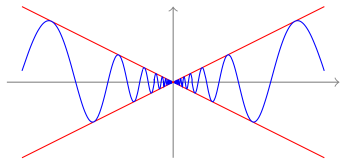


AN INTRODUCTION TO

SINGLE-VARIABLE CALCULUS

Luke Collins

maths.com.mt/notes

Version 0.1* (11th December, 2020)

Contents

1 Introduction **2**

2 Limits and Continuity **2**

2.1 Epsilons and Deltas 2

2.2 Theorems on Limits 11

2.3 Techniques for Evaluating Limits 15

2.4 Asymptotic Notation 16

3 Differentiation **19**

3.1 Calculus of Differences 19

3.2 Calculus of Differentials 24

*If you find any mathematical, grammatical or typographical errors whilst reading these notes, please let the author know via email: luke.collins@um.edu.mt.

1 Introduction

The topic of *infinitesimal calculus*, or just calculus as it's more commonly referred to, has quite the reputation for being "difficult" among callow students of mathematics. Although it is true that calculus may have, at some point, been deemed at the cutting-edge of mathematics education; today, this cannot be farther from the truth—due to its ubiquity, it has become an essential part of one's elementary mathematical skill set.

Calculus comprises two main areas, the first of which we shall consider is known as *differential calculus*. Let $f : A \to \mathbb{R}$ be a real-valued function. The principal goal in differential calculus is to approximate f "around" a point $x \in A$ using simpler functions. Surprisingly, this turns out to have deep connections to the other main part of calculus, the so-called *integral calculus*, which has to do with curved areas. But we will get to that in due time.

Before we can start exploring differential calculus itself, we need to establish some powerful mathematical machinery, the first of which is the notion of the limit of a real-valued function. This idea, or at least its modern formulation, is mainly the work of Augustin-Louis Cauchy.

2 Limits and Continuity

The limit is the essential idea underpinning all of calculus. Basically, it allows us to predict the value of a function at a point by looking at neighbouring points. Differentiation, integration and convergence of sequences/series all rely on the idea of a limit. It is definitely the most important (and probably most difficult) idea we will discuss in these notes.

2.1 Epsilons and Deltas

Consider the subset $A \subseteq \mathbb{R}$. We say a point $x \in \mathbb{R}$ is a *cluster point* of A if there are points in A which are arbitrary close to x (other than x itself, if it happens to be the case that $x \in A$). For example, if we take the set $A = (0, 1) \cup \{2\}$, then $2 \in A$ is not a cluster point because there are no points of A around 2 other than 2 itself. On the other hand, $1/2$ is a cluster point,

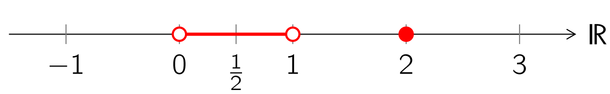


Figure 1: The set $(0, 1) \cup \{2\}$

since there are points in A within a distance of $\epsilon > 0$ from $1/2$, no matter how small ϵ is. 1 is also a cluster points because there are always points to its left, even though 1 itself is not a member of the set.

To make this idea of "arbitrarily close" precise, we state the definition as follows.

Definition 2.1 (Cluster point)**.** Let $A \subseteq \mathbb{R}$, and let $x \in \mathbb{R}$. Then x is said to be a *cluster point* or a *limit point of* A, if for any $\epsilon > 0$,

$$[(x-\epsilon, x+\epsilon) \cap A] \setminus \{x\} \neq \emptyset.$$

We need this idea because we can only make sense of limits at cluster points. Indeed, since the limit is meant to allow us to predict the value of $f(x)$ at $x = a$ by looking at neighbouring points, then a must have neighbouring points in the first place!

We can now state the (infamous) definition of the limit.

Definition 2.2 (Limit of a function)**.** Let $f\colon A \to \mathbb{R}$, and let a be a cluster point of A. Then $\ell \in \mathbb{R}$ is the *limit of f as x approaches a* or simply the *limit of f at a*, written

$$\lim_{x \to a} f(x) = \ell \qquad \text{or} \qquad f(x) \to \ell \text{ as } x \to a,$$

if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$,

$$x \neq a \quad \text{and} \quad a - \delta < x < a + \delta \implies \ell - \epsilon < f(x) < \ell + \epsilon.$$

Intuitively, what the definition tells us is that the limit is ℓ if for any $\epsilon > 0$, we can find a $\delta > 0$ such that all values in $(a-\delta, a+\delta)$ *apart from a itself* are mapped to values in $(\ell - \epsilon, \ell + \epsilon)$. [Keep re-reading this until you understand it! Refer to figure 2.]

Graphically, this corresponds to choosing $\delta > 0$ such that the outputs of the function f within the "input strip" $(a-\delta, a+\delta)$ (again excluding a itself) lie in the strip defined by the orange dashed lines in figure 2, corresponding to $(\ell - \epsilon, \ell + \epsilon)$. This is to hold for any $\epsilon > 0$: in particular, the ϵ-strip defined by the two dashed orange lines can be made arbitrarily small; which means that the corresponding chosen δ must also become small. This causes us to "narrow in" on the point (a, ℓ) (figure 3).

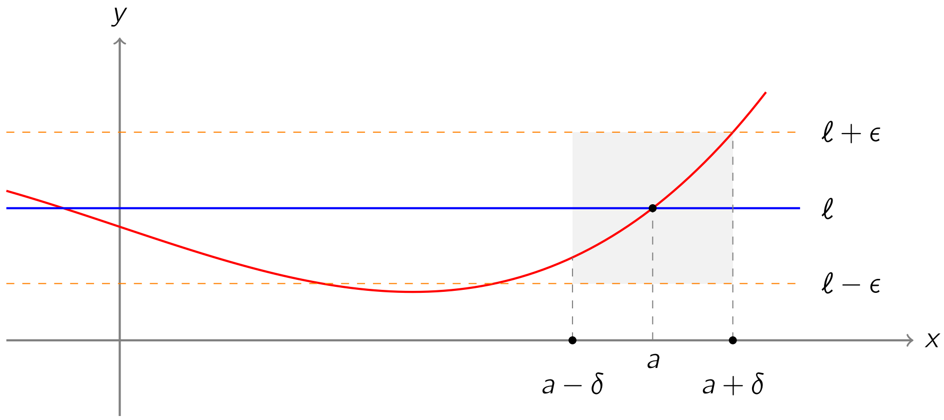


Figure 2: Limit of $f(x)$ at a

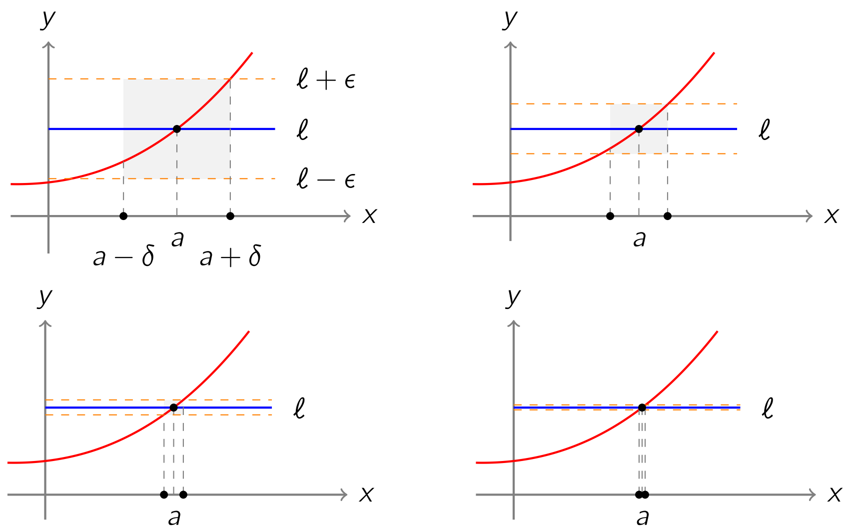


Figure 3: As $\epsilon > 0$ gets smaller, we are forced to choose $\delta > 0$ in such a way that the grey rectangle "narrows in" to (a, ℓ)

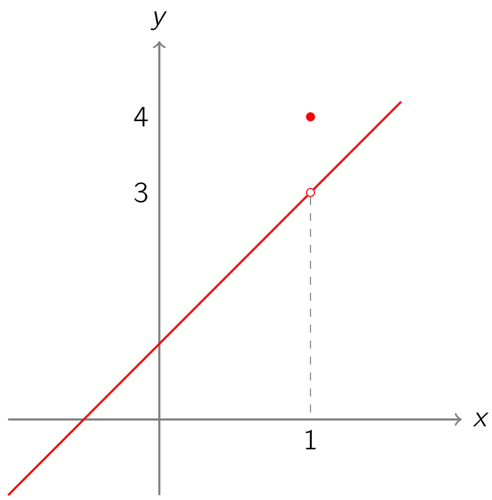


Figure 4: Sketch of $y = f(x)$

Example 2.3. What we are essentially doing here is predicting the value of a function at a point by studying where the neighbouring points are mapped (remember we are never considering where the point a itself is mapped!). Indeed, consider the strange function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 2x+1 & \text{if } x \neq 1 \\ 4 & \text{if } x = 1, \end{cases}$$

whose sketch is presented in figure 4. What is the limit of $f(x)$ at $x = 1$? Unfortunately definition 2.2 does not tell us how to find the limit; it only tells us what condition a number ℓ must satisfy in order to be the limit. We will therefore guess a value for ℓ, and check if it satisfies the defining condition.

Since the idea of a limit is to "predict" the value of a function at a point, we will guess that $\lim_{x\to 1} f(x) = 4$ as it is defined. What do we need to prove? Well we plug into the definition $a = 1$ and $\ell = 4$:

$$x \neq 1 \quad \text{and} \quad 1-\delta < x < 1+\delta \implies 4-\epsilon < f(x) < 4+\epsilon.$$

Noting that for $x \neq 1$, $f(x) = 2x + 1$, this becomes equivalent to

$$x \neq 1 \quad \text{and} \quad 1-\delta < x < 1+\delta \implies 4-\epsilon < 2x+1 < 4+\epsilon.$$

This is to hold for all $\epsilon > 0$, so we have no control over the value of ϵ. What we do have control over is the value of δ. Our goal is, for all $\epsilon > 0$, to find a $\delta > 0$ which makes the above true.

Let's say $\epsilon = 1.5$. Can we find δ in this case? We have

$$\begin{aligned} x \neq 1 \quad \text{and} \quad 1-\delta < x < 1+\delta &\implies 4-1.5 < 2x+1 < 4+1.5 \\ &\iff 2.5 < 2x+1 < 5.5 \end{aligned}$$

Can we take, say, $\delta = 0.5$? Well... no, because then we get

$$x \neq 1 \quad \text{and} \quad 0.5 < x < 1.5 \implies 2.5 < 2x+1 < 5.5$$

and $x = 0.6$ satisfies the antecedent but not the consequent. We need the inequality on x in the antecedent to be tighter. What if we take $\delta = 0.1$? In this case we get

$$x \neq 1 \quad \text{and} \quad 0.9 < x < 1.1 \implies 2.5 < 2x+1 < 5.5,$$

and this actually works! Indeed, if we massage the inequality on the right a bit, we have

$$\begin{aligned} x \neq 1 \quad \text{and} \quad 0.9 < x < 1.1 &\implies 2.5 < 2x+1 < 5.5 \\ &\iff 1.5 < 2x < 4.5 \\ &\iff 0.75 < x < 2.25, \end{aligned}$$

so we must ask ourselves, is it true that

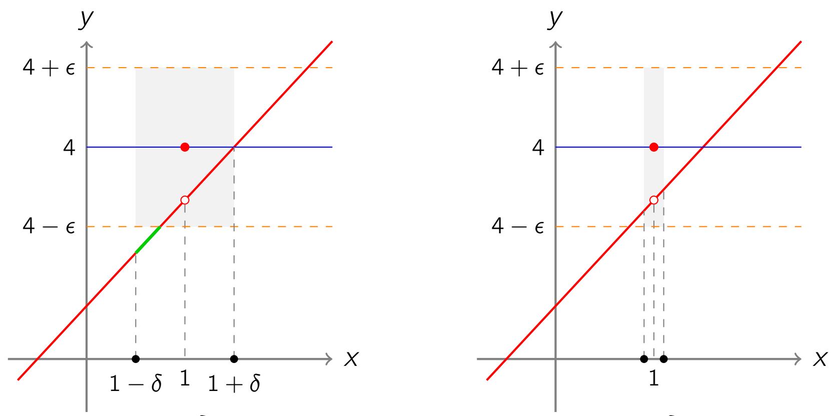
$$x \neq 1 \quad \text{and} \quad 0.9 < x < 1.1 \implies 0.75 < x < 2.25?$$

Clearly, it is. But what have we proved here? Have we shown that the limit is 4? No—the limit is 4 if we can find such a δ *for all* given $\epsilon > 0$, we only found one for $\epsilon = 1.5$. Refer to figure 5.

We can actually reduce the guesswork involved here. Let us again focus on the case $\epsilon = 1.5$, but leave δ unspecified. We can still carry out the same "massaging" on the consequent:

$$\begin{aligned} x \neq 1 \quad \text{and} \quad 1-\delta < x < 1+\delta &\implies 2.5 < 2x+1 < 5.5 \\ &\iff 1.5 < 2x < 4.5 \\ &\iff 0.75 < x < 2.25. \end{aligned}$$

Thus it is clear here that for $\epsilon = 1.5$ we can take any value of δ in the range $0 < \delta \leqslant 0.25$, because then the inequality on the left immediately incorporates the one on the right. If $\delta > 0.25$ though, then we can choose an x such that the inequality on the left holds, but the one on the right doesn't.



$\epsilon = 1.5$, $\delta = 0.5$
Points mapped to the green part of the line violate the definition, δ needs to be smaller

$\epsilon = 1.5$, $\delta = 0.1$
This δ works for this ϵ, the curve is mapped entirely within the ϵ-strip

Figure 5: Different values of δ for $\epsilon = 1.5$

Remember that what we are doing here is trying to fit the values of $f(x)$ for inputs close to $x = 1$ into the ϵ-strip around $\ell = 4$, where we get to decide what "close to $x = 1$" means by varying δ.

Now suppose instead that $\epsilon = 0.5$. Then this time we want

$$\begin{aligned} x \neq 1 \quad \text{and} \quad 1-\delta < x < 1+\delta &\implies 3.5 < 2x+1 < 4.5 \\ &\iff 2.5 < 2x < 3.5 \\ &\iff 1.25 < x < 1.75 \end{aligned}$$

But clearly no $\delta > 0$ we choose can ever make this implication true, since the value $x = 1 - \frac{\delta}{2}$ (for example) will always satisfy the antecedent but not the consequent.

What does this mean? It means that the limit is not 4! Why? Because for the limit to be 4, we need that "*for all* $\epsilon > 0$*, there exists* $\delta > 0$ ...", and we found an $\epsilon > 0$ for which no $\delta > 0$ does the job. Thus 4 fails to meet the definition.

Play around with different values of ϵ and δ for $f(x)$ at

https://maths.com.mt/calculus-eg-1.

This is actually not surprising at all, because remember, what the limit does is try and guess the value at $x = 1$ by looking at where all the *neighbouring* points are mapped to (not at $x = 1$ itself). Based off the neighbouring points, surely one would have to guess that the limit is 3!

Let us prove that the limit is indeed 3. But first, let us paraphrase definition 2.2 to make it easier to work with. ■

Remark 2.4. Recall that the absolute value function $|\cdot| \colon \mathbb{R} \to \mathbb{R}$ is defined by

$$|x| = \begin{cases} x & \text{if } x \geqslant 0 \\ -x & \text{otherwise,} \end{cases}$$

and that it obeys the following three properties for all $x, y \in \mathbb{R}$:

(i) $|x| = 0 \iff x = 0$,

(ii) $|xy| = |x||y|$,

(iii) $|x + y| \leqslant |x| + |y| \qquad$ (triangle inequality).

Consequently, if we have $-a < b < a$, we can write this more concisely as $|b| < a$. We can use this fact to write definition 2.2 more compactly. The condition there was

$$\begin{aligned} & x \neq a \quad \text{and} \quad a - \delta < x < a + \delta \implies \ell - \epsilon < f(x) < \ell + \epsilon \\ \iff & x \neq a \quad \text{and} \quad -\delta < x - a < \delta \implies -\epsilon < f(x) - \ell < \epsilon \\ \iff & x \neq a \quad \text{and} \quad |x - a| < \delta \implies |f(x) - \ell| < \epsilon \\ \iff & 0 < |x - a| < \delta \implies |f(x) - \ell| < \epsilon, \end{aligned}$$

where the last $\Leftrightarrow$ follows since $0 < |x - a|$ is true if and only if $x \neq a$ by (i) of the properties above. We shall prefer this version of definition 2.2 to work with, it makes proofs easier. ■

Therefore to restate the definition with the new condition:

Definition 2.2 (Limit of a function)**.** Let $f \colon A \to \mathbb{R}$, and let a be a cluster point of A. Then $\ell \in \mathbb{R}$ is the *limit of f at a*, written

$$\lim_{x \to a} f(x) = \ell \qquad \text{or} \qquad f(x) \to \ell \text{ as } x \to a,$$

if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$,

$$0 < |x - a| < \delta \implies |f(x) - \ell| < \epsilon.$$

Example 2.3 (continued). We now prove that $\lim_{x \to 1} f(x) = 3$.

Proof. Indeed, suppose we are given $\epsilon > 0$. We need to show that we can always find a value $\delta > 0$ such that for any $x \in \mathbb{R}$ (the domain of f), we have

$$\begin{aligned} 0 < |x-1| < \delta &\implies |f(x) - 3| < \epsilon \\ &\iff |(2x+1) - 3| < \epsilon \qquad \text{(since } x \neq 1 \text{ by antecedent)} \\ &\iff |2x - 2| < \epsilon \\ &\iff 2|x-1| < \epsilon \\ &\iff |x-1| < \epsilon/2. \end{aligned}$$

So what we want to show is that no matter what ϵ is, we can choose δ such that the implication

$$0 < |x-1| < \delta \implies |x-1| < \epsilon/2$$

is always true. But if we take $\delta = \epsilon/2$, then the implication is always true! $\square$

Thus we have proved that given any ϵ, an appropriate δ can be found for that ϵ, namely $\delta = \epsilon/2$. What this means is that values of x in the range $(1-\epsilon/2, 1+\epsilon/2)$ (as usual excluding 1 itself) are always mapped to $(4-\epsilon, 4+\epsilon)$, no matter what ϵ is!

The reason we divide by 2 is because the line $y = 2x + 1$ has gradient 2, so the sides of the usual "grey rectangle" of figure 2 will be in the ratio 2 : 1. Play with different values of ϵ: https://maths.com.mt/calculus-eg-2. ■

Example 2.5. As another example, we show that $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 5x - 3$ has

$$\lim_{x \to 1} f(x) = 2.$$

Indeed, suppose that $\epsilon > 0$ is given. Then we need to show that we can always find $\delta > 0$ such that for all $x \in \mathbb{R}$,

$$\begin{aligned} 0 < |x-1| < \delta &\implies |(5x-3) - 2| < \epsilon \\ &\iff |5x - 5| < \epsilon \\ &\iff 5|x-1| < \epsilon \\ &\iff |x-1| < \epsilon/5, \end{aligned}$$

take $\delta = \epsilon/5$ and we are done. ■

Exercise 2.6. Consider $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 10x + 1$ for all $x \in \mathbb{R}$. Guess the value of $\lim_{x\to 20} f(x)$ and prove that it is indeed the limit.

You may have encountered the term "continuous function" before, maybe stated loosely as "*a function which you can draw without removing your pencil from the paper*". Here we give a formal definition.

Definition 2.7 (Continuity)**.** Let $f\colon A \to \mathbb{R}$, and let $a \in A$. Then f is *continuous at a* if either a is not a cluster point of A, or a is a cluster point of A and

$$\lim_{x\to a} f(x) = f(a).$$

If f is continuous at each $a \in A$, we simply say f is *continuous*.

In other words, a function is continuous at a if its *predicted value* (i.e., the limit at a) is equal to its *assigned value* ($f(a)$). Thus a continuous function should not have unexpected jumps as f in example 2.3 had.

Example 2.8. Let us show that $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is continuous at $x = 2$, i.e. that

$$\lim_{x\to 2}(x^2) = 4.$$

Indeed, suppose we are given $\epsilon > 0$. We need to show that we can always find $\delta > 0$ such that for all $x \in \mathbb{R}$,

$$\begin{aligned} 0 < |x-2| < \delta \implies& |x^2 - 4| < \epsilon \\ \iff& |(x-2)(x+2)| < \epsilon \\ \iff& |x-2||x+2| < \epsilon \end{aligned}$$

As in previous examples, we should delight in the fact that the term $|x-2|$ appears on the right, because we know that we can control that with δ. The problem here is now the term $|x+2|$ — what can we do with this?

Well, we will do our best to transform into $|x-2|$! By the triangle inequality (remark 2.4) we have that

$$|x+2| = |x-2+4| \leqslant |x-2| + 4.$$

Now consider this: if some $\delta > 1$ works (say for example, $\delta = 2$ works), then any δ smaller than 1 would also work. Indeed, if any δ works, then anything smaller will also work. So if we add the perfectly reasonable assumption that $\delta \leqslant 1$, we get that

$$|x+2| \leqslant |x-2| + 4 < \delta + 4 \leqslant 1 + 4 = 5.$$

Thus if we set $\delta = \min\{\epsilon/5, 1\}$ (just in case the given ϵ is larger than 5), we get

$$0 < |x-2| < \delta \implies |x-2||x+2| < \epsilon/5(5) = \epsilon,$$

as required. ■

Exercise 2.9.
1. Show that $f(x) = 1$ is continuous at every point $x = c \in \mathbb{R}$.
2. Show that $f(x) = x$ is continuous at every point $x = c \in \mathbb{R}$.
3. Adapt example 2.8 to show that $f(x) = x^2$ is continuous at $x = 3$.
4. Show that $f(x) = x^3$ is continuous at $x = 1$.
5. Show that $f(x) = 1/x$ is continuous at $x = 2$.

2.2 Theorems on Limits

So far we have been using the terminology "*the* limit of f at x" (as opposed to "*a* limit"). Is it true that there is always only one number $\ell \in \mathbb{R}$ which is a limit of $f(x)$; i.e. which obeys the condition in definition 2.2?

The answer is:

Theorem 2.10 (Uniqueness of Limits)**.** *Let $f : A \to \mathbb{R}$, and let a be a cluster point of A. If there exists $\ell \in \mathbb{R}$ such that*

$$\lim_{x \to a} f(x) = \ell,$$

then ℓ is unique.

Proof. Suppose not. Suppose there are $\ell_1 \neq \ell_2$ such that $\lim_{x\to a} f(x) = \ell_1$ and $\lim_{x\to a} f(x) = \ell_2$. Define $\epsilon = |\ell_1 - \ell_2|$. Then for all $x \in A$, there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$0 < |x-a| < \delta_1 \implies |f(x) - \ell_1| < \epsilon/2, \quad (1)$$
$$\text{and} \quad 0 < |x-a| < \delta_2 \implies |f(x) - \ell_2| < \epsilon/2. \quad (2)$$

Now let $\delta = \min\{\delta_1, \delta_2\}$, and let $x \in (a-\delta, a+\delta) \cap A$ such that $x \neq a$.[1] Then x must satisfy both (1) and (2) above; i.e. we have both

$$|f(x) - \ell_1| < \epsilon/2 \quad \text{and} \quad |f(x) - \ell_2| < \epsilon/2.$$

Hence by the triangle inequality,

$$\begin{aligned} \epsilon = |\ell_1 - \ell_2| &= |\ell_1 - f(x) + f(x) - \ell_2| \\ &\leqslant |f(x) - \ell_1| + |f(x) - \ell_2| < \epsilon/2 + \epsilon/2 = \epsilon, \end{aligned}$$

in other words, $\epsilon < \epsilon$, a contradiction. □

Remark 2.11. Although the limit of a function is always unique, it does not always necessarily exist. Consider the function $f \colon \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{x}.$$

Although $0 \notin \operatorname{dom}(f)$, it is a cluster point, so we can still consider the limit as $x \to 0$. Indeed, suppose that $f(x) \to \ell$ as $x \to 0$. By definition 2.2, we have that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in \mathbb{R} \setminus \{0\}$,

$$0 < |x| < \delta \implies |1/x - \ell| < \epsilon. \tag{1}$$

Take $\epsilon = 1$, and let $x = \delta/(\delta + 1)$, where δ is the corresponding δ for $\epsilon = 1$. Clearly this value of x satisfies $0 < |x| < \delta$, so by (1) we get that

$$\begin{aligned} &\left|\frac{\delta+1}{\delta} - \ell\right| < 1 \\ \implies &\left|1 + 1/\delta - \ell\right| < 1 \\ \implies &-1 < 1 + 1/\delta - \ell < 1 \\ \implies &1/\delta < \ell < 1/\delta + 2. \end{aligned} \tag{2}$$

On the other hand, if we take $x = -\delta/(\delta + 1)$ this also satisfies $0 < |x| < \delta$, and in a similar fashion to the above we get that

$$-2 - 1/\delta < \ell < -1/\delta. \tag{3}$$

From (2) and (3), we get

$$1/\delta < \ell < -1/\delta \implies 1/\delta < -1/\delta \implies 2/\delta < 0 \implies 2\delta < 0 \implies \delta < 0,$$

contradicting that $\delta > 0$. ■

[1] There exists such an $x \in (a - \delta, a + \delta)$ since a is a cluster point of A.

There is a lot more about limits to be said, and here we've barely scratched the surface. However in the interest of time, we will have to state the most important results about them here without proof.

The important thing is that you have been exposed to working with ϵ's and δ's. The proof of each of these results involves very similar reasoning to what we've been doing so far. If you ever decide to study mathematics at university, the proofs of these theorems will be covered in a basic course on real analysis.

Theorem 2.12 (Algebraic Limit Theorem)**.** *Let $f, g\colon A \to \mathbb{R}$ be two real-valued functions, let a be a cluster point of A, and suppose $\lim_{x\to a} f(x) = \ell_f$ and $\lim_{x\to a} g(x) = \ell_g$ (i.e. they exist). Then*

(i) $\lim_{x\to a}(f(x) + g(x)) = \ell_f + \ell_g$,

(ii) $\lim_{x\to a}(f(x) - g(x)) = \ell_f - \ell_g$,

(iii) $\lim_{x\to a}(f(x)g(x)) = \ell_f \ell_g$,

(iv) If $\ell_g \neq 0$, then $\lim_{x\to a} {}^{f(x)}/_{g(x)} = {}^{\ell_f}/_{\ell_g}$.

Theorem 2.13 (Composition of Limits)**.** *Let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be two real valued functions such that $f(A) \subseteq B$, so that $g \circ f\colon A \to \mathbb{R}$ is defined. If $\lim_{x\to a} f(x)$ exists and g* is continuous there, *then*

$$\lim_{x\to a} f(x) = b \quad \textit{and} \quad \lim_{x\to b} g(x) = c \qquad \Longrightarrow \qquad \lim_{x\to a}(g \circ f)(x) = c,$$

or more concisely,

$$\lim_{x\to a}(g \circ f)(x) = g\big(\lim_{x\to a} f(x)\big),$$

since g is continuous at b.

Remark 2.14. Note the condition that g must be continuous at $b = \lim_{x\to a} f(x)$! It would be nice if in general

$$\lim_{x\to a} f(x) = b \quad \text{and} \quad \lim_{x\to b} g(x) = c \qquad \Longrightarrow \qquad \lim_{x\to a}(f \circ g)(x) = c \tag{1}$$

but unfortunately this is **not** true.

A classic counterexample is the following. Define $f, g\colon \mathbb{R} \to \mathbb{R}$ by

$$f(x) = g(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$

It is not hard to see that $\lim_{x\to 0} f(x) = 0 = \lim_{x\to 0} g(x)$. So (1) above would suggest that $\lim(g \circ f)(x) = 0$. But note that for all $x \neq 0$,

$$(g \circ f)(x) = g(0) = 1,$$

so we would actually get that $\lim_{x\to 0}(g \circ f)(x) = 1$. ■

Example 2.15. Theorems 2.12 and 2.13 are very useful, the allows us to prove, very easily, facts such as

$$\lim_{x\to a} \sin\left(\frac{x^3 + 7x^5}{x^2 + 1}\right) = \sin\left(\frac{a^3 + 7a^5}{a^2 + 1}\right)$$

without writing down a single ϵ or δ. One only needs to show that $\sin x$ and x^n are continuous (on $\mathbb{R}$), $\lim_{x\to a} 7 = 7$, $\lim_{x\to a} 1 = 1$ and $\lim_{x\to a} x = a$. The continuity of $\sin x$ is the only challenging one here, the rest could be set to you as easy exercises.

If we allow ourselves to use these facts, by theorems 2.12 and 2.13, we get

$$\lim_{x\to a} 7 = 7 \quad \text{and} \quad \lim_{x\to a} x^5 = a^5 \implies \lim_{x\to a} (7x^5) = 7a^5 \tag{1}$$

$$(1) \quad \text{and} \quad \lim_{x\to a} x^3 = a^3 \implies \lim_{x\to a} (x^3 + 7x^5) = a^3 + 7a^5 \tag{2}$$

$$\lim_{x\to a} x^2 = a^2 \quad \text{and} \quad \lim_{x\to a} 1 = 1 \implies \lim_{x\to a} (1 + x^2) = 1 + a^2 \tag{3}$$

$$(2), (3) \quad \text{and} \quad a^2 + 1 \neq 0 \implies \lim_{x\to a} \left(\frac{x^3 + 7x^5}{1 + x^2}\right) = \frac{a^3 + 7a^5}{a^2 + 1} \tag{4}$$

$$\sin x \text{ continuous} \quad \text{and} \quad (4) \implies \lim_{x\to a} \sin\left(\frac{x^3 + 7x^5}{1 + x^2}\right) = \sin\left(\frac{a^3 + 7a^5}{a^2 + 1}\right),$$

as required. ■

Theorems 2.12 and 2.13 yield easily the following facts about continuity.

Corollary 2.16 (Algebraic Continuity Theorem)**.** *Let $f, g\colon A \to \mathbb{R}$ be two real-valued functions continuous at $a \in A$. Then*

(i) *$f(x) + g(x)$ is continuous at a,*

(ii) *$f(x) - g(x)$ is continuous at a,*

(iii) *$f(x)g(x)$ is continuous at a,*

(iv) *If $g(a) \neq 0$, ${}^{f(x)}/_{g(x)}$ is continuous at a.*

Corollary 2.17 (Continuity of Composition)**.** *Let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be two real valued functions such that $f(A) \subseteq B$. If f is continuous at a, and g is continuous at $f(a)$, then $g \circ f$ is continuous at a.*

Next we state some important limits which crop up quite often and are not trivial to prove.

Theorem 2.18 (Special Limits)**.** *(i)* $\lim_{x\to 0} \frac{\sin x}{x} = 1$.

(ii) $\lim_{x\to 0} \frac{1-\cos x}{x} = 0$.

(iii) $\lim_{x\to 0} (1+x)^{1/x} = e$.

(iv) $\lim_{x\to\infty} \left(1+\frac{1}{x}\right)^x = e$. [2]

Finally, we give a list of familiar functions which are continuous.

Theorem 2.19 (List of continuous functions)**.** *The following functions are continuous everywhere on their domain.*

(i) *all constant functions,*

(ii) *all polynomials,*

(iii) *roots of x ($\sqrt{x}$, $\sqrt[3]{x}$, ...),*

(iv) *the absolute value $|x|$,*

(v) *all rational functions, i.e., all functions of the form*

$$f(x) = \frac{p(x)}{q(x)}$$

where p, q are polynomials (and $x \in \operatorname{dom}(f) \Leftrightarrow q(x) \neq 0$),

(vi) *the exponential function e^x,*

(vii) *the logarithm $\log x$,*

(viii) *the trigonometric functions* cos, sin, tan, sec, csc, cot

(ix) *the inverse trigonometric functions* $\cos^{-1}$, $\sin^{-1}$, $\tan^{-1}$.

2.3 Techniques for Evaluating Limits

TBA

[2] If for some $A \in \mathbb{R}$, the set $(A, \infty) \subseteq \operatorname{dom}(f)$, then when we write $\lim_{x\to\infty} f(x) = \ell$ to mean that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in \operatorname{dom}(f)$, we have $|x| > \delta \implies |f(x) - \ell| < \epsilon$.

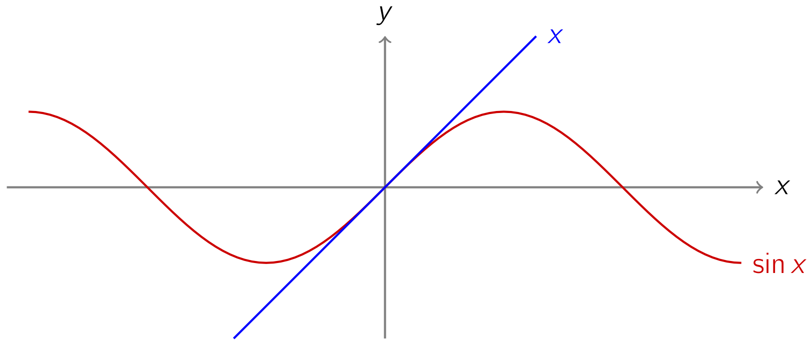


Figure 6: Plot of $\sin x$ and x on the same axes, notice that for small inputs x, they are very close

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

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

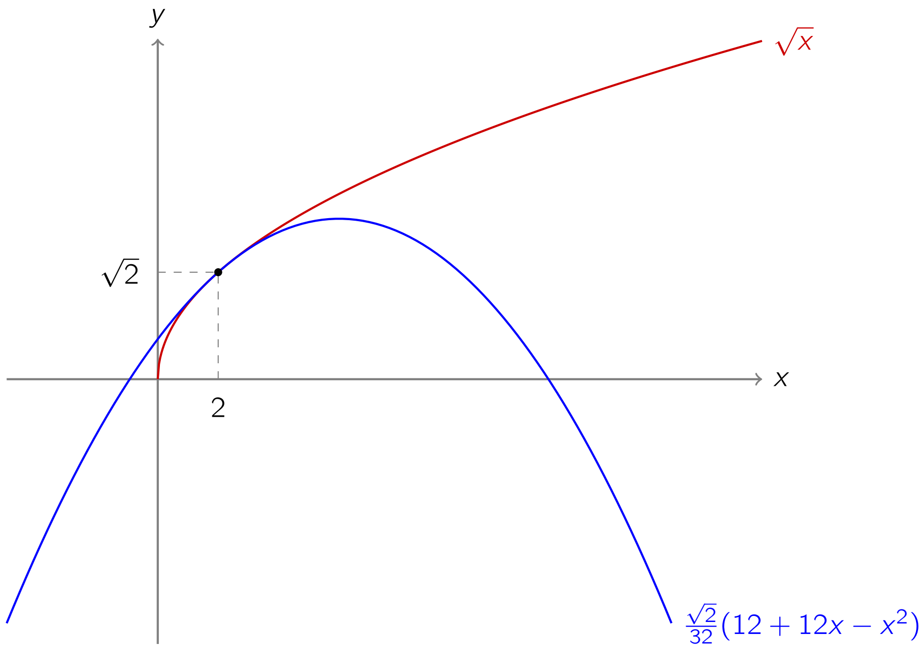


Figure 7: Plot of $\sqrt{x}$ and $\frac{\sqrt{2}}{32}(12+12x-x^2)$ on the same axes, notice that for x close to 2, they are good approximations of each other

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

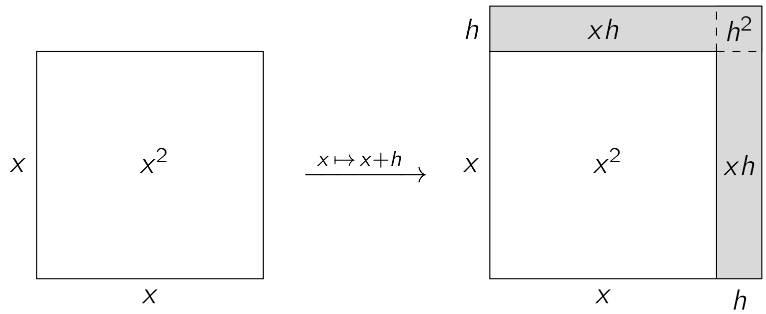


Figure 8: Interpreting example 3.2 as a change in area: $\Delta(x^2) = 2xh + h^2$

expressions in terms of x formally as functions. For instance, if $f\colon \mathbb{R} \to \mathbb{R}$ 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:

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

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 $f = f\circ \mathrm{id}$, where id denotes the *identity function* defined by $\mathrm{id}(x) = x$ for all $x \in \mathbb{R}$. If we apply proposition 3.9 to this composition, we see that

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

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 $\Delta\mathrm{id}$, since this is the expression defining $\mathrm{id}(x)$. Indeed,

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

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

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

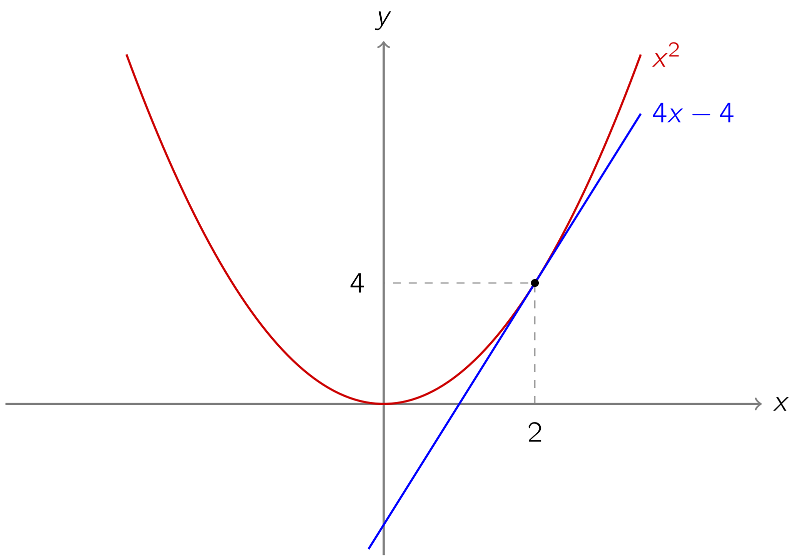


Figure 9: Plot of x^2 and $4x - 4$ on the same axes, notice that for points close to $x = 2$, they are very close.

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 $f : A \to \mathbb{R}$ 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 $f : A \to \mathbb{R}$ 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 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.)