So we can justifiably claim that linear approximation in this sense is what it means for $f(x)$ to have a derivative at a. This approach is quite important is several contexts. (Examples of this are (i) proof of the chain rule; (ii) justification of Newton's method for finding roots of $f(x) = 0$ accurately; (iii) use of linear approximation to replace difficult nonlinear problems by easier linear ones which we then hope are a good approximation to the original; and (iv) inspiration for what a derivative should be for functions of more than one variable. We will only see some of these in this module.)

Quite often the difference

$$f(x) - \left(f(a) + f'(a)(x - a) \right)$$

is roughly proportional to $(x - a)^2$ for small $x - a$, but this is not guaranteed to be so. (A theorem called Taylor's theorem can be used to justify this in many examples. We will not go into that here.)

Notice that if $x - a$ is very small (like 10^{-2} or 10^{-6} perhaps, then $(x - a)^2$ is a fraction of that again (so like 10^{-4} or 10^{-12}).

3.4   Rate of change interpretation

For $y = f(x)$ a function, in this interpretation we think of x changing and y changing accordingly.

Perhaps the easiest way to do this is to think of x are representing time (measure in seconds, say, from some starting time). Negative times then come from times before the starting time. To make this easier to think about, let's replace the variable x by t (for time) and look at $y = f(t)$. We can think of y as representing the position of a 'particle' moving along an axis. So we can use the number y to describe locations on the axis. The idea of a 'particle' is that it is a point like object, with no extent and so no doubt about exactly where it is on the axis.

Now, let us ask for the speed (or 'velocity' is a better word) of the particle. One way to work out the speed at time $t = a$ would be to wait for 1 unit of time to elapse until time $t = a + 1$. In that time, the particle will have moved from position $y = f(a)$ to position $y = f(a + 1)$. So a distance travelled (or displacement) of $f(a + 1) - f(a)$ in 1 unit if time.

Now, in fact, $f(a + 1) - f(a)$ could be positive or negative. It will be positive if $f(a + 1) < f(a)$, so that the movement was in the positive direction along the axis.

Now the speed, or really velocity since it has a sign giving a direction (as well as a magnitude), of the particle is

$$\frac{\text{displacement}}{\text{time}} = \frac{f(a + 1) - f(a)}{1}$$

If you think a bit more, you realise that this is the average velocity over the 1 second of time (or 1 unit of time if we measured time in minutes or hours). So that is ok if we know the velocity is constant. If we are not sure that the velocity is constant, we can take a

shorter time period h from $t = a$ to $t = a + h$ and then the displacement is $f(a+h) - f(a)$ in time h, So that gives

$$\text{average velocity over time } a \text{ to } a + h \ = \ \frac{\text{displacement}}{\text{time}} = \frac{f(a+h) - f(a)}{h}$$

The next idea is that if the velocity is changing, then we need a very short time h so as to get the average to be close to the instantaneous velocity at time $t = a$. The smaller we make h the better and this leads to the idea of taking

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

as the **definition** of the *instantaneous velocity* at time a.

You can also say it is the *instantaneous rate of change* of y with respect to t at $t = a$.

It is a little harder to think about but if x is something like temperature and $y = f(x)$ volume at temperature x, then $f'(a)$ will be the rate of change of volume with respect to temperature when the temperature is a. In another example, x could be the bank interest rate and $y = f(x)$ profit (for a real estate company say). Then $f'(a)$ gives the rate of change of profit with interest rate.

From linear approximation, it also means that a small change of h in the interest rate produces a change of $f(a + h) - f(a) \cong f'(a)h$ in profit.

4 Rules for differentiation

4.1 The derivative as a new function

If a function $f(x)$ has a derivative $f'(a)$ at each point a of its domain, then we have a new function whose value at x is $f'(x)$.

The so-called Leibniz notation writes

$$\frac{dy}{dx}$$

for the derivative of $y = f(x)$. That is instead of $f'(x)$, where it is clear that $f'(x)$ depends on x.

4.2 Basic formulae

(i) (constants)

For $f(x) = c = \text{constant}$, $\dfrac{dy}{dx} = f'(x) = 0$.

(ii) ($f(x) = x$)

For $f(x) = x$, $\dfrac{dy}{dx} = f'(x) = 1$.

(iii) (Derivative of a sum)

If $u = u(x)$ and $v = v(x)$ are differentiable and $y = f(x) = u(x) + v(x)$, then

$$\frac{dy}{dx} = \frac{d}{dx}(u(x) + v(x)) = u'(x) + v'(x) = \frac{du}{dx} + \frac{dv}{dx}.$$

(iv) (Derivative of a constant multiple)

If $u = u(x)$ is differentiable and $y = f(x) = ku(x)$ for a constant k, then

$$\frac{dy}{dx} = \frac{d}{dx}(ku(x)) = ku'(x) = k\frac{du}{dx}.$$

(v) (Derivative of a product)

If $u = u(x)$ and $v = v(x)$ are differentiable and $y = f(x) = u(x)v(x)$, then

$$\frac{dy}{dx} = \frac{d}{dx}(u(x)v(x)) = u'(x)v(x) + u(x)v'(x) = \frac{du}{dx}v + u\frac{dv}{dx}.$$

(vi) (positive integer powers)

For $f(x) = x^n$ (with $n \in \mathbb{N}$), $\dfrac{dy}{dx} = f'(x) = nx^{n-1}$.

(vii) (Derivative of a quotient)

If $u = u(x)$ and $v = v(x)$ are differentiable and $y = f(x) = u(x)v(x)$, then

$$\frac{dy}{dx} = \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{u'(x)v(x) - u(x)v'(x)}{(v(x))^2} = \frac{\dfrac{du}{dx}v - u\dfrac{dv}{dx}}{v^2}.$$

(viii) (negative integer powers)

For $f(x) = x^n$ (with $n \in \mathbb{Z}$, $n < 0$), $\dfrac{dy}{dx} = f'(x) = nx^{n-1}$ for $x \neq 0$.

(Note that $f(x) = x^{-|n|} = \dfrac{1}{x^{|n|}}$ does not make sense for $x \neq 0$.)

(ix) (Chain rule)

If $y = f(x)$ and $x = g(t)$, then we can consider y as depending on t, via $y = f(g(t)) = (f \circ g)(t)$. (Sometimes people call this "f after g"; it is formally called the composition of the two functions.)

The chain rule says that (provided both $f'(x)$ and $g'(t)$ make sense at all the values we need), then

$$(f \circ g)'(t) = f'(g(t))g'(t).$$

The usual way to remember it is via the equivalent statement in Leibniz notation:

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}.$$

In this version (where it says it is ok to "cancel the dx's"), the idea is that dy/dt is evaluated at t as is dx/dt, but dy/dx is evaluated at the x corresponding to t.

Proof. We won't prove these things, but they are laid out in a way that they could be proved one by one. The first two are really easy directly from the definition. The next 3 are not much harder. The result about x^n can be shown by induction using the product rule. The quotient rule is not hard to prove either and then the negative powers are an application. The chain rule is quite tricky to prove correctly. □

Examples 4.1. We have enough rules now to differentiate rational functions. No harm to practice. Try these for instance.

$$\frac{d}{dx}\frac{2x^6 - 5x^4 + x^3 - 1}{4x^3 - 2x^2 + 11x + 10}, \qquad \frac{d}{dx}(2x^3 - x^2 + x - 5)^{15}$$

Find the equation of the tangent line to the graph $y = (2x^3 - x^2 + x - 3)^5$ at the point (on the graph) where $x = 1$.