Differentiation (beta)

The `5-triple-star` Project

Compiled on July 21, 2019

Introduction

What is differentiation? Broadly speaking, differentiation is the study of the *linear approximation* of a function, or, more appetizingly, the tendency of the trend that some function $f(t)$ will take when t increases.
The studies of approximation, trends and optima are long-standing problems of many aspects of development like engineering. The first sight of differentiation can even be found in ancient Greek, Japan and China.
This note is divided into 3 parts. The first two sections are purely about the intuition behind the definition of derivative (the first principle), for those who think that this definition is weird. From the third section onwards, we focus on the techniques and laws that can help us differentiate a function without having to cast the first principle. This includes product rule, quotient rule, chain rule, etc.. Also, we will go through the derivatives of some special functions, like logarithm and trigonometry. The last few section are about some notions relating to derivatives, which you ought to know for one reason or another.
It is recommended that you follow the chapters in their given order. However, you may skip the supplementary information part at the end of some sections. It will be best if you finish all the exercises, which is not that many. Sometimes an exercise does not require knowledge from the section which it is in. They are there for sake of a better understanding or alternate proofs. Those exercises are marked with an asterisk (*).
Throughout this note, unless otherwise specified, all the functions take value in $\mathbb{R}$ and their domains are open sets (unions of some open intervals) in $\mathbb{R}$.

Contents

1 Intuition

We shall step into the world of calculus from common sense. Now imagine that you are shooting a gun (in a world with unknown physical rules) in a field. When you shoot it at an angle of 30°, the bullet flied for a distance of 100m, and when the angle is increased to 40°, it flied for a distance of 120m. Now the question: if I change the angle to, say, 35°, can you estimate the distance for which the bullet might fly this time?

You are guessing somewhere near 110m, am I right? But why? Why is everyone so keen on this approximation? Why don't you say 1200m? 240000m? 3m? Of course, this is somewhat like a common sense, but what can we see from such kind of intuition, when applied to the approximation to other functions?

The "common sense" is both natural and unnatural. It relies on a simple observation, or shall we say, mind experiment. When you magnify a curve (in this case, the graph of the function $f(\theta)$ mapping the incident angle to the distance of the bullet) indefinitely, you will expect, at the end of the day, something that looks like a straight line [1].

This is, as we call it, a *linear approximation*. Why linear? Partly because it is close to our common sense, but more importantly, we know about linear functions in a much more extensive mannar than we know other functions. From preliminary knowledges, we know that we can represent a line on Cartesian coordinate by the simple general from

$$y = mx + c$$

. Since we are only interested in the trend instead of the general position, the value of c is sometimes less important. Then, for a function f, if we know its

[1] This, however, is not always true. Bear in mind that there are plenty of functions on which this does not apply. You will be asked to construct such a function in the exercises

values $f(a), f(b)$ at a, b respectively, we can reasonably approximate what in between them by a line joining the two known points. Its slope, of course, is given by the equation

$$m = \frac{f(b) - f(a)}{b - a}$$

This is an approximation instead of the true pattern, but if f is a nice enough [2] function, then we can approximate f arbitrarily close when a, b are arbitrarily close.

Geometrically, with the above assumption, then when a, b are *really close*, the resulting line will just *touch* the function, i.e. it will become a tangent. Just like the slope of a linear equation tells us its steepness, this tangent gives us the trend of the function in a small region near a and b. How can we, in fact, make a and b "really close"? Of course, we have this wonderful tool called limit.

Exercises

1. Construct a function $f : \mathbb{R} \to \mathbb{R}$ such that there is a point on its graph such that however we magnify its surroundings, we cannot see anything near a straight line. (Hint: look at the edge of your desk)

2 The first principle

If we want to see the behaviour of a function $g(x)$ when a, b are really close, we take the limit

$$\lim_{b \to a} g(x)$$

Similarly, if we want to see the linear approximation of a function around a, we can consider

$$\lim_{b \to a} \frac{f(b) - f(a)}{b - a}$$

or

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

By the substitution $h = b - a$. Of course, this limit may or may not exist, but it does exist almost anywhere in almost all the functions ordinary people [3] will ever encounter.

Definition 2.1. If the above limit exists, we say that f is *differentiable* at a. If f is differentiable in a subset S of the domain, the function $f' : S \to \mathbb{R}$ defined by

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

is called the *derivative* of f.

Remark 2.2. The author prefer to denote the derivative of $y = f(x)$ as $f'(x)$, but there is often other notations like $\dot{y}$ or $\frac{\mathrm{d}y}{\mathrm{d}x}$. The difference of notations results from the rather complex history of the development of Calculus.

[2] Differentiable. Will cover later

[3] This excludes mathematicians, of course.

We certainly have the freedom to attempt evaluating the limit for some functions we like, and it is highly encouraged that you attempt evaluating this limit for some of your favourite functions.

Example 2.1. *We take $f(x) = x^2$, so*

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{2xh + h^2}{h} = \lim_{h\to 0}(2x+h) = 2x$$

Therefore,

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

Remark 2.3. When attempting evaluating a derivative, it is better NOT to write, in the beginning, something like

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

because we DO NOT YET know if the limit really exists. So one should attempt evaluating the limit first before concluding the derivative of the function.

Example 2.2. *We can also find the derivative of a function at some specific point. Take the function $f(x) = \exp(x) = e^x$, if we want to find its derivative at $x = 0$, we do*

$$\lim_{h\to 0} \frac{e^{0+h} - e^0}{h} = \lim_{h\to 0} \frac{e^h - 1}{h} = 1$$

So $f'(0) = 1$. [4]

Suppose $f(x) = mx+c$ for some constants m and c, then one may verify that f' is exactly the constant function m. What this gives us is exactly what we expect to be the geometrical meaning of a derivative: the slope of the tangent of a (differentiable) curve!

Exercises

1. Find the derivative of $f(x) = 1/x$ from first principle.
2. Find the derivative of $f(x) = \sqrt{x}$ from first principle.
3. Find the derivative of $f(x) = e^{ax}$ for a real number a.
4. (a) By using the sum-to-product formula, find the derivative of $f(x) = \sin(x)$ from first principle.
(b) Without evaluating the limit in the first principle, find the derivative of $f(x) = \cos(x)$ by using (a).
5. Suppose f is a differentiable function. Find the derivative of $g(x) = cf(x)$ where c is a constant in terms of $f'(x)$.

[4] Actually, the derivative of e^x is e^x itself, which you may wish to verify. And a more interesting thing is that the *only* functions whose derivatives are themselves are in the form $c \cdot e^x$ where c is a constant.

Supplementary Information

There is yet another intepretation of the derivative notion. If we want to directly conduct a linear approximation of the function near a point x, we are actually hoping to find a constant m such that $f(x)$ *behaves like* $mx+c$ for some onstant c. In other words, we want

$$f(x+h) = f(x) + mh + o(h)$$

for some small h. What is $o(h)$? It is what we call an *error term.* Since we can only *approximate* the function by a linear tangent, there should normally be some difference between this linear function and the original function. [5] We express this difference as $o(h)$. But there should be restrictions on the magnitude of $o(h)$ to produce a nice approximation, or m could just be any real number. Now, since we want the function to resemble that linear function as close as we please when h is small enough, we will want $o(h)$ to go to 0 faster than h does. In the language of limit, this means that

$$\lim_{h\to 0} \frac{o(h)}{h} = 0$$

So, if we rearrange the equation, we will find that

$$\frac{f(x+h) - f(x)}{h} = m + \frac{o(h)}{h} \to m \text{ as } h \to 0$$

This gives, again, the first principle. One may check that the existence of such an m is equivalent to the differentiability of f at point x.

3 Differentiating Sums, Products and Quotients

From this section on, we are going to introduce some rules that we can use to differentiate a function more easily. [6] We already know that we can take a multiplicative constant out of the differentiation sign from an exercise in the preceding section, but we can actually do something even more useful.
The first things that we will look at will be the derivatives of the sums, products and quotients (or fractions) of differentiable functions.
Now, suppose f and g are differentiable, then

$$\begin{aligned}
&\lim_{h\to 0} \frac{f(x+h) + g(x+h) - (f(x) + g(x))}{h} \\
=&\lim_{h\to 0} (\frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h}) \\
=&\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \\
=&f'(x) + g'(x)
\end{aligned}$$

Therefore, the function $f+g$ is differentiable as well, and its derivative is $f'+g'$.

[5] N.B. $o(h)$ depends on h. Also, for computer scientists, this $o(h)$ is exactly the little-o notation that you use when bounding time complexities

[6] Author of some Calculus testbook prefer the wording: "Construct more differentiable functions from existing ones".

Example 3.1. *Let $f(x) = x^2+3x$. We have already knwon how to differentiate x^2 and $3x$ alone respectively from the preceding section, so by what we have already got,*

$$f'(x) = (2x) + (3) = 2x + 3$$

We can do the same thing for products, of course. We of course start with our beloved first principle:

$$\begin{aligned}
&\lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
=&\lim_{h\to 0} f(x+h)\frac{g(x+h)-g(x)}{h} + \frac{f(x+h)-f(x)}{h}g(x) \\
=&\lim_{h\to 0} f(x+h) \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} + \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \lim_{h\to 0} g(x) \\
=&f(x)g'(x) + f'(x)g(x)
\end{aligned}$$

So $f \cdot g$ is differentiable and its derivative is $f \cdot g' + f' \cdot g$.

Example 3.2. *We now try to attempt to differentiate $f(x) = x^2e^x$. We know the derivatives of x^2 and e^x are $2x$ and e^x respectively, so*

$$f'(x) = (2x)e^x + x^2(e^x) = 2xe^x + x^2e^x$$

After sums and products, we arrive at quotients (i.e. fractions). However, in this case, the resulting rule is not as neat as the sum and product case. There are essentially many ways of getting the quotient rule. The simpliest of which is to think of f/g as $f \cdot 1/g$ and use the product rule to proceed. However, we have not covered the ways to differentiate the reciprocal of a function [7] yet. Another nice way to do it is to use implicit differentiation, which you will find in a later section. Here, we shall demonstrate the rule directly from first principle here.

$$\begin{aligned}
&\lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\
=&\lim_{h\to 0} \frac{1}{g(x+h)g(x)} \frac{f(x+h)g(x) - f(x)g(x+h)}{h} \\
=&\lim_{h\to 0} \frac{1}{g(x+h)g(x)} \left(\frac{f(x+h)-f(x)}{h}g(x) - f(x)\frac{g(x+h)-g(x)}{h} \right) \\
=&\frac{\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}g(x) - f(x)\lim_{h\to 0} \frac{g(x+h)-g(x)}{h}}{\lim_{h\to 0} g(x+h)g(x)} \\
=&\frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}
\end{aligned}$$

The quotient rule might not be as easy as the two preceding rules for memorization, but it is still essential. Just bear in mind the square in the denominator, minus sign in the nominator and the fact that f' goes before g'.

Theorem 3.1. *Let f and g be differentiable functions,*
1. $(f + g)' = f' + g'$
2. $(fg)' = f'g + fg'$
3. $(f/g)' = (f'g - fg')/(g^2)$

[7] You may consult the session about the chain rule later. It is, however, also possible to deduce it from first principle.

These three rules will be used quite frequently, especially in practical problems and examinations. So please do get yourself familiar with these before proceeding to the next sections.

Exercises

The following exercises may require knowledge that you can obtain from those in previous section(s).
1. Differentiate
$$\frac{\sqrt{x}}{e^x}$$
2. (a) Differentiate x^3 and x^4 by using the derivatives of x and x^2.
(b) Prove by mathematical induction that
$$\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}$$
for any positive integer n.
3. (a) Let f be a differentiable function. Find the derivative of $1/f$ from the first principle.
(b) Hence, prove the quotient rule from product rule.

Supplementary Information

The product rule is sometimes called *Leibniz' Rule*, named after Gottfried Wilhelm Leibniz, one of the inventors of Calculus. One of the reasons for it to have such a polished name is that, other than the usual way of multiplication, it can also be applied to other sort of multiplications.
Let $\langle x, y\rangle$ be a function which maps each pair of elements of a normed real vector space V to another normed real vector space W. [8] If it has the following properties:
(1) Bilinearity. I.e. $\langle ax + bz, y\rangle = a\langle x, y\rangle + b\langle z, y\rangle$ and $\langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle$ for any $a, b \in \mathbb{R}$ and $x, y, z \in V$.
(2) Continuity under the induced topology (by their norms).
Then, for two differentiable functions $f, g : \mathbb{R} \to V$, [9] then
$$\frac{\mathrm{d}}{\mathrm{d}x}\langle f(x), g(x)\rangle = \langle f(x), g'(x)\rangle + \langle f'(x), g(x)\rangle$$
Note that the commutative law, i.e. $\langle a, b\rangle = \langle b, a\rangle$, needs NOT hold in this case.

Proof. We can only give a somewhat vague proof here, since we have not rigorously covered the notion of continuity and limit in (normed) real vector spaces. But vaguely speaking, if f is a differentiable function from the set of real numbers to a (normed) real vector space, then
$$f'(x) = \lim_{h\to 0}\frac{f(x+h) - f(x)}{h}$$

[8] We assumed that the addition and scalar multiplication are already continuous under the respective induced topology of V and W.
[9] the domain of f can actually be extended to any normed real vector space as well, but that is far far beyond our scope

So we have

$$
\begin{aligned}
&\lim_{h\to 0}\frac{\langle f(x+h), g(x+h)\rangle - \langle f(x), g(x)\rangle}{h} \\
=&\lim_{h\to 0}\left\langle f(x+h), \frac{g(x+h)-g(x)}{h}\right\rangle + \lim_{h\to 0}\left\langle \frac{f(x+h)-f(x)}{h}, g(x)\right\rangle && \text{(Bilinearity)} \\
=&\left\langle \lim_{h\to 0} f(x+h), \lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\right\rangle + \left\langle \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}, g(x)\right\rangle && \text{(Continuity)} \\
=&\langle f(x), g'(x)\rangle + \langle f'(x), g(x)\rangle
\end{aligned}
$$

As desired. ☐

You may want to try writing the respective product rules for dot product, cross product and matrix product after you have learnt what they are when studying vectors and matrices.

4 Derivatives of Integral Powers

The most beloved family of functions for an algebraist is polynomials. It is quite right to say that polynomials are indeed the most important set of functions mathematicians study throughout the history of mathematics. So it seems only sensible that we would want to study their properties by every tool that we het in hand. In this case, that tool would be differentiation.
From the exercise from the previous section, you should have known

Theorem 4.1. *Let n be a positive integer, we have*

$$\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}$$

by using induction. There is another proof which is somewhat neater.

Proof. First principle always does the trick.

$$
\begin{aligned}
&\lim_{h\to 0}\frac{(x+h)^n - x^n}{h} \\
=&\lim_{h\to 0}\frac{1}{h}\sum_{m=1}^{n}\binom{n}{m}x^{n-m}h^m \\
=&\lim_{h\to 0}\sum_{m=1}^{n}\binom{n}{m}x^{n-m}h^{m-1} \\
=&\binom{n}{1}x^{n-1} = nx^{n-1}
\end{aligned}
$$

As desired. ☐

Our existing knowledge about evaluating the derivative of sums then tells us that

Theorem 4.2.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(a_0 + \sum_{i=1}^{n} a_i x^i\right) = \sum_{i=1}^{n} i a_i x^{i-1}$$

Corollary 4.3. *A non-constant polynomial's degree decreases by* 1 *upon differentiation.*

Obviously, by making use of the quotient rule, we can find our way through negative integer powers. In fact, if $f(x) = x^{-n}$ where n is a positive integer, we have

$$f'(x) = \frac{0 - nx^{n-1}}{x^{2n}} = (-n)x^{-n-1}$$

Combining with the fact that the constant function has a constantly zero derivative, we can see that

Theorem 4.4. *Let $f(x) = x^m$ where m is any integer, then $f'(x) = mx^{m-1}$.*

In the next section, we shall show that the above theorem holds for any rational powers as well.

Supplementary Information

The derivative of polynomials is not merely a tool we used to analysie the groowth of polynomial functions, but also their algebraic properties.
For example,

Definition 4.5. Let r be a real number and $P(x)$ a nonconstant polynomial. If r is a root of $P(x)$, then it is said to be with multiplicity m if and only if

$$P(x) = (x - r)^m Q(x)$$

where $Q(r) \neq 0$ (that is, r is not a root of Q).
Otherwise, we define the multiplicity of r in P to be 0.

Remark 4.6. It is obvious that the multiplicity is well-defined. It is, indeed, the minimal m such that $(x - r)^m$ divides P.

Theorem 4.7. *Let $P(x)$ be a nonconstant polynomial and r a root of it with multiplicity n. Then the multiplicity of r in $P'(x)$ is $n - 1$.*

Proof. Suppose that $P(x) = (x - r)^n Q(x)$, then

$$\begin{aligned} P'(x) =& n(x - r)^{n-1} Q(x) + (x - r)^n Q'(x) \\ =& (x - r)^{n-1}(nQ(x) + (x - r)Q'(x)) \end{aligned}$$

Now, since $Q(r) \neq 0$, r is not a root of $nQ(x) + (x - r)Q'(x)$ (simply plug it in). This means that r has multiplicity $n - 1$ in $P'(x)$. □

There are many other interesting algebraic or even number theorectical results relevant to the derivative of a polynomial. The Hensel's lemma, for example, might be one of the most fundamental results in building the p-adic number theory.

5 The Chain Rule and Derivative of Rational Powers

We begin the investigation of the chain rule by some elementary observations. There was an exercise in a previous section asking for the derivative of the reciprocal of a function from first principle. If you attempted that one, [10] you would remember that we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{f(x)} = -\frac{f'(x)}{f(x)^2}$$

Now, there is something strange here: if we substitute $y = f(x)$, the formula is essentially

$$-\frac{1}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x}$$

Let the derivative of y alone first, the rest is not something that we see every day. The only recent case where we get acquainted with something looks like $-1/y^2$ is when evaluating

$$\frac{\mathrm{d}}{\mathrm{d}y}\frac{1}{y} = -\frac{1}{y^2}$$

But, hey, $1/y$ is exactly $1/f(x)$, the function that we intended to differentiate at first! There is something fishy here, is it not?
If we write $g = 1/f$, then we arrive at

$$\frac{\mathrm{d}g}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x}$$

This is a nice formula indeed. The "dy"s "cancelled out" each other and greatly simplifies our evaluation. Clearly, we cannot *really* cancel out the d-stuff, because they are not real things. [11] However, that does not mean that we cannot prove it.
Now suppose that $g(x) = f(s(x))$, where f and s are both differentiable. Then, from first principle.

$$\begin{aligned}
&\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\\
=&\lim_{h\to 0}\frac{f(s(x+h))-f(s(x))}{h}\\
=&\lim_{h\to 0}\frac{f(s(x+h))-f(s(x))}{s(x+h)-s(x)}\frac{s(x+h)-s(x)}{h}
\end{aligned}$$

Since s is differentiable, when $h \to 0$, $s(x+h)-s(x) \to 0$ (otherwise its derivative would not exist). Hence

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

[10] If you did not, feel free to deduce this from the quotient rule.
[11] Well, not yet. In some contexts, we have similar things like that, called the *differential forms*, but the theory of that would not be relevant to you if you have not gone through a minimum of two years' study in Maths Major.

So

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

This means that g is differentiable and that its derivative is $f'(s(x))s'(x)$. One may check that this exactly what we have claimed. Conventionally, we denote g by $f \circ s$

Theorem 5.1. *If f and g are both differentiable, and the domain of f contains the image of g (just to ensure that $f(g(x))$ is always well-defined), then $f \circ g$ is differentible and its derivative is*

$$f'(g(x))g'(x)$$

or

$$\frac{\mathrm{d}(f\circ g)}{\mathrm{d}x} = \frac{\mathrm{d}(f\circ g)}{\mathrm{d}g}\frac{\mathrm{d}g}{\mathrm{d}x}$$

To avoid some common confusion, the author recommend the first form of the chain rule for all candidates.
We now can do some interesting stuff with the chain rule. A common application of it would be

Example 5.1. *Let $y = e^{x^2+2x}$. By the chain rule (setting $f(x) = e^x$ and $g(x) = x^2 + 2x$), we have*

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(g(x))g'(x) = e^{x^2+2x}(2x+2)$$

But not just that. Using the chain rule, we can also find the derivative of any rational power.

Theorem 5.2. *Let $f(x) = x^r$ where $r = p/q$ for some integers p, q, then $f'(x) = rx^{r-1}$.*

Proof. Let $h(x) = x^q$. So

$$px^{p-1} = (h\circ f)' = h'(f(x))f'(x) = qx^{r(q-1)}f'(x)$$

Rearrange to give

$$f'(x) = \frac{p}{q}x^{p-1-(p/q)(q-1)} = \frac{p}{q}x^{p/q-1} = rx^{r-1}$$

As desired. □

Note that in the above arguement we implicitly assumed the differentiability of f, which, though obvious, would make a nice exercise. A warm reminder is that you are making fool of yourself [12] if you directly attempt the first principle on f.
Now, we have extended our theorem to all rational powers, but how about irrational powers? If you possess the quality of mathematical rigour, you would

[12] It works, but if you do that, why bother the nice proof above?

have realized that we have not given a definition of irrational powers yet. Intuitively, for an irrational number $\mathfrak{I}$, if we find a sequence of rational numbers $\mathfrak{r}_n$ which converges to $\mathfrak{I}$, then we may with ton define

$$x^{\mathfrak{I}} = \lim_{n\to\infty} x^{\mathfrak{r}_n}$$

But there are two problems here:
Firstly, does the limit exist? This can be easily answered: yes. One may attempt that using $\epsilon - \delta$ or get an intuition that if rational numbers p, q are really close, so should be x^p and x^q, so the sequence $a_n = x^{\mathfrak{r}_n}$ intuitively converges.
Secondly, how do we know that the value of $x^{\mathfrak{I}}$ is independent of the choice of sequence $\mathfrak{r}_n$? There can be many totally different sequence that converges to the same value. How do we know that their difference does not affect the limit after we have put them on the exponent?
Even if we can answer the second question as well, how could we differentiate it given such conditions? These will be answered in the next section, along with the rigourous definition of irrational powers.

6 Derivatives of Logarithm and Irrational Powers

We already know that the derivative of the exponential function is itself. That is

$$\frac{\mathrm{d}e^x}{\mathrm{d}x} = e^x$$

It is natual for us to enquire the case for the opposite: the logarithm.
There is nothing very polished here, just the usual first principle.

$$\begin{aligned}
\lim_{h\to 0} \frac{\ln(x+h) - \ln(x)}{h} &= \lim_{h\to 0} \ln\left(\left(1 + \frac{1}{x}h\right)^{1/h}\right) \\
&= \ln\left(\lim_{h\to 0}\left(1 + \frac{1}{x}h\right)^{1/h}\right) \\
&= \ln(e^{1/x}) \\
&= \frac{1}{x}
\end{aligned}$$

Therefore the logarithm function is differentiable and its derivative is $1/x$.
Using the logarithmatic function, we could finally give a more genuine definition of irrational powers (or, in general, real powers). Observe first that if r is any rational number then

$$x^r = e^{r \ln x}$$

What is the advantage of formulating it like this? Well, one may notice that r is now the coefficient of $\ln x$ on the exponent, so the expression makes sense if we substitute r by simply any real number. The core reason why this can happen is because we have defined the exponentiation of the natural base in a way general enough to include all real numbers. So what we have tried to do is to let the exponentiation of any real number inherit this nice property of our

beloved e.
Of course, an expression that makes sense (well-defined) cannot let us to accept it as a notion. There are more important properties to be checked. One may attempt oneself that if we define a real power by the mannar

$$x^y = e^{y \ln x}$$

then the set of rules of exponentiation hold still. A more important property stated at the end of the last section can also be checked. For any sequence $\mathfrak{r}_n$ converging to a real number $\mathfrak{I}$, then the continuity of e^x and the multiplication operation guarantee that

$$\lim_{n\to\infty} x^{\mathfrak{r}_n} = \lim_{n\to\infty} e^{\mathfrak{r}_n \ln x} = e^{\ln x \lim_{n\to\infty} \mathfrak{r}_n} = e^{\mathfrak{I} \ln x} = x^{\mathfrak{I}}$$

So this way of defining the general power gives it properties that perfectly resemble what we expect it to have.
Now, how about its derivative?

Theorem 6.1. *Let y be any real number, we have*

$$\frac{\mathrm{d}x^y}{\mathrm{d}x} = yx^{y-1}$$

Proof. The chain rule works.

$$\frac{\mathrm{d}x^y}{\mathrm{d}x} = e^{y \ln x}\frac{y}{x} = yx^{y-1}$$

As desired. □

Our chasing of the derivaive of powers finally ends. But this is hardly the end of story. Many other sorts of functions exist. Trigonometric functions, for example, deserves attention. This introduces us the next section.

Supplementary Information

We now know how to differentiate any real powers, and they all have the same form. An interesting question, then, is if we can extend our journey to the field of complex numbers, $\mathbb{C}$.
The answer, however, is not promising. We can't. Why? Well, firstly the logarithmatic function might not be in a form which you will like it. Euler's theorem tells us that

$$z = |z|e^{i \arg z}$$

So in order to let the logarithmatic function makes sence, we must have

$$\ln z = \ln |z| + i \arg z$$

Can't see any problem here? Take a look at the latter part of the expression: $\arg z$. $\arg z$ is defined in a way that it only spans over a range of 2π, otherwise it will be ill-defined. But it simply does not work. Choose two complex numbers z, w such that the sum of their arguments is not in your chosen interval of length

2π. We can obviouly choose them no matter which interval we chose for the arg function. But then

$$\ln|w| + \ln|z| + i(\arg w + \arg z) = \ln w + \ln z = \ln wz = \ln|w| + \ln|z| + i\arg wz$$

But this is false: $\arg w + \arg z$ is simply not in the image of arg! So we cannot let the logarithmatic function have the same crucial properties that we will expect them to inherit from the real number case. A way this goes ill-defined is the evaluation of i^i.
But is there a way to well-define these powers? The original way does not work, but is there any other way that works? In fact, there is no way that well-defines complex exponentiation in the way we wish, which you may want to find out why yourself.

7 Derivatives of trigonometric functions

You have found that the derivative of $\sin(x)$ is $\cos(x)$ from a previous exercise. From the very same exercise which you are told so, you should also have proved that the derivative of $\cos(x)$ is $-\sin(x)$. The argument, not using the first principle or any other rule, is good to learn, so I am not spoiling your fun. The standard way of differentiating $\cos(x)$ is just sum-to-product all over again. Since it is not in any exercise (yet), I will demonstrate here:

$$\begin{aligned}\lim_{h\to 0}\frac{\cos(x+h)-\cos(x)}{h} &= \lim_{h\to 0}\frac{-2\sin(x+\frac{h}{2})\sin(\frac{h}{2})}{h}\\ &= \lim_{h\to 0} -\sin(x+\frac{h}{2})\frac{\sin(\frac{h}{2})}{\frac{h}{2}}\\ &= -\sin(x)\end{aligned}$$

Virtually all trigonometric functions are built from sines and cosines. So you will be expected to prove the following:

Theorem 7.1. *All trigonometric functions are differentiable in their respective domains. Plus,*

$$\frac{\mathrm{d}\sin(x)}{\mathrm{d}x} = \cos(x) \tag{1}$$
$$\frac{\mathrm{d}\cos(x)}{\mathrm{d}x} = -\sin(x) \tag{2}$$
$$\frac{\mathrm{d}\tan(x)}{\mathrm{d}x} = \sec^2(x) \tag{3}$$
$$\frac{\mathrm{d}\sec(x)}{\mathrm{d}x} = \sec(x)\tan(x) \tag{4}$$
$$\frac{\mathrm{d}\cot(x)}{\mathrm{d}x} = -\csc^2(x) \tag{5}$$
$$\frac{\mathrm{d}\csc(x)}{\mathrm{d}x} = -\csc(x)\cot(x) \tag{6}$$

From here on we can find the derivative of some rather strange functions by a combination of Theorem 7.1 and the chain rule.

Example 7.1. *Let* $f(x) = \tan(\ln(x))$*, we have* $f'(x) = \sec^2(\ln(x))/x$*.*

Example 7.2. *Let* $f(x) = \sin(\sin(x))\sin(x)$*, we have*

$$f'(x) = \sin(\sin(x))\cos(x) + \cos(\sin(x))\cos(x)\sin(x)$$

Exercises

1.* Prove (3) till (6) of Theorem 7.1 using quotient rule.
2.* Prove (3) till (6) of Theorem 7.1 using first principle.
3. Let $f(x) = \sin(x^2\cos(e^x))$, find $f'(x)$.

8 Implicit Differentiation

It is really not a very hard thing to believe, but I am not sure why people are buffled by it in their first sight of the idea. Therefore, I am going to say it here in case some people are confused by it: If two functions are identical, so are their derivatives.

Theorem 8.1. *If* $f = g$*, then* $f' = g'$*.*

Proof. Really? You need me to prove it? □

If it is really so trivial, why do we state it? One, somewhat obvious, reason, is that not all curves can be written in the form $y = f(x)$. Some curves simply cannot be described by a function. The circle being a noticable example. In this case, we cannot directly use first principle, our rules of differentiation, etc., but we must harness other tricks to obtain the formula of $\mathrm{d}y/\mathrm{d}x$ in terms of x and y. [13] This is how implicit differentiation comes in sight: We do want to know the slope of tangent to some point on a curve even if it cannot be described by a function.

Example 8.1. *The unit circle is the implicit curve*

$$x^2 + y^2 = 1$$

If we want to find its slope at a point (x, y) *on it, we can do this:*

$$\begin{aligned} x^2 + y^2 &= 1 \\ 2x + 2y\frac{\mathrm{d}y}{\mathrm{d}x} &= 0 \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= -\frac{x}{y} \end{aligned}$$

So, for example, if we want the slope of tangent at $(0.6, 0.8)$ *(please verify that this point is on the curve), we plug it in and find that it is* -0.75*. You may wish to verify that this is true.*

[13] This is important: Unlike the case of explicit curves, if we want to locate a point on an implicit curve, we almost surely need both the x and y coordinate of it.

Another way of using implicit differentiation is to evaluate derivatives of some bazzire functions.

Example 8.2. *We have looked into the derivative of any real power, but how about the power of itself? If $f(x) = x^x$, how do we find x^x? One* [14] *way is to use implicit differentiation.*

$$\begin{aligned} y =& x^x \\ \ln y =& x \ln x \\ \frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} =& \ln x + 1 \\ \frac{\mathrm{d}y}{\mathrm{d}x} =& y(\ln x + 1) \\ =& x^x \ln x + x^x \end{aligned}$$

Please look into the above example thoroughly and get familiar with this general idea: We try to simplify the aimed function f by another (simple) function g, which is the natual logarithm in this case. We want to do it in such a way that we know how to differentiate both $g(f(x))$ and $g(x)$. Assuming that we can, indeed, find such a function g, then we do the above procedure over again.

$$\begin{aligned} y =& f(x) \\ g(y) =& g(f(x)) \\ g'(y)\frac{\mathrm{d}y}{\mathrm{d}x} =& (g \circ f)'(x) \\ \frac{\mathrm{d}y}{\mathrm{d}x} =& \frac{(g \circ f)'(x)}{g'(y)} \\ =& \frac{(g \circ f)'(x)}{g'(f(x))} \end{aligned}$$

Of course, no one will be asking you to memorize the last formula, but you must get a gist of the idea. Now here is another example.

Example 8.3. *We take $f(x) = \arcsin(x) = \sin^{-1}(x)$ where $f : [0,1] \to [0, \pi/2]$. Following the above idea, we (secretly) take $g(x) = \sin(x)$, and this gives us*

$$\begin{aligned} y =& f(x) = \arcsin(x) \\ \sin y =& x \\ \cos y \frac{\mathrm{d}y}{\mathrm{d}x} =& 1 \\ \frac{\mathrm{d}y}{\mathrm{d}x} =& \frac{1}{\cos(\arcsin x)} \\ =& \frac{1}{\sqrt{1 - x^2}} \end{aligned}$$

The last equality is derived from the identity $\sin^2 + \cos^2 = 1$.

[14] There is another way. See exercises.

Exercises

1.* Actually, we do not have to use implicit differentiation to find the derivative of x^x. Please find its derivative by using the identity $x^y = e^{y \ln x}$.
2. Find the slope of the tangent to the curve defined by $ye^y = \ln x$ at $(1, 0)$.

9 Higher derivatives

The derivative of a function is a function. You already know that. So it comes right in mind this question: What happens if we differentiate a function over and over again?

Definition 9.1. Let f be a function, we define the n^{th} derivative of it inductively as follows:
(1) f' is the first derivative of f. [15]
(2) The derivative of the n^{th} derivative of f is the $(n+1)^{th}$ derivative of f.

Now, what are the meanings of higher derivatives? As derivatives identify the linear approximation of a function, if we take the argument of that function moving along the real line, it essentially gives the rate of change of that function. So, state it in a clumsy way, the second derivative is the rate of change of the rate of change of the function, and the n^{th} derivative is

$$\text{(the rate of change of)}^n\text{the function}$$

If that is not intuitive enough, recall what you have learnt from your physics lesson. You should have noted that if we let $s(t)$ to be the displacement function (on a straight line, if you are not familiar with higer dimensions), then s' is the velocity. Now, the second derivative of s, s'' is just the acceleration.
There is not very much additional skill to know to calculate the higher derivatives. Just rolling the process again and again is fine. But it is hugely important to know the concept.

A Credits

This work is part of the `5-triple-star` project. Below are the authors, maintainers and other contributers of this set of notes.

A.1 Authors

GitHub user @fivetriplestar is the initial author of this set of notes.

A.2 Maintainers

GitHub user @fivetriplestar has maintained this set of notes from June 2, 2019 to July 21, 2019.

[15] Or sometimes, f is the 0^{th} derivative of f itself.

B Answers to exercises

B.1 Section 1

1. Consider the bahaviour of the function $f(x) = |x|$ at $x = 0$.

B.2 Section 2

1.

$$\begin{aligned}\lim_{h\to 0}\frac{\frac{1}{x+h}-\frac{1}{x}}{h} &= \lim_{h\to 0}\frac{-h}{hx(x+h)}\\ &= \lim_{h\to 0} -\frac{1}{x(x+h)}\\ &= -\frac{1}{x^2}\end{aligned}$$

Therefore,

$$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = -\frac{1}{x^2}$$

2.

$$\begin{aligned}\lim_{h\to 0}\frac{\sqrt{x+h}-\sqrt{x}}{h} &= \lim_{h\to 0}\frac{h}{h\sqrt{x}+\sqrt{x+h}}\\ &= \lim_{h\to 0}\frac{1}{\sqrt{x}+\sqrt{x+h}}\\ &= \frac{1}{2\sqrt{x}}\end{aligned}$$

Therefore,

$$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \frac{1}{2\sqrt{x}}$$

3.

$$\begin{aligned}\lim_{h\to 0}\frac{e^{ax+ah}-e^{ax}}{h} &= \lim_{h\to 0} ae^{ax}\frac{e^{ah}-1}{ah}\\ &= (ae^{ax})(1)\\ &= ae^{ax}\end{aligned}$$

Therefore,

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

4. (a)

$$\begin{aligned}\lim_{h\to 0}\frac{\sin(x+h)-\sin(x)}{h} &= \lim_{h\to 0}\frac{2\cos(x+h/2)\sin(h/2)}{h}\\ &= \lim_{h\to 0}\cos(x+h/2)\frac{\sin(h/2)}{h/2}\\ &= \cos(x)\end{aligned}$$

Therefore,

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

(b)

$$\begin{aligned}
\lim_{h\to 0} \frac{\cos(x+h) - \cos(x)}{h} &= - \lim_{h\to 0} \frac{\sin(x - \pi/2 + h) - \sin(x - \pi/2)}{h} \\
&= \left.\frac{\mathrm{d}\sin(z)}{\mathrm{d}z}\right|_{z = x - \pi/2} \\
&= - \cos(x - \pi/2) \\
&= - \sin(x)
\end{aligned}$$

Therefore,

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

5.

$$\begin{aligned}
\lim_{h\to 0} \frac{g(x+h) - g(x)}{h} =& \lim_{h\to 0} \frac{cf(x+h) - cf(x)}{h} \\
=& c \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
=& cf'(x)
\end{aligned}$$

Therefore,

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

B.3 Section 3

1.

$$\frac{\mathrm{d}}{\mathrm{d}x} \frac{\sqrt{x}}{e^x} = \frac{\frac{1}{2\sqrt{x}} e^x - \sqrt{x} e^x}{e^{2x}} = \frac{1 - 2x}{2\sqrt{x} e^x}$$

2. (a)

$$\frac{\mathrm{d}}{\mathrm{d}x} x^3 = (1)x^2 + x(2x) = 3x^2$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^4 = (1)x^3 + x(3x^2) = 4x^3$$

(b) Let $P(n)$ be the proposition. When $n = 1$, L.H.S. = 1 = R.H.S., so $P(1)$ is true. Now assume that $P(k)$ is true for some positive integer k, then when $n = k + 1$,

$$\begin{aligned}
\text{L.H.S.} =& \frac{\mathrm{d}}{\mathrm{d}x} x^{k+1} \\
=& \left(\frac{\mathrm{d}}{\mathrm{d}x} x\right) x^k + x \left(\frac{\mathrm{d}}{\mathrm{d}x} x^k\right) \\
=& x^k + x(kx^{k-1}) \\
=& (k+1)x^k = \text{R.H.S.}
\end{aligned}$$

So $P(k+1)$ is true. By the principle of mathematical induction, $P(n)$ is true for all positive integer n.
3. (a)

$$\begin{aligned}\lim_{h\to 0}\frac{\frac{1}{f(x+h)}-\frac{1}{f(x)}}{h} &= -\lim_{h\to 0}\frac{1}{f(x)f(x+h)}\frac{f(x+h)-f(x)}{h}\\ &= -\frac{f'(x)}{f(x)^2}\end{aligned}$$

(b)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\frac{f(x)}{g(x)} =& \frac{\mathrm{d}}{\mathrm{d}x}f(x)\frac{1}{g(x)}\\ =& \left(\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right)\frac{1}{g(x)}+f(x)\left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{g(x)}\right)\\ =& \frac{f'(x)}{g(x)}-\frac{f(x)g'(x)}{g(x)^2}\\ =& \frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}\end{aligned}$$

B.4 Section 4

B.5 Section 5

B.6 Section 6

B.7 Section 7

1. (3)

$$\frac{\mathrm{d}}{\mathrm{d}x}\tan(x)=\frac{\mathrm{d}}{\mathrm{d}x}\frac{\sin(x)}{\cos(x)}=\frac{\sin^2(x)+\cos^2(x)}{\cos^2(x)}=\frac{1}{\cos^2(x)}=\sec^2(x)$$

(4)

$$\frac{\mathrm{d}}{\mathrm{d}x}\sec(x)=\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\cos(x)}=-\frac{1}{\cos^2(x)}(-\sin(x))=\sec(x)\tan(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\cot(x)=\frac{\mathrm{d}}{\mathrm{d}x}\frac{\cos(x)}{\sin(x)}=\frac{-\sin^2(x)-\cos^2(x)}{\sin^2(x)}=-\frac{1}{\sin^2(x)}=-\csc^2(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\csc(x)=\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sin(x)}=-\frac{1}{\sin^2(x)}(\cos(x))=-\csc(x)\cot(x)$$

2. (3)

$$\begin{aligned}\lim_{h\to 0}\frac{\tan(x+h)-\tan(x)}{h} =& \lim_{h\to 0}\frac{\sin(x+h)\cos(x)-\cos(x+h)\sin(x)}{h\cos(x)\cos(x+h)}\\ =& \lim_{h\to 0}\frac{\sin(h)}{h}\frac{1}{\cos(x)\cos(x+h)}\\ =& \sec^2(x)\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}\tan(x)}{\mathrm{d}x} = \sec^2(x)$$

(4)

$$\begin{aligned}
\lim_{h\to 0}\frac{\sec(x+h)-\sec(x)}{h} &= \lim_{h\to 0}\frac{\cos(x)-\cos(x+h)}{h\cos(x)\cos(x+h)} \\
&= \lim_{h\to 0}\frac{-2\sin(x+h/2)\sin(-h/2)}{h\cos(x)\cos(x+h)} \\
&= \lim_{h\to 0}\frac{\sin(x+h/2)}{\cos(x)\cos(x+h)}\frac{\sin(h/2)}{h/2} \\
&= \frac{\sin(x)}{\cos^2(x)} \\
&= \sec(x)\tan(x)
\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}\sec(x)}{\mathrm{d}x} = \sec(x)\tan(x)$$

(5)

$$\begin{aligned}
\lim_{h\to 0}\frac{\cot(x+h)-\cot(x)}{h} &= \lim_{h\to 0}\frac{\sin(x)\cos(x+h)-\cos(x)\sin(x+h)}{h\sin(x)\sin(x+h)} \\
&= \lim_{h\to 0}-\frac{\sin(h)}{h}\frac{1}{\sin(x)\sin(x+h)} \\
&= -\csc^2(x)
\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}\cot(x)}{\mathrm{d}x} = -\csc^2(x)$$

(6)

$$\begin{aligned}
\lim_{h\to 0}\frac{\csc(x+h)-\csc(x)}{h} &= \lim_{h\to 0}\frac{\sin(x)-\sin(x+h)}{h\sin(x)\sin(x+h)} \\
&= \lim_{h\to 0}\frac{2\cos(x+h/2)\sin(-h/2)}{h\sin(x)\sin(x+h)} \\
&= \lim_{h\to 0}-\frac{\cos(x+h/2)}{\sin(x)\sin(x+h)}\frac{\sin(h/2)}{h/2} \\
&= -\frac{\cos(x)}{\sin^2(x)} \\
&= -\csc(x)\cot(x)
\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}\csc(x)}{\mathrm{d}x} = -\csc(x)\cot(x)$$

3.

$$f'(x) = \cos(x^2\cos(e^x))(2x\cos(e^x) - x^2\sin(e^x)e^x)$$

B.8 Section 8

1. We have $x^x = e^{x \ln x}$, so

$$\frac{\mathrm{d}x^x}{\mathrm{d}x} = \frac{\mathrm{d}e^{x \ln x}}{\mathrm{d}x} = e^{x \ln x}(\ln x + x\frac{1}{x}) = x^x(\ln x + 1)$$

2.

$$\begin{aligned} ye^y &= \ln x \\ (ye^y + e^y)\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{1}{x} \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{1}{xye^y + xe^y} \end{aligned}$$

So when $x = 1, y = 0$, we have

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{(1,0)} = 1$$

B.9 Section 9