Calculus and the real world

April 5, 2014

Overview

These notes review the following main concepts:

- Definition of quantities as ratios
- Derivatives as ratios of increments in a function $f(x)$ over corresponding increments in x
- Differentiation as giving the slope of a graph
- Integrals as Riemann sums
- Integrals as areas under graphs
- Integrals as inverses of differentiation

These will be needed throughout the course.

Ratios

In physics and other areas of science, there are many quantities defined as ratios. Take the simple, high school definition of velocity

$$\text{velocity} = \frac{\text{displacement}}{\text{time elapsed}}$$

You will hopefully know that velocity is a a vector. We can keep things simple here by looking only at the velocity of something that moves along a straight line, for instance a car on a straight road. Let D be displacement — positive if motion is in one agreed-upon direction along the straight line, for instance North along a North-South road, and negative if in the other direction —and T time elapsed. Then the equation above can be written as

$$v = \frac{D}{T}. \tag{1}$$

Why does it make sense to define velocity as this particular ratio?

We presumably want velocity to be something that is bigger if the displacement D travelled in a fixed time interval T is bigger, for instance a car that covers 100 metres in 5 seconds is going faster than a car that covers only 50 metres in 5 seconds. We also want velocity to be bigger if the time T taken to cover a fixed distance D is shorter. For instance, a car that takes 5 seconds to cover 100 m is going faster than one that takes 10 seconds. So we want v to go up when D is made larger or when T is smaller. The definition above certainly does that. But so would many others: for instance

$$v = \frac{D}{T^2}$$

also increases with increasing D and decreasing T. So why specifically the ratio

$$v = \frac{D}{T}?$$

The real answer is that we expect distance covered to be proportional to time elapsed, at least for steady motion — something we will come back to shortly. In other words, wait twice the amount of time and the distance covered will be twice as large. Symbolically,

$$D \propto T$$

which means 'D equals T times some constant of proportionality.' Velocity is precisely that constant of proportionality

$$D = vT. \tag{2}$$

There are other examples we can easily think of. Why define density as

$$\rho = \frac{M}{V}$$

for a sample of material of mass M occupying volume V? We could go through the same argument as above — density increases when M is increased and decrease when V is increased, but there are many ways of combining M and V that do just that, besides M/V. The point is once more that, for a given material under constant conditions of pressure, temperature etc, we expect mass to be proportional to volume: twice the volume of the same stuff will contain twice the mass, so

$$M \propto V.$$

ρ is now the constant of proportionality, $M = \rho V$.

Differentiation

The argument that distance travelled D is proportional to time T is true only if the motion is steady: the velocity does not change. So how should we understand velocity when it is changing? If D and T are not proportional to each other, then

$$\frac{D}{T}$$

is not constant, but depends on the time T elapse. Figure **??** illustrates this. This also means we cannot sensibly define $v = D/T$ as velocity at a particular point in time, since the value of v then depends on the interval T.

The point is that, over a *short* time interval, velocity does not change by much. Even if there are forces acting on an object, these take time to change its velocity. This means that, provided the time elapse T is short, the distance D is still proportional to T, something which you might think of as an approximate proportionality that holds only when T is short.

Using D and T as symbols is not really the most useful approach now. We have just discussed that D *depends* on T, so we need to think of displacement as a *function* of time elapsed. For small T, that function can be approximated by

$$D \approx vT, \tag{3}$$

but that does not work in general for larger time intervals.

More useful still is to recognize that displacement is the difference between the *position* of the travelling object at some initial time t_0 and some later time t_1. This means we should define position as a function of time, for instance as

$$x = x(t).$$

Then

$$D = x(t_1) - x(t_0), \qquad T = t_1 - t_0$$

and (3) becomes

$$x(t_1) - x(t_0) \approx v \times (t_1 - t_0)$$

provided the time t_1 is close to the original time t_0. From this, the velocity can be approximated by

$$v \approx \frac{x(t_1) - x(t_0)}{t_1 - t_0}. \tag{4}$$

Basically, we are starting a stopwatch at an some given initial point in time t_0, having measured the initial position $x(t_0)$ and waiting until t_1 to measure a new position $x(t_1)$ in order to compute velocity.

We assume that the approximation involved in (4) becomes better and better as the time interval $t_1 - t_0$ is shrunk, which we do by allowing less time to elapse from the initial instant t_0. In other words, we 'move' the time t_1 at which we choose to stop the stopwatch closer and closer to the initial instant t_0 at which the stopwatch is started. In practice, the choice of time interval is limited by the accuracy with which we can keep time and with which we can measure position. The question of how short the tune interval $t_1 - t_0$ needs to be in order for (4) to give a good approximation depends on the particular physical situation: if we try to measure the velocity of a car, we may need to make the time interval a small fraction of a second (cars can brake and accelerate significantly in the space of a second) while when measuring the velocity of a planet like the Earth, a time interval of a day may suffice (because a day is short compared with the length of time it takes the Earth to complete an orbit, during which its velocity changes significantly).

Mathematically, we can sidestep that issue by saying that we take the *limit* of t_1 approaching the fixed initial time t_0, and put

$$v = \lim_{t_1 \to t_0} \frac{x(t_1) - x(t_0)}{t_1 - t_0}. \tag{5}$$

Note that this is a matter of *definition*: we simply agree that this is a sensible way of defining the concept of 'velocity' mathematically. Mathematically, of course, the issues of being able to *measure* the time interval $t_1 - t_0$ and the positions $x(t_1)$ and $x(t_0)$ do not occur, so we can take a limit.

Note 1 *One thing that is important about the definition* (5) *is that it is* consistent *with the earlier definiton* (1). *If velocity is constant, the two definitons give the same answer. With constant velocity, we can compute the same value of v through $v = D/T = [x(t_1) - x(t_0)]/[t_1 - t_0]$, regardless of how big $t_1 - t_0$ is. The limit on the right-hand side of* (5) *becomes obsolete, but doesn't change the answer. Retaining the limit simply* generalizes *the original definition* (1).

Note 2 *You may have read or heard about the mathematical definition of a limit. This course is not concerned with formal mathematics, but with making conceptual and practical connections between mathematics and physical sciences, in particular geophysics, and especially in using mathematics to formulate geophysics questions and finding answers to them. Nonetheless, it is still worth knowing about formal definitions and why they work. The explanation below only scratches the surface of what a limit is and how to deal with limits in practice, but you may find it useful nevertheless — although you will not need it to follow the rest of the course.*

(5) *formally says that, given a fixed initial time t_0 and any positive number δ, no matter how small, we can find another positive number ϵ that is small enough so that*

$$\left| v - \frac{x(t_1) - x(t_0)}{t_1 - t_0} \right| < \delta \qquad \textit{whenever } |t_1 - t_0| < \epsilon \tag{6}$$

Now, what does that actually mean?

Remember that

$$\frac{x(t_1) - x(t_0)}{t_1 - t_0} \tag{7}$$

is displacement travelled over time elapsed for some finite time interval $t_1 - t_0$; this is a quantity we could actually calculate from measurements, assuming we can make measurements accurately enough. We see that this is used as an approximation to velocity in (4). But what does it mean to say it is an 'approximation'?

The first inequality in (6),

$$\left| v - \frac{x(t_1) - x(t_0)}{t_1 - t_0} \right| < \delta$$

tells us that the difference between the actual *velocity v and its approximation by $[x(t_1) - x(t_0)]/(t_1 - t_0)$ — which is the same as D/T — is less than δ. The modulus sign simply takes care of the possiblity that the difference between actual velocity and its approximation might be negative. Think of δ as an error tolerance. We measure $[x(t_1) - x(t_0)]/(t_1 - t_0)$, and we want that to be close to the actual velocity. In this case, we are saying we want it to be closer to v than the amount δ; 'closeness' is quantified as the difference between the two quantities. What the definition of the limit says is that, no matter how small or 'tight' we make our error tolerance δ, we can always shrink our measurement interval $t_1 - t_0$ to a level so that we meet that error tolerance. ϵ simply tells us how small the time interval needs to be in order to meet the error tolerance. Generally, of course, the smaller the error tolerance δ, the smaller the allowable time interval ϵ.*

Now, (5) is nothing more than the definition of a derivative. It just takes getting used to the fact that a derivative can be written in a number of different ways, by re-writing (5) using simple algebra.

Perhaps the most common way to write a derivative, at least in physics, is this: instead of saying I have an initial time t_0 and a final time t_1, I can say I start at time t_0 and wait a time interval δt. This is nothing more than defining the time interval δt trough

$$\delta t = t_1 - t_0, \tag{8}$$

There is nothing magic about combining the letters δ and t; δt is a number, just like t_1 and t_0. The δ part of the notaiton is meant to remind you that δt is small — the amount of time elapsed is short.

But then

$$t_1 = t_0 + \delta t,$$

and, substituting for t_1,

$$x(t_1) = x(t_0 + \delta t).$$

We can therefore rewrite (5) as

$$v = \lim_{t_1 \to t_0} \frac{x(t_0 + \delta t) - x(t_0)}{\delta t}. \tag{9}$$

However, the limit as $t_1 \to t_0$ is the limit of $t_1 - t_0 \to 0$, so $\delta t \to 0$. Therefore

$$v = \lim_{\delta t \to 0} \frac{x(t_0 + \delta t) - x(t_0)}{\delta t}. \tag{10}$$

You hopefully recognize this as the definition of the derivative of x with respect to t, at the point in time $t = t_0$. Mathematicians often do not like writing δt; instead they will use 'h'. Replacing δt by h throughout, we get

$$v = \lim_{h \to 0} \frac{x(t_0 + h) - x(t_0)}{h}.$$

Note 3 *You will probably know a derivative as the 'slope of a graph'. This comes from the slope of a straight line, which we define as 'rise over run'. If x against t were a straight line, then 'rise' would be change in x and 'run' would be change in t, given by $x(t_1) - x(t_0)$ and $t_1 - t_0$, respectively. This suggests that slope is again the ratio*

$$\frac{x(t_1) - x(t_0)}{t_1 - t_0}.$$

When x plotted against t is not a straight line, then this is the slope of the straight line connecting the point $(t_0, x(t_0))$ to the point $(t_1, x(t_1))$. This straight line intersects the actual curve $x(t)$ in two points. We can move the two points closer together by letting t_1 approach t_0. As we do this, the angle of the straight line changes, and the line more and more appears to just 'touch' the curve. In the limit of $t_1 \to t_0$, we say that the straight line becomes a tangent to the curve $x(t)$; the slope of that tangent is then the limit of $[x(t_1) - x(t_0)]/(t_1 - t_0)$, or in other words, the derivative.

Of course v is itself a function of time; the definition (10) gives v evaluated at a point in time t_0, which we took to be fixed as δt is made smaller, but which could actually be any point in time: If we make measurements of $[x(t_0 + \delta t) - x(t_0)]/\delta t$ at different times t_0, always using very small time intervals δt over which the displacement $x(t_0 + \delta t) - x(t_0)$ is measured, we will get different answers depending on the starting time t_0 for each measurement. So v depends on t_0.

Generally, if we simply relabel t_0 by t (these are just *labels*, there is nothing special about using the symbol t_0 to denote an instant in time instead of using t), we have

$$v(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{\delta t \to 0} \frac{x(t + \delta t) - x(t)}{\delta t}.$$

The notation $\mathrm{d}x/\mathrm{d}t$ is often used for a derivative. This is because the derivative is the limit of a change in x (i.e. the displacement $x(t_0+\delta t) - x(t_0)$ over the corresponding change in t (i.e., δt). That notation becomes clearer if we define δx in the same way we defined δt above, by writing

$$\delta x = x(t + \delta t) - x(t).$$

In that case, (5) becomes

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{\delta t \to 0} \frac{\delta x}{\delta t}.$$

Note 4 *It is worth remembering that the notation* $\mathrm{d}x/\mathrm{d}t$ *is meant to* remind *you of the ratio* $\delta x/\delta t$ *but is not quite the same; there is no finite* $\mathrm{d}x$ *being divided by a finite* $\mathrm{d}t$ *here,* $\mathrm{d}x/\mathrm{d}t$ *refers to the limit if the ratio* $\delta x/\delta t$*, in which there is neither a finite* δt *nor a finite* δx*. If you were to try to define* $\mathrm{d}t = \lim_{\delta t \to 0} \delta t$*, you would simply get* $\mathrm{d}t = 0$*, and similarly* $\mathrm{d}x = 0$*; the ratio of these would be* $0/0$*, which is not defined. So you have to take the limit of the ratio, not take the ratio of the limits of* δx *and* δt*.*

In this course, the formalism of taking limits is not the main point. What you will need to develop is understanding that, for very small δt, ratio

$$\frac{\delta x}{\delta t}$$

becomes a better and better approximation of the actual derivative

$$v(t) = \frac{\mathrm{d}x}{\mathrm{d}t},$$

when δt is made smaller and smaller, and that therefore the displacement travelled δx can be approximated as being proportional to δt, with the constant of proportionality being the instantaneous velocity:

$$\delta x \approx v(t)\delta t. \tag{11}$$

This is what we mean by 'rate of change of x with respect to t': the small change δx that happens in time δt is approximately proportional to δt, with the constant of proportionality being the 'rate of change'.

Note 5 *Rephrasing the contents of note 2, what we mean by saying that* $\delta x/\delta t$ *becomes a better and better approximation to the actual derivative* $v(t)$ *as* δt *is made smaller and smaller is that the difference between the approximation* $\delta x/\delta t$ *and the actual derivative* $v = \mathrm{d}x/\mathrm{d}t$ *goes to zero as* δt *goes to zero:*

$$\lim_{\delta t \to 0} \left(\frac{\delta x}{\delta t} - v(t) \right) = \lim_{\delta t \to 0} \left(\frac{x(t+\delta t) - x(t)}{\delta t} - v(t) \right) = 0.$$

This is straightforward to see from the definition of the derivative

$$v(t) = \lim_{\delta t \to 0} \frac{x(t+\delta t) - x(t)}{\delta t}$$

if we rearrange to

$$\lim_{\delta t\to 0}\left(\frac{x(t+\delta t)-x(t)}{\delta t}\right)-v(t)=0$$

and recognize that $v(t)$ does not depend on δt (v is simply a function of time t, but not of the measurement time interval δt). This means $v(t)=\lim_{\delta t\to 0} v(t)$. Then the fact that the sum of limits of two functions is the limit of the sum of the functions gives us

$$\lim_{\delta t\to 0}\left(\frac{x(t+\delta t)-x(t)}{\delta t}\right)-\lim_{\delta t\to 0} v(t)=\lim_{\delta t\to 0}\left(\frac{x(t+\delta t)-x(t)}{\delta t}-v(t)\right)=0$$

as expected.

Exercise 1 *Let $f(x)=\exp(x)$.*

1. *Plot $f(x)$ against x (use your favourite software package)*
2. *What is $\mathrm{d}f/\mathrm{d}x$ at $x=1$?*
3. *Calculate $\delta f/\delta x$ for $x=1$ and for (a) $\delta x=1$, (b) $\delta x=0.5$, (c) $\delta x=0.1$, (d) $\delta x=0.05$, (e) $\delta x=0.01$, (h) $\delta x=0.005$, (g) $\delta x=10^{-7}$. Below which value of δx does $\delta f/\delta x$ reproduce $\mathrm{d}f/\mathrm{d}x$ to two significant figures?*
4. *Plot the straight lines connecting the points $(x,f(x))$ and $(x+\delta x, f(x)+\delta f)$ on the graph.*
5. *Plot the tangent to the graph at $x=1$.*

Integration

Imagine now that you know the velocity $v(t)$ changes over time, for instance from some measurement that does not directly tell you position. Such a measurement could be the rotation rate of an axle in a vehicle or similar. How can you use this information to figure out where the object was at a particular point in time?

If velocity were a constant, you would simply invert the defintion of velocity (1) to get (2), or in words, compute displacement as velocity times time elapsed. This obviously does not work when v changes over time. However, we can still compute small displacements δxx that occur over short periods of time δt using (11),

$$\delta x = v(t)\delta t. \tag{12}$$

This works because, over a short time period δt we can treat the velocity $v(t)$ as approximately constant.

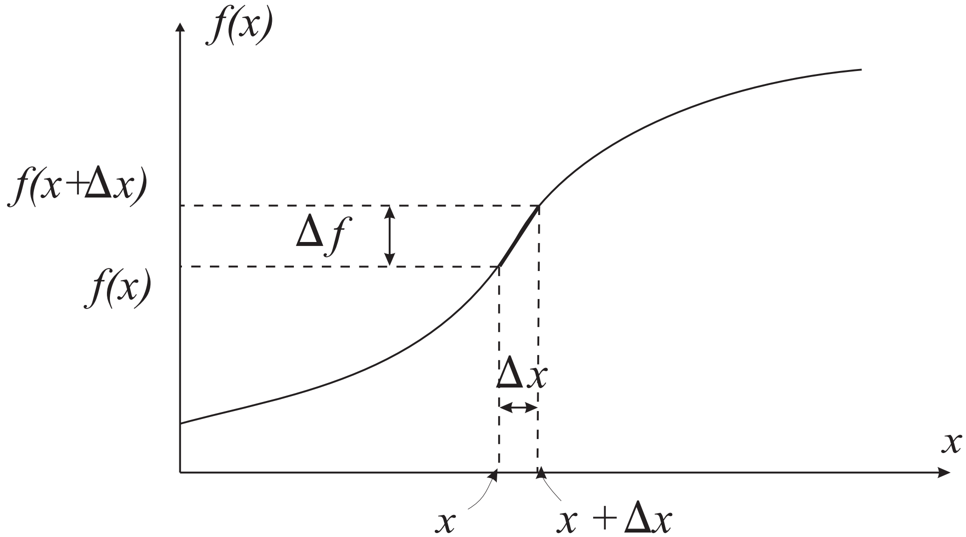

Figure 1: The derivative of a function is the slope of the graph $f(x)$.

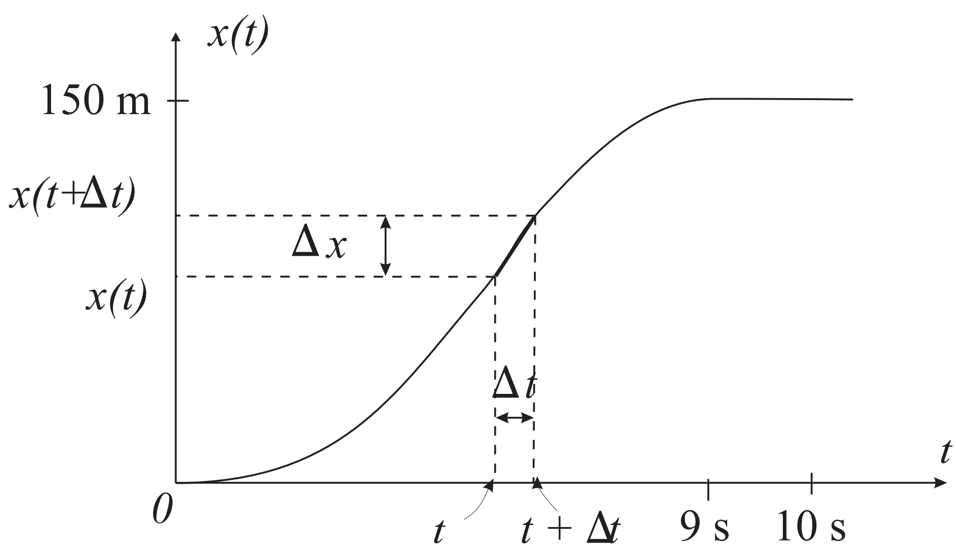

Figure 2: Position $x(t)$ of a travelling object.

To calculate the displacement that occurs over a longer time interval, say from an initial time t_{in} to a final time t_f, we do not try to compute the whole displacement at once. Instead, we split the long time interval into a lot of short ones, for each of which we will use (12). This basically means picking many intermediate points in time between t_{in} and t_f, which we will label t_1, t_2, t_3, $\ldots$, t_{N-1}, so we have picked N points, where N is a large number. We assume these points in time are arranged in order, so that t_2 is greater than t_1, t_3 is greater than t_2 etc. In short,

$$t_{in} < t_1 < t_2 < t_3 < \ldots < t_{N-1} < t_f.$$

These points in time split the interval from t_{in} to t_f into N smaller intervals, with lengths that we can write as

$$\delta t_1 = t_1 - t_{in}, \qquad \delta t_2 = t_2 - t_1, \qquad \text{etc,}$$

or more succinctly,

$$\delta t_n = t_n - t_{n-1}$$

if we write $t_0 = t_{in}$ and $t_N = t_f$.

If we space the intermediate points out fairly evenly, then, for a large number N of these points, all the time intervals δt_n between them will be small, and we can write

$$\delta x_n \approx v(t_n)\delta t_n$$

where δx_n is the displacement travelled in the nth interval. This must of course be the difference between position at times t_{n-1} and t_n,

$$\delta x_n = x(t_n) - x(t_{n-1}).$$

We expect that the total displacement between the first and last point (between t_{in} and t_f) is given by summing the displacements travelled in all the short time intervals, or

$$x(t_f) - x(t_{in}) = \sum_{n=1}^{N} \delta x_n.$$

It is actually easy to show that this is the case:

$$\begin{aligned}\sum_{n=1}^{N} \delta x_n &= \sum_{n=1}^{N} [x(t_n) - x(t_{n-1}]\\ &= [x(t_1) - x(t_0)] + [x(t_2) - x(t_1)] + [x(t_3) - x(t_2)] + \ldots + [x(t_N) - x(t_{N-1})]\end{aligned}$$

In the sum, all terms except $x(t_0)$ and $x(t_N)$ appear exactly twice, once with a plus sign and once with a minus sign. They therefore cancel, leaving only $x(t_0)$ and $x(t_N)$, and

$$\sum_{n=1}^{N} \delta x_n = x(t_N) - x(t_0) = x(t_f) - x(t_{in})$$

because $t_0 = t_{in}$ and $t_N = t_f$.

But we have $\delta x_n \approx v(t_n)\delta t_n$, so that

$$x(t_f) - x(t_{in}) \approx \sum_{n=1}^{N} v(t_n)\delta t_n. \tag{13}$$

This becomes a better and better approximation if all the time intervals δt_n are made smaller and smaller, which of course simultaneously means the number of intervals N must be made larger. Of course, the sum on the right-hand side of (13) is a *Riemann sum*, and in the limit of small δ_n's, large N, it becomes a definite intergral,

$$x(t_f) - x(t_{in}) = \int_{t_{in}}^{t_f} v(t)\,\mathrm{d}t = \lim_{N\to\infty,\,\delta_n\to 0} \sum_{n=1}^{N} v(t_n)\delta t_n. \tag{14}$$

Note 6 *Above, we replaced the approximation sign in* (13) *with an equality in* (14). *In a more formal course, this would require a mathematical course that involves the definition of a limit etc. This is where we depart from formality: the purpose of this course is not to provide you with mathematical proofs, but to give you a practical understanding of how the mathematical techniques presented apply to physics problems.*

Equation (14) invovles the definite integral over $v(t)$, which however was the derivative of $x(t)$ to begin with. In other words, we have one version of the *fundamental theorem of calculus*, stating that the integral over the derivative of x gives the difference between x at the upper and lower limits of integration,

$$x(t_f) - x(t_{in}) = \int_{t_{in}}^{t_f} \mathrm{d}x\,\mathrm{d}t\,\mathrm{d}t.$$

There is a second version of the theorem. Suppose we again put $v(t) = \mathrm{d}x/\,\mathrm{d}t$, so that

$$x(t_f) - x(t_{in}) = \int_{t_{in}}^{t_f} v(t)\,\mathrm{d}t.$$

Now differentiate with respect to t_f while holding t_{in} constant:

$$\frac{\mathrm{d}}{\mathrm{d}t_f} x(t_f) = \frac{\mathrm{d}}{\mathrm{d}t_f} \int_{t_{in}}^{t_f} v(t)\,\mathrm{d}t. \tag{15}$$

The $x(t_{in})$ term goes away under differentiation.

The notation on the left looks a bit awkward: we would much rather be differentiating with respect to t than t_f. The reason why we cannot is that we are already using 't' to denote the integration variable on the right, and t must vary from t_{in} to

t_f. As a result, we cannot simultaneously use t to be the limit of integration.[1] The derivative on the left is in fact $v(t_f)$; this becomes clear if we write

$$\frac{\mathrm{d}}{\mathrm{d}t_f}x(t_f) = \lim_{\delta t\to 0}\frac{x(t_f+\delta t)-x(t_f)}{\delta t}.$$

This is however nothing more than the definition of v in (9) with t_f replacing t_0. Therefore (15) really states that

$$v(t_f) = \frac{\mathrm{d}}{\mathrm{d}t_f}\int_{t_{in}}^{t_f} v(t)\,\mathrm{d}t,$$

or in words, that differentiating the with respect to the upper limit gives the integrand, evaluated at the upper limit.

Note 7 *An integral is also the 'area under a curve'. More specifically, the integral*

$$\int_a^b f(x)\,\mathrm{d}x$$

is the area of the region bordered by the x-axis, the line $x = a$, the line $x = b$ and the curve $y = f(x)$. Any part of that region lying below the x-axis is considered to have a negative area.

To see that this is the case, split the interval from a to b in the same way as above by inserting points x_1, x_2, x_3, $\ldots x_{N-1}$, with

$$a < x_1 < x_2 < \ldots < x_{N-1} < b,$$

The area bordered by the curve $y = f(x)$, the x-axis, and the vertical lines $x = x_{n-1}$ and $x = x_n$ is a thin strip that can be approximated as a rectangle of width $\delta x_n = x_n - x_{n-1}$ and height $y = f(x_n)$. The size of that area is therefore given by base times height, or

$$\delta A_n = f(x_n)\delta x_n.$$

To get the total area, we need to sum all the δA's, which gives

$$\sum_{n=1}^{N}\delta A = \sum_{n=1}^{N} f(x_n)\delta x_n.$$

[1] t is in fact a dummy variable and could be replaced by another symbol that has not yet been used; it would for instance be legitimate to write

$$\int_{t_{in}}^{t_f} v(t)\,\mathrm{d}t = \int_{t_{in}}^{t_f} v(t')\,\mathrm{d}t'$$

or even

$$\int_{t_{in}}^{t_f} v(t)\,\mathrm{d}t = \int_{t_{in}}^{t_f} v(u)\,\mathrm{d}u$$

.

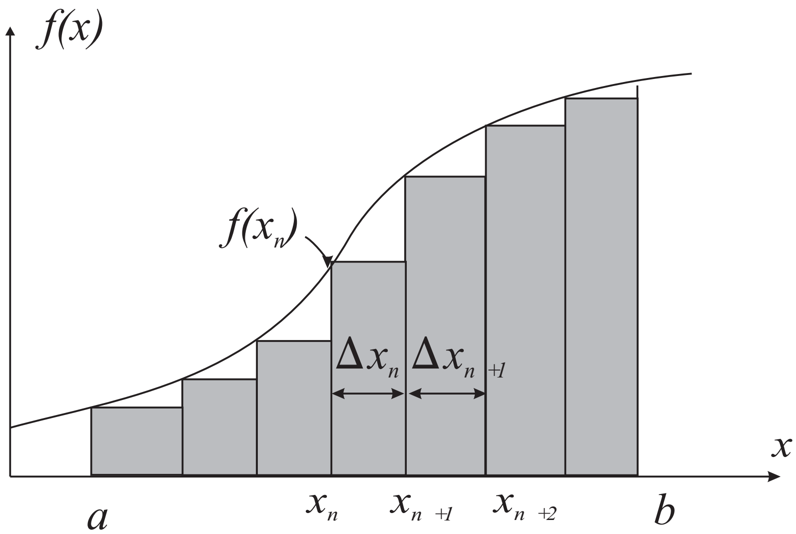

Figure 3: The integral of $f(x)$ is the area under its graph, which is approximately the sum of the area of the grey rectangles. If we increase the number of rectangles, the approximation becomes better.

Turning this Riemann sum into an integral, it is clear that x goes from a to b, so the area becomes

$$\int_a^b f(x)\,\mathrm{d}x.$$

The area computed for the rectangles is obviously negative if $f(x_n) < 0$. If regions below the x-axis are not to be treated as having negative area, the integral that needs to be taken is

$$\int_a^b |f(x)|\,\mathrm{d}x.$$