + \Delta x) = f(x + \Delta x) \cdot g(x + \Delta x)$$

$$= [f(x) + f'(x) \cdot \Delta x]\,[g(x) + g'(x) \cdot \Delta x]$$

$$= f(x) \cdot g(x) + [f'(x) \cdot g(x) + f(x) \cdot g'(x)]\,\Delta x$$

$$+[\Delta x]^2 f'(x) \cdot g'(x)$$

Divide this through by Δx and we have

$$\frac{y(x + \Delta x)}{\Delta x} = \frac{y(x)}{\Delta x} + f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

$$+\Delta x \cdot f'(x) \cdot g'(x)$$

Note that $y(x + \Delta x) - y(x) = \Delta y$ and let Δx shrink to zero, and all that remains is

$$\boxed{\frac{\Delta y}{\Delta x} \underset{\Delta x \to 0}{\longrightarrow} y'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)\ .}$$

So it's true, whichever way you express it!

6.4 Examples

- **Constant times a Function:** Since the derivative of a *constant* is always *zero* (it *doesn't change*), the PRODUCT RULE gives

$$\frac{d}{dx}\,[a \cdot y(x)] \;=\; a \cdot \frac{dy}{dx}$$

where a is any constant (*i.e.* not a function of x). This is sometimes referred to as "pulling the constant factor outside the derivative."

- **Power Law:** The simplest class of derivatives are those of power-law functions, $y(x) = x^p$. We have derived the result for $p = 2$ earlier; for $p = 3$ we have $y(x) = x \cdot x^2$, and since $dx/dx = 1$, the PRODUCT RULE gives

$$\frac{d}{dx}\left[x^3\right] \;=\; x \cdot 2\,x + 1 \cdot x^2 \;=\; 3x^2$$

Using the same trick, you can easily show that $dy/dx = 4x^3$ for $y(x) = x^4$, $dy/dx = 5x^4$ for $y(x) = x^5$, and so on for all integer values of p. It turns out that the general result

$$\boxed{\frac{d}{dx}\,[x^p] \;=\; p\,x^{p-1}}$$

is valid for *all* powers p, whether positive, negative, integer, rational, irrational, real, imaginary or complex. That's a little harder to prove, but you can look it up on *Wikipedia*.

- **Function of a Function:** Suppose y is a function of x and x is in turn a function of t. Then if t changes by Δt, x changes by

$$\Delta x = \frac{dx}{dt} \cdot \Delta t$$

and y changes by

$$\Delta y = \frac{dy}{dx} \cdot \Delta x = \frac{dy}{dx} \cdot \frac{dx}{dt} \cdot \Delta t.$$

Dividing both sides by Δt gives

$$\frac{\Delta y}{\Delta t} = \frac{dy}{dx} \cdot \frac{dx}{dt}$$

and if we let $\Delta t \to 0$ we get

$$\boxed{\frac{d}{dt}\,\{y[x(t)]\} \;=\; \frac{dy}{dx} \cdot \frac{dx}{dt}}$$

(CHAIN RULE)

Chapter 7

The Exponential Function

Having dabbled in *derivatives*,[1] before we go on to *integrals*, I'd like to introduce a peculiar function that figures centrally in most of Physics and is (IMNERHO) a Mathematical Miracle. I'm going to approach it from a decidedly **non**-Physicist perspective.

Suppose the newspaper headlines read, "The cost of living went up 10% this year." Can we translate this information into an *equation*? Let "V" denote the value of a dollar, in terms of the "real goods" it can buy — whatever economists mean by that. Let the elapsed time t be measured in years (y). Then suppose that V is a function of t, $V(t)$, which function we would like to know explicitly. Call *now* "$t = 0$" and let the initial value of the dollar (now) be V_0, which we could take to be \$1.00 if we disregard inflation at earlier times.[2]

Then our news item can be written

$$V(0) = V_0$$

and

$$V(1\,\mathrm{y}) = (1 - 0.1)\, V_0 = 0.9\, V_0.$$

This formula can be rewritten in terms of the *changes* in the dependent and independent variables, $\Delta V = V(1\,\mathrm{y}) - V(0)$ and $\Delta t = 1\,\mathrm{y}$:

$$\frac{\Delta V}{\Delta t} = -0.1\, V_0, \tag{1}$$

where it is now to be *understood* that V is measured in "1998 dollars" and t is measured in years. That is, the average *time rate of change* of V is proportional to the value of V at the beginning of the time interval, and the constant of proportionality is $-0.1\ \mathrm{y}^{-1}$. (By y^{-1} or "inverse years" we mean the *per year* rate of change.)

This is almost like a derivative. If only Δt were infinitesimally small, it would *be* a derivative. Since we're just trying to describe the qualitative behaviour, let's make an *approximation*: assume that $\Delta t = 1$ year is "close enough" to an infinitesimal time interval, and that the above formula (1) for the inflation rate can be turned into an *instantaneous* rate of change:[3]

$$\frac{dV}{dt} = -0.1\, V. \tag{2}$$

This means that the dollar in your pocket right now will be worth only \$0.99999996829 in one second.

Well, this is interesting, but we cannot go any further with it until we ask a crucial question: "What will happen if this goes on?" That is, suppose we assume that equation (2) is not just a temporary situation, but *represents a consistent and ubiquitous property* of the function $V(t)$, the "real value" of your dollar bill as a function of time.[4]

[1] (the mathematical kind, not the bogus financial "instruments" that brought on the Global Financial Crisis of 2008!)

[2] Since our dollar will be worth *less* a year from now, we should really call it **de**flation!

[3] The error introduced by this approximation is not very serious.

[4] Banks, insurance companies, trade unions, and governments all pretend that they don't assume this, but they would all go bankrupt if they *didn't* assume it.

Applying the d/dt "operator" to both sides of Eq. (2) gives

$$\frac{d}{dt}\left(\frac{dV}{dt}\right) = \frac{d}{dt}(-0.1\,V)$$

or

$$\frac{d^2V}{dt^2} = -0.1\,\frac{dV}{dt}. \tag{3}$$

But dV/dt is given by (2). If we substitute that formula into the above equation (3), we get

$$\frac{d^2V}{dt^2} = (-0.1)^2\,V = 0.01\,V. \tag{4}$$

That is, the rate of change of the rate of change is always positive, or the (negative) rate of change is getting *less* negative all the time.[5] In general, whenever we have a *positive second derivative* of a function (as is the case here), the *curve is concave upwards*. Similarly, if the second derivative were *negative*, the curve would be concave *downwards*.

So by noting the initial value of V, which is formally written V_0 but in this case equals \$1.00, and by applying our understanding of the "graphical meaning" of the first derivative (slope) and the second derivative (curvature), we can visualize the function $V(t)$ pretty well. It starts out with a maximum downward slope and then starts to level off as time increases. This general trend continues indefinitely. Note that while the function always decreases, it *never reaches zero*. This is because, the closer it gets to zero, the slower it decreases [see Eq. (2)]. This is a very "cute" feature that makes this function especially fun to imagine over long times.

We can also apply our analytical understanding to the formulas (2) and (4) for the derivatives: every time we take still another derivative, the result is still proportional to V — the constant of proportionality just picks up another factor of (-0.1). This is a *really neat* feature of this function, namely that we can write down *all its derivatives* with almost no effort:

$$\begin{aligned}
\frac{dV}{dt} &= -0.1\,V & (5)\\
\frac{d^2V}{dt^2} &= (-0.1)^2\,V = +0.01\,V & (6)\\
\frac{d^3V}{dt^3} &= (-0.1)^3\,V = -0.001\,V & (7)\\
\frac{d^4V}{dt^4} &= (-0.1)^4\,V = +0.0001\,V & (8)\\
&\vdots \\
\frac{d^nV}{dt^n} &= (-0.1)^n\,V \qquad \text{for } any\ n. & (9)
\end{aligned}$$

This is a pretty nifty function. What *is* it? That is, can we write it down in terms of familiar things like t, t^2, t^3, and so on?

First, note that Eq. (9) can be written in the form

$$\frac{d^nV}{dt^n} = k^n\,V, \qquad \text{where} \qquad k = -0.1 \tag{10}$$

A simpler version would be where $k = 1$, giving

$$\frac{d^nW}{dt^n} = W, \tag{11}$$

$W(t)$ being the function satisfying this criterion. We should perhaps try figuring out this simpler problem first, and then come back to $V(t)$.

Let's try expressing $W(t)$, then, as a linear combination[6] of such terms. For starters we will try a "third order polynomial" (*i.e.* we allow terms up to t^3):

$$W(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

Then

$$\frac{dW}{dt} = a_1 + 2a_2 t + 3a_3 t^2$$

follows by simple "differentiation" [a single word for "taking the derivative"]. Now, these

[5] A politician trying to obfuscate the issue might say, "The rate of decrease is decreasing."

[6] "Linear combination" means we multiply each term by a simple constant and add them up.

two equations have similar-looking right-hand sides, provided that we pretend not to notice that term in t^3 in the first one, and provided the constants a_n obey the rule $a_{n-1} = n a_n$ [*i.e.* $a_0 = a_1$, $a_1 = 2a_2$ and $a_2 = 3a_3$]. But we can't really neglect that t^3 term! To be sure, its "coefficient" a_3 is smaller than any of the rest, so if we had to neglect anything it might be the best choice; but we're trying to be precise, right? How precise? Well, precise enough. In that case, would we be precise enough if we added a term $a_4 t^4$, preserving the rule about coefficients [$a_3 = 4a_4$]? No? Then how about $a_5 t^5$? And so on. No matter how precise an agreement with Eq. (11) we demand, we can always take enough terms, using this procedure, to achieve the desired precision. Even if you demand infinite precision, we just [just?] take an *infinite* number of terms:

$$W(t) = \sum_{n=0}^{\infty} a_n t^n, \quad \text{where} \quad a_{n-1} = n\, a_n \tag{12}$$

$$\text{or} \quad a_n = \frac{a_{n-1}}{n}. \tag{13}$$

Now, suppose we give $W(t)$ the initial value 1. [If we want a different initial value we can just multiply the whole series by that value, without affecting Eq. (11).] Well, $W(0) = 1$ tells us that $a_0 = 1$. In that case, $a_1 = 1$ also, and $a_2 = \frac{1}{2}$, and $a_3 = \frac{1}{2} \times \frac{1}{3}$, and $a_4 = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{4}$, and so on. If we define the *factorial* notation,

$$n! \;\equiv\; n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1 \tag{14}$$

(read, "*n factorial*") and define $0! \equiv 1$, we can express our function $W(t)$ very simply:

$$W(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \tag{15}$$

We could also write a more abstract version of this function in terms of a generalized variable "x":

$$W(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{16}$$

Let's do this, and then define $x \equiv k\,t$ and set $V(t) = V_0\, W(x)$. Then, by the CHAIN RULE for derivatives,

$$\frac{dV}{dt} = V_0 \frac{dW}{dx} \frac{dx}{dt} \tag{17}$$

and since $\frac{d}{dt}(k\,t) = k$, we have

$$\frac{dV}{dt} \;=\; k\, V_0\, W \;=\; k\, V. \tag{18}$$

By repeating this we obtain Eq. (10). Thus

$$V(t) \;=\; V_0\, W(kt) \;=\; V_0 \sum_{n=0}^{\infty} \frac{(kt)^n}{n!} \tag{19}$$

where $k = -0.1$ in the present case.

This is a nice description; we can always calculate the value of this function to any desired degree of accuracy by including as many terms as we need until the change produced by adding the next term is too small to worry us.[7] But it is a little clumsy to keep writing down such an unwieldy formula every time you want to refer to this function, especially if it is going to be as popular as we claim. After all, mathematics is the art of precise abbreviation. So we give $W(x)$ [from Eq. (16)] a special name, the "**exponential**" function, which we write as either[8]

$$\exp(x) \qquad \text{or} \qquad e^x. \tag{20}$$

In FORTRAN it is represented as `EXP(X)`. It is equal to the number[9]

$$e \;=\; 2.71828182845904509\cdots \tag{21}$$

raised to the x^{th} power.[10] In our case we have $x \equiv -0.1\,t$, so that our "answer" is

$$V(t) \;=\; V_0\, e^{-0.1\,t} \tag{22}$$

[7]This is exactly what a "scientific" hand calculator does when you push the function key whose name will be revealed momentarily.

[8]Now you know which key it is on a calculator.

[9] I'm betting you can easily figure out how to calculate the value of e to any desired precision. Am I wrong?

[10] You are probably wondering what it could possibly *mean* to raise a constant to a power that is not an integer. Stay tuned.... It gets a *lot* weirder!

which is a lot easier to write down than Eq. (19).

Now, the choice of notation e^x is not arbitrary. There are a lot of rules we know how to use on a number raised to a power. One is that

$$e^{-x} \equiv \frac{1}{e^x} \tag{23}$$

You can easily determine that this rule also works for the definition in Eq. (16).

The "inverse" of this function (the power to which one must raise e to obtain a specified number) is called the "**natural logarithm**" or "ln" function. We write

$$\text{if} \quad W = e^x, \quad \text{then} \quad x = \ln(W)$$

or

$$x = \ln(e^x) \tag{24}$$

A handy application of this definition is the rule

$$y^x = e^{x \ln(y)} \quad \text{or} \quad y^x = \exp[x \ln(y)]. \tag{25}$$

Before we return to our original function, is there anything more interesting about the "natural logarithm" than that it is the inverse of the "exponential" function? And what is so all-fired special about e, the "base" of the natural log? Well, it can easily be shown[11] that the *derivative* of $\ln(x)$ is a very simple and familiar function:

$$\frac{d[\ln(x)]}{dx} = \frac{1}{x}. \tag{26}$$

This is perhaps the most useful feature of $\ln(x)$, because it gives us a direct connection between the exponential function and a function whose derivative is $1/x$. [The handy and versatile rule $\frac{d(x^r)}{dx} = rx^{r-1}$ is valid for any value of r, including $r = 0$, but it doesn't help us with this task. Why?] It also explains what is so special about the number e.

[11] Watch for this phrase! Whenever someone says "It can easily be shown. . . ," they mean, "This is possible to prove, but I haven't got time; besides, I might want to assign it as homework."

Summary: Exponential Functions

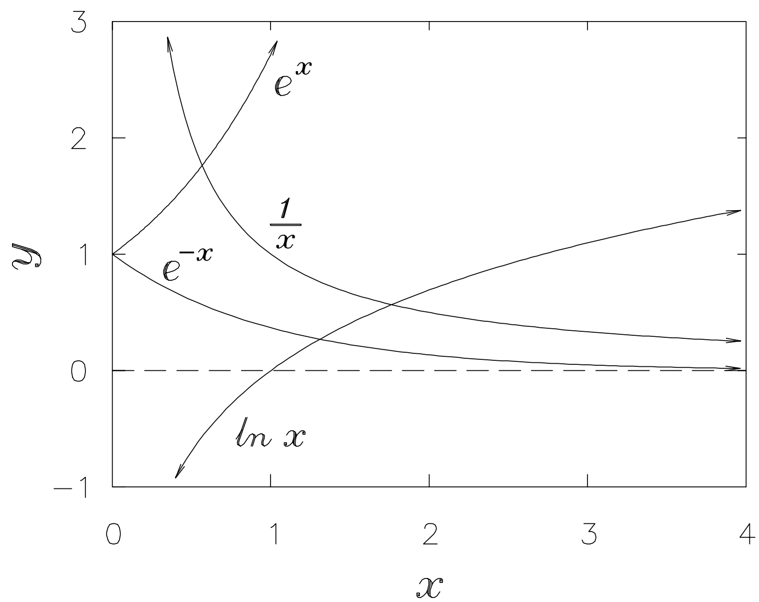


Figure 7.1 The functions e^x, e^{-x}, $\ln(x)$ and $1/x$ plotted on the same graph over the range from $x = 0$ to $x = 4$. Note that $\ln(0)$ is undefined. [There is no finite power to which we can raise e and get zero.] Similarly, $1/x$ is undefined at $x = 0$, while $1/(-x) = -1/x$. Also, $\ln(1) = 0$ [because any number raised to the zeroth power equals 1 — you can easily check this against the definitions] and $\ln(\xi)$ [where ξ any positive number less than 1] is negative. However, there is no such thing as the natural logarithm of any *negative* number.

Our formula (22) for the real value of your dollar as a function of time is the *only* function which will satisfy the differential equation (2) from which we started. The *exponential* function is one of the most useful of all for solving a wide variety of differential equations. For now, just remember this:

> Whenever you have $\frac{dy}{dx} = k\,y$, you can be sure that $y(x) = y_0\, e^{kx}$ where y_0 is the "initial value" of y [when $x = 0$]. Note that k can be either positive or negative.

Chapter 8

Integrals

8.1 Antiderivatives

One way to think of an *integral* is as sort of *the opposite of a derivative* — sometimes called (to the dismay of Real Mathematicians) "*antiderivatives*". We can "solve" antiderivatives the same way we "solve" long division problems: by trial and error guessing! Suppose we are given an explicit function $f(x)$ [for example, $f(x) = x^2$] and told that $f(x)$ *is the derivative of* a function $y(x)$ which we would like to know — that is,

$$\text{If} \quad \frac{dy}{dx} = x^2, \quad \text{what is} \quad y(x)\,?$$

Well, we know that

$$\frac{d}{dx}\left[x^3\right] \;=\; 3\,x^2,$$

so we must divide by 3 to get

$$y(x) = \tfrac{1}{3}\,x^3 \;+\; y_0$$

where the constant term y_0 (the value of y when $x = 0$) cannot be determined from the information given — the derivative of any constant is zero, so such an *integral* is always undetermined to within such a *constant of integration*.

But what about the *range* of integration? The procedure described above actually defines the INDEFINITE INTEGRAL,

$$y(x) = \int x^2 dx.$$

What about the DEFINITE INTEGRAL,

$$A = \int_{x_0}^{x_1} x^2 dx.$$

The prescription in that case is to just "*plug in*" the values of x at the upper and lower limits to the expression obtained as the ANTIDERIVATIVE and calculate their *difference*:

$$A = \left(\frac{1}{3}x_1^3 + y_0\right) - \left(\frac{1}{3}x_0^3 + y_0\right) = \frac{1}{3}\left(x_1^3 - x_0^3\right)$$

(Note how the constant of integration cancels out in the definite integral.)

8.2 Constant times a Function

$$\int a \cdot f(x)dx = a \cdot \int f(x)dx$$

(the integral of a constant times a function is the constant times the integral of the function). This is easy to see in terms of the *area under the curve*: if the curve is raised by a factor of a, the area under it is raised by that same factor.

8.3 Power of x

The same reasoning that gave us the integral of x^2 can be extended to the integral of x^p,

$$\boxed{\text{If} \;\; y(x) = x^p, \quad \int y(x)dx = \frac{x^{p+1}}{p+1} + y_0}$$

with one important exception: $p \neq -1$. What does x^{-1} mean? It's the thing you multiply $x^{+1} = x$ by to get $x^0 = 1$. (Any number raised to the zeroth power is 1.) So $x^{-1} = 1/x$ and in fact $x^{-p} = 1/x^p$. What, then, is

$$\int \frac{dx}{x} \; ?$$

You can make a plot of the curve $y(x) = 1/x$ and see by inspection that the area under that curve is not zero; so the result we're specifically excluding is

$$\int x^{-1} dx \neq \frac{x^{-1+1}}{-1+1} = \frac{x^0}{0} = \frac{1}{0} \to \infty.$$

So what *is* $\int x^{-1} dx$? As stated in the chapter on Exponentials, it is the NATURAL LOGARITHM of x,

$$\boxed{\int_{x_0}^{x_1} \frac{dx}{x} = \ln x_1 - \ln x_0 = \ln\left(\frac{x_1}{x_0}\right)}$$

8.4 Exponentials

The derivative of $\exp(kx)$ (where k is a constant) is

$$\frac{d}{dx} e^{kx} = k e^{kx}$$

so the *antiderivative* of $\exp(kx)$ is just

$$\boxed{\int_{x_0}^{x_1} e^{kx} dx = \frac{1}{k}\left[e^{kx_1} - e^{kx_0}\right]}$$

8.5 Substitution of Variables

Suppose $u(x)$ is a familiar function and $u'(x)$ is its familiar derivative.[1] Then **if** $y(x)dx$ can be expressed in the form $f[u(x)]u'(x)dx$, we can replace $u'(x)dx$ by du so that[2]

$$\boxed{\int_{x_0}^{x_1} y(x)\; dx = \int_{u(x_0)}^{u(x_1)} f(u)\; du}$$

A trivial example is when $u(x) = kx$ (a constant times x). Then we substitute u/k for x and du/k for dx. This is helpful when integrating (for example)[3]

$$\int e^{kx} dx = \frac{1}{k} \int e^u du = \frac{1}{k} e^u = \frac{1}{k} e^{kx}$$

8.6 Integration by Parts

Sometimes there are two functions of x, $u(x)$ and $v(x)$, with familiar derivatives such that $\int f(x)dx$ can be expressed in the form $\int u\; dv$ where $dv \equiv v'(x)\; dx$. Then

$$\boxed{\int_{x_0}^{x_1} f(x) dx = [u\; v]_{x_0}^{x_1} - \int_{v(x_0)}^{v(x_1)} v\; du}$$

where $[u\; v]_{x_0}^{x_1} \equiv u(x_1)v(x_1) - u(x_0)v(x_0)$.

See if you can use INTEGRATION BY PARTS to find the definite integral

$$\int_0^1 x e^{-kx} dx$$

[1] Remember, $u'(x)$ is Mathematician's notation for du/dx.

[2] Note the use of the *differential* $du \equiv u'(x)\; dx$. It looks almost as if du and dx were regular *quantities* that we could do algebra with at will. We Physicists play fast and loose with differentials, while Real Mathematicians wince the way you might when observing someone riding a bicycle "no hands" down a busy street, blindfolded. (We're not really unable to see where we're going; our blindfolds are just translucent, not opaque. :-)

[3] Remember, $\exp(u)$ is its own derivative and therefore also its own integral!

Chapter 9

Calculus "Cheat Sheet "

9.1 Derivatives

- **Definition:** The rate of change [slope] of a function at a point is the limiting value of its *average* slope over an interval including that point, as the width of the interval shrinks to zero:

$$\frac{dy}{dx} \equiv \lim_{\Delta x \to 0} \frac{y(x+\Delta x) - y(x)}{\Delta x}$$

All the remaining Laws and Rules can be proven by algebraic manipulation of this definition.

- **Operator Notation:** The symbol $\frac{d}{dx}$ (read "derivative with respect to x") can be thought of as a mathematical "verb" (called an *operator*) which "operates on" whatever we place to its right. Thus

$$\frac{d}{dx}[y] \equiv \frac{dy}{dx}$$

- **Power Law:** The simplest class of derivatives are those of power-law functions:

$$\frac{d}{dx}[x^p] = p\, x^{p-1}$$

valid for *all* powers p, whether positive, negative, integer, rational, irrational, real, imaginary or complex.

- **Product Law:** The derivative of the product of two functions is *not* the product of their derivatives! Instead,

$$\frac{d}{dx}[f(x) \cdot g(x)] = \frac{df}{dx} \cdot g(x) + f(x) \cdot \frac{dg}{dx}$$

- **Constant times a Function:** The *Product Law* gives

$$\frac{d}{dx}[a \cdot y(x)] = a \cdot \frac{dy}{dx}$$

where a is a constant (*i.e.* not a function of x). This is often referred to as "pulling the constant factor outside the derivative."

- **Function of a Function:** Suppose y is a function of x and x is in turn a function of t.

Then $$\frac{d}{dt}\{y[x(t)]\} = \frac{dy}{dx} \cdot \frac{dx}{dt}$$

(**Chain Rule**)

- **Exponentials:**

$$\frac{d}{dx}\left[e^{kx}\right] = k \cdot e^{kx}$$

where k is any constant.

- **Natural Logarithms:**

$$\frac{d}{dx}[\ln x] = \frac{1}{x}$$

9.2 Integrals

I'm going to mix definite and indefinite integrals together, but not really indiscriminately. Some are more obvious as indefinite integrals and others as definite ones. My goal is to be as obvious as possible! :-)

- **Constant times a Function:**

$$\int a \cdot f(x)dx = a \cdot \int f(x)dx$$

- **Power of x:**

$$\int x^p dx = \frac{x^{p+1}}{p+1} + \text{const}$$

- **Exponentials:**

$$\int_{x_0}^{x_1} e^{kx}dx = \frac{1}{k}\left[e^{kx_1} - e^{kx_0}\right]$$

- **Substitution of Variables:**

$$\int_{x_0}^{x_1} f[u(x)]u'(x)\ dx = \int_{u(x_0)}^{u(x_1)} f(u)\ du$$

- **Integration by Parts:**

$$\int_{x_0}^{x_1} f(x)dx = [u\ v]_{x_0}^{x_1} - \int_{v(x_0)}^{v(x_1)} v\ du$$

Chapter 10

Differential Equations

The defining property of the EXPONENTIAL FUNCTION $\exp(x) \equiv e^x$ is that *it is its own derivative*:

$$\frac{d}{dx}\, e^x = e^x \tag{1}$$

and therefore its own n^{th} derivative and its own *integral*,

$$\int e^x \, dx = e^x \ + \text{const.}$$

If we add a factor to the exponent, like $x \to u = -\lambda x$, we can use the CHAIN RULE

$$\frac{d}{dx} f[u(x)] = \frac{df}{du} \times \frac{du}{dx}$$

to "bring down a factor of $-\lambda$" when we take the derivative:

$$\frac{d}{dx}\, e^{-\lambda x} = -\lambda \, e^{-\lambda x} \ ;$$

otherwise it's still the same "derivative of itself".

We actually used this DIFFERENTIAL EQUATION (1) to find the necessary coefficients in the expansion of $\exp(x)$ as a power series,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{2}$$

and thereby introduced what may be the most *useful* function in all of science. (A sreong statement, but one I stand by!)

Having solved what is perhaps the *simnplest* nontrivial DIFFERENTIAL EQUATION, we can go on to some trickier cases. Unsurprisingly, most of the examples I can offer come from Physics. The first and most obvious case deserves its own Chapter. . . .

Chapter 11

Simple Harmonic Motion

In the previous chapter we observed that the *defining property* of the EXPONENTIAL FUNCTION $\exp(x) \equiv e^x$, the fact that *it is its own derivative*,

$$\frac{d}{dx} e^x = e^x ,$$

is a simple, but very important, example of a DIFFERENTIUAL EQUATION. We now turn to Physics to add another paradigm to our repertoire.

11.1 Periodic Behaviour

Nature shows us many "systems" which return periodically to the same initial state, passing through the same sequence of intermediate states every period. Life is so full of periodic experiences, from night and day to the rise and fall of the tides to the phases of the moon to the annual cycle of the seasons, that we all come well equipped with "common sense" tailored to this paradigm.[1] It has even been suggested that the concept of *time* itself is rooted in the *cyclic* phenomena of Nature.

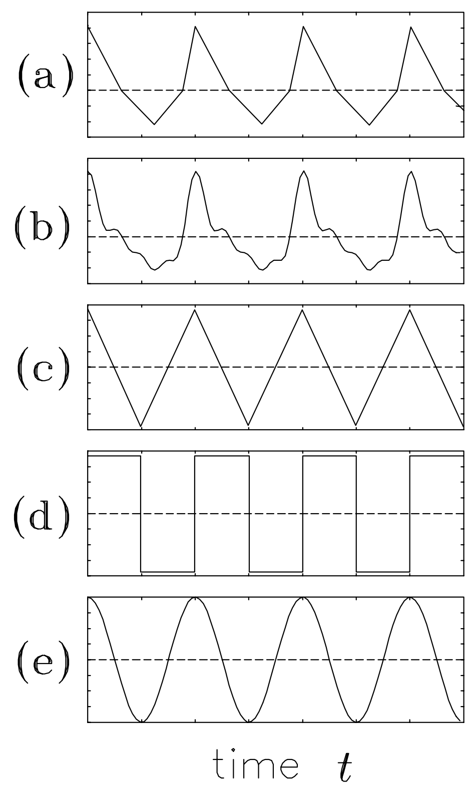


Figure 11.1 Some periodic functions.

In Physics, of course, we insist on narrowing the definition just enough to allow precision. For instance, many phenomena are *cyclic* without being *periodic* in the strict sense of the word.[2]

[1]Many people are so taken with this paradigm that they apply it to all experience. The *I Ching*, for instance, is said to be based on the ancient equivalent of "tuning in" to the "vibrations" of Life and the World so that one's awareness resonates with the universe. By New Age reckoning, cultivating such resonances is supposed to be the fast track to enlightenment. Actually, Physics relies very heavily on the same paradigm and in fact supports the notion that many apparently random phenomena are actually superpositions of regular cycles; however, it offers little encouragement for expecting "answers" to emerge effortlessly from such a tuning of one's mind's resonances. Too bad. But I'm getting ahead of myself here.

[2]Examples of *cyclic* but not necessarily *periodic* phenomena are the mass extinctions of species on Earth that seem to have occurred roughly every 24 million years, the "seven-year cycle" of sunspot activity, the return of salmon to the river of their origin and recurring droughts in Africa. In some cases the basic reason for the cycle is understood and

Here *cyclic* means that the same general pattern keeps repeating; *periodic* means that the system passes through the same "phase" at *exactly* the same time in every cycle and that all the cycles are *exactly* the same length. Thus if we know all the details of *one full cycle* of true periodic behaviour, then we know the subsequent state of the system at *all* times, future and past. Naturally, this is an idealization; but its utility is obvious.

Of course, there is an infinite variety of possible *periodic* cycles. Assuming that we can reduce the "state" of the system to a single variable "q" and its time derivatives, the graph of $q(t)$ can have any shape as long as it *repeats* after one full period. Fig. 11.1 illustrates a few examples. In (a) and (b) the "displacement" of q away from its "equilibrium" position [dashed line] is not symmetric, yet the phases repeat every cycle. In (c) and (d) the cycle is symmetric with the same "amplitude" above and below the equilibrium axis, but at certain points the slope of the curve changes "discontinuously." Only in (e) is the cycle everywhere smooth and symmetric.

11.2 Dot Notation

In Physics we are obsessed with *time*. We are most likely to think of every quantity as a *funtion* of time, t. Therefore, being lazy, we thought of a compact way to express *time rates of change — derivatives* with respact to t: instead of dx/dt, we just write $\dot{x}$ and instead of d^2x/dt^2 we write $\ddot{x}$:

$$\dot{x} \equiv \frac{dx}{dt} \qquad \text{and} \qquad \ddot{x} \equiv \frac{d^2x}{dt^2} \tag{1}$$

Now, you might ask, "What about higher derivatives? Do you just keep adding more dots?" No, for three reasons: first, there is no LaTeXcommand for putting 3 or more dots over a symbol; second, we *very* rarely have to deal with derivatives higher than the second derivative in Physics;[3] and third, it would begin to look ridiculous, like more than two "primes" in the Mathematician's notation $y'''(x)$.

11.3 Sinusoidal Motion

There is one sort of periodic behaviour that is mathematically the *simplest possible* kind. This is the "sinusoidal" motion shown in Fig. 11.1(e), so called because one realization is the sine function, $\sin(x)$. It is easiest to see this by means of a crude mechanical example.

11.3.1 Projecting the Wheel

Imagine a rigid wheel rotating at constant angular velocity about a fixed central axle. A bolt screwed into the rim of the wheel executes *uniform circular motion* about the centre of the axle.[4] For reference we scribe a line on the wheel from the centre straight out to the bolt and call this line the *radius vector*. Imagine now that we take this apparatus outside at high noon and watch the motion of the *shadow* of the bolt on the ground. This is (naturally enough) called the *projection* of the circular motion onto the horizontal axis. At some instant the radius vector makes an angle $\theta = \omega t + \phi$ with the horizontal, where ω is the angular frequency of the wheel (2π times the number of full revolutions per unit time) and ϕ is the initial angle

it is obvious why it only repeats approximately; in other cases we have no idea of the root cause; and in still others there is not even a consensus that the phenomenon is truly cyclic — as opposed to just a random fluctuation that just happens to mimic cyclic behaviour over a short time. Obviously the resolution of these uncertainties demands "more data," *i.e.* watching to see if the cycle continues; with the mass extinction "cycle," this requires considerable patience. When "periodicity debates" rage on in the absence of additional data, it is usually a sign that the combatants have some other axe to grind.

[3] this falls out from Newton's Second Law, which "guides our thinking", but it does seem almost mystical sometimes.

[4] Note the frequency with which we periodically recycle our paradigms!

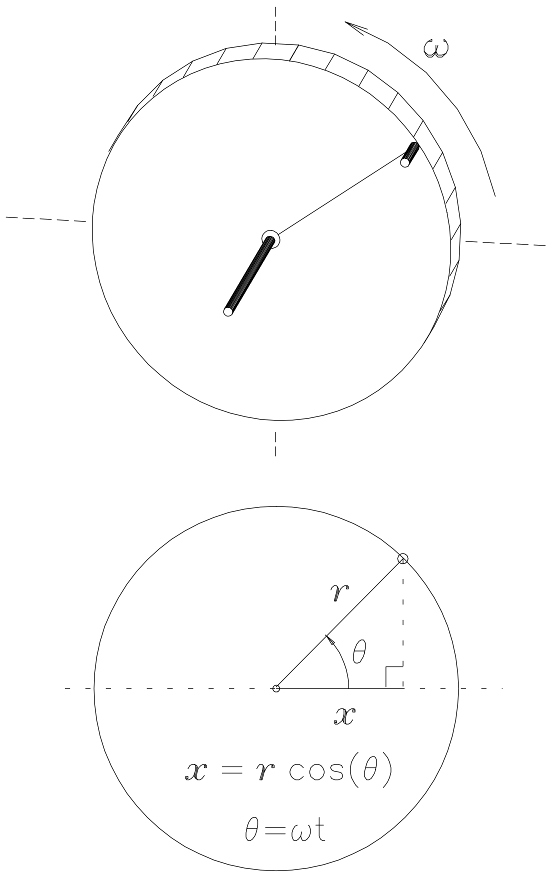


Figure 11.2 Projected motion of a point on the rim of a wheel.

(at $t = 0$) between the radius vector and the horizontal.[5] From a side view of the wheel we can see that the distance x from the shadow of the central axle to the shadow of the bolt [*i.e.* the *projected horizontal displacement* of the bolt from the centre, where $x = 0$] will be given by trigonometry on the indicated right-angle triangle:

$$\cos(\theta) \equiv \frac{x}{r}$$

$$\Rightarrow \quad x = r\cos(\theta) = r\cos(\omega t + \phi) \qquad (2)$$

The resultant *amplitude* of the displacement as a *function of time* is shown in Fig. 11.3.

[5]The inclusion of the "initial phase" ϕ makes this description completely general.

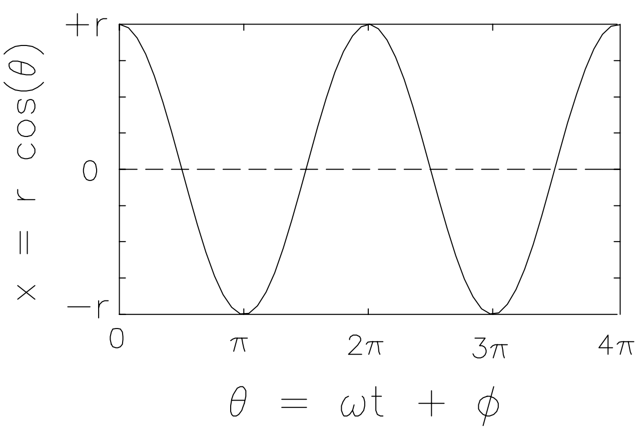


Figure 11.3 The cosine function.

The *horizontal velocity* v_x of the projected shadow of the bolt on the ground can also be obtained by trigonometry if we recall that the vector velocity $\vec{v}$ is always perpendicular to the radius vector $\vec{r}$. I will leave it as an exercise for the reader to show that the result is

$$v_x = -v\sin(\theta) = -r\,\omega\sin(\omega t + \phi) \qquad (3)$$

where $v = r\omega$ is the constant speed of the bolt in its circular motion around the axle. It also follows (by the same sorts of arguments) that the *horizontal acceleration* a_x of the bolt's shadow is the projection onto the $\hat{x}$ direction of $\vec{a}$, which we know is back toward the centre of the wheel — *i.e.* in the $-\hat{x}$ direction; its value at time t is given by

$$a_x = -a\cos(\theta) = -r\,\omega^2\cos(\omega t + \phi) \qquad (4)$$

where $a = \dfrac{v^2}{r} = r\omega^2$ is the magnitude of the centripetal acceleration of the bolt.

11.4 Simple Harmonic Motion

The above mechanical example serves to introduce the idea of $\cos(\theta)$ and $\sin(\theta)$ as *functions* in the sense to which we have (I hope) now become accustomed. In particular, if we realize that (by definition) $v_x \equiv \dot{x}$ and $a_x \equiv \ddot{x}$, the formulae for $v_x(t)$ and $a_x(t)$ represent the *derivatives* of $x(t)$:

$$\begin{aligned} x &= r\cos(\omega t + \phi) & (5)\\ \dot{x} &= -r\,\omega\sin(\omega t + \phi) & (6)\\ \ddot{x} &= -r\,\omega^2\cos(\omega t + \phi) & (7) \end{aligned}$$

— which in turn tell us the derivatives of the sine and cosine functions:

$$\begin{aligned} \frac{d}{dt}\cos(\omega t + \phi) &= -\omega\sin(\omega t + \phi) & (8)\\ \frac{d}{dt}\sin(\omega t + \phi) &= \omega\cos(\omega t + \phi) & (9) \end{aligned}$$

So if we want we can calculate the n^{th} derivative of a sine or cosine function almost as easily as we did for our "old" friend the exponential function. I will not go through the details this time, but this feature again allows us to express these functions as *series expansions*:

$$\begin{aligned} \exp(z) &= 1 + z + \tfrac{1}{2}z^2 + \tfrac{1}{3!}z^3 + \tfrac{1}{4!}z^4 + \cdots\\ \cos(z) &= 1 - \tfrac{1}{2}z^2 + \tfrac{1}{4!}z^4 - \cdots\\ \sin(z) &= z - \tfrac{1}{3!}z^3 + \cdots \end{aligned} \qquad (10)$$

where I have shown the exponential function along with the sine and cosine for reasons that will soon be apparent.

It is definitely worth remembering the SMALL ANGLE APPROXIMATIONS

$$\text{For } \theta \ll 1, \quad \cos(\theta) \approx 1 - \tfrac{1}{2}\theta^2 \quad \text{and} \quad \sin(\theta) \approx \theta. \qquad (11)$$

11.4.1 The Spring Pendulum

Another mecahnical example will serve to establish the paradigm of SIMPLE HARMONIC MOTION (*SHM*) as a solution to a particular type of *equation of motion*.[6]

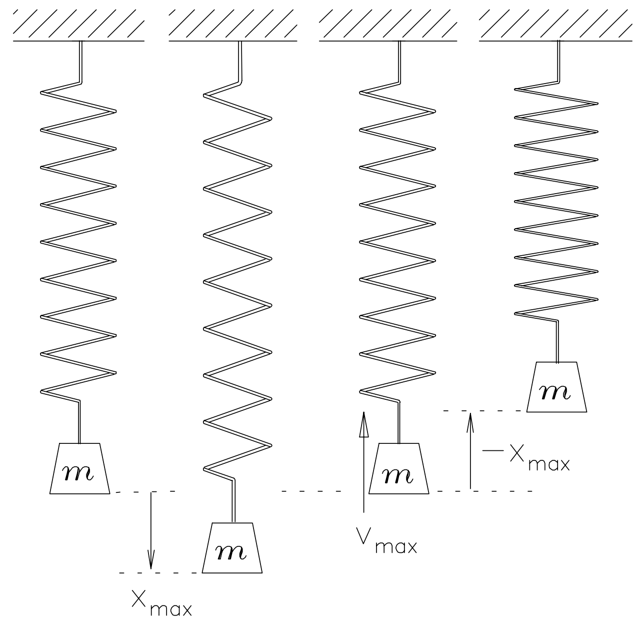


Figure 11.4 Successive "snapshots" of a mass bouncing up and down on a spring.

As discussed in a previous chapter, the *spring* embodies one of Physics' premiere paradigms, the *linear restoring force*. That is, a force which disappears when the system in question is in its "equilibrium position" x_0 [which we will define as the $x = 0$ position $(x_0 \equiv 0)$ to make the calculations easier] but increases as x moves away from equilibrium, in such a way that the *magnitude* of the force F is proportional to the displacement from equilibrium [F is *linear* in x] and the *direction* of F is such as to try to *restore* x to the original position. The constant of proportionality is called the *spring constant*, always written k. Thus $F = -kx$ and the resultant equation of motion is

$$\ddot{x} = -\left(\frac{k}{m}\right)x \qquad (12)$$

Note that the *mass* plays a rôle just as essential

[6] Although we have become conditioned to expect such *mathematical* formulations of relationships to be more removed from our intuitive understanding than easily visualized *concrete* examples like the projection of circular motion, this is a case where the mathematics allows us to draw a sweeping conclusion about the detailed behaviour of *any* system that exhibits certain simple properties. Furthermore, these conditions *are actually satisfied* by an incredible variety of *real* systems, from the atoms that make up any solid object to the interpersonal "distance" in an intimate relationship. Just wait!

as the *linear restoring force* in this paradigm. If $m \to 0$ in this equation, then the acceleration becomes infinite and in principle the spring would just return instantaneously to its equilibrium length and stay there!

In the leftmost frame of Fig. 11.4 the mass m is at rest and the spring is in its equilibrium position (*i.e.* neither stretched nor compressed) defined as $x = 0$. In the second frame, the spring has been gradually pulled down a distance $x_{\max}$ and the mass is once again at rest. Then the mass is released and accelerates upward under the influence of the spring until it reaches the equilibrium position again [third frame]. This time, however, it is moving at its maximum velocity $v_{\max}$ as it crosses the centre position; as soon as it goes higher, it *compresses* the spring and begins to be *decelerated* by a linear restoring force in the opposite direction. Eventually, when $x = -x_{\max}$, all the kinetic energy has been been stored back up in the compression of the spring and the mass is once again instantaneously at rest [fourth frame]. It immediately starts moving downward again at maximum acceleration and heads back toward its starting point. In the absence of friction, this cycle will repeat forever.

I now want to call your attention to the acute similarity between the above differential equation and the one we solved for exponential decay:

$$\dot{x} = -\kappa\, x \tag{13}$$

and, by extension,

$$\ddot{x} = \kappa^2\, x \tag{14}$$

the solution to which equation of motion (*i.e.* the function which "satisfies" the differential equation) was

$$x(t) = x_0\, e^{-\kappa t} \tag{15}$$

Now, if only we could equate κ^2 with $-k/m$, these equations of motion (and therefore their solutions) would be exactly the same! The problem is, of course, that both k and m are intrinsically positive constants, so it is tough to find a real constant κ which gives a negative number when squared.

Imaginary Exponents

Mathematics, of course, provides a simple solution to this problem: just have κ be an *imaginary* number, say

$$\kappa \equiv i\,\omega \qquad \text{where} \qquad i \equiv \sqrt{-1}$$

and ω is a positive real constant. Let's see if this trial solution "works" (*i.e.* take its second derivative and see if we get back our equation of motion):

$$x(t) = x_0\, e^{i\,\omega\, t} \tag{16}$$

$$\dot{x} = i\,\omega\, x_0\, e^{i\,\omega\, t} \tag{17}$$

$$\ddot{x} = -\omega^2\, x_0\, e^{i\,\omega\, t} \tag{18}$$

$$\text{or} \qquad \ddot{x} = -\omega^2\, x \tag{19}$$

$$\text{so} \qquad \omega \equiv \sqrt{\frac{k}{m}} \tag{20}$$

OK, it works. But what does it *describe*? For this we go back to our series expansions for the exponential, sine and cosine functions and note that *if we let* $z \equiv i\theta$, the following *mathematical identity* holds:[7]

$$e^{i\,\theta} = \cos(\theta) + i\,\sin(\theta) \tag{21}$$

Thus, for the case at hand, if $\theta \equiv \omega\, t$ [you probably knew this was coming] then

$$x_0\, e^{i\,\omega\, t} = x_0\,\cos(\omega t) + i\, x_0\,\sin(\omega t)$$

— *i.e.* the formula for the projection of uniform circular motion, with an imaginary part "tacked on." (I have set the initial phase ϕ to zero just to keep things simple.) What does *this* mean?

[7] You may find this unremarkable, but I have never gotten over my astonishment that functions so ostensibly unrelated as the *exponential* and the *sinusoidal* functions could be so intimately connected! And for once the mathematical oddity has profound *practical applications*.

I don't know.

What! How can I say, "I don't know," about the premiere paradigm of Mechanics? We're supposed to know *everything* about Mechanics! Let me put it this way: we have happened upon a nice tidy mathematical representation that *works* — *i.e.* if we use certain rules to manipulate the mathematics, it will faithfully give correct answers to our questions about how this thing will behave. The rules, by the way, are as follows:

> Keep the imaginary components through all your calculations until the final "answer," and then *throw away* any remaining *imaginary parts* of any actual *measurable quantity*.

The point is, there is a difference between understanding how something *works* and knowing what it *means*. Meaning is something we put into our world by act of will, though not always conscious will. How it works is there before us and after we are gone. No one asks the "meaning" of a screwdriver or a carburetor or a copy machine; some of the conceptual tools of Physics are in this class, though of course there is nothing to prevent anyone from *putting* meaning into them.[8]

11.5 Damped Harmonic Motion

Let's take stock. In the previous chapter we found that

$$x(t) \;=\; [\text{constant}] \;-\; \frac{v_0}{\kappa}\, e^{-\kappa t}$$

satisfies the basic differential equation

$$\ddot{x} = -\kappa \dot{x} \qquad \text{or} \qquad a = -\kappa v$$

defining *damped* motion (*e.g.* motion under the influence of a frictional force proportional to the velocity). We now have a solution to the equation of motion defining *SHM*,

$$\ddot{x} = -\omega^2 x \quad \Rightarrow \quad x(t) = x_0\, e^{i\omega t},$$

where

$$\omega = \sqrt{\frac{k}{m}}$$

and I have set the initial phase ϕ to zero just to keep things simple. Can we put these together to "solve" the more general (and realistic) problem of *damped harmonic motion*? The differential equation would then read

$$\ddot{x} \;=\; -\omega^2 x \;-\kappa\, \dot{x} \tag{22}$$

which is beginning to look a little hard. Still, we can sort it out: the first term on the *RHS* says that there is a linear restoring force and an inertial factor. The second term says that there is a damping force proportional to the velocity. So the differential equation itself is not all that fearsome. How can we "solve" it?

As always, by trial and error. Since we have found the *exponential* function to be so useful, let's try one here: *Suppose* that

$$x(t) = x_0\, e^{Kt} \tag{23}$$

where x_0 and K are unspecified constants. Now plug this into the differential equation and see what we get:

$$\dot{x} \;=\; K\, x_0\, e^{Kt} \;=\; K\, x$$

and

$$\ddot{x} \;=\; K^2\, x_0\, e^{Kt} \;=\; K^2\, x$$

The whole thing then reads

$$K^2\, x \;=\; -\omega^2\, x \;-\kappa\, K\, x$$

which can be true "for all x" only if

$$K^2 = -\omega^2 - \kappa\, K \qquad \text{or} \qquad K^2 + \kappa K + \omega^2 = 0$$

[8] I am reminded of a passage in one of Kurt Vonnegut's books, perhaps *Sirens of Titan*, in which the story of creation is told something like this: God creates the world; then he creates Man, who sits up, looks around and says, "What's the meaning of all this?" God answers, "What, there has to be a meaning?" Man: "Of course." God: "Well then, I leave it to you to think of one."

which is in the standard form of a general quadratic equation for K, to which there are two solutions:

$$K = \frac{-\kappa \pm \sqrt{\kappa^2 - 4\omega^2}}{2} \qquad (24)$$

Either of the two solutions given by substituting Eq. (24) into Eq. (23) will satisfy Eq. (22) describing *damped SHM.* In fact, generally any *linear combination* of the two solutions will also be a solution. This can get complicated, but we *have* found the answer to a rather broad question.

11.5.1 Limiting Cases

Let's consider a couple of "limiting cases" of such solutions. First, suppose that the linear restoring force is extremely weak compared to the "drag" force — *i.e.*[9] $\kappa \gg \omega = \sqrt{\frac{k}{m}}$. Then $\sqrt{\kappa^2 - 4\omega^2} \approx \kappa$ and the solutions are $K \approx 0$ [*i.e.* $x \approx$ constant, plausible only if $x = 0$] and $K \approx -\kappa$, which gives the same sort of damped behaviour as if there were no restoring force, which is what we expected.

Now consider the case where the linear restoring force is very strong and the "drag" force extremely weak — *i.e.* $\kappa \ll \omega = \sqrt{\frac{k}{m}}$. Then $\sqrt{\kappa^2 - 4\omega^2} \approx 2\,i\,\omega$ and the solutions are $K \approx -\frac{1}{2}\kappa \pm i\,\omega$, or[10]

$$\begin{aligned} x(t) &= x_0\, e^K && (25)\\ &\approx x_0\, \exp(\pm i\,\omega t\; - \gamma t) && (26)\\ &= x_0\, e^{\pm i\,\omega t} \cdot e^{-\gamma t} && (27) \end{aligned}$$

where $\gamma \equiv \frac{1}{2}\kappa$. We may then think of $[i\,K]$ as a *complex frequency*[11] whose real part is $\pm\omega$ and whose imaginary part is γ. What sort of situation does this describe? It describes a *weakly damped harmonic motion* in which the usual sinusoidal pattern damps away within an "envelope" whose shape is that of an exponential decay. A typical case is shown in Fig. 11.5.

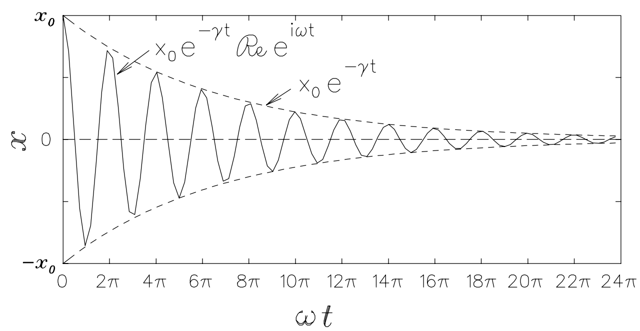


Figure 11.5 Weakly damped harmonic motion. The initial amplitude of x (whatever x is) is x_0, the angular frequency is ω and the damping rate is γ. The cosine-like oscillation, equivalent to the *real* part of $x_0\, e^{i\,\omega\,t}$, decays within the *envelope function* $x_0\, e^{-\gamma\,t}$.

11.6 Generalization of SHM

As for all the other types of *equations of motion*, *SHM* need not have anything to do with masses, springs or even Physics. Even within Physics, however, there are so many different kinds of examples of *SHM* that we go out of our way to generalize the results: using "q" to represent the "coordinate" whose displacement from the equilibrium "position" (always

[9] The "≫" symbol means "... is *much* greater than..." — there is an analogous "≪" symbol that means "... is much *less* than...."

[10] There is a general rule about *exponents* that says, "A number raised to the sum of two powers is equal to the product of the same number raised to each power separately," or

$$a^{b+c} = a^b \cdot a^c.$$

[11] The word "complex" has, like "real" and "imaginary," been ripped off by Mathematicians and given a very explicit meaning that is not entirely compatible with its ordinary dictionary definition. While a *complex* number in Mathematics may indeed be complex — *i.e.* complicated and difficult to understand — it is *defined* only by virtue of its having both a *real* part and an *imaginary* part, such as $z = a + i\,b$, where a and b are both *real.* I hope that makes everything crystal clear....

taken as $q = 0$) engenders some sort of restoring "force" $Q = -k\,q$ and "μ" to represent an "inertial factor" that plays the rôle of the mass, we have

$$\ddot{q} = -\left(\frac{k}{\mu}\right) q \tag{28}$$

for which the solution is the real part of

$$q(t) = q_0\, e^{i\omega t} \qquad \text{where} \qquad \omega = \sqrt{\frac{k}{\mu}} \tag{29}$$

When some form of "drag" acts on the system, we expect to see the qualitative behaviour pictured in Fig. 11.5 and described by Eqs. (23) and (24). Although one might expect virtually every real example to have some sort of frictional damping term, in fact there are numerous physical examples with no damping whatsoever, mostly from the microscopic world of solids.

11.7 The Universality of SHM

If two systems satisfy the same equation of motion, their behaviour is the same. Therefore the motion of the mass on the spring is *in every respect identical* to the horizontal component of the motion of the peg in the rotating wheel, even though these two systems are physically quite distinct. In fact, *any* system exhibiting both a LINEAR RESTORING "FORCE" and an INERTIAL FACTOR analogous to MASS will exhibit *SHM*.[12] Moreover, since these arguments may be used equally well in reverse, the horizontal component of the *force* acting on the peg in the wheel must obey $F_x = -kx$, where k is an "effective spring constant."

[12] Examples are plentiful, especially in view of the fact that *any* potential energy minimum is approximately quadratic for small enough displacements from equilibrium. A prime example from outside Mechanics is the *electrical circuit*, in which the charge Q on a capacitor plays the rôle of the displacement variable x and the inertial factor is provided by an inductance, which resists changes in the current $I = dQ/dt$.

Why is *SHM* characteristic of such an enormous variety of phenomena? Because *for sufficiently small* displacements from equilibrium, *every* system with an equilibrium configuration satisfies the first condition for *SHM*: the linear restoring force. Here is the simple argument: a linear restoring force is equivalent to a potential energy of the form $U(q) = \frac{1}{2}k\,q^2$ — *i.e.* a "quadratic minimum" of the potential energy at the equilibrium configuration $q = 0$. But if we "blow up" a graph of $U(q)$ near $q = 0$, *every* minimum looks quadratic under sufficient magnification! That means *any* system that *has* an equilibrium configuration also has some analogue of a "potential energy" which is a minimum there; if it also has some form of *inertia* so that it tends to stay at rest (or in motion) until acted upon by the analogue of a *force*, then it will automatically exhibit *SHM* for small-amplitude displacements. This makes *SHM* an extremely powerful paradigm.

11.7.1 Equivalent Paradigms

We have established previously that a LINEAR RESTORING FORCE $F = -kx$ is completely equivalent to a QUADRATIC MINIMUM IN POTENTIAL ENERGY $U = \frac{1}{2}kx^2$. We now find that, with the inclusion of an INERTIAL FACTOR (usually just the MASS m), either of these *physical paradigms* will guarantee the *mathematical paradigm* of *SHM* — *i.e.* the displacement x from equilibrium will satisfy the equation of motion

$$x(t) \;=\; x_{\max}\, \cos(\omega t + \phi) \tag{30}$$

where $x_{\max}$ is the *amplitude* of the oscillation. Any $x(t)$ of this form automatically satisfies the definitive equation of motion of *SHM*, namely

$$\frac{d^2x}{dt^2} \;=\; -\omega^2\, x \tag{31}$$

and *vice versa* — whenever x satisfies Eq. (31), the explicit time dependence of x will be given by Eq. (30).

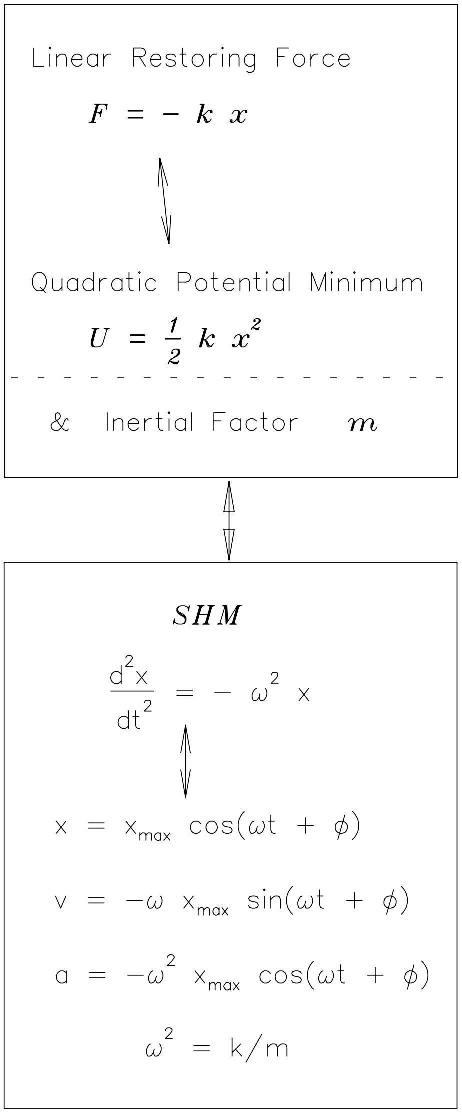


Figure 11.6 Equivalent paradigms of *SHM*.

11.8 Resonance

No description of *SHM* would be complete without some discussion of the general phenomenon of *resonance*, which has many practical consequences that often seem very counterintuitive.[13] I will, however, overcome my zeal for demonstrating the versatility of Mathematics and stick to a simple qualitative description of resonance. Just this once.

The basic idea is like this: suppose some system exhibits all the requisite properties for *SHM*, namely a linear restoring "force" $Q = -k\,q$ and an inertial factor μ. Then *if once set in motion* it will oscillate forever at its "resonant frequency" $\omega = \sqrt{k/\mu}$, unless of course there is a "damping force" $D = -\kappa\mu q$ to dissipate the energy stored in the oscillation. As long as the damping is weak $\left[\kappa \ll \sqrt{k/m}\right]$, any oscillations will persist for many periods. Now suppose the system is initially at rest, in equilibrium, ho hum. What does it take to "get it going?"

The *hard* way is to give it a great whack to start it off with lots of kinetic energy, or a great tug to stretch the "spring" out until it has lots of potential energy, and then let nature take its course. The *easy* way is to give a tiny push to start up a small oscillation, then wait exactly one full period and give another tiny push to increase the amplitude a little, and so on. This works because *the frequency ω is independent of the amplitude q_0*. So if we "drive" the system *at its natural "resonant" frequency ω*, no matter how small the individual "pushes" are, we will slowly build up an *arbitrarily large oscillation.*[14]

Such resonances often have dramatic results. A vivid example is the famous movie of the collapse of the Tacoma Narrows bridge, which had a torsional [twisting] resonance[15] that was excited by a steady breeze blowing past the bridge. The engineer in charge anticipated all the other more familiar resonances [of which there are many] and incorporated devices specifically designed to safely damp their

[13] It is, after all, one of the main purposes of this book to dismantle your intuition and rebuild it with the faulty parts left out and some shiny new paradigms added.

[14] Of course, this assumes $\kappa = 0$. If damping occurs at the same time, we must put at least as much energy *in* with our driving force as friction takes *out* through the damping in order to build up the amplitude. Almost every system has some limiting amplitude beyond which the restoring force is no longer linear and/or some sort of losses set in.

[15] (something like you get from a blade of grass held between the thumbs to create a loud noise when you blow past it)

oscillations, but forgot this one. As a result, the bridge developed huge twisting oscillations [mistakes like this are usually painfully obvious when it is too late to correct them] and tore itself apart.

A less spectacular example is the trick of getting yourself going on a playground swing by leaning back and forth with arms and legs in synchrony with the natural frequency of oscillation of the swing [a sort of pendulum]. If your kinesthetic memory is good enough you may recall that it is important to have the "driving" push exactly $\frac{\pi}{2}$ radians [a quarter cycle] "out of phase" with your velocity — *i.e.* you *pull* when you reach the *motionless* position at the top of your swing, if you want to achieve the maximum result. This has an elegant mathematical explanation, but I promised. . . .

Chapter 12

Vectors

12.1 Cartesian Vectors

A vector quantity is one that has both *magnitude* and *direction*. Another (equivalent) way of putting it is that a vector quantity has several components in *orthogonal* (perpendicular) directions. The idea of a vector is very abstract and general; one can define useful vector spaces of many sorts, some with an infinite number of orthogonal basis vectors, but the most familiar types are simple 3-dimensional quantities like position, speed, momentum and so on.

12.1.1 Vector Notation

The conventional notation for a vector is $\vec{\boldsymbol{A}}$, sometimes written $\vec{A}$ or $\boldsymbol{A}$ but most clearly recognizable when *both* in **boldface** *and* a little arrow over the top.

12.1.2 Unit Vectors

In Cartesian coordinates (x, y, z) a vector $\vec{\boldsymbol{A}}$ can be expressed in terms of its three scalar *components* A_x, A_y, A_z and the corresponding unit vectors $\hat{\imath}, \hat{\jmath}, \hat{k}$ (sometimes written as $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ or occasionally as $\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \hat{\boldsymbol{x}}_3$) thus:

$$\vec{\boldsymbol{A}} = \hat{\imath} A_x + \hat{\jmath} A_y + \hat{k} A_z \tag{1}$$

where the little "hat" over a symbol means (in this context) that it has unit magnitude and thus imparts *only direction* to a scalar like A_x.[1]

[1] There are many choices of coordinates and unit vectors, such as *cylindrical* (r, θ, z) and *spherical* (r, θ, ϕ) coordinates, but only in the simple Cartesian coordinates are the directions of the unit vectors permanently fixed.

A unit vector $\hat{\boldsymbol{a}}$ can be formed from any vector $\vec{\boldsymbol{a}}$ by dividing it by its own *magnitude* a:

$$\hat{\boldsymbol{a}} = \frac{\vec{\boldsymbol{a}}}{a} \qquad \text{where} \quad a = |\vec{\boldsymbol{a}}| = \sqrt{a_x^2 + a_y^2 + a_z^2}\,. \tag{2}$$

Already we have used a bunch of concepts before defining them properly, the usual chicken-egg problem with mathematics. Let's try to catch up:

12.1.3 Multiplying a Vector by a Scalar

Multiplying a vector $\vec{\boldsymbol{A}}$ by a scalar b has no effect on the *direction* of the result (unless $b = 0$) but only on its *magnitude* and/or the *units* in which it is measured — if b is a pure number, the units stay the same; but multiplying a velocity by a mass (for instance) produces an entirely new quantity, in that case the momentum.

This type of product always commutes: $\vec{\boldsymbol{A}}b = b\vec{\boldsymbol{A}}$.

12.1.4 Dividing a Vector by a Scalar

Dividing a vector by a scalar c is the same as *multiplying* it by $1/c$.

12.1.5 Adding or Subtracting Vectors

In two dimensions one can draw simple diagrams depicting "tip-to-tail" or "parallelogram law" vector addition (or subtraction); this is not so easy in 3 dimensions, so we fall back

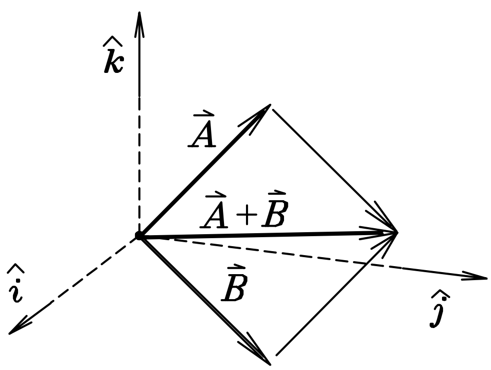


Figure 12.1 Vector addition.

on the algebraic method of *adding components*. Given $\vec{A}$ from Eq. (1) and

$$\vec{B} = \hat{\imath}B_x + \hat{\jmath}B_y + \hat{k}B_z \tag{3}$$

we write

$$\vec{A}+\vec{B} = \hat{\imath}(A_x+B_x)+\hat{\jmath}(A_y+B_y)+\hat{k}(A_z+B_z) \; . \tag{4}$$

Subtracting $\vec{B}$ from $\vec{A}$ is the same thing as adding $-\vec{B}$ (a vector of the same length as $\vec{B}$ but pointing in the opposite direction).

12.1.6 Multiplying Two Vectors ...

... to get a Scalar: we just add together the products of the components,

$$\begin{aligned} \vec{A}\cdot\vec{B} &= A_xB_x \\ &+ A_yB_y \\ &+ A_zB_z \; , \end{aligned} \tag{5}$$

also known as the "dot product", which commutes: $\vec{A}\cdot\vec{B} = \vec{B}\cdot\vec{A}$. Note that the square of the *magnitude* of $\vec{A}$ is just the scalar product of $\vec{A}$ with *itself*:

$$A^2 = \vec{A}\cdot\vec{A} \; .$$

... to get a Pseudovector:

$$\begin{aligned} \vec{A}\times\vec{B} &= \hat{\imath}(A_yB_z - A_zB_y) \\ &+ \hat{\jmath}(A_zB_x - A_xB_z) \\ &+ \hat{k}(A_xB_y - A_yB_x) \; . \end{aligned} \tag{6}$$

This "cross product" is actually a pseudovector (or, more generally, a tensor), because (unlike the nice dot product) it has the unsettling property of *not commuting* $\left(\vec{A}\times\vec{B} = -\vec{B}\times\vec{A} \right)$ — but we often treat it like just another vector.

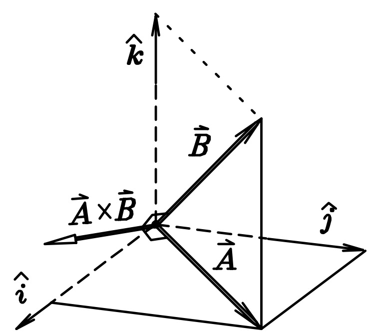


Figure 12.2 Vector Product.

12.2 Orthogonality

The definition of a *vector* as an entity with both magnitude and direction can be *generalized* if we realize that "direction" can be defined in more dimensions than the usual 3 spatial directions, "up-down, left-right, and back-forth," or even other dimensions *excluding* these three. The more general definition would read,

> A **vector** quantity is one which has *several independent attributes* which are *all measured in the same units* so that "transformations" are possible. (This last feature is only essential when we want the advantages of mathematical manipulation; it is not necessary for the concept of multi-dimensional entities.)

We can best illustrate this generalization with an example of a vector that has nothing to do with 3-D space:

> Example: the **Cost of Living**, $\vec{\mathbf{C}}$, is in a sense a true vector quantity (although the Cost of Living *index* may be properly thought of as a *scalar*, as we can show later).

To construct a simple version, the Cost of Living can be taken to include:

- C_1 = **housing** (e.g., monthly rent);
- C_2 = **food** (e.g., cost of a quart of milk);
- C_3 = **medical** service (e.g., cost of a bottle of aspirin);
- C_4 = **entertainment** (e.g., cost of a movie ticket);
- C_5 = **transportation** (e.g., bus fare);
- $C_6 \ldots C_7 \ldots$ etc. (a finite number of "*components*.")

Thus we can write $\vec{\mathbf{C}}$ as an ordered sequence of numbers representing the values of its respective "components":

$$\vec{\mathbf{C}} = (C_1, C_2, C_3, C_4, C_5, \ldots) \tag{1}$$

We would normally go on until we had a reasonably "complete" list – i.e., one with which the cost of any additional item we might imagine could be expressed *in terms of* the ones we have already defined. The technical mathematical term for this condition is that we have a "*complete basis set*" of components of the Cost of Living.

Now, we can immediately see an "inefficiency" in the way $\vec{\mathbf{C}}$ has been "composed:" As recently as 1975, it was estimated to take approximately one pound of gasoline to grow one pound of food in the U.S.A.; therefore the cost of **food** and the cost of **transportation** are obviously *not independent*! Both are closely tied to the cost of **oil**. In fact, a large number of the components of the cost of living we observe are intimately connected to the cost of oil (among other things). On the other hand (before we jump to the fashionable conclusion that these two components should be replaced by oil prices alone), there is *some* measure of independence in the two components. How do we deal with this quantitatively?

To reiterate the question more formally, how do we quantitatively describe the extent to which certain components of a vector are superfluous (in the sense that they merely represent combinations of the other components) *vs.* the extent to which they are truly "independent?" To answer, it is convenient to revert to our old standby, the (graphable) analogy of the **distance** vector in *two dimensions*.

Suppose we wanted to describe the position of any point P in the "$x - y$ plane." We could draw the two axes "a" and "b" shown above. The position of an arbitrary point P is uniquely determined by its (a, b) coordinates, defined by the prescription that to change a we move parallel to the a-axis and to change b we move parallel to the b-axis. This is a unique and quite legitimate way of specifying the position of any point (in fact it is often used in crystallography where the orientation of certain crystal axes is determined by nature); yet there is something vaguely troubling about this choice of coordinate axes. What is it? Well, we have an intuitive sense of "up-down" and "sideways" as being *perpendicular*, so that if something moves "up" (as we normally think of it), in the above description the values of both a and b will change. But isn't our intuition just the result of a well-entrenched convention? If we got used to thinking of "up" as being in the "b" direction shown, wouldn't this cognitive dissonance dissolve?

No. In the first place, nature provides us with an unambiguous characterization of "down:" it is the direction in which things fall when released; the direction a string points when tied to a plumb bob. "Sideways," similarly, is the direction defined by the surface of an undisturbed liquid (as long as we neglect the curvature of the Earth's surface). That is, gravity fixes our notions of "appropriate" geometry. But is this in turn arbitrary (on nature's part) or is there some good reason why "independent" components of a vector should be perpendicular? And what exactly do we *mean* by "perpendicular," anyway? Can we define the concept in a way which might allow us to generalize it to other kinds of vectors besides space vectors?

The answer is bound up in the way Euclid found to express the geometrical properties of the world we live in; in particular, the "metric" of space – the way we define the **magnitude** (*length*) of a vector. Suppose you take a ruler and turn it at many angles; your idea of the *length* of the ruler is *independent* of its *orientation*, right? Suppose you used the ruler to make off distances along two perpendicular axes, stating that these were the horizontal and vertical components (x, y) of a distance vector. Then you use the usual "parallelogram rule" to locate the tip of the vector, draw in a line from the origin to that point, and put an arrowhead on the line to indicate that you have a vector. Call it "$\vec{\mathbf{r}}$". *You can use the same ruler, held at an angle, to measure the length r of the vector.* Pythagoras gave us a formula for this length. It is

$$r = \sqrt{x^2 + y^2}. \qquad (2)$$

This formula is the key to Euclidean geometry, and is the working definition of perpendicular axes: x and y are perpendicular if and only if Eq. (2) holds. It does *not* hold for "a" and "b" described earlier!

You may feel that this "metric" is obvious and necessary from first principles; it is not. If you treat this formula as correct using the Earth's surface as the "$x - y$ plane" you will get good results until you start measuring off distances in the thousands of miles; then you will be 'way off! Imagine for instance the perpendicular lines formed by two longitudes at the North Pole: these same "perpendicular" lines *cross again* at the South Pole!

Well, of course, you say; that is because the Earth's surface is *not* a plane; it is a sphere; it is *curved*. If we didn't feign ignorance of that fact, if we did our calculations in *three* dimensions, we would always get the right answers. Unfortunately not. The space we live in is actually *four*-dimensional, and it is *not* flat, *not* "Euclidean," in the neighborhood of large masses. Einstein helped open our eyes to this fact, and now we are stuck with a much more cognitively complex understanding.

But we have to start somewhere, and the space we live in from day to day in "pretty Euclidean," and it is only in the violation of sensible approximations that modern physics is astounding, so we will pretend that only Euclidean vector spaces are important. (Do you suppose there is a way to generalize our definition of "perpendicularity" to include non-Euclidean space as well?)

Finally returning to our original example, we would like to have $\vec{\mathbf{C}}$ expressed in an "orthogonal, complete basis", $\vec{\mathbf{C}} = (C_1, C_2, C_3, C_4, C_5, \dots)$, so that we can define the **magnitude** of $\vec{\mathbf{C}}$ by

$$C = |\vec{\mathbf{C}}| = \sqrt{C_1^2 + C_2^2 + C_3^2 + \dots} \qquad (3)$$

("Orthogonal" and "normal" are just synonyms for "perpendicular.") We could call $\vec{\mathbf{C}}$ the "Cost of Living Index" if we liked. There is a problem now. Our intuitive notion of "independent" components is tied up with the idea that one component can change without affecting another; yet as soon as we attempt to be specific about it, we find that we cannot even define a criterion for formal and exact independence (orthogonality) without generating a new notion: the idea of a magnitude as defined by

Eq. (3). Does this definition agree with out intuition, the way the "ruler" analogy did? Most probably we *have* no intuition about the "magnitude" of the "cost of living vector." So we have created a new concept – not an arbitrary concept, but one which is guaranteed to have a large number of "neat" consequences, one we will be able to do calculations with, make transformations of, and so on. In short, a "rich" concept.

There is another problem, though; while we can easily test our space vectors with a ruler, there is no unambiguous "ruler" for the "cost of living index." Furthermore, we may make the *approximation* that the cost of tea bags is orthogonal to the cost of computer maintenance, but in so "messy" a business as economics we will never be able to prove this rigorously. There are too many "hidden variables" influencing the results in ways we do not suspect. This is too bad, but we can still live with the imperfections of an approximate model if it serves us well.

Chapter 13

Complex Numbers

Imaginary numbers were mentioned in the Algebra chapter and again in the chapters on Differential Equations and Simple Harmonic Motion, but here I'd like to build a bridge to a nominally "different" discipline: LITERARY CRITICISM.

Back in the mid-1960s when I was in college, Along with Physics and Mathematics courses I took one on "Structural Conventiuons in Poetry" where I learned about Northrup Frye and his theory of GENRES, famously represented by a diagram something like this:

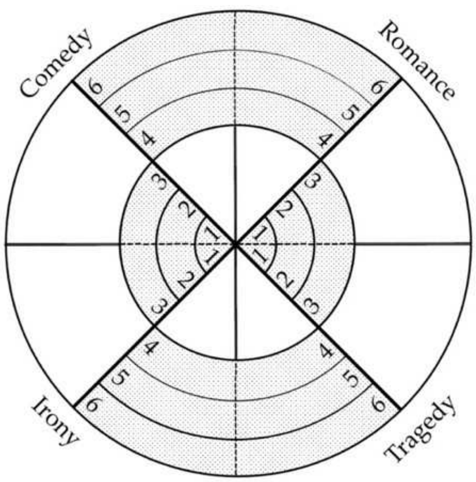
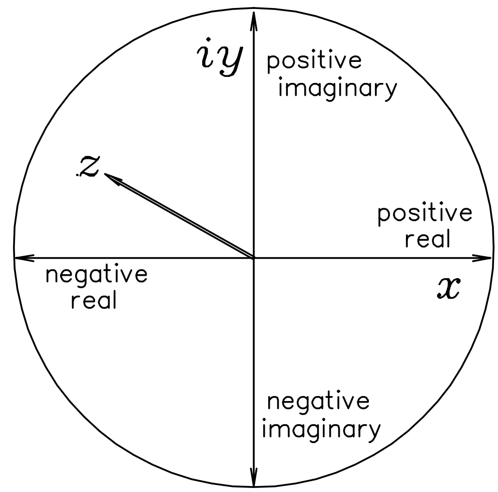


Figure 13.1 Diagram of Frye's Genres.

I immediately translated this into the *complex plane*, which I'd just learned about in Physics and Math courses (see below). So I re-drew Frye's diagram in the Literary Complex Plane:



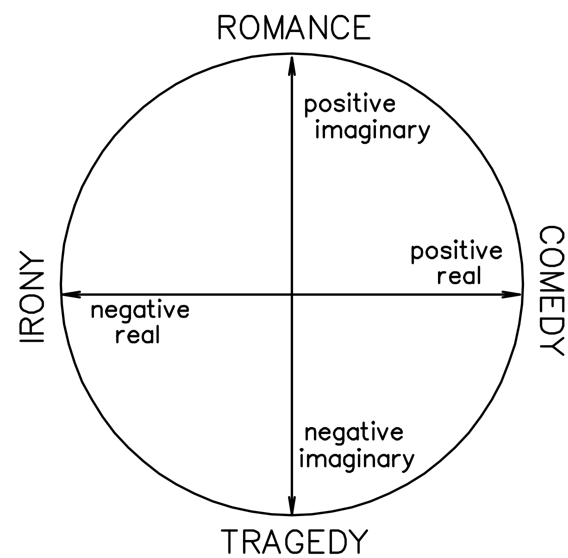


Figure 13.2 TOP: The complex plane. BOTTOM: Genres.