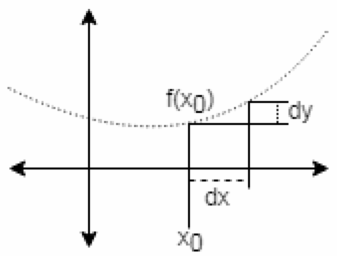
Analysis 1, Week 19
COMS10003 Lecture Notes for Spring 2020

Problem statement: where is my bus?

You board a bus at the Triangle at 10am exactly. Your bus will travel up Whiteladies Road towards Stoke Bishop. Where do you expect it to be at 10:02? To answer this question, we can assume that buses do not randomly disappear and reappear in other places, so the bus will be somewhere near the triangle at 10:01 and a bit further away – but not too far – at 10:02. If we know that the speed limit on Whiteladies Road is 20mph, that gives us a limit on how far it can go in two minutes; if we assume that the roads are free flowing then we can give a pretty good estimate of where we expect the bus to be.

Mathematically, the situation is that we have a function $y = f(x)$ which represents the position of the bus at time y, where we model time along the x–axis and Whiteladies Road in one dimension as the y–axis. We know that the function takes the value $y_0 = f(x_0)$ at a particular point x_0 and we want to find the function value at another point x_1 "not too far away". You might ask why we don't just evaluate $f(x_1)$, but the reason is that analysis was created to solve problems in physics, evaluating a function means making a measurement and measurements can be expensive. In a way, analysis asks how well we can understand a function without having to evaluate it everywhere.

If we set $dx := x_1 - x_0$ to be the distance between the point x_0 where we know the value and the point x_1 we are interested in, then we have the equation $f(x_1) = f(x_0 + dx)$. Here dx is simply a variable whose name has two letters, which as Computer Scientists should not confuse us. Similarly, we define $dy := f(x_1) - f(x_0) = f(x_0 + dx) - f(x_0)$ to be the change in the quantity we are interested in. As a picture, the situation looks like this:



If the function is not too "jumpy" and dx is "small" then we can imagine the value $f(x_0 + dx)$ will not be too far away from $f(x_0)$ either, e.g. dy will be "small" too. Let's stop focusing on a particular point x_0 and think of the function in general. If we think of x as a time and $f(x)$ as the bus's position at time x, then the quantity

$$\frac{dy}{dx} = \frac{f(x + dx) - f(x)}{dx}$$

represents the change in position divided by the change in time, which is what we call the speed of the bus: if x = 10:00, dx = 2 minutes, f(x) = 0 [miles from the Triangle] and f(x+dx) = 0.5 miles then the bus has been going at dy/dx = 0.5/(2/60) = 15 miles per hour.

Differentials, the original way

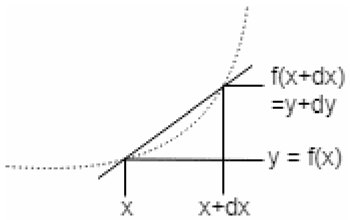
It is well known that differential calculus was discovered by both Isaac Newton and Gottfried Wilhelm Leibniz. What is less well known is that they did not use limits and the epsilon-delta method as it is taught today to define differentials. Instead, they made the following observation. If you know $f(x) = y$ and evaluate the function at a nearby spot $x + dx$, then you can sometimes write the difference in terms of a polynomial in dx. Consider the function $f(x) = 3x^3 + x$:

$$dy = f(x + dx) - f(x) = 3x^3 + 9x^2dx + 9x(dx)^2 + 3(dx)^3 + x + dx - 3x^3 - x$$
$$= 9x^2dx + 9x(dx)^2 + 3(dx)^3 + dx$$

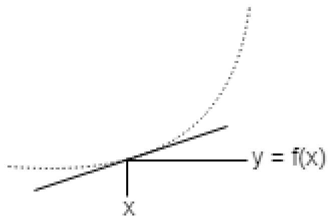
The original idea was that if dx is "small, but just small enough to care about" then $(dx)^2$ is "really small" and $(dx)^3$ "really really small" and so on – what if, Newton and Leibniz thought, we simply ignored all powers of d that are higher than one? For example, if we care about an accuracy of dx = 0.01 then $(dx)^2$ = 0.0001 and so on, which is much smaller than we care about. This also answers the question "how small?": if $dx < 1$ then $dx > (dx)^2 > (dx)^3$ ... so we at least need $dx < 1$ for this approach to have a chance of working.

If we do this to the function above, we get $dy = (9x^2 + 1)dx$. We can interpret this as saying that if we move a small distance dx away from x then our function value changes by dy. The value of dy depends on x as well as dx, which makes sense as the original function is not a straight line.

What we are really doing by dropping all higher powers of dx is approximating the function by a straight line. If dy and dx have small but nonzero values, without dropping any powers we would get a line that goes though the points (x, y) and $(x + dx, y + dy)$:



However, by dropping the higher powers of dx in the calculation, we produce a line that is the tangent to the function at the point x:



The product rule, the original way

Suppose we have a function $y = f(x) = a(x)b(x)$ which is a product of two functions a and b. What is dy/dx? The original argument goes like this[1]: we start with $y = f(x)$, that is we have a point x in mind where $f(x)$ has the value y. We can also evaluate the other functions at this point: let $u = a(x)$ and $v = b(x)$. If we increase x a little bit to $x + dx$ then the function value of f

[1] See: S. P. Thompson, Calculus Made Easy, 1913. A PDF of this work, now out of copyright, can be found by searching for the author and title online.

increases to $y + dy$, whereas the value of the function a will have increased too, so let's call its new value $u + du$ and the new value of b we call $v + dv$. This gets us:

$$dy = f(x + dx) - f(x) = a(x + dx)b(x + dx) - a(x)b(x) = (u + du)(v + dv) - uv$$

We can multiply out the brackets at the right to get

$$dy = u(dv) + v(du) + (du)(dv)$$

If both du and dv are small, then $(du)(dv)$ is "really small" again, so we can just drop that term and get $dy = u(dv) + v(du)$. Dividing everything by dx gives us the usual product rule:

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

The sum rule for $f(x) = a(x) + b(x)$ can be derived the same way.

Note on notation

This note is for pedants and functional programmers, and not part of the learning outcomes for this unit, but it helps to clarify a common problem with the notation.

We are talking about a function with type $f: \boldsymbol{R} \to \boldsymbol{R}$ (where $\boldsymbol{R}$ is the real numbers). We can often describe a function abstractly without a variable, e.g. "the squaring function" or "the sine function", but to write down an expression for a function we need to invent a variable. Writing $f(x) = x^2$ or $f(t) = t^2$ means exactly the same thing: changing the name of the variable does not change the function. However, if x_0 is a point in $\boldsymbol{R}$ then $f(x_0)$ means the point in $\boldsymbol{R}$ that you get by evaluating f at x_0. In other words, the notation $f(t)$ means a different thing depending on if t refers to a variable or a point, which is ugly – and the notation for differentials is even worse.

Differentiation is an operation on functions, which we could write abstractly with the symbol ∇ (we'll see why later on) and has type $\nabla: (\boldsymbol{R} \to \boldsymbol{R}) \to (\boldsymbol{R} \to \boldsymbol{R})$. For example, if f is the function that maps $x \mapsto x^2$ then ∇f is the function that maps $x \mapsto 2x$.

Sticking with $f(x) = x^2$, the common notation for the derivative is $\frac{df}{dx}$ or $\frac{d}{dx}f(x)$ where the d/dx part means "take the derivative" (what I called ∇ above). In other words, d/dx is formally a function of type $(\boldsymbol{R} \to \boldsymbol{R}) \to (\boldsymbol{R} \to \boldsymbol{R})$ which we are applying here to a function f of type $(\boldsymbol{R} \to \boldsymbol{R})$ to get another function of type $(\boldsymbol{R} \to \boldsymbol{R})$**.** The notation for the derivative is a bit unfortunate in that a variable $x$is mentioned, even though it has no point in being there. The second form is a bit clearer: putting d/dx in front of any expression means "differentiate this expression, treating it as a function in one variable x and assuming that all other letters are constants".

Unfortunately, this notation offers no nice way to say "differentiate the function and then evaluate the new function at a point x_0". You can write $\frac{d}{dx}f(x_0)$ but this can look like x_0 is the variable defining the function in the first place, or that you first evaluate then differentiate – functional purists could write this as $\left(\frac{d}{dx}f\right)(x_0)$ but no-one does that in practice. Another notation for differentiate-then-evaluate, which is unambiguous and common but clunky, is

$$\left.\frac{df}{dx}\right|_{x=x_0}$$

An alternative notation is to use a prime to denote differentiation, which is understood to mean with respect to whatever variable you are using: if $f(x) = x^2$ then $f'(x) = 2x$. This notation lets you evaluate nicely in one go, e.g. $f'(x_0)$ means evaluate the derivative at the point x_0.

For higher derivatives (the derivative of the derivative and so on) the first notation uses $d^n f/dx^n$ for deriving n times; in the prime notation, you can go some way with repeated primes e.g. $f'''(x)$ after which you write the number of derivatives as a bracketed exponent: $f^{(n)}(x)$.

Some differentials

The following rules allow you to compute the differential of a lot of functions:

- Power rule: if $f(x) = x^n$ for n an integer, then $f'(x) = nx^{n-1}$.
- Constant rule: if $f(x) = c \cdot g(x)$ where c is a constant or any term that does not depend on x then $f'(x) = c \cdot g'(x)$. If $f(x) = c$ is a constant function, then $f'(x) = 0$.
- Sum rule: if $f(x) = f_1(x) + f_2(x)$ is a sum of two functions then $f'(x) = f_1{}'(x)\_f_2{}'(x)$, as long as the differentials of both summands exist.

This lets us differentiate polynomials: if $f(x) = 3x^2 - 5x + 6$ then $f'(x) = 6x - 5$. The sum rule lets us treat each term individually, the constant rule means we can deal with the coefficients after we've done the x^n and the power rule deals with the x^n terms. Two more useful rules:

- Product rule: if $f(x) = f_1(x)f_2(x)$ then, as long as everything exists, $f'(x) = f_1(x)f_2'(x) + f_1'(x)f_2(x)$.
- Chain rule: if we can write a function as a composition of two others, $y(x) = f(g(x))$ then $y'(x) = f'\big(g(x)\big)g'(x)$.

For example, consider $y(x) = 1/(x + 1)$. We can write $y(x) = f\big(g(x)\big)$ with $g(x) = x + 1$ and $f(x) = 1/x$. Now $f'(x) = (-1)/x^2$ by the power rule ($1/x = x^{-1}$) and $g'(x) = 1$, so $y'(x) = (-1)/(x + 1)^2$.

As a special case of this rule we get the division rule: $(1/g(x))' = -g'(x)/g(x)^2$.

Some useful functions are worth memorising (though we will see why later on):

- $\sin'(x) = \cos(x)$ and $\cos'(x) = -\sin(x)$.
- $(e^x)' = e^x$, the exponential function does not change when you differentiate it.
- $(\ln x)' = 1/x$.

The modern approach: limits

The reason that calculus is not usually taught the original way anymore is that mathematicians discovered this approach does not work for all functions. The modern methods with limits and epsilon-delta analysis answers the question: which functions are we allowed to do this to? But a lot of functions that we encounter in practice (physics, engineering, other sciences) work just fine under the original approach.

In our example with $f(x) = 3x^3 + x$ again, if we directly try and calculate the differential we get

$$\frac{dy}{dx} = \frac{9x^2dx + 9x(dx)^2 + 3(dx)^3 + dx}{dx} = 9x^2 + 1 + 9x(dx) + 3(dx)^2$$

In this case we could simply ignore any remaining dx terms (we've already divided by dx once so these were the original $(dx)^2$ and higher power terms). But what if the fraction doesn't disappear? That's why the modern definition is

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

where we replaced "dx" with "h" to get the usual form of the definition back. This gives you the answer when you are allowed to differentiate a function at a particular point: it is allowed if the limit exists.

Reminder – quantifiers

Maths books will tell you that ∀ means "for all" and ∃ means "there exists", and that is correct, but it is helpful to think of quantifier definitions and proofs as a two-player game. The first player claims a statement. The second player challenges the first: whenever a "for all" quantifier appears, the challenger can choose any number that meets the conditions. Whenever a "there exists" quantifier appears, the first player must reply with a suitable answer. If the first player can always win this game by making the statement true, the statement is true.

For example, the following statement is true: for every integer there is an even larger integer

$$\forall\, n \in \mathbf{N}\; \exists\, m \in \mathbf{N} \quad m > n$$

If you pick for example n = 30, I can reply with m = 31; my strategy is to set m = n + 1. A strategy that always wins is nothing other than a proof of the statement.

Conversely, it is false that there is an integer larger than all others ($\exists\, m \in \mathbf{N}\; \forall\, n \in \mathbf{N}\; m > n$): with the quantifiers this way round, I have to pick m first and you can falsify the statement by picking n = m + 1 again.

Sequences, Series and Limits

You have an empty bucket. On day one, you add one unit (one pint, or one litre if you prefer) of water to the bucket. On each following day, you add some more water – consider three different possibilities:

1. Every day you add one more unit of water.
2. On day n you add 1/n units of water.
3. Every day you add half as much water as on the previous day.

Mathematically, we can model this as a sequence of numbers $a_1, a_2, a_3, \ldots$ where a_n is the amount of water that you add on day n. For example, the first sequence is simply $a_n = 1$ for all n whereas the second one is $b_n = 1/n$ and the third one is $c_n = 1/2^{n-1}$ (which you can see from the format $c_1 = 0, c_{n+1} = c_n/2$.

The amount of water in the bucket can be modelled as a series, which is a sequence built out of another sequence by adding up terms. For the first example we have

$$s_1 = a_1, \qquad s_2 = a_1 + a_2, \qquad s_3 = a_1 + a_2 + a_3 \qquad \ldots \qquad s_n = a_1 + \cdots + a_n$$

What happens to our sequences and series as n gets large? For the first sequence a_n, the sequence element is always 1 however large n becomes. The associated series however is $s_n = n$

which "goes to infinity" as n does. For the second sequence, b_n becomes smaller and smaller but the associated series (let's call it t_n) still becomes infinitely large; the third sequence c_n also grows smaller but at a much faster rate, and the series (let's call it u_n) approaches but never exceeds 2. You can see this fact by observing that each day halves the amount of water missing to complete 2 units, that is

$$u_n = \sum_{k=1}^{n} \frac{1}{2^{k-1}} = 2 - \frac{1}{2^{n-1}}$$

So, if the bucket can hold at least 2 units of water, it will never overflow.

Definition of the limit

Mathematically, a sequence a_n of real numbers has a limit a if the terms get arbitrarily close:

$$\lim_{n\to\infty} x_n = x \quad := \quad \forall\, \varepsilon > 0\ \exists n_0 \forall n > n_0\ |x - x_n| < \varepsilon$$

Read this as follows:

1. For every epsilon that you choose – it can be as small as you want, as long as it is greater than zero,
1. I can find you an integern_0,
2. such that for any $n > n_0$ you choose, the difference between a and a_n is at most your epsilon.

Looking at our example sequences:

- For the sequence $a_n = 1$ clearly $\lim_{n\to\infty} a_n = 1$ which holds for $n_0 = 1$ for any epsilon: the limit of a constant sequence is simply the constant.
- For $b_n = 1/n$ the limit is 0. This sequence has the property that $|b_n| < |b_{n_0}|$ for any $n > n_0$, in other words the sequence elements are decreasing. So for a given ε we can pick any $n_0 > 1/\varepsilon$, e.g. for $\varepsilon = 0.01$, $n_0 = 101$ will make the statement true.
- For $c_n = 1/2^{n-1}$ and any ε we need $2^{n_0-1} > 1/\varepsilon$ so a number above $\log_2 1/\varepsilon$ will do. Again, the sequence elements are decreasing.

Consider the sequence $s_n = n$ that we got as the series of sums over a_n. For $\varepsilon = 1/3$ if $|s_{n_0} - s| < 1/3$ then $|s_{n_0+1} - s|$ will be greater than $1/3$ again as all the sequence elements are one apart. Therefore, this series cannot have a limit.

The sequence $t_n = \sum_{k=1}^{n} 1/k$ also does not have a limit, though this is harder to show. A quick plot will show however that it "gets infinitely big". However, we have just argued that $u_n = \sum_{k=1}^{n} 1/2^{k-1} = 2 - 2^{n-1}$ so the limit of this sequence must be 2; given any $\varepsilon > 0$ we want to solve the equation $2^{-(n-1)} < \varepsilon$ which we can write as $2^{-n} < \varepsilon/2$ and solve with another logarithm. Note that while the elements of $c_n = 2^{1-n}$ decrease as n grows, the elements of u_n are increasing, but the difference $|2 - u_n|$ is decreasing. In general, to show that x is the limit of a sequence x_n we usually need to show two things: that $|x - x_n|$ is decreasing and that for a given ε we can find one particular n_0 that makes the inequality $|x - x_{n_0}| < \varepsilon$ work.

Some sequences have no limit, others have exactly one. You will see in the exercises that a sequence can never have more than one limit, so it makes sense to talk about "the limit" of a sequence if one exists.

Continuous functions

We can define limits for real-valued functions and at arbitrary points:

$$\lim_{x \to a} f(x) = b \quad := \quad \forall \varepsilon > 0 \exists \delta > 0 \forall x \; |x - a| < \delta \Rightarrow |f(x) - b| < \varepsilon$$

Read this definition as follows. Suppose you have an unknown function f given as a piece of code and you are allowed to evaluate f at any point except a, but you want to know f(a). You try points x getting closer and closer to a (but never a itself). If, the closer you get to a with your x, the closer f(x) gets to b then f(x) has a limit of b at point a. Again, a function might or might not have a limit at a particular point a.

In terms of a game: the limit is b if, for every ε you pick (however small) I can find a δ such that if you pick any x which is close enough to a (namely $|x - a| < \delta$) then $|f(x) - b|$ is smaller than the value of ε that you chose.

Does this mean that f(a) = b? The answer is sometimes, namely when the function is continuous:

A function $f: \mathbf{R} \to \mathbf{R}$ is called *continuous* if at for every real number a, the limit $\lim_{x \to a} f(x)$ exists and is equal to $f(a)$.

Continuous implies that the function does not "jump": if f(x) is 0 for x<2 and f(x) is 1 for x greater or equal to 2, then the limit of f(x) at 2 cannot exist: $f(2 - \delta) = 0, f(2 + \delta) = 1$ for any positive δ so neither 0 nor 1 can be the limit at this point (and nor can anything else).

The good news is that polynomial functions, exponential and logarithm functions, powers, roots and sines and cosines are all continuous. The limit of a function can exist even if the function is not defined there: the function $f(x) = (x + 2)/(x + 2)$ is not defined at (–2) but it has a limit of 1 there (the function is in fact 1 everywhere else). Dividing can cause functions to become not continuous however: $\sin(1/x)$ is not continuous at 0 (and it is not defined there either).

Differentials, the standard way

Once you know limits, the differential $f'(x)$ of a function $f(x)$ is usually defined as:

$$f'(x) := \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

More precisely, we say that f is *differentiable* at x if this limit exists there, and we simply say that f is differentiable if it is differentiable at every point where it is defined.

For example if $f(x) = ax^2$ where a is a constant then

$$f'(x) = \lim_{h \to 0} \frac{a(x + h)^2 - ax^2}{h} = \lim_{h \to 0} \frac{x^2 + 2hx + h^2 - x^2}{h} = \lim_{h \to 0} a(2x + h) = 2ax$$

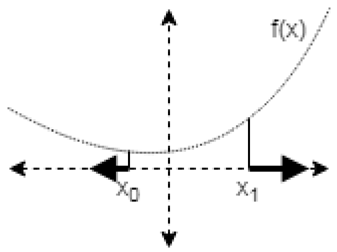
since once the denominator is gone, we can simply set $h = 0$ in the last equation.

Finding extremal points

The differential (a.k.a. derivative) of a function measures the function's slope at a point:

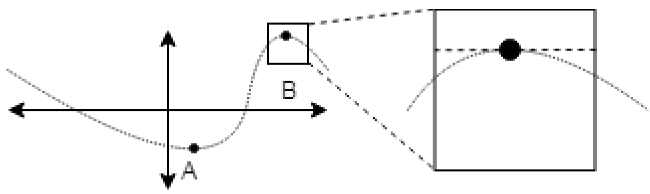
- If $f'(x_0) > 0$ then f is rising at the point x_0 (and the value of the differential tells us how steeply it is rising there).
- If $f'(x_0) < 0$ then f is falling at the point x_0.
- If $f'(x_0) = 0$ then f is flat at x_0, neither rising nor falling.

We can think of the derivative at a particular point as a 1-dimensional vector along the x-axis, starting at the point in question and pointing in the direction you have to move to go "up" along the function's curve; the length of this vector tells you how steep the function is at that point:



In this example, f is falling at x_0 so the derivative vector points in the minus direction; the function is fairly shallow there, so the vector is fairly short. On the other hand, at x_1 the function is rising and a bit steeper, so the vector points in the positive direction and is a bit longer. We will come back to this notion of the derivative as a vector when we do multidimensional functions.

A *local minimum* of a function is a point where the function is lower than any point nearby, the "bottom of a valley". A *local maximum* is the opposite – the "top of a peak", a point where the function is higher than any point nearby:



In this diagram, A is a local minimum and B a local maximum of the function indicated. A point that is a local minimum or maximum is called a *local extremum* (or: *local optimum*).

On a sufficiently "nice" function, a minimum is a point x_0 where the function is falling immediately before the point, flat at the point and rising immediately afterwards; a local maximum is the opposite. In other words, at a local extremum, the derivative is zero. This lets us find such points, with one caveat: the derivative can be zero even at a non-extremal point. This includes "saddles" (points where the function rises, flattens out and then rises again like $x_0 = 0$ for $f(x) = x^3$) or functions like $f(x) = 2$ where the function is simply flat everywhere.

To check what kind of point we have, we can introduce higher derivatives, which simply means taking the derivative more than once. This is written $d^2 f(x)/dx^2$ for the second derivative and $d^n f(x)/dx^n$ for the n-th derivative.

- A point where the first derivative is not zero is definitely not a minimum or maximum.
- A point where the first derivative is zero and the second derivative is greater than zero is a local minimum. For example, $f(x) = x^2$ has $f'(0) = 0$ and $f''(0) = 2$.
- A point where the first derivative is zero and the second derivative is less than zero is a local maximum. For example, $f(x) = -x^2$ at $x_0 = 0$.
- A point where the first two derivatives are both zero could be anything – you can look up if you want how to carry on checking the third derivative etc. until you get an answer.