Chapter 1
Dimensions

Mathematics developed alongside and took inspiration from the physical sciences. Famous mathematicians of the 17th and 18th centuries – for example, Newton, Euler, and the Bernoulli brothers – are equally famous for their contributions to physics. Today, however, mathematics is taught separately from physics, chemistry, engineering, economics, and other fields that use mathematics. Ideas developed in these fields must scale a high wall before they can return to benefit mathematics. To nudge a few ideas over the wall – that is the purpose of this book.

Dimensions are one of the useful ideas from a foreign land. This chapter introduces the method of dimensions – more formally known as dimensional analysis. The first example analyzes a flawed argument that is frequent in the media. Succeeding examples apply the method to derivatives and integrals.

1.1 Power of multinationals

Critics of globalization (the kindler, gentler name for imperialism) often make the following argument:

> In Nigeria, a relatively economically strong country, the GDP [gross domestic product] is $99 billion. The net worth of Exxon is $119 billion. 'When multinationals have a net worth higher than the GDP of the country in which they operate, what kind of power relationship are we talking about?' asks Laura Morosini.[1]

Before reading further, try the next problem.

▶ *What is the most egregious fault in the comparison between Exxon and Nigeria? It's a competitive field, but one fault stands out.*

The comparison between Exxon and Nigeria has many problems. First, the comparison perhaps exaggerates Exxon's power by using its worldwide assets (net worth) rather than its assets only in Nigeria. On the other hand, Exxon can use its full international power when negotiating with Nigeria, so perhaps the worldwide assets are a fair basis for comparison.

A more serious problem is the comparison of net worth and GDP. The $99 billion for Nigeria's GDP is shorthand for a flow of $99 billion per year. Because economic flows are a social phenomenon, valid economic arguments should not depend on an astronomical phenomenon such as the time the earth needs to travel around the sun. The comparison under study, however, has this defect.

Suppose, for example, that economists used the decade as the unit of time in measuring GDP. Then Nigeria's GDP (assuming it remains steady from year to year) would be roughly $1 trillion per decade and would be conventionally reported as $1 trillion. Now Nigeria towers over the Exxon, whose puny assets are a mere one-tenth of Nigeria's GDP. To produce the opposite conclusion, suppose the week were the unit of time for measuring GDP. Nigeria's GDP becomes $2 billion per week conventionally reported as $2 billion: Now puny Nigeria stands helpless before the mighty Exxon, 50-fold larger than Nigeria. Either conclusion can be produced merely by changing the units of time – a symptom of a fundamentally sick comparison.

The disease is in comparing unlike quantities. Net worth is an amount of money – money is its dimensions – and is typically measured in units of dollars. GDP, however, is a flow (or rate) with dimensions of money per time, and typical units of dollars per year. Comparing net worth to GDP means comparing money to money per time. Because the dimensions of these two quantities are not the same, the comparison is nonsense!

---

[1] Source: 'Impunity for Multinationals', *ATTAC*, 11 Sept 2002, http://www.globalpolicy.org/socecon/tncs/2002/0911impunity.htm, retrieved 11 Sept 2006.

A similarly and more obviously flawed comparison is to compare length per time (speed) with length. Listen to how ridiculous it sounds: 'I walk 1.5 meters per second – much smaller than the Empire State building in New York, which is 300 meters high.' To produce the opposite conclusion, measure time in hours: 'I walk 5000 meters per hour – much larger than the Empire State building, which is 300 meters high.' Because the dimensions of these two quantities are not the same, the comparison is nonsense.

The Nigeria and Exxon comparison illustrates several ideas:

1. *Necessary condition for a valid comparison.* In a valid comparison, the compared quantities have identical dimensions. Therefore, compare not apples to oranges (except as in, 'I prefer eating apples to oranges.')
2. *Dimensions versus units.* Dimensions are general and generic, such as money per time or length. Units are dimensions made flesh in a system of measurement – for example, measuring net worth in dollars or measuring GDP in dollars per year. The worldwide standard measurement system is the System International (SI), based on fundamental (base) units of kilograms, meters, and seconds.
3. *Lots of rubbish.* Bad arguments are ubiquitous, so keep your eyes open!
4. *Bad argument, reasonable conclusion.* The conclusion of the article, that large oil companies can exert massive power over poorer countries, seems reasonable. However, as a mathematician, I find the reasoning distressing. The example illustrates a lesson about theorems and proofs: Examine the whole proof not just the theorem statement, for a correct theorem is not justified by a faulty proof.

The necessary condition given above, that the quantities have the same dimensions, is not a sufficient condition. A costly illustration is the 1999 Mars Climate Orbiter (MCO), which probably crashed into the surface of Mars or got launched back into space, instead of staying in orbit around Mars. The Mishap Investigation Board found that the cause was a mismatch between English and metric units [2, p. 6]:

> The MCO MIB has determined that the root cause for the loss of the MCO spacecraft was the failure to use metric units in the coding of a ground software file, 'Small Forces', used in trajectory models. Specifically, thruster performance data in English units instead of metric units was used in the software application code titled SM_FORCES (small forces). A file called Angular Momentum Desaturation (AMD) contained the output data from the SM_FORCES software. The data in the AMD file was required to be in metric units per existing software interface documentation, and the trajectory

> modelers assumed the data was provided in metric units per the requirements.

So, pay attention to dimensions and units!

The next examples apply dimensional reasoning to several mathematical problems.

Problem 1.1 More bad comparisons
Find other everyday comparisons – *e.g.* on the news, or in the newspaper – that are dimensionally faulty.

1.2 Free fall

The fall of an object without air resistance – free fall – illustrates how to, and how not to use dimensions in differential calculus. On the negative example, here is how many calculus textbooks present free-fall problems:

> A ball at rest falls from a height of h feet. Find its speed v in feet per second when it hits the ground, given a gravitational acceleration of g feet per second squared and neglecting air resistance.

The ball has the simplest initial velocity – zero – so that the solution does not obscure learning about the tool of dimensions. (For the general case with an initial velocity, try **Problem 2.27** on **page 58**.)

In the free-fall problem, the units of height (feet) have been separated from the variable h, making h a pure number or a dimensionless constant. Similarly, g and v are also pure numbers. Since g, h, and v are pure numbers – are dimensionless – any comparison of v with quantities derived from g and h is a comparison between dimensionless quantities. It is therefore a dimensionally valid comparison.

In short, purging the units makes it difficult to use dimensions in checking an answer, or in guessing answers. It is like fighting with one hand tied behind your back. Thus handicapped, you have to solve several equations involving derivatives – *i.e.* solve a differential equation:

$$\frac{d^2y}{dt^2} = -g, \text{ with } y(0) = h \text{ and } \dot{y}(0) = 0,$$

where $y(t)$ is the ball's height at t seconds, $\dot{y}(t)$ is the ball's velocity, and g is the gravitational acceleration. [Mathematics textbooks indicate derivatives with a prime, as in $y'(x)$. Furthermore, physics textbooks, whose convention I also use, indicate a time derivative using a dot, as in $\dot{y}(t)$.]

Problem 1.2 Calculus solution
Show, using calculus, that this second-order differential equation has the following solution:

$$\dot{y}(t) = -gt,$$
$$y(t) = -\frac{1}{2}gt^2 + h.$$

▶ *Use the result of* **Problem 1.2** *to show that the impact speed is* $\sqrt{2gh}$.

The height $y(t)$ determines the time of impact; the time, together with the velocity $\dot{y}(t)$, determines the impact speed. In particular, when $y(t) = 0$ the ball hits the ground. The impact happens at $t_0 = \sqrt{2h/g}$. The velocity at time t_0 is

$$\dot{y}(t_0) = -gt_0 = -g\sqrt{2h/g} = -\sqrt{2gh}.$$

The impact speed (the unsigned velocity) is therefore $\sqrt{2gh}$.

This derivation contains several spots for algebra mistakes: for example, not taking the square root when solving for t_0, or dividing rather than multiplying by g when finding the impact velocity. Enough practice prevents such mistakes on simple problems, but I want methods for complex problems with numerous and subtle pitfalls.

One method is dimensions. So here is the free-fall problem rewritten so that dimensions can help solve it:

> A ball at rest falls from a height h. Find its speed v when it hits the ground, given a gravitational acceleration g and neglecting air resistance.

Now the dimensions of height and gravitational acceleration belong, respectively, to the quantities h and g. Not only is this version simpler to state and more general than the unit-scrubbed version on **page 8**, but the dimensions help find the impact speed – without needing to solve differential equations.

In the free-fall problem, the dimensions of height h are length or L for short. The dimensions of gravitational acceleration g are length per time squared or LT^{-2}, where T stands for the dimension of time. The impact speed is a function of g and h, so it must be a combination of g and h with the dimensions of speed, which are LT^{-1}.

Problem 1.3 Dimensions of familiar quantities
In terms of length L, mass M, and time T, what are the dimensions of energy, power, and torque?

▶ *Show that $\sqrt{gh}$ has dimensions of velocity.*

The combination $\sqrt{gh}$ has the correct dimensions:

$$\sqrt{LT^{-2} \times L} = \sqrt{L^2T^{-2}} = LT^{-1},$$

which are the dimensions of speed.

▶ *Show that $\sqrt{gh}$ is the only combination of g and h with dimensions of speed (except for multiplication by a dimensionless constant).*

One way to show that $\sqrt{gh}$ is the unique combination is to propagate constraints. The method of constraint propagation is discussed in many references, of which my favorite is [3]. If the constraint lead to only one solution, then the solution is unique. Here, the starting, and only constraint is that any valid combination of g and h must have dimensions of speed. That constraint contains two smaller constraints: that the powers of length L and of time T match on both sides of the equation

$$v_i = \text{combination of } g \text{ and } h.$$

Time T shows up on the left side as T^{-1}. On the right side, it appears only in g, which contains T^{-2}. So, to satisfy the T constraint, a valid combination of g and h must have the form

$$v_i = \sqrt{g} \times \text{function of } h.$$

The second constraint is that the powers of L match on both sides. On the left side, speed contains L^1. On the right side, $\sqrt{g}$ contains $L^{1/2}$. The missing $L^{1/2}$ must come from the function of h, so that function is $\sqrt{h}$ times a dimensionless constant.

The two constraints, therefore, require that the speed have this form:

$$v_i = \sqrt{gh} \times \text{dimensionless constant}.$$

The idiom of $\times$dimensionless constant occurs so frequently that it deserves a compact notation:

$$v_i \sim \sqrt{gh}.$$

The symbol $\sim$ is a cousin of $=$. The cousins of $=$ include, in order of increasing exactness:

$\propto$: equal, except perhaps for a factor that may have dimensions
$\sim$: equal, except perhaps for a factor with no dimensions
$\approx$: equal, except perhaps for a factor near 1

Now compare the dimensions result $v \sim \sqrt{gh}$ to the exact result:

$$v_i = \sqrt{2} \times \sqrt{gh}.$$

The dimensions result lacks only the dimensionless factor $\sqrt{2}$. This factor is generally the least important factor. Consider the error if one instead omits the drop height h:

$$v \propto \sqrt{2g},$$

where the $\propto$ indicates that the dimensions are not be identical on the left and right sides. Imagine computing the impact speed for drop heights that range from 20 cm (a flea high jump) to 5 m (platform diving). This factor-of-25 variation in height contributes a factor-of-5 variation in impact speed. In contrast, the dimensionless factor contributes, independent of h, a mere factor of roughly 1.4. Most of the useful variation in impact speed comes not from the dimensionless factor but rather from the symbolic factors – just the factors calculated exactly by the method of dimensions.

By emphasizing the important information, the dimensions solution splits the exact result into meaningful, memorable pieces:

1. the dimensioned factor of $\sqrt{gh}$, the only dimensionally valid combination of g and h.
2. the dimensionless factor of $\sqrt{2}$. This factor results from integration or from otherwise solving the differential equation.

To recreate the exact solution, remember a short dimensions argument for $\sqrt{gh}$ and only one random piece of information: the dimensionless factor $\sqrt{2}$. This partition is emphasized by the following dimensionless form where the left side contains the dimensioned factors and the right side contains the dimensionless factor:

$$\frac{v}{\sqrt{gh}} = \sqrt{2}.$$

But the method requires quantities with dimensions! So:

Do not separate a quantity from its dimensions.

Their dimensions help you to check proposed solutions and to guess new ones.

The two examples – Exxon/Nigeria and free fall – show several benefits of the method of dimensions. First, it is easy to use, making algebra mistakes less likely and nonsense quicker to spot. Second, the method does *not* produce exact answers. Exact answers can contain too much information, for example less important factors such as $\sqrt{2}$ that obscure important factors such as $\sqrt{gh}$. As the great psychologist William James put it, 'The art of being wise is the art of knowing what to overlook' [4, Ch. 22]. Dimensions are an important tool in the box of a wise problem solver.

Problem 1.4 Vertical throw
You throw a ball directly upward with speed v. Roughly how long does it take to return to your hand? What is the dimensionless factor unaccounted for by dimensional analysis?

1.3 Integration

The free-fall example shows the value of not separating dimensioned quantities from their units. However, what if the quantities are dimensionless, such as the 5 and x in

$$\int_0^\infty e^{-5x^2}\,dx?$$

Or the quantities may have unspecified dimensions – a common case in mathematics because it is the universal language of science. Probability, for example, uses Gaussian integrals such as

$$\int_{x_1}^{x_2} e^{-x^2/2\sigma^2}\,dx,$$

where x could be height, detector error, or much else. Thermal physics, via the Boltzmann factor, uses similar integrals such as

$$\int e^{-\frac{1}{2}mv^2/kT},$$

where v is a speed. Mathematics, as the common language, studies the shared features of these integrals, and ends up with the general form

$$\int e^{-\alpha x^2},$$

where α and x have unspecified dimensions.

That vagueness gives mathematics its power of abstraction. How then, without losing the benefits of mathematical abstraction, to retain the benefit of dimensions? The answer is to assign a consistent set of dimensions to quantities with unspecified dimensions.

Here is the approach illustrated using this general Gaussian integral:

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\,dx.$$

This general form, unlike the specific variant with $\alpha = 5$, does not determine the dimensions of x or α, and that uncertainty provides the freedom needed for the method of dimensions.

The method's core idea is that any equation must be dimensionally valid. So, in the following equation the left and right sides have identical dimensions:

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\,dx = \text{something}.$$

To study the right side's dimensions, it helps to know what the right side looks like, beyond just knowing that it is 'something'.

▶ *Is the right side a function of x? It is a function of α? Does it have a constant of integration?*

The integration problem contains no symbolic quantities other than x and α. But x is the integration variable and the integral is over a definite range, so x disappears upon integration (and no constant of integration appears). Therefore the right side is a function only of α:

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\,dx = f(\alpha).$$

The function f might incorporate pure (dimensionless) numbers like 2/3 or $\sqrt{\pi}$, but α is its only input having dimensions.

For the equation to be dimensionally valid, the integral must have the same dimensions as $f(\alpha)$. But the dimensions of $f(\alpha)$ depend on the dimensions of α. So the procedure will have three steps:

1. assigning dimensions to α,
2. finding the dimensions of the integral, and
3. making an $f(\alpha)$ with identical dimensions.

1.3.1 Assigning dimensions to α

Before deciding the dimensions of α, let's analyze the constraint on those dimensions. The constraint arises because the α appears in the exponent of $e^{-\alpha x^2}$, and an exponent, such as n in 2^n, says how many times to multiply a quantity by itself:

$$2^n = \underbrace{2 \times 2 \times \cdots \times 2}_{n \text{ terms}}.$$

The notion of 'how many times' is a pure number. The number might be a positive integer – the most intuitive case – a negative integer, a fraction, or even a complex number; but it is always a pure number. Therefore:

> An exponent is dimensionless.

Hence the product αx^2 must be dimensionless. So, assigning dimensions to x is equivalent to assigning dimensions to α.

An effective but complicated choice is not to assign dimensions to x or α. Instead, leave their dimensions unknown and manipulate them as $[x]$ and $[\alpha]$, where the bracket notation [quantity] is a convenient shorthand for the dimensions of the quantity. Then the requirement that αx^2 be dimensionless means

$$[\alpha] \times [x]^2 = 1,$$

so

$$[\alpha] = [x]^{-2}.$$

In this approach, with its plethora of brackets, the notation risks burying the reasoning.

A too-simple alternative is to make x dimensionless. That choice makes α and $f(\alpha)$ dimensionless. However, dimensions then cannot restrict the form of $f(\alpha)$. As with the sin of separating a quantity from its units (**Section 1.2**), making x and α dimensionless ties one hand behind your back – not wise preparation for a street fight.

An almost-as-simple but effective alternative is to make x a length. This choice is natural if you imagine the x axis lying on the floor. Then x is the (signed) distance from the origin, which has dimensions of length. To make αx^2 dimensionless, the dimensions of α must be L^{-2}. In symbols:

$$[x] = L,$$
$$[\alpha] = L^{-2}.$$

1.3.2 Dimensions of the integral

The dimensional assignments enable the next step: finding the dimensions of the integral:

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\, dx.$$

The dimensions of an integral depend on three factors:

1. the dimensions of an integral sign – or how the integral sign affects dimensions;
2. the dimensions of the integrand $e^{-\alpha x^2}$; and
3. the dimensions of the dx factor.

Look first at the dimensions of an integral sign. It originated as an elongated 'S' standing for *Summe*, the German word for sum: Vertically stretch an 'S' or a summation sign $\sum$, and they turn into an integral sign. Indeed, an integral is the limit of a sum:

$$\lim_{\Delta x \to 0} \sum_{n=-\infty}^{\infty} e^{-\alpha (n\Delta x)^2}\, \Delta x = \int_{-\infty}^{\infty} e^{-\alpha x^2}\, dx.$$

This historical information is useful in conjunction with the fundamental principle of dimensions:

> Do not add apples to oranges.

Lengths can be added only to lengths, and times only to times. Therefore all terms in a sum, and the sum itself, have identical dimensions. Similarly, the dimensions of the integral are the dimensions of one term in the integral, namely $e^{-\alpha x^2}\,dx$. This argument is summarized in the following principle:

Integration does not affect dimensions:

$$\left[\int f(x)\,dx\right] = [f(x)\,dx] = [f(x)] \times [dx]\,.$$

An equivalent but slightly cryptic and short statement is:

The integration sign $\int$ does not affect dimensions.

Problem 1.5 Integrating velocity
Position is the integral of velocity, but position and velocity have different dimensions. How can this difference be consistent with the maxim that integration does not affect dimensions?

Since the integration sign does not affect dimensions, the dimensions of the integral are the dimensions of the exponential factor $e^{-\alpha x^2}$ multiplied by the dimensions of dx. The exponential, despite its fierce exponent $-\alpha x^2$, is merely several copies of the e multiplied together. Since e has no dimensions, e^{anything} also has no dimensions. Therefore the exponential factor $e^{-\alpha x^2}$ contributes no dimensions to the integral.

▶ *What are the dimensions of* dx*?*

The factor of dx, however, is another story. To find the dimensions of dx, follow the venerable advice of Silvanus Thompson [5, p. 1]: Read d to mean 'a little bit of'. Then dx means 'a little bit of x'. Since a little length is still a length, dx is a length. In general:

dx has the same dimensions as x.

Another route to this conclusion is via the definition of dx. It comes from Δx where Δx is, for example, $x_{n+1} - x_n$. The difference of two x's has the same dimensions as x, so Δx, dx, and x have identical dimensions. A slightly cryptic but compact statement is the differential counterpart to the maxim that the integration sign does not affect dimensions:

> The d operation does not affect dimensions.

Assembling all the pieces, the whole integral has dimensions of length:

$$\underbrace{\left[\int e^{-\alpha x^2}\,dx\right]}_{L} = \underbrace{\left[e^{-\alpha x^2}\right]}_{1} \times \underbrace{[dx]}_{L}.$$

1.3.3 Making an $f(\alpha)$ with correct dimensions

The next step is to make $f(\alpha)$ have the same dimensions as the integral. The dimensions of α are L^{-2}, so the only way to turn α into a length is to form the power $\alpha^{-1/2}$, perhaps multiplied by a dimensionless constant. So

$$f(\alpha) \sim \alpha^{-1/2}.$$

To determine the dimensionless constant hidden in the $\sim$ symbol, set $\alpha = 1$:

$$f(1) \equiv \int_{-\infty}^{\infty} e^{-x^2}\,dx.$$

This classic integral is evaluated in **Section 2.1.2** and shown to be $\sqrt{\pi}$. So that $f(\alpha) \sim \alpha^{-1/2}$ and $f(1) = \sqrt{\pi}$,

$$f(\alpha) = \sqrt{\frac{\pi}{\alpha}}.$$

In other words,

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\,dx = \sqrt{\frac{\pi}{\alpha}}.$$

This method of doing an integral suggests the following maxim:

> Assign dimensions to quantities with unspecified dimensions, then use the method of dimensions to guess the answer.

Problem 1.6 Change of variable
Use a variable substitution to show that $f(\alpha) = \alpha^{-1/2}$.

Setting $\alpha = 1$ is an example of easy-cases reasoning, the topic of **Chapter 2**, and it slightly contradicts the theme of this chapter. For **Section 1.3.1** exhorted you to assign dimensions to α, which turned out to be L^{-2}. Setting $\alpha = 1$ violates that specification. However, this temporary contradiction is useful because it helps show, in conjunction with the calculation in **Section 2.1.2**, that the dimensionless constant in $f(\alpha)$ is $\sqrt{\pi}$.

Problem 1.7 More difficult exponential
Use dimensions to investigate

$$\int_0^\infty e^{-\alpha t^3}\, dt.$$

1.4 Differentiation using dimensions

Because the reverse of integration is differentiation, dimensions also help analyze differentiation. Start with the idea that a derivative is a ratio of two differentials df and dx:

$$f'(x) = \frac{df}{dx}.$$

Since d means 'a little bit of',

$$f'(x) = \frac{df}{dx} = \frac{\text{a little bit of } f}{\text{a little bit of } x}.$$

The dimensions of $f'(x)$ are therefore:

$$\left[f'(x)\right] = \frac{[\text{a little bit of } f]}{[\text{a little bit of } x]}.$$

Since a little bit of a quantity has the same dimensions as the quantity itself,

$$\left[f'(x)\right] = \frac{[\text{a little bit of } f]}{[\text{a little bit of } x]} = \frac{[f]}{[x]}.$$

Differentiating with respect to x is, in its effect on dimensions, equivalent to dividing by x.

In other words, $f'(x)$ and f/x have identical dimensions.

This strange conclusion is worth testing with a familiar example. Take distance s as the function to differentiate, and time as the independent variable. The derivative of $s(t)$ is $\dot{s}(t) = ds/dt$. Are the dimensions of $\dot{s}(t)$ the same as the dimensions of s/t? Yes: The derivative $\dot{s}(t)$ is velocity, which has dimensions LT^{-1}, as does the quotient s/t.

Problem 1.8 Dimensions of a second derivative
What are the dimensions of $f''(x)$?

1.4.1 Estimating the magnitude of a derivative: The secant method

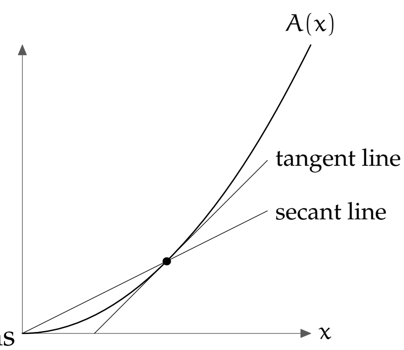
Since df/dx and f/x have identical dimensions, perhaps they have identical or similar magnitudes? They often do. As an easy test, pick a function with an easy-to-compute derivative – for example, a function such as $A(x) = x^2$, which could represent the area of a square with side x.

The dimensions of $A'(x)$ are

$$[A'] = \frac{[A]}{[x]}.$$

To estimate A' itself, erase the brackets:

$$A'(x) \sim \frac{A}{x}.$$



The twiddle $\sim$ reflects the possible dubiousness of removing the brackets, and means that the equation might lack a dimensionless factor. This approximation is illustrated in the figure. The approximation's left side $A'(x)$ is the slope of the tangent line. The approximation's right side $A(x)/x$ is the slope of the secant line. With $A(x) = x^2$, the two slopes differ by only a factor of 2 (a dimensionless constant):

$$A'(x) = 2x \qquad \text{and} \qquad \frac{A(x)}{x} = x.$$

Problem 1.9 Higher powers
Investigate the secant approximation for $A(x) = x^n$.

Problem 1.10 Second derivatives
Use the secant approximation to estimate $A''(x)$ with $A(x) = x^2$. How does the approximation compare to the exact second derivative?

▶ *How well does the secant approximation work for* $A(x) = x^2 + 100$*?*

The secant approximation to the tangent is useful. However, as shown when $A(x) = x^2 + 100$, it is limited. At $x = 1$, the secant and tangent lines then have these slopes:

$$\begin{cases} A'(1) = 2 & \text{(tangent);} \\ \dfrac{A(1)}{1} = 101 & \text{(secant).} \end{cases}$$

Although the ratio of these two slopes is dimensionless, it is also distressingly large – more than 50.

Problem 1.11 Investigating the discrepancy
With $A(x) = x^2 + 100$, *sketch* the ratio

$$\frac{\text{secant slope}}{\text{tangent slope}}$$

as a function of x.

Confirm that the ratio is dimensionless. However, it is not constant! But the argument from dimensions was that $A'(x) \sim A(x)/x$, indicating that there is a missing dimensionless factor. How can the dimensionless factor not be constant (tricky!)?

The large discrepancy must lie in the approximations made by the secant method. It replaces the derivative $A'(x)$ with $A(x)/x$. Those forms look significantly different and therefore hard to compare in general. However, the secant slope $A(x)/x$ can be cast into a form similar to the derivative. The derivative, or tangent slope, has this definition:

$$A'(x) = \lim_{\Delta x \to 0} \frac{A(x) - A(x - \Delta x)}{\Delta x}.$$

The secant slope can be massaged into a similar form by subtracting 0 from the numerator and denominator:

$$\frac{A(x)}{x} = \frac{A(x) - 0}{x - 0}.$$

Setting $\Delta x = x$ in definition of derivative makes it almost the same as the secant slope:

$$A'(x) \approx \frac{A(x) - A(0)}{x}.$$

Of course, $\Delta x = x$ rather than $\Delta x \to 0$ gives us only an approximation to a derivative. But that is not the main problem with the secant approximation. Rather, the main problem is that the secant approximation replaces $A(0)$ with 0. Here are the two forms side by side to facilitate comparison:

$$\underbrace{\frac{A(x) - 0}{x - 0}}_{\text{secant slope}} \quad \text{and} \quad \underbrace{\frac{A(x) - A(0)}{x}}_{\text{approximate tangent slope}}.$$

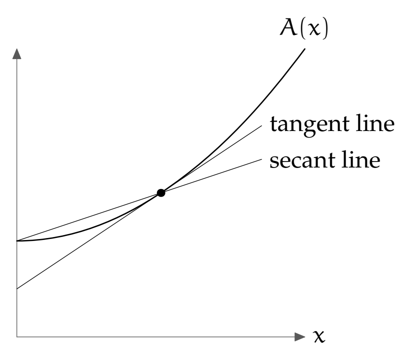
So, the flaw in the secant method is that the line to the point $(x, A(x))$ starts from the origin $(0, 0)$, which might not be a point worthy of such distinction. Let's correct this overemphasis.

1.4.2 An improved secant method

To improve the secant approximation, start the secant line not from $(0, 0)$ but rather from $(0, A(0))$.

▶ *Using* $(0, A(0))$ *as the starting point for the secant, compare the secant and tangent slopes for* $A(x) = x^2 + 0.5$.

Let's call the new approach the $x = 0$ secant approximation. Here is a picture of it with $A(x) = x^2 + 0.5$. The slope of the tangent is now always double the slope of the secant, as with the vanilla secant method applied to $A(x) = x^2$. The $x = 0$ secant approximation is not bothered by our adding a constant to $A(x)$. In fancy words, it is unaffected by vertical translation. In steadily fancier words, it is robust against vertical translation or is invariant to vertical translation.



▶ *Experiment with horizontal translation.*

The next question is whether the improved method is ruined by horizontal translation. To find out, translate the original $A(x) = x^2$ to the right by 100 units to get $A(x) = (x - 100)^2$. The $x = 0$ secant approximation starts from $(0, 10^4)$. So when $x = 100$, for example, the $x = 0$ secant line goes from $(0, 10^4)$ to $(100, 0)$, and has a slope of -100. The tangent, however, has zero slope.

So, even the improved secant approximation has problems. The original secant approximation hardcoded $(0, 0)$ into the method. The improved approximation hardcoded $(0, A(0))$ into the method. The improvement corrects only half of the lack of invariance in the original method: 0 is still singled out as the special value of x.

1.4.3 Significant-change method

A method that is invariant to vertical and horizontal translation is to use criteria that 'move with the curve', in other words, that depend on how $A(x)$ behaves around the point of interest. One example is to replace the exact definition of derivative

$$A'(x) = \lim_{\Delta x \to 0} \frac{A(x + \Delta x) - A(x)}{\Delta x}$$

with the approximation of letting Δx remain finite. For example, stop the limit at $\Delta x = 0.01$, which seems tiny enough.

That choice has two related defects. First, it cannot work if x has dimensions. For example, if x is a length, what length is small enough? The choice depends on the problem: In a problem about the solar system, 1 mm would be sufficiently small; in a problem about fog droplets, 1 mm could be too large.

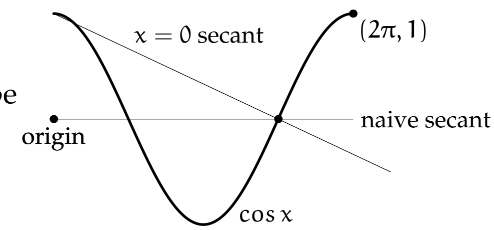
The second defect extends the first using fancy words: The choice $\Delta x = 0.01$, or any fixed choice, is not scale invariant. Although $\Delta x = 0.01$ might produce reasonably accurate derivative estimates for the function $A(x) = \sin x$, it would fail if x is rescaled to $1000x$ to make $A(x) = \sin 1000x$.

These problems suggest the following method that is invariant to vertical and horizontal translations and to changes of scale:

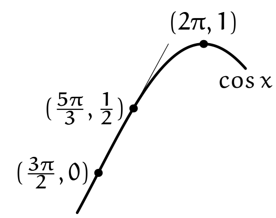
$$A'(x) \sim \frac{\text{a significant change in } A \text{ near } x}{\text{change in } x \text{ for } A \text{ to change significantly}}.$$

The term 'significant change' is vague, but an example will illustrate its sense and how to use this approximation. Choose $A(x) = \cos x$ and estimate $A'(3\pi/2)$ using the three methods: the naive secant approximation, the $x = 0$ secant approximation, and the significant-change approximation.

In the naive secant method, the secant line passes through $(0, 0)$ and $(3\pi/2, 0)$. So it has zero slope – a poor approximation to the exact slope $A'(3\pi/2) = 1$. In the $x = 0$ secant method, the secant line passes through $(0, 1)$ and $(3\pi/2, 0)$. So it has slope $-2/3\pi$ – a terrible approximation to the true slope, since it has the wrong sign!



Let's hope that the significant-change approximation rescues us. Starting at $x = 3\pi/2$, the function $A(x) = \cos x$ increases from 0 toward 1 (at $x = 2\pi$); so, slightly arbitrarily, call 1/2 a significant change in $A(x)$. Since $\cos x$ reaches 1/2 when $x = 3\pi/2 + \pi/6$, the derivative estimate is



$$A'(x) \sim \frac{\text{a significant change in } A \text{ near } x}{\text{change in } x \text{ for } A \text{ to change significantly}} \sim \frac{1/2}{\pi/6} = \frac{3}{\pi}.$$

Numerically, $A'(x) \approx 0.955$ – amazingly close to the true slope of 1.

This example shows the significant-change approximation in a particularly accurate light. However, the naive and the $x = 0$ secant approximations remain useful. They are easy to apply, especially the naive secant approximation: It replaces $A'(x)$ by simply $A(x)/x$ – with no need to estimate any quantity. When you know so little about the function that you cannot decide upon a significant change in its value, fall back to the secant approximations. Examples are given in **Problem 1.23** and **Problem 1.24**.

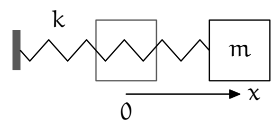
Problem 1.12 Derivative of a quadratic
With $A(x) = x^2$, estimate $A'(5)$ using the three methods – naive secant, $x = 0$ secant, and significant change – and compare those estimates with the true slope.

Problem 1.13 Derivative of the logarithm
Use the significant-change method to estimate the derivative of $A(x) = \ln x$ at $x = 10$, and compare the estimate with the true slope.

1.5 Differential equations: The spring–mass system

Having used dimensions to simplify integration (**Section 1.3**) and differentiation (**Section 1.4**), the next peak to scale is using dimensions to understand differential equations.

The example is the differential equation that describes the motion of a block connected to an ideal spring:



$$m\frac{d^2x}{dt^2} + kx = 0,$$

where m is the mass of the block, x is its position relative to equilibrium, t is time, and k is the spring's stiffness or spring constant. Let's analyze the motion after pulling the block a distance x_0 to the right and, at $t = 0$, releasing it.

1.5.1 Checking its dimensions

Upon encountering an equation, a healthy reflex is to check its dimensions. If the dimensions of all terms are not identical, the equation is not worth solving – a great saving of effort. Even when the dimensions match, the effort of checking dimensions is worthwhile because it prompts reflection on the meaning of the terms in the equation – an important step in understanding and solving it.

▶ *What are the dimensions of the two terms in the spring equation?*

The first term contains the second derivative d^2x/dt^2, which is an acceleration, so the first term is mass times acceleration. That product is the right side of Newton's second law $F = ma$ and therefore has dimensions of force.

That second derivative was a familiar one. However, many differential equations contain unfamiliar derivatives. As an example, here are the Navier–Stokes equations of fluid mechanics (discussed in **Section 2.4**):

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\boldsymbol{\nabla})\mathbf{v} = -\frac{1}{\rho}\boldsymbol{\nabla} p + \nu\boldsymbol{\nabla}^2\mathbf{v}.$$

What are the dimensions of those derivatives?! To prepare for such occasions, let's practice by finding the dimensions of d^2x/dt^2 from first principles.

▶ *Which choice is correct:*

$$\left[\frac{d^2x}{dt^2}\right] = \begin{cases} [x^2] / [t^2] \\ \textit{or} \\ [x] / [t^2]? \end{cases}$$

Both choices are plausible. Their denominators contain $[t^2]$, matching the dt^2 in the denominator of the second derivative. The difference between the choices is in the numerator: $[x^2]$ versus $[x]$. Each option has merit. In favor of $[x^2]$, the exponent 2 matches the exponent in d^2x. In favor of $[x]$, the exponent 2 in d^2x belongs to the d rather than to the x, which appears only to the first power. Which consideration wins?

To decide, use the principle from **Section 1.3.2** that d means 'a little bit of'. Then dx means 'a little bit of x', and $d^2x = d(dx)$ means 'a little bit of a little bit of x'. Thus d^2x has the same dimensions as x rather than as x^2. Equivalently, the derivative and therefore the second derivative are linear operations. So the dimensions of d^2x/dt^2 contain the dimensions of x rather than x^2.

The other half of the second derivative is the denominator dt^2 – a convenient notation but also a lazy notation because it could mean $(dt)^2$ or $d(t^2)$. It turns out to mean $(dt)^2$; however, even if it meant $d(t^2)$, its dimensions would still be the dimensions of t^2. So

$$\left[\frac{d^2x}{dt^2}\right] = \frac{[x]}{[t^2]}.$$

This result applies to d^2x/dt^2 no matter the meaning of x and t. In the spring equation, x means distance and t means time, so the dimensions of d^2x/dt^2 are LT^{-2}. As expected, it is an acceleration, and the equation's first term $m(d^2x/dt^2)$ has dimensions of force.

▶ *What are the dimensions of the second term kx?*

An ideal spring exerts a force proportional to its extension x (Hooke's law). The constant of proportionality is the stiffness k:

$$k \equiv \frac{\text{force}}{\text{extension}}.$$

So kx is a force, and the two terms of the spring equation have identical dimensions.

Problem 1.14 Dimensions of spring constant
What are the dimensions of k?

Problem 1.15 Dimensions of surface tension
Surface tension, often denoted γ, is the energy required to create a unit area of new surface. What are the dimensions of γ?

Problem 1.16 Getting the sign right
In the spring equation, should the kx term be $+kx$, as it is written now, or should it be $-kx$?

1.5.2 Estimating the magnitudes of the terms

Since the spring equation has correct dimensions, it is not obviously nonsense, and it may be worthy of further analysis. One method of analysis, for any differential equation, is to replace derivatives by their approximate magnitudes. This step turns differential equations into algebraic equations – a great simplification.

One way to estimate these magnitudes is the significant-change method of **Section 1.4.3**. In that section, the examples used first derivatives; however, the method extends to higher derivatives:

$$\frac{d^2x}{dt^2} \sim \frac{\text{significant change in } x}{(\text{change in } t \text{ for } x \text{ to change significantly})^2}.$$

Problem 1.17 Explanation
Why is the numerator only the first power of a significant change in x, whereas the denominator is the second power of the corresponding change in t?

First estimate the numerator: a significant change in x. The initial position is x_0; as the spring oscillates, the position varies between $-x_0$ and $+x_0$. So x_0 is a significant change; many other choices, such as $0.5x_0$, $1.2x_0$, are also significant changes, but x_0 is the simple choice.

Now estimate the denominator: the time for the block to move a distance comparable to x_0. It is difficult to estimate directly without solving the spring equation – the task that these approximate methods are supposed to avoid. So for the moment leave that time unknown, calling it τ. It is the time constant or characteristic time of the spring–mass system and is related to the oscillation period. The relation will become apparent after solving for τ approximately and comparing the approximate and exact solutions.

With x_0 as the significant change in x and τ as the corresponding time, the block's acceleration d^2x/dt^2 is roughly x_0/τ^2. So the spring equation's first term $m(d^2x/dt^2)$ is roughly mx_0/τ^2.

▶ *What does 'is roughly' mean in this context?*

'Is roughly' cannot mean that mx_0/τ^2 and $m(d^2x/dt^2)$ are within a fixed tolerance, say within a factor of 2: $m(d^2x/dt^2)$ sweeps from positive values through zero to negative values and back again, whereas mx_0/τ^2 remains constant. Rather, 'is roughly' means that a typical value of $m(d^2x/dt^2)$ – for example, its root-mean-square value or the mean of its absolute value – is comparable to mx_0/τ^2. So let's extend the twiddle notation $\sim$ to include this frequent usage, and write

$$m\frac{d^2x}{dt^2} \sim \frac{mx_0}{\tau^2}.$$

Now use the same meaning of 'is roughly' to estimate the spring equation's second term kx. For the position x, a typical value is comparable to x_0, so $kx \sim kx_0$.

The magnitudes of the two terms are connected by the spring equation:

$$m\frac{d^2x}{dt^2} + kx = 0$$

The zero right side means that the two magnitudes are comparable:

$$\frac{mx_0}{\tau^2} \sim kx_0.$$

The amplitude x_0 divides out! That kind of cancellation always happens with a linear differential equation. With x_0 gone, it cannot affect the time constant τ. This independence has a physical interpretation:

> In ideal spring motion – so-called simple harmonic motion – the oscillation period is independent of amplitude.

The argument for this conclusion involved several approximations, but the conclusion is exact.

After cancelling the x_0, the leftover is $m/\tau^2 \sim k$, so the characteristic time is $\tau \sim \sqrt{m/k}$: The block needs that much time to move a significant distance. The reciprocal of time is frequency, so the position changes significantly at a frequency $1/\tau \sim \sqrt{k/m}$.

▶ *Is this frequency an angular frequency or a circular frequency? (They differ by a factor of 2π.)*

To decide whether $\sqrt{k/m}$ is a circular or angular frequency, compare it to the exact solution of the spring equation. (Having an exact solution for comparison is one reason to use the spring equation to develop these approximation methods.) The exact solution is

$$x = x_0 \cos \omega t,$$

where $\omega = \sqrt{k/m}$, which is also $1/\tau$. Since ω appears in ωt without any factor of 2π, it is an angular frequency. Therefore, the period is

$$T = 2\pi\sqrt{\frac{m}{k}}.$$

Problem 1.18 Checking dimensions in the alleged solution
What are the dimensions of ωt? What are the dimensions of $\cos \omega t$? Check the dimensions of the proposed solution $x = x_0 \cos \omega t$, and the dimensions of the proposed period $2\pi\sqrt{m/k}$.

Problem 1.19 Verification
Substitute $x = x_0 \cos \omega t$ into the spring differential equation to check that $\omega = \sqrt{k/m}$.

The approximate analysis of the spring equation shows how to turn differential equations into algebraic equations. Although this conversion sacrifices accuracy, algebra is simpler than calculus. Therefore:

> Turn differential equations into algebraic equations by replacing derivatives with their approximate magnitudes or typical values.

We developed this technique using the spring equation partly because its exact solution is well known. That solution provides a second interpretation of the characteristic time τ. Look at how the position changes as a function of time. When the phase angle ωt is zero, the mass is at the release position $x = x_0$. When $\omega t = \pi/2$, the mass reaches the equilibrium position $x = 0$, at which point x has changed very significantly. Since $\pi/2$ is slightly larger than 1, a reasonable candidate for when x has made a significant change is when $\omega t = 1$: in other words, when the phase ωt has advanced by 1 rad. This result applies to many systems:

> A system usually changes significantly when cosines or sines change by 1 rad.

The corresponding advance in t is $\sqrt{k/m}$, a reasonable estimate for the characteristic time and in agreement with the estimate made by turning the spring differential equation into an algebraic equation.

1.6 Summary and problems

This chapter introduced several rules of thumb and problem-solving methods:

1. Do not add apples to oranges: Every term in an equation or sum must have identical dimensions.
2. Preserve the dimensions for quantities that have them. Instead of 'g meters per second squared', simply write g.
3. The differential symbol d, as in dx, means 'a little bit of' and is dimensionless. Therefore, the dimensions of a derivative $f'(x)$ are the dimensions of f/x.
4. Exponents are dimensionless – as are the arguments of most elementary functions (such as sine and cosine).
5. The dimensions of an integral are the dimensions of its contents, *including* the differential (e.g. dx or dA). The integral sign itself is dimensionless.
6. Choose dimensions for quantities with unspecified dimensions, such as x and α in
$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\, dx.$$
7. The magnitude of df/dx is roughly

$$\frac{\text{significant change in } f \text{ near } x}{\text{change in } x \text{ over which } f \text{ changes significantly}}$$

This approximation is a great simplifier: It turns a differential equation into an algebraic equation.

The following problems apply and extend these ideas.

Problem 1.20 Arcsine integral
Given this integral:

$$\int \sqrt{1-x^2}\,dx = \frac{\arcsin x}{2} + \frac{x\sqrt{1-x^2}}{2},$$

use dimensions to find $\int \sqrt{1-3x^2}\,dx$.

Problem 1.21 Integral of an exponential
Using dimensions, find

$$\int_0^\infty e^{-ax}\,dx.$$

Problem 1.22 Arctangent
Using dimensions, find

$$\int \frac{dx}{x^2+a^2}.$$

A useful result is the following:

$$\int \frac{dx}{x^2+1} = \tan^{-1} x + C.$$

Problem 1.23 Lennard–Jones potential
The Lennard–Jones potential is a model of the interaction energy between two nonpolar molecules. It has the form

$$V(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Use the (naive) secant method to find r_0, the distance r at which $V(r)$ is a minimum. Compare the answer to the r_0 found using honest methods (calculus).

Problem 1.24 Maxima and minima on the cheap
Let $f(x)$ be an increasing function and $g(x)$ a decreasing function. Use the naive secant approximation to show that $h(x) = f(x) + g(x)$ has a minimum where $f(x) = g(x)$. This problem generalizes Problem 1.23.

This balancing heuristic is simple, quick, and has a physical interpretation. If $f(x)$ and $g(x)$ represent energy contributions from two physical processes, such as gravitational and thermal energy, then the whole system is in equilibrium when the two processes balance (when $f(x) = g(x)$).

Problem 1.25 Kepler's third law

Newton's law of universal gravitation – the famous inverse-square law – says that the gravitational force between two masses is

$$F = \frac{Gm_1 m_2}{r^2},$$

where G is Newton's constant, m_1 and m_2 are the two masses, and r is their separation. Combine that law with Newton's second law for a planet orbiting the sun:

$$m\frac{d^2\mathbf{r}}{dt^2} = \frac{GMm}{r^2}\hat{\mathbf{r}},$$

where M is the mass of the sun and m the mass of the planet.

Use approximation methods to find how the orbital period τ varies with orbital radius r, and compare your result with Kepler's third law.

Problem 1.26 Don't integrals compute areas?

A common belief is that integration computes areas. Areas have dimensions of L^2. So how can the Guassian integral in Section 1.3.2 have dimensions of L?