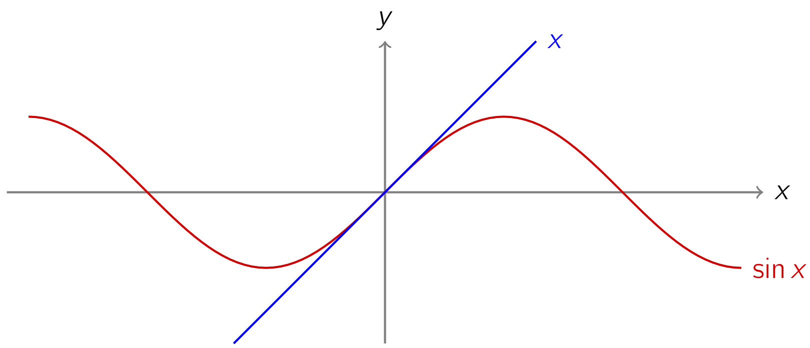

Figure 6: Plot of $\sin x$ and x on the same axes, notice that for small inputs x, they are very close

2.4 Asymptotic Notation

Let $f, g\colon A \to \mathbb{R}$ be functions, and let a be a cluster point of A. If

$$\lim_{x\to a} \frac{f(x)}{g(x)} = 1,$$

what could we deduce? If the ratio of two numbers is 1, then they are equal. Since the *limit* of the ratio is 1, it must means that for points close to a, the values of $f(x)$ and $g(x)$ are approximately the same. This is written as

$$f(x) \sim g(x) \qquad \text{as } x \to a,$$

and we say that $f(x)$ is *asymptotic to* $g(x)$ *as* $x \to a$. For instance, (i) of theorem 2.18 tells us that

$$\sin x \sim x$$

as $x \to 0$. This must mean that for points close to 0, $\sin x$ and x are approximately equal. Indeed, if we look at a sketch of their graphs (figure 6), we see that this is the case. In fact, $\sin(0.01) = 0.009999$ (for example). Another example, we have that

$$\sqrt{x} \sim \frac{\sqrt{2}}{32}(12 + 12x - x^2) \qquad \text{as } x \to 2.$$

Indeed, since both functions are continuous at 2, we can just plug in $x = 2$ directly to get that

$$\lim_{x\to 2} \frac{\sqrt{x}}{\frac{\sqrt{2}}{32}(12 + 12x - x^2)} = \frac{\sqrt{2}}{\frac{\sqrt{2}}{32}(12 + 12(2) - 2^2)} = 1,$$

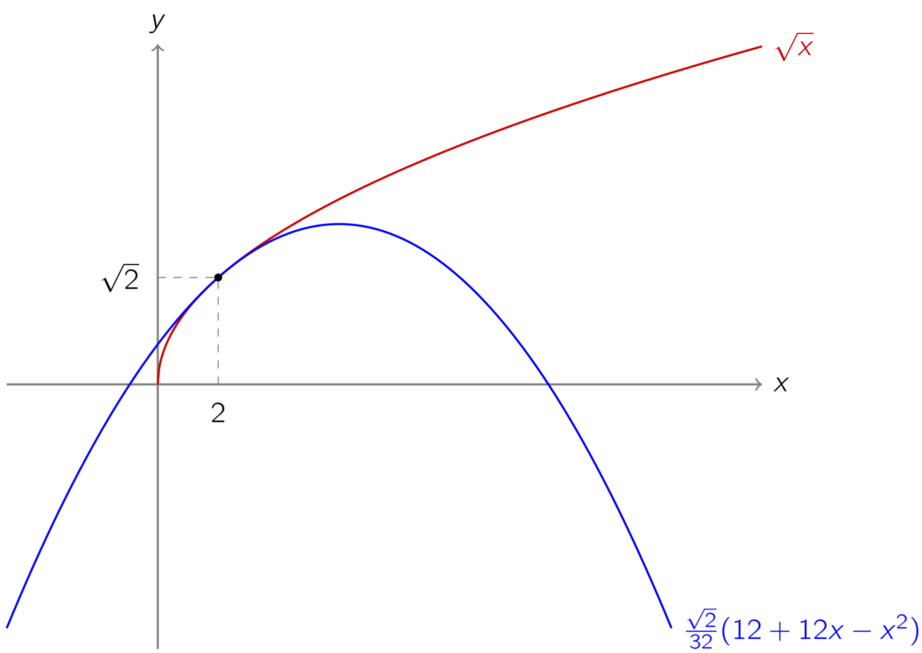

Figure 7: Plot of $\sqrt{x}$ and $\frac{\sqrt{2}}{32}(12+12x-x^2)$ on the same axes, notice that for x close to 2, they are good approximations of each other

and we can see in figure 7 that they good approximations to each other for inputs close to $x = 2$.

What if, on the other hand, we have that

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 0?$$

In this case, for points close to a, the values of $f(x)$ must be much smaller (in size) than those of $g(x)$; since the ratio of two numbers is approximately zero if the numerator is much smaller than the denominator.

Definition 2.20 (Little-o notation)**.** Let $f, g \colon A \to \mathbb{R}$ be functions, and let a be a cluster point of A. we say that f is *little-oh of g as $x \to a$*, or that f *is dominated asymptotically by g as $x \to a$*, written

$$f(x) = o(g(x)) \qquad \text{as } x \to a,$$

if $\lim_{x \to a} f(x)/g(x) = 0$.

For instance, theorem 2.18(ii) tells us that

$$1 - \cos x = o(x)$$

as $x \to 0$. In other words, $1 - \cos x$ is smaller than x as we approach 0. Even though they are both 0 when $x = 0$, for points *close to* 0, the value of $1 - \cos x$ is smaller than that of x, which means that it goes to 0 more rapidly. Indeed, if we tabulate different values as x approaches 0, we see this is the case.

x	1	0.1	0.001	0.0001	0.0001
$1 - \cos x$	0.4597	0.0049	0.000049	0.00000049	0.00000049

So what this means in terms of approximations, is that if we have some expression involving both x and $1 - \cos x$, such as

$$7 + 3x - 2(1 - \cos x),$$

we can say that this is approximately $7 + 3x$ if x is small enough (i.e., close enough to 0). Indeed, when $x = 0.1$,

$$7 + 3x = 7.3 \qquad \text{and} \qquad 7 + 3x - 2(1 - \cos x) = 7.290008.$$

Another example of the notation: let $n > 1$. Then

$$x^n = o(x)$$

as $x \to 0$, since

$$\lim_{x\to 0} \frac{x^n}{x} = \lim_{x\to 0} x^{n-1} = 0^{n-1} = 0,$$

where we used the fact that x^{n-1} is continuous at 0. In other words, x dominates x^n when the input x is close to 0 ($x \to 0$), so something like

$$1 + 3x - 2x^2 + 5x^3$$

is well approximated by $1 + 3x$ when x is small.

If $f(x) - g(x) = o(h(x))$, we also write that $f(x) = g(x) + o(h(x))$. For instance, we said that $1+3x-2x^2+5x^3$ is well approximated by $1+3x$ when x is small; this is because their difference is $o(x)$, i.e.,

$$(1 + 3x - 2x^2 + 5x^3) - (1 + 3x) = o(x) \qquad \text{as } x \to 0.$$

Instead, we can write this as

$$1 + 3x - 2x^2 + 5x^3 = 1 + 3x + o(x) \qquad \text{as } x \to 0,$$

which means that the expression on the left is equal to $1 + 3x$ plus something which is not as significant as x when we are close to 0.

Remark 2.21. Even though we are using = here, this is a bit of an abuse of notation. Indeed, if $f(x) = o(g(x))$ and $h(x) = o(g(x))$, it doesn't mean that $f(x) = h(x)$, which is not usual behaviour of equality. ■

3 Differentiation

3.1 Calculus of Differences

Definition 3.1 (Difference)**.** Let $f : A \to \mathbb{R}$ be a function. The *difference* or *change of f at x by h*, denoted by $\Delta f(x, h)$, is the quantity defined by

$$\Delta f(x, h) = f(x + h) - f(x).$$

This is so that when we change the input x to f by h, we get

$$\underbrace{f(x+h)}_{\text{new value}} = \underbrace{f(x)}_{\text{old value}} + \underbrace{\Delta f(x,h)}_{\text{change}}.$$

We will often abuse notation slightly, writing truncated versions of this function such as $\Delta f(x)$ or just Δf when things are clear from context. We will also treat

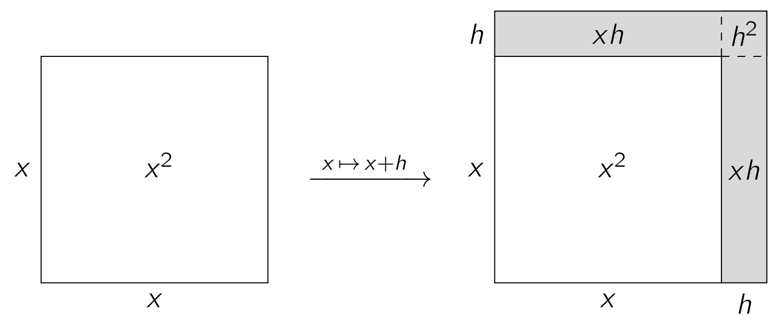

Figure 8: Interpreting example 3.2 as a change in area: $\Delta(x^2) = 2xh + h^2$

expressions in terms of x formally as functions. For instance, if $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2$, we would write $\Delta(x^2)$ for Δf. In other words, where we have otherwise been very careful about distinguishing between the notations f and $f(x)$, here we will not be so strict about it (otherwise some of the theorems we will do end up looking needlessly more complicated).

Example 3.2. We compute the change of the function x^2. We have

$$\begin{aligned}\Delta(x^2) &= (x+h)^2 - x^2\\ &= x^2 + 2xh + h^2 - x^2\\ &= 2xh + h^2.\end{aligned}$$

Thus when $x = 2$ (for instance), we have $2^2 = 4$, and if we change the input by $h = 0.1$ to get 2.1^2, we just need to add

$$\Delta(x^2)(2, 0.1) = 2(2)(0.1) + 0.1^2 = 0.41$$

to the value of 2^2, which gives us that $2.1^2 = 4.41$.

We can interpret what we've computed here as the change of area when we extend the sides of a square, as shown in figure 8. ■

Example 3.3. Another example, we find $\Delta(x^3 - 2x + 5)$.

$$\begin{aligned}\Delta(x^3 - 2x + 5) &= (x+h)^3 - 2(x+h) + 5 - (x^3 - 2x + 5)\\ &= x^3 + 3x^2h + 3xh^2 + h^3 - 2x - 2h + 5 - x^3 + 2x - 5\\ &= (3x^2 - 2)h + 3xh^2 + h^3\end{aligned}$$

Notice that when $x = 5$, the $f(x) = 5^3 - 2\cdot 5 + 5 = 120$ (where $f = x^3 - 2x + 5$). If we want to find the value of $f(105)$, we can finding the change with $x = 5$

and $h = 100$, and then just add that to 120. Indeed, the change is

$$\begin{aligned}\Delta f(5, 100) &= (3 \cdot 5^2 - 2) \cdot 100 + 3(5) \cdot 100^2 + 100^3\\ &= 7\,300 + 150\,000 + 1\,000\,000\\ &= 1\,157\,300,\end{aligned}$$

so the function at $x = 105$ equals $1\,157\,420$. (Obviously we could have just plugged in 150 into $f(x)$ directly, but then we wouldn't be using Δf, and this wouldn't be much of an example.)

Perhaps an example which illustrates a bit better why it is worthwhile to study changes: notice that when h is small, we can approximate the change just by taking the first term $(3x^2 - 2)h$, since terms in higher powers of h are less significant (they are $o(h)$). So for instance, to approximate the value of the function at 105.1, we just work out the change with $x = 105$ and $h = 0.1$:

$$\Delta f(105, 0.1) \approx 33\,073 \cdot 0.1 = 3\,307.3.$$

We found that the function at 105 equals $1\,157\,420$, so adding the approximate change above, we get that the function at $105.1 \approx 1\,160\,727.3$. (The actual value of $f(105.1)$ is about $1\,160\,730.45$, so this only introduces a relative error of 0.0002%). ■

Even though most of the results here do not make any restrictions on the size of h, it will be instructive to think of h as "small", since when we get to differentials (which are the main object of differential calculus), we will be thinking about the case where h is small, just as we saw in the last example. With this in mind, we have the following result.

Proposition 3.4 (Change of x^n)**.** *Let $n \in \mathbb{N}$, and let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^n$. Then we have*

$$\Delta f(x, h) = n\,x^{n-1}h + o(h)$$

as $h \to 0$.

Proof. This follows immediately by the binomial theorem. Indeed, we have

$$\begin{aligned}\Delta f(x, h) &= (x + h)^n - x^n\\ &= \sum_{k=0}^{n} \binom{n}{k} x^{n-k} h^n - x^n\end{aligned}$$

$$
\begin{aligned}
&= x^n + n\,x^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \cdots + h^n - x^n \\
&= n\,x^{n-1}h + \Big(\frac{n(n-1)}{2}x^{n-2}h + \cdots + h^{n-1}\Big)h \\
&= n\,x^{n-1}h + o(h),
\end{aligned}
$$

since $\frac{n(n-1)}{2}x^{n-2}h + \cdots + h^{n-1}$ is a polynomial in h (so it is continuous in h), and it equals 0 when $h = 0$. □

Example 3.5. We have $50^3 = 125\,000$. By the above, we have

$$\Delta(x^3)(50, h) = 3 \cdot 50^2 h = 7\,500h + o(h),$$

as $h \to 0$, and so

$$50.05^3 = 50^3 + \Delta(x^3)(50, 0.05) \approx 125\,000 + 7\,500(0.05) = 125\,375.$$

The precise value is 125 375.375 125, so the error is 0.00029%. ■

Remark 3.6 (Arithmetic of functions). Let $f, g \colon A \to \mathbb{R}$ be functions, and let $\lambda \in \mathbb{R}$ be a constant. Then when we write $f + g$ or λf, we refer the functions defined in the obvious way, i.e.,

$$(f + g)(x) = f(x) + g(x) \qquad \text{and} \qquad (\lambda f)(x) = \lambda\, f(x)$$

for all $x \in A$. For instance, $3\sin^2 + 2\log$ is the function such that

$$(3\sin^2 + 2\log)(x) = 3\sin^2 x + 2\log x.$$

We similarly infer the meaning of expressions such as fg, f/g, and so on; e.g.

$$\Big(\frac{\sin\cos + \sqrt{\cdot}}{\log^3}\Big)(x) = \frac{\sin x \cos x + \sqrt{x}}{\log^3 x}.$$

Notice particularly that the juxtaposition $\sin\cos$ became the product $\sin x \cos x$, and not the composition $\sin(\cos x)$ (for which we would instead write $\sin\circ\cos$). This convention will allow us to state properties about Δ in a concise way. ■

Proposition 3.7 (Linearity of Δ)**.** *Let $f, g \colon A \to \mathbb{R}$ be two real-valued functions, and let a, b be two constants. Then*

$$\Delta(af + bg) = a\,\Delta f + b\,\Delta g$$

Proof. We have

$$\begin{aligned}\Delta(af+bg) &= (af+bg)(x+h)-(af+bg)(x)\\ &= a\,f(x+h)+b\,g(x+h)-a\,f(x)-b\,g(x)\\ &= a(f(x+h)-f(x))+b(g(x+h)-g(x))\\ &= a\,\Delta f+b\,\Delta g.\end{aligned}$$ □

Proposition 3.8 (Product Rule for Δ)**.** *Let $f,g\colon A\to\mathbb{R}$ be two real-valued functions. Then*

$$\Delta(fg)=\Delta f\,g+f\,\Delta g+\Delta f\,\Delta g$$

Proof. This is another straightforward proof:

$$\begin{aligned}\Delta(fg) &= (fg)(x+h)-(fg)(x)\\ &= f(x+h)\,g(x+h)-f(x)\,g(x)\\ &= f(x+h)\,g(x+h)-f(x)\,g(x+h)\\ &\qquad\qquad+f(x)\,g(x+h)-f(x)\,g(x)\\ &= [f(x+h)-f(x)]g(x+h)+f(x)[g(x+h)-g(x)]\\ &= \Delta f\,g(x+h)+f\,\Delta g\\ &= \Delta f\cdot(g(x)+\Delta g(x,h))+f\,\Delta g\\ &= \Delta f\,g+f\,\Delta g+\Delta f\,\Delta g.\end{aligned}$$ □

Proposition 3.9 (Chain Rule for Δ)**.**

$$\Delta(f\circ g)(x,h)=\Delta f(g(x),\Delta g(x,h)).$$

Proof. Just by expanding the definition, we have

$$\begin{aligned}\Delta(f\circ g)(x,h) &= (f\circ g)(x+h)-(f\circ g)(x)\\ &= f(g(x+h))-f(g(x))\\ &= f(g(x)+\Delta g(x,h))-f(g(x))\\ &= \Delta f(g(x),\Delta g(x,h)).\end{aligned}$$ □

Remark 3.10 (Δx). Notice that for any function $f\colon A\to\mathbb{R}$, we have $f=f\circ\mathsf{id}$, where id denotes the *identity function* defined by $\mathsf{id}(x)=x$ for all $x\in\mathbb{R}$. If we apply proposition 3.9 to this composition, we see that

$$\Delta f=\Delta(f\circ\mathsf{id})=\Delta f(\mathsf{id}(x),\Delta\mathsf{id}(x,h))=\Delta f(x,\Delta\mathsf{id}).$$

Now just as we informally write $\Delta(x^2)$ in place of Δf (when $f(x) = x^2$), here we can write $\Delta(x)$, or just Δx, for $\Delta\mathsf{id}$, since this is the expression defining $\mathsf{id}(x)$. Indeed,

$$\Delta x = \Delta(x)(x, h) = \mathsf{id}(x + h) - \mathsf{id}(x) = x + h - x = h,$$

so we have

$$\Delta f(x, h) = \Delta f(x, \Delta x).$$

Going forward, we will be writing Δx for the change of the input (instead of h which we have been using so far). We can just think of Δx as an independent variable just as we thought of h, but if we instead interpret it as a difference in the sense of definition 3.1, (i.e., we think of Δx as $\Delta(x)$), then by our reasoning above, everything ends up being the same. ■

3.2 Calculus of Differentials

We've already seen that it can be useful to take the principal part of a difference to approximate a change Δf when Δx is small, as we did in the second part of example 3.3. When the difference is essentially proportional to Δx, we say that f is *differentiable*.

Definition 3.11 (Differentiable)**.** Let $f : A \to \mathbb{R}$ be a function, and let $a \in A$. Then f is said to be *differentiable at $x = a$* if there exists a constant A (which may depend on a) such that

$$\Delta f(a, \Delta x) = A\,\Delta x + o(\Delta x)$$

as $\Delta x \to 0$. This constant is called the *derivative of f at $x = a$*, and we denote it by $f'(a)$.

If f is differentiable at every $a \in A$, we just say that f is differentiable.

Example 3.12. We saw in example 3.2 that when $f(x) = x^2$,

$$\Delta f(a, \Delta x) = 2a\,\Delta x + \Delta x^2 = 2a\,\Delta x + o(\Delta x),$$

so x^2 is differentiable at each a in its domain, and its derivative at $x = a$ is $f'(a) = 2a$. ■

A function being differentiable at a point captures the idea of being "smooth" there. Essentially, a function is differentiable at a if it can be approximated by a line there. Indeed, if f is differentiable at a, then

$$f(a + \Delta x) = f(a) + A\,\Delta x + o(\Delta x)$$

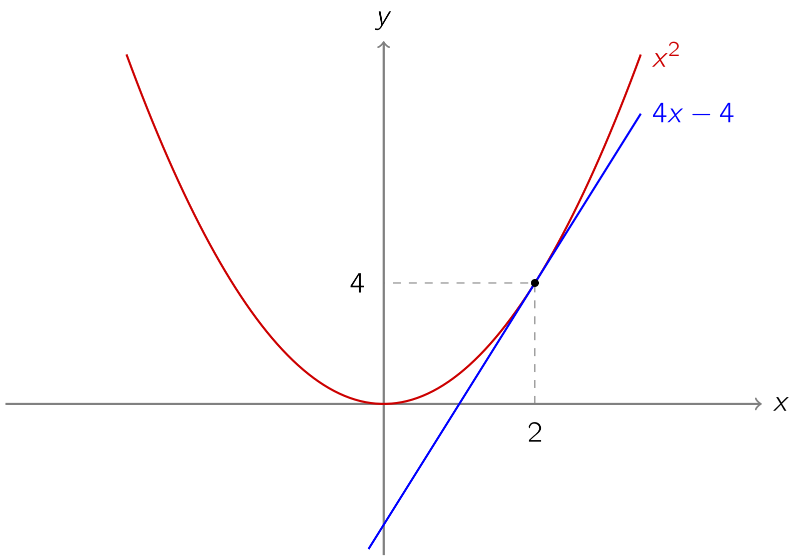

Figure 9: Plot of x^2 and $4x - 4$ on the same axes, notice that for points close to $x = 2$, they are very close.

and if we let $x = a + \Delta x$, this becomes

$$f(x) = f(a) + A \cdot (x - a) + o(x - a).$$

In other words, when x is close to a (or equivalently, when the difference $x - a = \Delta x$ is small), we have

$$f(x) \approx f(a) + A \cdot (x - a).$$

For instance, the derivative of x^2 at $x = 2$ is $A = 4$. Thus, for points close to 2, we have

$$x^2 \approx f(2) + A \cdot (x - 2) = 4 + 4(x - 2) = 4x - 4.$$

Indeed, if we plot these on the same axes, we can see that this gives us a good approximation for points close to $x = 2$ (figure 9). In general the line

$$y = f(a) + f'(a)\,(x - a)$$

is called the *tangent line* of f at $x = a$. Notice the derivative is precisely the gradient of this line.

Thus you should have the following intuitive understanding of what it means to be differentiable at the point $x = a$: if you keep zooming in to the function at $x = a$, looking really close, it should resemble a line. If it does, then the function is differentiable at that point, and the derivative of the function there is the gradient of this line.

An example of something which is not differentiable is $|x|$ at $x = 0$ (we will prove this formally later). But intuitively, if you keep zooming in towards the point where $x = 0$, it never looks like a line, it retains its V-shape.

Remark 3.13 (Uniqueness of Derivatives). If a function is differentiable at $x = a$, then its derivative $f'(a)$ is unique. In other words, we cannot find two different constants, A and B, such that

$$\Delta f = A\,\Delta x + o(\Delta x) \qquad \text{and} \qquad \Delta f = B\,\Delta x + o(\Delta x).$$

Indeed, if f is differentiable at $x = a$, then there exists A such that

$$\Delta f(a, \Delta x) = A\,\Delta x + o(\Delta x),$$

which by definition of little-o, means that

$$\Delta f(a, \Delta x) - A\,\Delta x = o(\Delta x),$$

i.e., that

$$\lim_{\Delta x \to 0} \left(\frac{\Delta f(a, \Delta x) - A\,\Delta x}{\Delta x} \right) = 0,$$

which is equivalent to saying that

$$\lim_{\Delta x \to 0} \left(\frac{\Delta f(a, \Delta x)}{\Delta x} - A \right) = 0.$$

Now clearly $\lim_{\Delta x \to 0} A = A$ since A is constant with respect to Δx, so applying the rule $\lim(f(x) + g(x)) = \lim f(x) + \lim g(x)$ (theorem 2.12(i)), we have that

$$\begin{aligned} \lim_{\Delta x \to 0} \frac{\Delta f(a, \Delta x)}{\Delta x} &= \lim_{\Delta x \to 0} \left(\left(\frac{\Delta f(a, \Delta x)}{\Delta x} - A \right) + A \right) \\ &= \left(\frac{\Delta f(a, \Delta x)}{\Delta x} - A \right) + \lim_{\Delta x \to 0} A \\ &= 0 + A = A. \end{aligned}$$

In other words, we have shown that if f is differentiable at $x = a$, then the derivative $A = f'(a)$ is equal to

$$\lim_{\Delta x \to 0} \frac{\Delta f(a, \Delta x)}{\Delta x},$$

and since limits are unique theorem 2.10, then this number is unique. ■

We can summarise the reasoning of remark 3.13 in following proposition:

Proposition 3.14. *Let $f\colon A \to \mathbb{R}$ be a function, and let $a \in A$. Then*

$$f \text{ is differentiable at } x = a \text{ with derivative } A \iff \lim_{\Delta x \to 0} \frac{\Delta f(x, \Delta x)}{\Delta x} \text{ exists and equals } A.$$

Proof. The direction $\Rightarrow$ follows from what we said in remark 3.13. To see why the other direction is true, we can basically reverse the steps we applied. Indeed, suppose that the limit

$$\lim_{\Delta x \to 0} \frac{\Delta f(a, \Delta x)}{\Delta x}$$

exists and equals A. Since $\lim_{\Delta x \to 0} A = A$, applying theorem 2.12(ii), we get that

$$\lim_{\Delta x \to 0} \left(\frac{\Delta f(a, \Delta x)}{\Delta x} - A \right) = \lim_{\Delta x \to 0} \frac{\Delta f(a, \Delta x)}{\Delta x} - \lim_{\Delta x \to 0} A = A - A = 0,$$

i.e.,

$$\lim_{\Delta x \to 0} \left(\frac{\Delta f(a, \Delta x) - A\,\Delta x}{\Delta x} \right) = 0,$$

i.e.,

$$\Delta f(a, \Delta x) - A\,\Delta x = o(\Delta x),$$

which rearranges to give $\Delta f(a, \Delta x) = A\,\Delta x + o(\Delta x)$, as required. □

The principal part of the difference Δf is called the *differential* of f.

Definition 3.15 (Differential)**.** Let $f\colon A \to \mathbb{R}$ be differentiable at $x = a$ with derivative $f'(a)$. The *differential of f at a* is the function defined by

$$df(a, h) = f'(a)\,h.$$

Example 3.16. Let $f(x) = x^2$. From example 3.12, we saw that $f'(a) = 2a$, so

$$df(a, h) = 2a\,h.$$

We have $\Delta f(a, h) = 2ah + h^2 = df + o(h)$. ■

In general, if f is differentiable at a, then we have that

$$\Delta f(a, h) = df(a, h) + o(h).$$

In particular, notice that the what makes Δf different from df is the "error term" which insignificant compared to h when h is small (i.e., it is $o(h)$ as $h \to 0$).

Just as we abused functional notation with Δ, here we do the same, writing things like $\Delta(x^2) = 2x\,h$. Moreover, just as in remark 3.10, we note that

$$\Delta(x)(a, h) = (a + h) - h = h = 1 \cdot h + 0 = 1 \cdot h + o(h),$$

so $\mathbf{id}$ is differentiable with derivative 1 for all a, and the differential

$$dx = d(x)(a, h) = h.$$

Consequently, we can either interpret dx as an independent variable (just as we were doing with h), or as the differential of the identity, it doesn't make any difference, and we will subsequently be writing

$$df(a, dx) \qquad \text{instead of} \qquad df(a, h).$$

Thus in summary, for differentiable f, we have that

$$\Delta f = df + o(dx).$$

Let's do an example.

Example 3.17. Let $f(x) = 3x^3 - 2x + 1$. Let us show that this is differentiable at every point x in its domain. Indeed,

$$\begin{aligned}\Delta f(x, dx) &= 3(x + dx)^3 - 2(x + dx) + 1 - (3x^3 - 2x + 1)\\ &= 3x^3 + 9x^2\,dx + 9x\,dx^2 + 3dx^3 - 2x - 2dx + 1 - 3x^3 + 2x - 1\\ &= (9x^2 - 2)\,dx + 9x\,dx^2 + 3\,dx^3,\end{aligned}$$

thus we have that f is differentiable with derivative $f'(x) = 9x^2 - 2$, and

$$\Delta f = \underbrace{(9x^2 - 2)\,dx}_{df} + \underbrace{9x\,dx^2 + 3\,dx^3}_{o(dx)},$$

so the differential df is $(9x^2 - 2)\,dx$. ■

Now we will translate some of the properties of Δ into properties of d.

Proposition 3.18 (Linearity of d)**.** *Suppose $f, g\colon A \to \mathbb{R}$, let $a, b \in \mathbb{R}$ and $x \in A$, and suppose that f and g are both differentiable at x. Then $af + bg$ is also differentiable at x, and*

$$d(af + bg) = a\,df + b\,dg.$$

Proof. Since f and g are differentiable at x, we have that $\Delta f = df + o(dx)$ and $\Delta g = dg + o(dx)$. By proposition 3.7,

$$\begin{aligned}\Delta(af + bg) &= a\Delta f + b\Delta g \\ &= a(df + o(dx)) + b(dg + o(dx)) \\ &= a\,df + b\,dg + o(dx) \\ &= (a\,f'(x) + b\,g'(x))\,dx + o(dx)\end{aligned}$$

so we agree with definition 3.11, and the differential is $a\,df + b\,dg$. □

Let us give the differential of an important class of functions, the powers of x.

Proposition 3.19. *Let $n \in \mathbb{N}$. Then $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^n$ is differentiable, and moreover,*

$$d(x^n) = n\,x^{n-1}\,dx.$$

This is precisely the statement of proposition 3.4. Combining this fact with proposition 3.18, we can find the differential of any polynomial.

Example 3.20. We have

$$\begin{aligned}d(4x^3 - 2x^2 + 5x - 9) &= 4d(x^3) - 2d(x^2) + 5d(x) - 9d(1) \\ &= 4(3x^2\,dx) - 2(2x\,dx) + 5\,dx - 9(0) \\ &= (12x^2 - 4x + 5)\,dx,\end{aligned}$$

where it is straightforward to check that $d1 = 0$. ■

Remark 3.21 (Leibniz Notation). Notice that in general,

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

In particular, the value $\frac{df}{dx}$ does not depend on the value of dx; it's just $f'(a)$. Consequently, the notation

$$\frac{df}{dx}(a)$$

is sometimes used as an alternative to $f'(a)$. In a similar spirit, the notation $\frac{d}{dx}$ denotes the "derivative operator", i.e.,

$$\frac{d}{dx}(f)(a) = \frac{df}{dx}(a),$$

so that we would write things like the previous example as

$$\frac{d}{dx}(4x^3 - 2x^2 + 5x - 9) = 12x^2 - 4x + 5,$$

where the derivative is the subject of the equation. ■

We will continue phrasing things in the notes in terms of differentials rather than using Leibniz notation, it will be advantageous to do so when it comes to integrals. (If you encounter dy/dx in the wild, you can just interpret it literally, where the dx's cancel out.)

It turns out that a more general version of proposition 3.19 is true.

Theorem 3.22 (Power Rule)**.** *Let $r \in \mathbb{R}$. Then x^r is differentiable, and*

$$d(x^r) = r\,x^{r-1}\,dx.$$

We will not give the proof here, but it is essentially a consequence of the fact that $x^r = \exp(r \log x)$.

Example 3.23. We determine the equation of the tangent line to the curve $y = f(x)$, where $f(x) = \frac{10\sqrt{x}-5}{x}$, at the point $x = 4$. Hence, we approximate the value of $\frac{1}{3}(10\sqrt{3} - 5)$.

We have

$$f = \frac{10\sqrt{x} - 5}{x} = 10x^{-1/2} - 5x^{-1} \implies df = (10(-\tfrac{1}{2})x^{-3/2} - 5(-1)x^{-2})\,dx$$
$$= \left(\frac{5}{x^2} - \frac{5}{x\sqrt{x}}\right) dx,$$

in particular, at $x = 4$, $df = -\frac{5}{16}\,dx$, so the derivative is $f'(4) = -\frac{5}{16}$, which gives the tangent line

$$\begin{aligned} y &= f(4) + f'(4)\,(x - 4) \\ &= \tfrac{15}{4} - \tfrac{5}{16}(x - 4) \\ \implies 5x + 16y &= 80. \end{aligned}$$

Using the tangent line, $\frac{1}{3}(10\sqrt{3} - 5) \approx y(3) = \frac{1}{16}(80 - 5 \cdot 3) = 4.0625$. ■

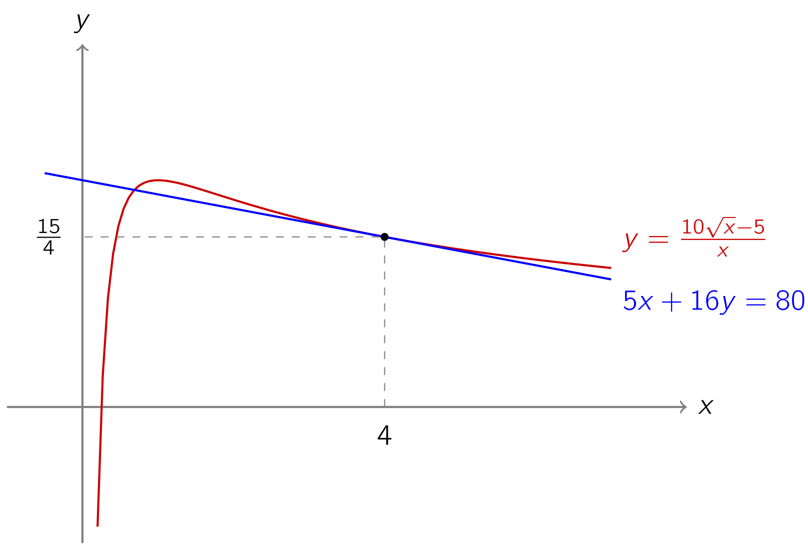

Figure 10: Plot of $y = \frac{10\sqrt{x}-5}{x}$ and the tangent line $5x + 16y = 80$ on the same axes, notice that for points close to $x = 4$, they are very close.