5 Differentiation

Differentiation is intended to give a measure of the rate of change of a function.

Suppose you are measuring the position $x(t)$ at time t of a particle moving a long a line. Measuring its position at times t_0 and t say, allows you to define an *average velocity* between t and t_0 as

$$\frac{x(t) - x(t_0)}{t - t_0}$$

If you make the difference $\Delta t = t - t_0$ smaller and smaller the quotient of the distance Δx travelled in Δt

$$\frac{\Delta x}{\Delta t}$$

approximate the “true” velocity at time t_0 better and better. In other words

$$v(t_0) = \lim_{t \to t_0} \frac{x(t) - x(t_0)}{t - t_0}$$

is the velocity at time t_0. It measures the “rate of change” of the distance travelled at a particular time. If you graph the position $x(t)$ over the time t, then $v(t_0)$ is the slope of the *tangent* to the graph of $x(t)$ at the point $(t_0, x(t_0))$.

5.1 Differentiable functions

5.1.1 Definition and examples

Definition.
Let I be an interval and f a function defined on I. We say f is **differentiable** at $x_0 \in I$ if
Equation 1

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists and is finite. In this case we denote this limit by $f'(x_0)$ and call it the **derivative** of f at x_0. You will often also find the notation $\frac{df}{dx}(x_0)$ or $\left.\frac{d}{dx}\right|_{x_0} f$ for the derivative of f at x_0. EOD.

Examples

1. Any *constant* function is differentiable everywhere it is defined, and its derivative is 0.
2. Any *linear* function is differentiable everywhere: $f(x) = mx + b$, then $f'(x_0) = m$.
3. It is a little more work is to show that *monomials* of the form $f(x) = x^n$ are differentiable everywhere.
   Indeed, consider
   $$\frac{x^n - x_0^n}{x - x_0} = x^{n-1} + x^{(n-2)}x_0 + \cdots + x x_0^{n-2} + x_0^{n-1} = \sum_{k=0}^{n-1} x^{(n-k-1)}\, x_0^k$$
   The right hand side is a sum of continuous functions, all defined at x_0, and we conclude
   $$\frac{dx^n}{dx}(x_0) = n x_0^{n-1}$$

4. The function $f(x) = \sin x$ is differentiable at $x_0 = 0$, and
$$f'(0) = \lim_{x\to 0} \frac{\sin x}{x} = 1$$
Indeed, by T4, we know that for any zero sequence $x_n \neq 0$, we have $\lim \frac{\sin x_n}{x_n} = 1$. But by the definition of limits this means $\sin'(0) = 1$.
5. The function $f(x) = |x|$, defined on all of $\mathbb{R}$, is differentiable everywhere except at $x_0 = 0$. Indeed, if $x_0 \neq 0$, it is easy to see that
$$f'(x_0) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$
But if $x_0 = 0$, then $\lim_{x\to 0} \frac{|x|}{x}$ does not exist (the quotient alternates between ± 1 around 0).
6. The function $f: (-\infty, 0] \to \mathbb{R}$, defined by $f(x) = |x|$ *is* differentiable at 0), as then
$$\lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0^-} \frac{|x|}{x} = -1$$
In that sense the *domain* of the function in question is important.

5.1.2 Equivalent definitions

There are several equivalent statements expressing that f is differentiable at x_0:

1. f is differentiable at x_0.
2. $f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ exists and is finite.
3. There is a number $c = f'(x_0)$ such that $f(x_0 + h) = f(x_0) + ch + r(h)$ and $\lim_{h\to 0} \frac{r(h)}{h} = 0$. Here $r(h)$ is the function $f(x_0 + h) - f(x_0) - ch$ (which as a function of h which is always defined on an interval containing 0).

The second statement is often convenient, and the last statement is important as it allows for a direct generalization to functions in more than one variable.

It is a good exercise to show that these three statements are indeed equivalent. For the second, note that if $x_n \neq x_0$ converges to x_0, then $h_n \coloneqq x_n - x_0 \neq 0$ converges to 0 (and vice versa). This shows that f is differentiable at x_0 with derivative $f'(x_0)$ if and only if $g(h) \coloneqq f(x_0 + h)$ is differentiable at $h_0 = 0$ with derivative $g'(0) = f'(x_0)$. If f is defined on an interval I containing x_0, g is defined on the interval $I' = -x_0 + I = \{h \in \mathbb{R} \mid h + x_0 \in I\}$.

For the third, note there was a typo in a previous version. $r(h)$ is defined as $f(x_0 + h) - f(x_0) - ch$ which is defined for every $c \in \mathbb{R}$. If and only if $f'(x_0)$ exists and is equal to c does $r(h)$ satisfy that $\frac{r(h)}{h} \to 0$ for $h \to 0$. Again $r(h)$ is defined on an interval containing 0 (the same interval on which $g(h)$ above is defined).

5.1.3 Differentiability and continuity

Lemma (Differentiable Means Continuous; DMC).
Let f be defined on an interval I and differentiable at $x_0 \in I$. Then f is continuous at x_0. EOL.

Proof. For the limit of $\frac{f(x)-f(x_0)}{x-x_0}$ for $x \to x_0$ to exist and be finite, it is necessary that $f(x) \to f(x_0)$. But that means f is continuous at x_0. QED.

Note the converse is false. For example, $|x|$ is a continuous function but not differentiable (see Example 5 in 5.1.1 above) at $x_0 = 0$. One can find examples of functions that are continuous everywhere and nowhere differentiable.

Before we study differentiable function in greater detail, we list some immediate consequences of the definition.

5.1.4 Some rules of differentiation

Lemma (Useful Differentiation Theorem; UDT)
Let f, g be functions defined on an interval I and differentiable at some $x_0 \in I$.

1. $f + g$ is differentiable at x_0 and $(f+g)'(x_0) = f'(x_0) + g'(x_0)$.
2. For any $c, d \in \mathbb{R}$, $cf + dg$ is differentiable at x_0 and $(cf+dg)'(x_0) = cf'(x_0) + dg'(x_0)$.
3. fg is differentiable at x_0 and $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$.
4. If $g(x) \neq 0$ for all $x \in I$, then $\frac{f}{g}$ is differentiable at x_0 and
$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$$

Part 2 is often called the **Product** or **Leibniz Rule**. Part 3 is usually referred to as the **Quotient Rule**.

Proof.

1. This is a consequence of the fact that the limit of a sum is the sum of limits (if they exist) (ULT1).
2. This is a consequence of ULT3.
3. By 5.1.2 above we may write
$$f(x_0+h) = f(x_0) + f'(x_0)h + r(h)$$
$$g(x_0+h) = g(x_0) + g'(x_0)h + s(h)$$
where $\frac{r(h)}{h}, \frac{s(h)}{h} \to 0$ as $h \to 0$. Then
$$f(x_0+h)g(x_0+h) = f(x_0)g(x_0) + (f'(x_0)g(x_0) + f(x_0)g'(x_0))h + t(h)$$
where $t(h) = f'(x_0)g'(x_0)h^2 + (f(x_0) + f'(x_0)h + r(h))s(h) + (g(x_0) + g'(x_0)h)r(h)$.
Then $\lim_{h\to 0}\frac{t(h)}{h} = 0$ because $\frac{r(h)}{h}, \frac{s(h)}{h} \to 0$ and the coefficients converge to finite numbers (namely $g(x_0)$ and $f(x_0)$ respectively).
4. We first treat the case $f = 1$. We must show that $\frac{1}{g}$ is differentiable at x_0. To this end consider
$$\frac{g(x)^{-1} - g(x_0)^{-1}}{x - x_0} = \frac{g(x_0) - g(x)}{g(x)g(x_0)(x-x_0)} = \frac{-1}{g(x)g(x_0)}\frac{g(x)-g(x_0)}{x-x_0} \to \frac{-1}{g(x_0)^2}g'(x_0)$$
for $x \to x_0$. This uses the fact that g is continuous at x_0 (5.1.3).
Now the general case follows from 3:
$$\left(\frac{f}{g}\right)'(x_0) = \left(f\left(\frac{1}{g}\right)\right)'(x_0) = \frac{f'(x_0)}{g(x_0)} + f(x_0)\left(-\frac{g'(x_0)}{g(x_0)^2}\right) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$$

QED.

The second statement of the proposition applies in particular in case one of the functions is constant.

Thus $(cf)'(x_0) = cf'(x_0)$.

This shows that the set $\mathcal{D}(I, x_0)$ of functions defined on I that are differentiable at x_0 form a **subspace** of the set $\mathcal{F}(I)$ of all functions on I.

So far, we discussed only derivatives at a single point.

Example

1. Any polynomial function is differentiable everywhere.
2. For any $n \in \mathbb{N}$, x^{-n} is differentiable everywhere and $(x^{-n})' = (-n)x^{-n-1}$.
3. Any rational function is differentiable everywhere in its domain.

5.1.5 Linearity of differentiation

Let I be an interval and $x_0 \in I$. If we define

$$\mathcal{D}(I, x_0) = \{\, f \in \mathcal{F}(I) \mid \exists f'(x_0)\}$$

then, $\mathcal{D}(I, x_0)$ is a subspace of $\mathcal{F}(I) = \{f: I \to \mathbb{R}\}$, as it is nonempty (it contains for example, the constant functions), and it is closed under addition and scalar multiplication (the multiplication of a function with a constant).

In fact, by 5.1.4, this means that $\mathcal{D}(I, x_0)$ is a *subalgebra* of $\mathcal{F}(I)$ (because it is also closed under the multiplication operation on $\mathcal{F}(I)$.

In addition, the maps $\mathcal{D}(I, x_0) \to \mathbb{R}$, defined by $f \mapsto f'(x_0)$, is a linear transformation (or, a *derivation*[1] of the algebra $\mathcal{D}(I, x_0)$).

5.1.6 Differentiability of some functions

To gain further experience with derivatives we consider the following statement:

Suppose f^2 is differentiable at x_0. Does it follow that f is differentiable at x_0?

The short answer is no: we have seen that $f(x) = |x|$ is not differentiable at 0, yet f^2 is.

So we modify the statement to

Suppose f^2 is differentiable at x_0 and f is continuous at x_0 with $f(x_0) \neq 0$. Then f is differentiable at x_0, and $f'(x_0) = \frac{(f^2)'(x_0)}{2f(x_0)}$.

Statement 5-1

To see this observe that $f(x)^2 - f(x_0)^2 = \big(f(x) + f(x_0)\big)\big(f(x) - f(x_0)\big)$. Then

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{\big(f^2(x) - f^2(x_0)\big)}{(x - x_0)(f(x) + f(x_0))}$$

which is defined if x is close enough to x_0 (such that $f(x) + f(x_0) \neq 0$). This works because f is

[1] A *derivation* of an algebra is a linear transformation that also satisfies the Leibniz Rule.

continuous and $f(x_0) \neq 0$. But the limit for $x \to x_0$ of the right hand side exists (again using the continuity of f and the fact that f^2 is differentiable at x_0).

Example

The function $f(x) = \sqrt{x}$ defined on $[0, \infty)$ is differentiable on $(0, \infty)$ and not differentiable at $x_0 = 0$ (which we will see below).

By the above, as $f^2 = x$ is differentiable, we conclude that whenever $f(x_0) \neq 0$, then $f'(x_0) = \frac{1}{2f(x_0)} = \frac{1}{2\sqrt{x_0}}$. Note this neatly fits into the scheme of computing derivatives of powers of x: $\left(x^{\frac{1}{2}}\right)' = \frac{1}{2}x^{-\frac{1}{2}}$. EOE.

This can be extended to all functions of the form $g(x) = x^{\frac{1}{p}}$ (for $x > 0$).

Indeed, for any function g we have

$$g^p(x) - g^p(x_0) = (g(x) - g(x_0)(g(x)^{p-1} + g(x)^{p-2}g(x_0) + g(x)^{p-3}g(x_0)^2 + \cdots + g(x_0)^{p-1})$$

If $g(x_0) \neq 0$ and g is continuous at x_0, then the second factor on the right is nonzero for x close to x_0, and converges to $pg(x_0)^{p-1}$.

We find that

$$\frac{g(x) - g(x_0)}{x - x_0} = \frac{g^p(x) - g^p(x_0)}{x - x_0} \cdot \frac{1}{\sum_{i=0}^{p-1} g(x)^{p-1-i}g(x_0)^i} \to \frac{1}{pg(x_0)^{p-1}}(g^p)'(x_0)$$

for $x \to x_0$.

So

$$g'(x_0) = \frac{(g^p)'(x_0)}{pg(x_0)^{p-1}}$$

In the case of $g(x) = x^{\frac{1}{p}}$, we get

$$g'(x_0) = \frac{1}{p\sqrt[p]{x_0}^{p-1}} = \frac{1}{p{x_0}^{\frac{p-1}{p}}} = \frac{1}{p}x_0^{\frac{1}{p}-1}$$

Guided exercise

We use the previous example to show that $\cos x$ is differentiable at 0 and $\cos' 0 = 0$.

1. Show that if f, g are functions (defined on an interval I) such that $f^2 + g^2 = c$ is constant on I, then g^2 is differentiable at $x_0 \in I$ if and only if f^2 is.
2. Conclude that $\cos^2 x$ is differentiable at 0 and compute its derivative.
3. Conclude that $\cos x$ is differentiable at 0 with derivative 0.

The fact that $\cos x$ is differentiable everywhere will be deduced soon.
EOE.

Example

The exponential function $\exp x$ is differentiable at 0. Recall that we defined $\exp x = E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. We have seen that $E(0) = 1$. Therefore $E(x) - E(0) = \lim_{N\to\infty} \sum_{n=1}^{N} \frac{x^n}{n!} = \sum_{n=1}^{\infty} \frac{x^n}{n!}$, keeping in mind that the symbol $\sum_{n=1}^{\infty} \frac{x^n}{n!}$ may mean both a series and the limit of the series. Thus, $E(x) - E(0) = x \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}$. Note that $F(x) \coloneqq \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}$ is absolutely convergent. Indeed, for any $x \in \mathbb{R}$, and any $N \in \mathbb{N}$,

$$\sum_{n=0}^{N} \frac{|x|^n}{(n+1)!} \leq \sum_{n=0}^{N} \frac{|x|^n}{n!} \leq E(|x|)$$

Thus $|F(x)| \leq F(|x|) \leq E(|x|)$. We conclude that for any $x \neq 0$,

$$\frac{E(x) - 1}{x} = F(x)$$

We will now show that $\lim_{x\to 0} F(x) = 1$. For this observe that $F(x) - 1 = \lim_{N\to\infty} \sum_{n=1}^{N} \frac{x^n}{(n+1)!} = x \lim_{N\to\infty} \sum_{n=0}^{N-1} \frac{x^n}{(n+2)!} = x \sum_{n=0}^{\infty} \frac{x^n}{(n+2)!}$. This limit exists (and is finite) because $G(x) \coloneqq \sum_{n=0}^{\infty} \frac{x^n}{(n+2)!}$ is absolutely convergent. (Again $G(|x|) \leq E(|x|) < \infty$.) For $x \in (-1,1)$ we have

$$|G(x)| \leq G(|x|) \leq E(|x|) \leq E(1)$$

By BTZ (for functions), we conclude that $\lim_{x\to 0} xG(x) = 0$. This means $lim_{x\to 0} F(x) = 1$. EOE.

Example

The exponential function $\exp x$ is differentiable everywhere.

Indeed, let $x_0 \in \mathbb{R}$. Then

$$\frac{\exp(x_0 + h) - \exp(x_0)}{h} = \exp(x_0) \left(\frac{\exp h - \exp(0)}{h} \right) \to \exp(x_0) \exp' 0 = \exp(x_0)$$

for $h \to 0$. EOE.

Remark

The exponential function $f(x) = \exp x$ has the remarkable property that $f'(x) = f(x)$ for all x. This explains parts of its importance. We will see soon that it is essentially the only function defined on $\mathbb{R}$ that has this property (up to multiplication with a constant). EOR.

5.1.7 The chain rule of differentiation

The chain rule deals with composition of functions.

Proposition

Suppose f is defined on some interval I, g is defined on some interval J, and $g(J) \subseteq I$. If g is differentiable at $x_0 \in J$, and f is differentiable at $y_0 = g(x_0) \in I$, then $h = f \circ g$ is differentiable at x_0 with $h'(x_0) = f'\big(g(x_0)\big)g'(x_0)$. EOP.

Proof. As before let $g(x_0 + h) = g(x_0) + g'(x_0)h + r(h)$ and consider

$$h(x_0 + h) = f\big(g(x_0 + h)\big) = f(g(x_0) + g'(x_0)h + r(h))$$

Then if $f(y_0 + k) = f(y_0) + f'(y_0)k + s(k)$, we find that

$$h(x_0 + h) = f(y_0) + f'(y_0)\big(g'(x_0)h + r(h)\big) + s(g'(x_0)h + r(h))$$

It remains to show that $\lim_{h\to 0}\left(\frac{f'(y_0)r(h)+s\left(g'(x_0)h+r(h)\right)}{h}\right) = 0$.

Note that $\frac{f'(y_0)r(h)}{h} \to 0$, so we must show that $\lim_{h\to 0}\frac{s\left(g'(x_0)h+r(h)\right)}{h} = 0$. Let $\sigma(h) = \frac{s(h)}{h}$ for $h \neq 0$ and $\sigma(0) = 0$. Then σ is continuous at 0, and

$$\frac{s\left(g'(x_0)h + r(h)\right)}{h} = \frac{g'(x_0)h + r(h)}{h}\sigma\left(g'(x_0)h + r(h)\right) \to g'(x_0)\sigma(0) = 0$$

for $h \to 0$. This uses that for all $h \neq 0$, $\frac{s\left(g'(x_0)h+r(h)\right)}{h} = \frac{g'(x_0)h+r(h)}{h}\frac{s\left(g'(x_0)h+r(h)\right)}{g'(x_0)h+r(h)}$ if $g'(x_0)h + r(h) \neq 0$. If $g'(x_0)h + r(h) = 0$, then both $s\left(g'(x_0)h + r(h)\right) = \sigma\left(g'(x_0)h + r(h)\right) = 0$. QED.

Examples

1. $\frac{1}{x}$ has derivative $-\frac{1}{x^2}$. So for any f that is differentiable and nonzero at x_0 we conclude that
$$\left(\frac{1}{f}\right)'(x_0) = \frac{d}{dx}\left(\frac{1}{x}\circ f\right)(x_0) = -\frac{1}{f(x_0)^2}f'(x_0)$$
in line with the quotient rule above.
2. We have seen that $f = \sqrt[p]{x}$ is differentiable for $x > 0$. Then $f^p = x$. So the chain rule would demand that
$$(f^p)' = pf^{p-1}f' = 1$$
It then follows that $f'(x_0) = \frac{1}{pf^{p-1}(x_0)}$, from which we conclude that f cannot be differentiable at $x_0 = 0$. For $x_0 > 0$ this suggests $f'(x_0) = \frac{1}{p\sqrt[p]{x_0^{p-1}}} = \frac{1}{p}x_0^{\frac{1}{p}-1}$ consistent with what we obtained before. But note that this does *not* directly prove that $f'(x_0)$ exists for $x_0 > 0$.
This procedure is known as *implicit differentiation*.
3. Let f be defined on $\mathbb{R}_{>0}$ as $f(x) = \sqrt[p]{x^q}$. Then
$$f'(x_0) = \frac{1}{p}\left(x_0^q\right)^{\frac{1-p}{p}}qx_0^{q-1} = \frac{q}{p}x_0^{\frac{q(1-p)+p(q-1)}{q}} = \frac{q}{p}x_0^{\frac{q}{p}-1}$$
This shows that for any rational number $r > 0$ we have x^r is differentiable at any $x_0 > 0$ and $\left.\frac{d}{dx}\right|_{x_0} x^r = rx^{r-1}$.

EOE.

The second item begs the following question: let f, g be functions such that $g \circ f$ is defined and differentiable. Suppose g is differentiable at $y_0 = f(x_0)$. Does it follow that f is differentiable at x_0?

The chain rule would say $(g \circ f)'(x_0) = g'(y_0)f'(x_0)$ suggesting that $f'(x_0) = \frac{(g\circ f)'(x_0)}{g'(y_0)}$ exists as long as $g'(y_0) \neq 0$.

This is actually true:

Let g be differentiable at $y_0 = f(x_0)$ and let $g \circ f$ be differentiable at x_0. If $g'(y_0) \neq 0$, then f is differentiable at x_0 and $f'(x_0) = \frac{(g \circ f)'(x_0)}{g'(y_0)}$.

To see why this holds, let

$$q(y) = \vdash \begin{cases} \dfrac{g(y) - g(y_0)}{y - y_0} & y \neq y_0 \\ g'(y_0) & y = y_0 \end{cases}$$

Then $q(y)$ is continuous at y_0. Moreover, for any $x \neq x_0$ we have

$$\frac{g\big(f(x)\big) - g\big(f(x_0)\big)}{x - x_0} = q\big(f(x)\big)\frac{f(x) - f(x_0)}{x - x_0}$$

(because both sides are equal to 0 if $f(x) = f(x_0) = y_0$. As $x \to x_0$, the left hand side converges to $g \circ f'(x_0)$, thus the right hand side converges as well. If $g'(y_0) \neq 0$ this means $\lim\limits_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$ must exist.

5.1.8 Invertible functions

Recall that a function $f: I \to J$ is called invertible with inverse g if there is a function $g: J \to I$ such that $f \circ g = \mathrm{id}_J$ and $g \circ f = \mathrm{id}_I$.

In this case, the *inverse* g is denoted f^{-1}.

Remark
For a function f to be invertible it must be both, injective and surjective. However, if f is injective, we often identify it with the function with codomain the range of f, and then it is automatically bijective (as a function to its range). This is not really an issue if f is continuous as then the range is again an interval, if the domain is an interval.

Proposition
Suppose f is continuous on an interval I and invertible and suppose $x_0 \in D_f$ and $f'(x_0) \neq 0$ for some $x_0 \in I$. Then $f^{-1}: f(I) \to I$ is differentiable at $y_0 = f(x_0)$, and $(f^{-1})'(y_0) = \frac{1}{f'(x_0)}$.
If $f'(x_0) = 0$, then f^{-1} is not differentiable at x_0. EOP.

Proof. As f is continuous, $f(I) = J$ is an interval. Let $g = f^{-1}$. If $f'(x_0) \neq 0$, we must show that

$$\lim_{y \to y_0} \frac{g(y) - g(y_0)}{y - y_0}$$

exists and is finite. Note that as f is invertible, it is injective, and so is $g = f^{-1}$. This means if $y \neq y_0$, then $x = f^{-1}(y) \neq x_0 = f^{-1}(y_0)$. Let $y_n \in J$ be a sequence with $y_n \neq y_0$ but $y_n \to y_0$. Then $x_n := f^{-1}(y_n) \neq x_0$ is a sequence in I. Moreover, $x_n \to x_0$. This uses the fact that f^{-1} is continuous at y_0.
Then, if $f'(x_0) \neq 0$

$$\frac{g(y_n) - g(y_0)}{y_n - y_0} = \frac{x_n - x_0}{f(x_n) - f(x_0)} \to \frac{1}{f'(x_0)}$$

for $n \to \infty$. This shows that g is differentiable if $f'(x_0) \neq 0$.

If $f'(x_0) = 0$, and f^{-1} were differentiable at $y_0 = f(x_0)$, then the chain rule asserts that $1 = (f^{-1} \circ f)'(x_0) = f^{-1\prime}(y_0) f'(x_0) = 0$, which is impossible. QED.

Example

1. We have seen that $\exp x$ is differentiable everywhere. Moreover $\exp'(x_0) = \exp x_0$ for all $x \in \mathbb{R}$.

5.2 Local properties of differentiable functions

5.2.1 Positive and negative derivative

Suppose f is differentiable at some $x_0 \in I$. It is a natural question to ask whether we can draw any conclusions from properties of $f'(x_0)$.

For example, does it mean something whether $f'(x_0) > 0$ or $f'(x_0) < 0$?

Lemma
Let I be an interval and $x_0 \in I$. Suppose f is defined on I and differentiable at x_0. Suppose $f'(x_0) > 0$. Then there is $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0) \cap I$ and all $y \in (x_0, x_0 + \delta) \cap I$ we have

$$f(x) < f(x_0) < f(y)$$

If $f'(x_0) < 0$, there is $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0) \cap I$ and all $y \in (x_0, x_0 + \delta) \cap I$ we have

$$f(x) > f(x_0) > f(y)$$

Proof. We prove the first assertion. The second follows by applying the first to $-f$.

Let $\delta > 0$ be small enough such that $\frac{f(x) - f(x_0)}{x - x_0}$ is close enough to $f'(x_0)$ as to be positive as well whenever $0 < |x - x_0| < \delta$ and $x \in I$.

As $x - x_0 > 0$ if and only if $x > x_0$, this means $f(x) > f(x_0)$ if $x > x_0$ and $f(x) < f(x_0)$ if $x < x_0$. QED.

5.2.2 Local extrema

Definition
Let f be a function defined on an interval I and let $x_0 \in I$. Then f has a

- **local maximum** at x_0, if there is $\delta > 0$ such that $\forall x \in I \cap (x_0 - \delta, x_0 + \delta): f(x) \leq f(x_0)$.
- **local minimum** at x_0, if there is $\delta > 0$ such that $\forall x \in I \cap (x_0 - \delta, x_0 + \delta): f(x) \geq f(x_0)$.

f has a **local extremum** at x_0 if it has a local maximum or minimum at x_0.

A local extremum need not be a global extremum (that is, $f(x_0) = \sup\{ f(x)\}$ or $\inf\{f(x)\}$). Unless I is a closed interval and f is continuous, local or global extreme need not exist.

Lemma
Let f be a function defined on some interval I. Suppose f has a local extremum at $x_0 \in I^\circ$ and suppose f is differentiable at x_0. Then $f'(x_0) = 0$. EOL.

Proof. This is an immediate consequence from the definition and the lemma in 5.2.1 which shows that the derivative cannot be positive or negative at a local extremum in the interior of I.

For example, if $f'(x_0) > 0$ at the inner point x_0, then there are sequences $x_n < x_0$, $y_n > x_0$ in I such that $x_n, y_n \to x_0$ and $f(x_n) < f(x_0) < f(y_n)$ contradicting the definition of a local maximum or minimum. The proof for negative derivative at x_0 is similar. QED.

This is one of the few cases where it is important that x_0 is an interior point: Consider $f(x) = x$ defined on $[0,1]$. Then f has a local (in fact global) extremum at $x_0 = 1$. But of course $f'(1) = 1 \neq 0$.

Example
Consider $f: \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x^3 - x$. Then $f'(x) = 3x^2 - 1$, and we find candidates for local extrema as $x_{1/2} = \pm\frac{1}{\sqrt{3}}$. Let us focus on $x_2 = \frac{1}{\sqrt{3}}$. Consider the interval $I = [0, b]$. Then f attains its infimum on this interval (Min/Max Principle). That is, there is $x_0 \in I$ such that $f(x_0) = \inf_I f$. There are three possible cases: $x_0 = 0$, $x_0 = b$, or $x_0 \in I^\circ$. In the last case, we must have $f'(x_0) = 0$ and thus, $x_0 = x_1$. $f(0) = 0 > f(x_1) = \frac{1}{3\sqrt{3}} - \frac{1}{\sqrt{3}}$. So $0 \neq x_0$. Next, for b large enough, $f(b) > 0$ and hence in this case $x_0 \neq b$ either. It follows f has a local minimum at $x_0 = x_1$.
This shows that quite a bit of ad-hoc reasoning is necessary even in the case of a relatively simple function. EOE.

5.2.3 Rolle's Theorem

Theorem
Let f be a continuous function defined on $[a, b]$ and differentiable on at least (a, b) such that $f(a) = f(b)$. Then there is $c \in (a, b)$ such that $f'(c) = 0$. EOT.

The theorem should be intuitively reasonable: if $f(a) = f(b)$ it is not a stretch that this means it must have a horizontal tangent somewhere in (a, b).

Proof. As f is continuous, the maximum principle shows that f must attain its minum and maximum somewhere on $[a, b]$. If f is constant then $f'(c) = 0$ for all c and we are done.

Otherwise $f(a) = f(b)$ cannot be equal to both the maximum and minimum of f. Therefore f has a global extremum at an interior point $x_0 \in I$. It is necessarily also a local extremum, and by the lemma in 5.2.2 this means $f'(x_0) = 0$. QED.

Rolle's Theorem has only one purpose: proving the Mean Value Theorem (below). However it *does* help finding local extrema:

Example
Let f be a continuous function on the interval I, differentiable on at least I°. For any $a < b \in I$ with $f(a) = f(b)$ there must be $c \in (a, b) \subseteq I$ such that $f'(c) = 0$ *and* f has a local extremum at c (as this is what the proof of Rolle's Theorem actually shows).

So in the example above: for $f(x) = x^3 - x$, we know $f(0) = 0$, $f(x_1) < 0$, and $\lim_{x\to\infty} f(x) = \infty$. So there must be $b > 0$ with $f(b) = 0$, and then necessarily x_1 (as the only zero of f' in $(0, b)$ must be the place of a local minimum. EOE.

Note that if f is not differentiable everywhere, there are simple counterexamples to the statement of the theorem: consider $f(x) = |x|$ on $[-1,1]$. Then $f(-1) = f(1)$ but there is no $c \in (-1,1)$ where $f'(c)$ is defined and equal to 0.

5.2.4 The Mean Value Theorem

Rolle's Theorem is a special case of the so called Mean Value Theorem, which relates the derivative to the "average" slope of f.

Example

Suppose you are tasked with enforcing speeding laws along a long stretch of highway (with not much inbetween the two endpoints of the stretches). Rather than bother measuring the instantaneous speed of vehicles you install cameras at both ends of the stretch which record a time coded picture of the licence plate of each passing car. If the stretch of highway covers a distance Δd, and the time difference of the two pictures of a licence plate Δt is so small that $\frac{\Delta d}{\Delta t} > L$ where L is the speed limit (which we assume to be constant), you send the owner of the car a ticket. Now an owner decides to challenge the ticket, and they claim never to have driven faster than L. What is your response in court?
While it is intuitively clear that the owner must be lying, it is a bit technical to prove that this is actually the case, and the proof below relies on the fact that the travelled distance is a continuous function in time and in fact differentiable everywhere except possibly at the two endpoints. EOE.

Theorem

Let $f: [a, b] \to \mathbb{R}$ be continuous and suppose f is differentiable on (a, b).
Then there is $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a)$$

EOT.

Proof. Consider the function $g(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x - a)$ defined in $[a, b]$.

Then $g(a) = g(b) = f(a)$. As g is differentiable on (a, b), Rolle's Theorem implies that there is $c \in (a, b)$ where

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

QED.

Now we got the speeding driver: As $\Delta d = d(b) - d(a) > L(b - a)$ we must have that $v(c) > L$ for at least one $c \in (a, b)$ where $v(t) = d'(t)$ is the car's speed at time t.

Applied to this situation the Mean Value Theorem says that you must hit the *average* speed at least once during a trip.

Geometrically this means the slope of the graph of f must be equal to the slope of the line connecting $(a, f(a))$ and $(b, f(b))$ *somewhere*. The slope of this line is in this sense the *average* slope of the graph.

Corollary (Zero derivative means constant)

Let f be continuous on the interval I, and differentiable on at least I°. If $f' = 0$ on I°, then f is constant on all of I. EOC.

Proof. Let $a < b \in I$. We must show that $f(a) = f(b)$. Now $[a, b] \subseteq I$ and f is differentiable on (a, b). By the MVT (applied to f on $[a, b]$) we have $f(b) - f(a) = f'(c)(b - a)$ for some $c \in (a, b)$. Since $f'(c) = 0$ it follows that $f(b) = f(a)$. QED.

In particular, this means if two continuous f, g are defined on an interval I and $f' = g'$ on $I°$, then $f = g + C$ for some constant C. This is the corollary applied to the difference $f - g$ (which has zero derivative everywhere on $I°$).

Example
If $f' = m$ is constant on (a, b), then $f = mx + C$ on $[a, b]$. Indeed, $(f - mx)' = 0$ and hence $f - mx = C$ is constant. EOE.

Warning
Note that the fact that the domain here is an *interval* is important:
Let f be defined on $(-1,0)$ as $f(x) = -1$, and on $(0,1)$ as $f(x) = 1$. Then $f'(x) = 0$ everywhere on $D = (-1,0) \cup (0,1) = (-1,1) \setminus \{0\}$ but f is **not** constant.
So, the more general statement would be that f must be constant on any *interval contained in its domain*.
The precise condition for the corollary to hold, which can be generalized to functions in more variables, is that the domain of f must be *connected*. This is a term borrowed from topology and means that the domain cannot be written as a disjoint union of relative open subsets. For $\mathbb{R}$, a subset A is connected if and only if for every $a \neq b \in A$, the line segment connecting them (that is, $[a, b]$) is contained in A.
EOW.

5.2.5 Monotone functions and the derivative

Proposition
Let f be continuous on an interval I and differentiable on at least the interior $I°$. Then

- f is *monotone increasing* iff $f' \geq 0$ on $I°$.
- f is *monotone decreasing* iff $f' \geq 0$ on $I°$.

If $f' > 0$ on $I°$, then f is strictly monotone increasing. If $f' < 0$ on $I°$ then f is strictly monotone decreasing. EOP.

Proof. If $f$is monotone increasing, we cannot have $f'(c) < 0$ for any $c \in I°$ (see 5.2.1). Conversely, if $f' \geq 0$ on $I°$, by the MVT 5.2.4 for any $x \leq y \in I$, we have $f(y) - f(x) = f'(c)(y - x) \geq 0$ for some $c \in (x, y)$. Thus, $f(x) \leq f(y)$ and f is monotone increasing.

The proof in the second case is similar.

Finally, suppose $f' > 0$ on $I°$. Then the argument just given shows that $f(y) - f(x) > 0$ for all $x < y \in I$. Again, the argument is similar if $f' < 0$. QED.

Note that if f is stricyly increasing, it does not follow that $f' > 0$. As an example consider $f(x) = x^3$ defined on $[-1,1]$. Then f is strictly increasing, but $f'(0) = 0$. It is still true however, that $f' \geq 0$.

Example
Suppose I is an interval of any type. Recall that a function $f: I \to \mathbb{R}$ is called *invertible* (as a function to its range $J \coloneqq f(I)$) if it is one-one (injective), since it is then *bijective* onto its range J. If f is continuous

then J is an interval. We also know that in this case f is invertible if and only if f is strictly monotone on I.

Find a maximal[2] *interval I such that $\sin x$ is invertible on I.*

Our approach is simple: first identify an interval where $\cos x = \sin' x > 0$. We know that is true on $K = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and can conclude that $\sin x$ is strictly monotone on K, and in fact on $I = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. We also know that $\cos x < 0$ for any $x > \frac{\pi}{2}$ and close to $\frac{\pi}{2}$. And similarly, $\cos x < 0$ for $x < -\frac{\pi}{2}$ and close to $-\frac{\pi}{2}$. Thus $I = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is a maximal interval where $\sin x$ is invertible. Then $J = \sin I = [-1,1]$, and the inverse of $\sin$ is often called $\arcsin$, or, obviously, $\sin^{-1}$.

Similar reasoning shows that $\tan x$ is invertible on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ with range $\mathbb{R}$ (and as $\tan x$ is undefined for $x = \pm\frac{\pi}{2}$ this is a maximal interval). The inverse of $\tan$ is denoted $\arctan$; it is defined on all of $\mathbb{R}$ and has range $(-\frac{\pi}{2}, \frac{\pi}{2})$. EOE.

5.2.6 The first derivative test

Using the definition of a local extremum it may be hard to figure out whether any given function has one at a point c.

We have seen that the derivative vanishes at local extrema. The converse is false as again $f(x) = x^3$ shows at $x_0 = 0$. This means to find local extrema, we need a way to decide whether f has a local extremum at its **critical points**. A critical point here is any x_0 where $f'(x_0) = 0$ or where f' is not defined.

Theorem (First Derivative Test)
Let I be an interval, f continuous on I and $c \in I$ a critical point.

- If there is a relative open subinterval $J \subseteq I$ such that $c \in J$ and $f' \leq 0$ on $J \cap (-\infty, c)$ and $f' \geq 0$ on $J \cap (c, \infty)$, then f has a **local minimum** at c.
- If there is a relative open subinterval $J \subseteq I$ such that $c \in J$ and $f' \geq 0$ on $J \cap (-\infty, c)$ and $f' \leq 0$ on $J \cap (c, \infty)$, then f has a **local maximum** at c.

EOT.

To clarify the "relative open" business: if c is an interior point of I, all this says is that there is $\delta > 0$ such that $J = (c - \delta, c + \delta) \subseteq I$ and, in the first scenario, $f' \leq 0$ on $(c - \delta, c)$ and $f' \geq 0$ on $(c, c + \delta)$.

In the second scenario, $f' \geq 0$ on $(c - \delta, c)$ and $f' \leq 0$ on $(c, c + \delta)$.

In case $c \in \partial I$, then one of the two (or both if I is just consisting of a point) is vacuous as there are no points in $J \cap (-\infty, c)$ for example if $I = [c, b)$ for some $b > 0$.

---

[2] For the purposes here, a *maximal* I is an interval that is not properly contained in any interval with the same property.

Proof. We discuss the first scenario. The proof of the second is similar (or simply apply the first scenario to the function $-f$).

Then f is monotone decreasing on $J \cap (-\infty, c]$ and monotone increasing on $J \cap [c, \infty)$. In particular, there is $\delta > 0$ such $f(x) \geq f(c)$ for all $x \in I \cap (c - \delta, c + \delta)$. But that means f has a local minimum at c.

Example
Let $f: \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = \begin{cases} x \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Then f is not differentiable at $x = 0$. Does it have a local extremum at $x = 0$?
The answer is no, as there are zero sequences $x_n < 0$ and $y_n > 0$ such that $f(x_n) < 0$ and $f(y_n) > 0$.
But not if we put $g = f^2$, then g has a local minimum at $x = 0$. However, there is no interval of the form $[0, \delta)$ with $\delta > 0$ such that f is monotone increasing, simply because there are $x, y \in [0, \delta)$ such that $x < y$ and $g(x) > g(y) = 0$. For this oberserve that $\frac{1}{k\pi}$ for almost all $k \in \mathbb{N}$ is in $[0, \delta)$ and is a zero of g. This shows that a function need not be monotone close to one side of a local extremum.
EOE.

5.2.7 The second derivative test

Recall that if g is differentiable, then $g'(c) > 0$ tells us that $g(x) < g(c) < g(y)$ for all $x < c < y$ "close enough" to c. Applying this reasoning to $g = f'$ we obtain the second derivative test.

Theorem (Second Derivative Test)
Let f be defined on an interval I and twice differentiable at a critical point x_0. Then

- If $f''(x_0) < 0$, f has a local maximum at x_0.
- If $f''(x_0) > 0$, f has a local minimum at x_0.
- If $f''(x_0) = 0$, f may or may not have a local extremum of either type at x_0.

Proof. We prove the first case. The second is similar, and the last is proven by the examples $x^3, \pm x^4$ at $x_0 = 0$.

By assumption f' is defined on a relative open interval containing x_0. As $f''(x_0) < 0$, the results of 5.2.1 show that $f' > 0$ on some set of the form $(x_0 - \delta, x_0) \cap I$, and $f' < 0$ on $(x_0, x_0 + \delta) \cap I$ (for a suitably small chosen $\delta > 0$).

But then f is strictly monotone decreasing on $(x_0 - \delta, x_0] \cap I$ and strictly monotone decreasing on $[x_0, x_0 + \delta) \cap I$ by 5.2.5.

Thus, f has a local maximum at x_0. QED.

From the proof we see this criterion cannot possibly capture all instances where f has a local extremum, as it forces f to be strictly monotone (of opposing direction) in a neighbourhood of c. For example the theorem says nothing about constant functions, which have local (and global) extrema everywhere.

Example
Find all local extrema of $\sin x$.

Then $\sin' x = \cos x = 0$ means $x = \frac{\pi}{2} + k\pi$ for some $k \in \mathbb{Z}$.
And $\sin'' x = -\sin\, x$. Now $\sin\left(\frac{\pi}{2} + k\pi\right) = \vdash \begin{cases} 1 & k \text{ is even} \\ -1 & k \text{ is odd} \end{cases}$
This shows that $\sin x$ has a local maximum at $\frac{\pi}{2} + k\pi$ if k is even, and a local minimum if k is odd. Of course we know that also directly.

Example

Find all local extrema of $f(x) = x^3 - \frac{3}{2}x^2 - 6x - 1$ (defined on $\mathbb{R}$).
First, $f'(x) = 3x^2 - 3x - 6$. Then $f'(x) = 0$ iff $x^2 - x - 2 = 0$. The solution to this equation is $x_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + 2} = 2$ and $x_2 = -1$.
As $f''(x) = 6x - 3$, we find that $f''(x_1) > 0$ and $f''(x_2) < 0$ so f has a local minimum at x_1 and a local maximum at x_2.
Does f have global extrema? EOE.

Example

Let f be a polynomial function of degree 3 such that f' has two distinct roots. Show that f has a local minimum at one and a local maximum at the other. If $f(x) = ax^3 + \cdots$ which is which depending on $a > 0$ or $a < 0$?

We know there are $x_1 < x_2$ such that $f'(x_1) = f'(x_2) = 0$. Rolle's Theorem asserts that then there is $x_3 \in (x_1, x_2)$ with $f''(x_3) = 0$. As f'' is a degree one polynomial, it has exactly one root. This means $f''(x_1) \neq 0$ and $f''(x_2) \neq 0$. Moreover, as f'' is a linear function, the signs of $f''(x_1), f''(x_2)$ are opposed. f has a local maximum at one and a local minimum at the other. If $a > 0$, then f'' is strictly increasing, which means x_1 is a local maximum and x_2 is a local minimum for f. If $a < 0$ the places are interchanged and f has a local minimum at x_1 and a local maximum at x_2.

5.3 The derivative as a function

Let f be defined on an interval I. The **derivative** of f, denoted f', is the function$f' : D_f \to \mathbb{R}$ defined by $x \mapsto f'(x_0)$ where $D_f = \{x_0 \in I \mid f'(x_0) \text{ exists}\}$.

Of course, the derivative is only interesting if $D_f \neq \emptyset$, and in fact, when we discuss f' as a function, we usually want D_f to comprise a relative open interval (in I) containing a given point in question.

If $x_0 \in D_f$ and D_f contains a relative open interval J[3] with $x_0 \in J$, then it makes sense to ask whether f' is again differentiable at or continuous at x_0. This leads to the following recursive definition:

Definition

Let I be an interval, and f a function on I differentiable at x_0.
We say f is **continuously differentiable** at x_0 if $f'(x)$ is defined for x in a relative open interval containing x_0 and f' is continuous at x_0.

For any natural number n we say f is **$\boldsymbol{n}$-times differentiable at $\boldsymbol{x_0}$** if:

1. $f^{(n-1)}$ is defined on a relative open interval containing x_0.

[3] This means that J is just an open interval if I is an open interval itself. If $I = [a, b]$, then J is open or of the form $[a, c)$, $(c, b]$, or $[a, b]$. And similarly, if I is a half-open interval, J is open or contains the boundary point of I.

2. $f^{(n-1)}$ is differentiable at x_0.

Here for any n the function $f^{(n)}$ is defined as $(f^{(n-1)})'$, where $f^{(n-1)}$ is defined on a relative open neighbourhood of x_0.

A function is called **smooth** if it is n times differentiable everywhere in its domain for every n. EOD.

These definitions sound more complicated than they actually are.

If f is differentiable at x_0, and also on a set of the form $(x_0 - \varepsilon, x_0 + \varepsilon)$, it makes sense to define $f''(y_0)$ for $y_0 \in (x_0 - \varepsilon, x_0 + \varepsilon)$ as $f''(y_0) = (f')'(y_0)$.

Remark

Suppose f, g are defined and differentiable everywhere on an interval I. Then, as functions on I, we have $(f+g)' = f' + g'$, $(fg)' = f'g + g'f$. If g is nowhere 0 on I, then $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$.

Example

Let $f: \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = \vdash \begin{cases} -\frac{1}{2}x^2; & x < 0 \\ \frac{1}{2}x^2; & x \geq 0 \end{cases}$$

Then f is differentiable everywhere $f'(x) = |x|$. Thus f is twice differentiable for all $x \neq 0$ (in fact it is n-times differentiable for all n there), but not at $x = 0$.

Exercise

Let I be an interval and f be a rational function on I. That is $f = \frac{p}{q}$ where p, q are polynomial functions defined on I and $q(x) \neq 0$ for all $x \in I$. Then f' is defined on I.

Let $f(x_0) = f'(x_0) = 0$. Show that $(x - x_0)^2$ must divide p as a polynomial (that is $p = (x - x_0)^2 g$ where g is a polynomial function on I).

Show that if f is a polynomial such that $f(x_0) = f^{(1)}(x_0) = \cdots = f^{(k)}(x_0) = 0$ then $(x - x_0)^{k+1}$ divides f.

Example

We have $\sin' x = \cos x$ and $\cos' x = -\sin x$.

Indeed, $\sin(x+h) = \sin(x)\cos(h) + \cos(x)\sin(h)$. Therefore $\sin(x+h) - \sin(x) = \sin(x)(\cos(\mathrm{h}) - 1) + \cos(x)\sin(h)$, and

$$\frac{1}{h}(\sin(x+h) - \sin(x)) = \frac{\sin(x)(\cos(h) - 1)}{h} + \frac{\cos(x)\sin(x)}{h} \to 0 + \cos(x)$$

for $h \to 0$.

Similarly,

$$\begin{aligned}\cos(x+h) - \cos(x) &= \cos(x)\cos(h) - \sin(x)\sin(h) - \cos(x) \\ &= \cos(x)(\cos(h) - 1) - \sin(x)\sin(h)\end{aligned}$$

and the result follows because $\frac{\cos h - 1}{h} \to 0$ and $\frac{\sin h}{h} \to 1$ for $h \to 0$.

This shows that the trigonometric functions are smooth. In fact, they are solutions for the **differential equation**

$$f'' = -f$$

This equation is fundamentally important in classical mechanics (it is essentially describing the *harmonic oscillator*). One can show that *any* twice differentiable function f that satisfies this equation on an interval I is of the form $a \sin x + b \cos x$ for some constants a, b. Of couse any such function satisfies this equation because

$$(a \sin x + b \cos x)'' = a \sin'' x + b \cos'' x = -a \sin x - b \cos x$$

It is possible to *define* the trigonometric functions sine and cosine as **solutions to the differential equation** $f'' = -f$ with the prescribed values $\sin 0 = 0$, $\sin' 0 = 1$ and $\cos 0 = 1$, $\cos' 0 = 0$. This would require that we know such solutions exist and are unique (if enough prescribed values are given). However, it would have the advantage that we would not rely on somewhat murky concepts from Euclidean geometry: the definition of $\sin x$, for example, relies on the definition of the *arc length* of a segment of a unit circle. To define this precisely quite a bit of work is required. We will hopefully see a rigorous definition of the trigonometric functions towards the end of the term. EOE.

5.4 Interlude: Exponential functions and logarithms

Proof. The uniqueness assertion follows from the previous lemma, as all exponential functions are continuous. We have to postpone the existence part for now.

Remark

As mentioned in class one *can* define the exponential function a^x as $a^x = \lim_{n\to\infty} a^{r_n}$ where $r_n \to x$ is a sequence of rational numbers. One then has to show that this limit always exists and is independent of the sequence chosen, and finally that the so defined function is differentiable. All of this is possible.
A more conceptual way is to figure out properties of exponential functions and then in the end show that there is a function satisfying the desired properties. There are essentially two ways to do this, and we will soon see how.

5.4.1 The derivative of an exponential function

Proposition

Let f be the exponential function with base $a > 0$. Then f is differentiable and $f' = cf$ where $c = \ln a$.

Proof. We know that$f(x) = \exp(x \ln a)$ is the unique exponential function with base a. Computing its derivative using the chain rule gives $f'(x) = \ln a \ f(x)$ as claimed. QED.

Remark

Functions satisfying an equation of the form $f' = cf$ are important in many applications. They describe systems experiencing exponential growth (if $c > 0$) or decay (if $c < 0$) and appear in many contexts. Not every such function is an exponential function. But every such function is a constant multiple of an exponential function, motivating the terminology *exponential* growth or decay. EOR.

Note that for all $x, y > 0$ we have $\ln(xy) = \ln(x) + \ln(y)$. This is shown simply by applying exp on both sides of the equation. (It also follows by the fact that $\ln$ is the inverse of a group homomorphism and hence itself a group homomorphism.) As a consequence, we also have that $\ln x^n = n \ln x$ for all $n \in \mathbb{N}$.

Moreover, as $\exp$ is strictly monotone increasing, so is its inverse $\ln$. Its range must be all of $\mathbb{R}$ and we conclude that $\lim_{x\to 0} \ln x = -\infty$ and $\lim_{x\to\infty} \ln x = \infty$.

As with exponential functions, we can also define logarithms for different bases. If $a > 0$ is any real number except 1, the exponential function with base a has range $(0, \infty)$ and is invertible. Its inverse function is denoted by $\log_a x$. If $a \neq 1$ then $\ln a \neq 0$.

Note that since $y = a^x = \exp(x \ln a)$, we get $\log_a y = x = \frac{\ln a^x}{\ln a} = \frac{\ln y}{\ln a}$.

Exercise
Show that for any $a > 0$, $a \neq 1$, and any positive x, y, we have

1. $\log_a(xy) = \log_a x + \log_a y$
2. $\log_a x^n = n \log_a x$
3. $\log_a$ is strictly increasing if $a > 1$ and strictly decreasing if $a < 1$.

EOE.

We conclude the section with the following observation on arbitrary differentiable functions satisfying an equation of the form $f' = cf$.

Proposition
Let f be a differentiable function on $\mathbb{R}$ such that $f' = cf$. Then there is a unique constant d such that $f(x) = d\exp(cx)$ for all x. EOP.

Proof. Let $g(x) = \exp(-cx)\, f(x)$. Then $g'(x) = -c\exp(-cx)\, f(x) + \exp(-cx)\, cf(x) = 0$. So g is constant, $g = \frac{f}{\exp(cx)} = d$. QED.

This shows that up to multiplication by a constant, the functions of exponential growth or decay are exponential functions.

5.5 Local properties of differentiable functions continued

Here is a sometimes useful theorem (which we haven't discussed in class).

Theorem*
Let I be an interval and $f: I \to \mathbb{R}$ be a differentiable function and suppose f' is continuous. Let $x_0 \in I$ and suppose $f'(x_0) \neq 0$. Then there is a relative open neighbourhood J of $x_0 \in J \subseteq I$, and an interval $K \subseteq \mathbb{R}$ such that $f(J) = K$ and $f: J \to K$ is invertible. Moreover, f^{-1} is also continuously differentiable on K.

This is a version of a much more difficult analogous result for functions in several variables (called the Implicit Function Theorem).

Proof. We prove the case $f'(x_0) > 0$ and leave the case $f'(x_0) < 0$ as an exercise.

As f' is continuous, there is a relative open subset $J \subseteq I$ with $x_0 \in J$ such that $f' > 0$ on J. We denote the restriction of f to J by f as well. Then f is strictly monotone on J, and $f(J) = K$ is an interval. Finally, f^{-1} is differentiable everywhere on K by the previous proposition, from which we obtain the formula that ${f^{-1}}'(y) = \frac{1}{f'(f^{-1}(y))}$. As the multiplicative inverse of a composition of two continuous functions, ${f^{-1}}'$ is continuous as well. QED.

The theorem essentially says that any continuously differentiable function is *locally* invertible, as long as its derivative is nonzero.

Warning
As before, it is important to note that $f' > 0$ or $f' < 0$ only guarantees that f is injective (because it is strictly monotone) if the domain of f is an interval.
Consider $D = (0,1) \cup (1,2) = (0,2) \setminus \{1\}$ and $f(x) = x$ on $(0,1)$ and $f(x) = x - 1$ on $(1,2)$. $f' = 1 > 0$ on D, but f is not injective. EOW.

Exercise
Suppose f is a smooth invertible function, and $f'(x) \neq 0$ for all x. Show that f^{-1} is also smooth. EOE.

Example

1. Consider $f(x) = \sin x$. Then f is not invertible unless we restrict its domain to a small enough set, say $I = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Then $f(I) = [-1,1]$ and $f: I \to [-1,1]$ is invertible.
   Its inverse function is often denoted $\sin^{-1} x$ or $\arcsin x$. We find that $\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\cos \sin^{-1} x}$, which is defined as long as $x \neq \pm 1$. Here f^{-1} is differentiable on $(-1,1)$. Moreover $\cos t$ is positive on the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Now observe that $\cos \sin^{-1} x = \sqrt{1 - (\sin \sin^{-1} x)^2} = \sqrt{1 - x^2}$. The upshot is: $f'(x) = \frac{1}{\sqrt{1-x^2}}$ on $(-1,1)$.
2. Consider $\tan x$ which has domain $\mathbb{R} \setminus \left(\frac{\pi}{2} + \mathbb{Z}\pi\right)$. Recall that $\frac{d}{dx}\tan x = \frac{d}{dx}\frac{\sin x}{\cos x} = \frac{\cos x \cos x + \sin x \sin x}{(\cos x)^2} = \frac{1}{(\cos x)^2}$. This function is positive on $I = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, so $\tan x$ is strictly increasing on that interval. As $\sin x$ approaches ± 1 and $\cos x$ approaches 0 towards the boundaries of I, $\tan(I) = (-\infty, \infty)$. And $\tan x$ has an inverse, denoted $\tan^{-1} x$ or $\arctan x$, defined on all of $\mathbb{R}$.
   Then $\frac{d}{dx}\arctan x = (\cos \arctan x)^2$. Now note that $(\tan x)^2 = \frac{(\sin x)^2}{(\cos x)^2} = \frac{1}{(\cos x)^2} - 1$ or
   $$(\cos x)^2 = \frac{1}{1 + (\tan x)^2}$$
   Using this we get $\frac{d}{dx}\arctan x = \frac{1}{1+x^2}$.
3. Consider the exponential function $f(x) = \exp(x)$. As $f' = f > 0$, f is invertible. We have seen that the range of f is all of $\mathbb{R}_{>0}$. Let $g(x) = \ln x$ be the inverse function. Then for all $y_0 \in \mathbb{R}_{>0}$ $g'(y_0) = \frac{1}{f'(g(y_0))} = \frac{1}{f(g(y_0))} = \frac{1}{y_0}$.
4. Let $y \in \mathbb{R}$ and for $x > 0$ Consider $f(x) = x^y$. Then $f(x) = \exp\,(y \ln x)$, so the chain rule gives $f'(x) = \frac{\exp(y \ln x) y}{x} = \frac{x^y y}{x} = \frac{\exp(y \ln x - \ln x\,)}{x} = y\, x^{y-1}$.
5. Let $f(x) = x^x$ defined on $x > 0$. Then $f(x) = \exp\,(x \log x)$ and $f'(x) = \exp(x \log x)\,(x \log x)' = x^x(\log x + 1) = (\log x + 1)x^x$

Example
Consider the function $f: \mathbb{R} \to \mathbb{R}$ defined as
$$f(x) = \begin{cases} 0. & ; x = 0 \\ x^2 \sin \frac{1}{x} & ; x \neq 0 \end{cases}$$
is differentiable. However, f' is not continuous at 0.

The differentiation rules immediately show that f is smooth on $\mathbb{R} \setminus \{0\}$.
Now consider $x_0 = 0$.
Then for $x \neq 0$, $\frac{f(x)-f(x_0)}{x-x_0} = x \sin\frac{1}{x}$. Now $\lim_{x\to 0} x \sin\frac{1}{x} = 0$ by BTZ. We conclude that $f'(0) = 0$.
But for $x \neq 0$, we have $f'(x) = 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)$. Again $2x \sin\left(\frac{1}{x}\right) \to 0$ for $x \to 0$. However, $\lim_{x\to 0} \cos\left(\frac{1}{x}\right)$ does not exist. For this consider the sequences $x_n = \frac{1}{n\pi}$, Then $x_n \to 0$ but $\cos\left(\frac{1}{x_n}\right) = (-1)^n$ is not convergent.

Remark
We emphasize here that the derivative of a differentiable function need not be continuous.

5.6 L'Hôpital's rule

When computing limits of quotients of functions, we run into trouble when both functions have $\pm\infty$ or 0 as a limit. Here we will find a rule that is surprisingly useful. We have seen examples of this: for example, the limit of $\frac{\sin x}{x}$ for $x \to 0$. Sometimes one can attack these limits directly. But sometimes this is hard.

Consider the following situation: suppose f, g are differentiable on $[a, b]$ and $f(a) = g(a) = 0$. Suppose also that $g(x) \neq 0$ for $x > a$. We want to find $\lim_{x\to a} \frac{f(x)}{g(x)}$ if it exists. If $g'(a) \neq 0$, then this limit is equal to $\frac{f'(a)}{g'(a)}$.

Indeed $\frac{f(x)}{g(x)} = \frac{f(x)-f(a)}{g(x)-g(a)} = \frac{f(x)-f(a)}{x-a}\frac{x-a}{g(x)-g(a)} \to \frac{f'(a)}{g'(a)}$ for $x \to a$.

If f, g are *continuously differentiable* at a, this means $\lim_{a\to x} \frac{f(x)}{g(x)} = \lim_{a\to x} \frac{f'(x)}{g'(x)}$.

This works for example in case $f = \sin x$ and $g = x$. But of course, this is cheating because $\lim_{x\to 0} \frac{\sin x}{x} = \sin' 0$.

Derivatives can be useful when computing some limits of quotients of functions. A different version of a similar rule as the above is

Theorem (L'Hôpital's rule)
Let I be an interval and $a \in I$. Let f, g be functions defined and differentiable on $J = I \setminus \{a\}$. Suppose that $g'(x) \neq 0$ for all $x \in J$. Furthermore, suppose one of the following two statements holds

1. $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$.
2. $\lim_{x\to a} g(x) = \pm\infty$.

Suppose $L = \lim_{x\to a} \frac{f'(x)}{g'(x)}$ exists. Then $\lim_{x\to a} \frac{f(x)}{g(x)} = L$.

Remark

HISTORY

Many attributions in mathematics are wrong or at least imprecise. This rule is named after Guillaume de l'Hôpital (1661-1704) (a.k.a. Guillaume-François-Antoine Marquis de l'Hôpital, Marquis de Sainte-Mesme, Comte d'Entremont and Seigneur d'Ouques-la-Chaise) who published it in his 1696 book "Analyse des infiniment petits pour l'intelligence des lignes courbes", apparently the first text-book on the differential calculus. It is possible that the rule actually originated with Johann Bernoulli (1667-1748).

The point of first choosing I and then defining $J = I \setminus \{a\}$ is the following: In the majority of applications I is of the form $[a, b)$ or $(b, a]$, that is a is a boundary point of I. Then J is again an interval (a, b) or (b, a), respectively. The limit that the l'Hôpital rule then computes is a one-sided limit. But in some cases, we want to compute a two sided limit, and then a is an interior point of I. In that case J is not an interval itself but a union of two intervals: if $I = (c, d)$ and $a \in (c, d)$, then $J = (c, a) \cup (a, d)$. You should not worry too much about this.

The proof will require a stronger version of the Mean Value Theorem.

Lemma (Cauchy Mean Valuey Theorem)
Let f, g be continuous on $[a, b]$ and differentiable on (a, b). Then there is $c \in (a, b)$ such that

$$f'(c)\big(g(b) - g(a)\big) = g'(c)(f(b) - f(a))$$

In particular if $g'(x) \neq 0$ on (a, b), then

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}$$

Proof. Consider h defined on $[a, b]$ as $h(x) = (f(x) - f(a))\big(g(b) - g(a)\big) - (g(x) - g(a))(f(b) - f(a))$. Then h is continuous on $[a, b]$ and differentiable on (a, b). Also $h(a) = h(b) = 0$. By Rolle's Theorem there must be $c \in (a, b)$ such that $h'(c) = 0$. But hat is equivalent to the first assertion.

Now if $g'(x) \neq 0$ for all $x \in (a, b)$, then $g(b) \neq g(a)$ by the usual Mean Value Theorem. Thus, the second assertion follows. QED.

The usual MVT is the special case where $g(x) = x$.

Proof of l'Hôpital's Rule. The proof follows the one in "Heuser, *Lehrbuch der Analysis*". It has the advantage that it covers most cases with a minimum of case by case considerations. We will show that $\lim\limits_{x\to a^+} \frac{f(x)}{g(x)} = L = \lim\limits_{x\to a^-} \frac{f(x)}{g(x)}$ whenever this makes sense. Both limits exist and are equal if a is an interior point of I, and only one of them exists if a is a boundary point of I.

Suppose $(-\infty, a) \cap J \neq \emptyset$. Then we must show that $\lim\limits_{x\to a^-} \frac{f(x)}{g(x)} = L$.

Suppose first hat $L \in \mathbb{R} \cup \{-\infty\}$, and let $M > L$ arbitrary. We will show that there is $\delta_1 > 0$ such that $(a - \delta_1, a \subseteq J$ and $\frac{f(x)}{g(x)} < M$ for all $x \in (a - \delta_1, a)$.

To start let N be a number such that $L < N < M$. there is $\delta > 0$ such that $(a - \delta, a) \subseteq J$ and we have $\frac{f'(x)}{g'(x)} < N$ for all $x \in (a - \delta, a)$.

Let $y \neq x \in (a - \delta, a)$. Then $g(x) \neq g(y)$ by the MVT applied to $[x, y]$ (if $y > x$) or $[y. x]$ (if $y < x$). The generalized MVT then shows there is z between x, y such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(z)}{g'(z)} < N < M$$

(note that $z \in (a - \delta, a)$). This holds for all $y \neq x \in (a - \delta, a)$.

If 1. holds then this means $\lim_{y\to a}\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f(x)}{g(x)} \leq N < M$. For this we use that $g(x) \neq 0$ for x close to a (because $g(x) = 0$ for at most one $x \in (a-\delta, a)$ by the MVT).

If 2. holds, then $g(x), g(y) \neq 0$ for x, y close enough to a as well. Let us fix $y \in (a-\delta, a)$ such that $g(y) \neq 0$. If x is close enough to a, then $\frac{g(x)-g(y)}{g(x)} = 1 - \frac{g(y)}{g(x)} > 0$ by 2. The above shows that

$$\frac{f(x)-f(y)}{g(x)-g(y)}\frac{g(x)-g(y)}{g(x)} = \frac{f(x)-f(y)}{g(x)} < N\frac{g(x)-g(y)}{g(x)} = N - N\frac{g(y)}{g(x)}$$

And so

$$\frac{f(x)}{g(x)} < N - N\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)}$$

This holds for all x close enough to a. Now $\lim_{x\to a}\left(N - N\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)}\right) = N$. Thus, there is $0 < \delta' < \delta$ such that for all $x \in (a-\delta', a)$, we have $N - \frac{g(y)}{g(x)} + \frac{f(y)}{g(x)} < M$ and then for such x also $\frac{f(x)}{g(x)} < M$.

The upshot is for every $M > L$ we have $\delta_1 = \delta'$ such that for all $x \in (a-\delta_1, a)$

$$\frac{f(x)}{g(x)} < M$$

Now if $L \in \mathbb{R} \cup \{\infty\}$ we can mimic these arguments to construct δ_2 such that for all $M' < L$ and all $x \in (a-\delta_2, a)$, we have

$$\frac{f(x)}{g(x)} > M'$$

Indeed, by the generalized MVT for a chosen $M' < L$ and any N' with $M' < N' < L$

$$\frac{f(y)-f(x)}{g(y)-g(x)} = \frac{f'(z)}{g'(z)} > N' > M'$$

as long as x, y are close enough to a. In Case 1. the limit for $y \to a^-$ of the left hand side

$$\frac{f(x)}{g(x)} \geq N' > M'$$

And in Case 2. this means again for y fixed and x close enough to a

$$\frac{f(x)-f(y)}{g(x)} = \frac{f(x)-f(y)}{g(x)-g(y)}\frac{g(x)-g(y)}{g(x)} > N' - N'\frac{g(y)}{g(x)}$$

and so

$$\frac{f(x)}{g(x)} > N' - N'\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)}$$

The limit of the right hand side is N' and so the right hand side is $> M'$ for y close enough to a.

We therefore find δ_2 as claimed.

Now we are done: if L is infinite, $L = \infty$ say, then we have for any M' a δ_2 such that $\frac{f(x)}{g(x)} > M'$ for all $x \in (a - \delta_2, a)$. If $L = -\infty$ we have for every M a δ_1 such that $\frac{f(x)}{g(x)} < M$ on $(a - \delta_1, a)$.

For finite L we combine the two statements above: for $\varepsilon > 0$ let $M = L + \varepsilon$ and $M' = L - \varepsilon$ and put $\delta = \min\{\delta_1, \delta_2\}$. Then for all $x \in (a - \delta, a)$ we have

$$L - \varepsilon < \frac{f(x)}{g(x)} < L + \varepsilon$$

and it follows that $\lim_{x \to a^-} \frac{f(x)}{g(x)} = L$.

The case that $(a, \infty) \cap J \neq \emptyset$ is similar. Finally, if a is an inner point of I the arguments (applied to $J \cap (a, \infty)$ and $(-\infty, a) \cap J$ separately) given show that $\lim_{x \to a^+} \frac{f(x)}{g(x)} = \lim_{\mathrm{x} \to \mathrm{a}^-} \frac{f(x)}{g(x)}$. QED.

Corollary (L'Hôpital's Rule for $x \to \pm\infty$)
The theorem also holds for intervals $(-\infty, b)$ and (b, ∞) and $a = \pm\infty$. EOL.

Poof. We only must modify our argument slightly. We will discuss the case $a = \infty$. As before $g(x) \neq 0$ for x large enough. Indeed, $g'(x) \neq 0$ everywhere, means g is injective. So $g(x) = 0$ for at most one x. We may also assume that $b > 0$.

Consider the functions $F = f \circ \frac{1}{x}$ and $G = g \circ \frac{1}{x}$ on the interval $(0, b^{-1})$.

Then $\lim_{x \to 0} F(x) = \lim_{x \to \infty} f(x) = 0$. Similarly, $\lim_{x \to 0} G(x) = 0$.

Next, $F'(x) = f'\left(\frac{1}{x}\right)\frac{-1}{x^2}$ and $G'(x) = -\frac{g'\left(\frac{1}{x}\right)}{x^2} \neq 0$ on $(0, b^{-1})$. Finally,

$$\lim_{x \to 0} \frac{F'(x)}{G'(x)} = \lim_{x \to 0^+} \frac{f'\left(\frac{1}{x}\right)}{g'\left(\frac{1}{x}\right)} = L$$

We may now apply the theorem to conclude the result. QED.

Examples

1. We have seen $\lim_{x \to 0} \frac{\sin x}{x} = 1$. Using l'Hôspital's Rule we can verify this:
$$\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\cos x}{1} = 1$$
2. Let $f =$ exp be the exponential function. Let $g(x)$ be any polynomial, of degree $n \geq 0$, say. Then $\lim_{x \to \infty} \frac{f(x)}{g(x)} = \pm\infty$ with the sign equal to the sign of $g(x)$ for large x.
Indeed, by induction on n the result holds for smaller degree polynomials (and the result is trivial for degree 0). Now $\lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{x \to \infty} \frac{f(x)}{g'(x)} = \pm\infty$. As the sign of $g(x)$ is the same as the sign of $g'(x)$ for large x, the result follows.
This is an example where one has to apply the rule repeatedly.

3. With f as in 2., $f(-x) = f(x)^{-1}$. This immediately shows that $f(x) \to 0$ for $x \to -\infty$. Moreover, for every polynomial g, we find $\lim_{x\to-\infty} f(x)g(x) = 0$.
   Indeed, this follows as $\lim_{x\to-\infty} f(x)g(x) = \lim_{x\to\infty} \frac{g(-x)}{f(x)} = 0$ by 2.
   Similarly, $\lim_{x\to\infty} g(x)f(-x) = 0$ for every polynomial g.
4. Consider $h(x) = x\log\left(1+\frac{1}{x}\right)$ defined for $x > 0$. Then we can write $h(x) = \frac{f(x)}{g(x)}$ with $f(x) = \log\left(1+\frac{1}{x}\right)$ and $g(x) = \frac{1}{x}$. Note that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = 0$.
   We have $f'(x) = \frac{1}{1+\frac{1}{x}}\frac{-1}{x^2}$ and $g'(x) = -\frac{1}{x^2}$. Thus $\frac{f'(x)}{g'(x)} = \frac{1}{1+\frac{1}{x}} \to 1$.
   We find that $\lim_{x\to\infty} x\log\left(1+\frac{1}{x}\right) = 1$.
   Note, a similar argument shows $\lim_{x\to 0^+}\left(\frac{\log(1+x)}{x}\right) = 1$.
5. Compute $\lim_{x\to\infty}\frac{e^x+e^{-x}}{e^x-e^{-x}}$. Here the rule does not help at all. However,
$$\frac{e^x+e^{-x}}{e^x-e^{-x}} = \frac{1+e^{-2x}}{1-e^{-2x}} \to 1$$

Application

Part 4. of the example above shows that $\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e$.
Indeed, $\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = \lim_{x\to\infty}\exp\left(x\log\left(1+\frac{1}{x}\right)\right) = \exp\left(\lim_{x\to\infty}\left(x\log\left(1+\frac{1}{x}\right)\right)\right) = \exp(1)$.
Since $x_n = n$ is a particular sequence with limit ∞, We conclude that $\exp(1) = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$. EOA.

An interesting consequence of l'Hôspital's Rule is the following lemma (which can also be proved using the MVT as in a homework problem):

Lemma
Let I be an interval, $a \in I$, and let $f: I \to \mathbb{R}$ be a function differentiable on (at least) $I \setminus \{a\}$ and continuous on all of I.
Suppose $\lim_{x\to a} f'(x) = L$ exists and is finite. Then f is differentiable at a and $f'(a) = L$.
In particular, f is continuously differentiable at a. EOL.

Proof. By l'Hôspital's Rule,
$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a} = \lim_{x\to a}\frac{f'(x)}{1} = L$$

QED.

Example
Let again
$$f(x) = \begin{cases} 0; & x = 0 \\ e^{-\frac{1}{x^2}}; & x \neq 0 \end{cases}$$
Then f is continuous as $\lim_{x\to 0} e^{-\frac{1}{x^2}} = 0$. Moreover for $x \neq 0$, $f'(x) = -\frac{2}{x^3}e^{-\frac{1}{x^2}}$.
We will show that $\lim_{x\to 0} f'(x) = 0$, and conclude by the lemma that $f'(0) = 0$.

For this note that $-\frac{2}{x^3}e^{-\frac{1}{x^2}} = -\frac{2}{x^3 e^{\frac{1}{x^2}}}$. So it suffices to show that $\lim_{x\to 0}\left|x^3 e^{\frac{1}{x^2}}\right| = \infty$.

But $\lim_{x\to 0^+} x^3 e^{\frac{1}{x^2}} = \lim_{x\to\infty}\frac{e^{x^2}}{x^3} = +\infty$, and $\lim_{x\to 0^-} x^3 e^{\frac{1}{x^2}} = \lim_{x\to\infty} -\frac{e^{x^2}}{x^3} = -\infty$. Together $\lim_{x\to 0}\left|x^3 e^{\frac{1}{x^2}}\right| = \infty$ follows. EOE.

5.7 Convex functions

Definition

Let I be an interval. A function f defined on I is called **convex**, if for all $a < b \in I$ we have

Equation 2

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

for all $x \in (a, b)$. It is called **strictly convex** if the inequality above is strict for all $x \in (a, b)$.
f is called **concave** if $-f$ is convex, and strictly concave, if $-f$ is strictly convex. EOD.

Convexity is fundamentally a geometric notion.

Recall that $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$ can be identified with the Euclidean plane.

A subset $C \subseteq \mathbb{R}^2$ is called **convex** if for every pair $p, q \in C$, the *line segment* $\{ap + (1 - a)q \mid 0 \leq a \leq 1\}$ is again a subset of C.

If you are familiar with linear algebra, the line through two points p, q is defined by the vector equation $x = p + t(q - p) = (1 - t)p + tq$ where t ranges over all real numbers.

The right-hand side of Equation 2 is the function whose graph is the line segment joining $(a, f(a))$ and $(b. f(b))$.

We can reparametrize this by putting $s = \frac{x-a}{b-a}$, then $x = (1 - s)a + sb$ and $s \in (0,1)$ whenever $x \in (a, b)$. Equation 2 then becomes

$$f\big(a + s(b - a)\big) \leq f(a) + \big(f(b) - f(a)\big)s = (1 - s)f(a) + sf(b)$$

for all $s \in (0,1)$. Rewriting it slightly again, we get

$$f\big((1 - s)a + sb\big) \leq (1 - s)f(a) + sf(b)$$

f is convex if this inequality holds for all $s \in (0,1)$ and strictly convex, if it is a strict inequality for all $s \in (0,1)$.

For a concave function the inequalities are simply reversed.

Exercise

How that f is convex on an interval I if and only if for all $n \geq 2$ and all $t_1, t_2, \dots, t_n \in (0,1)$ with $t_1 + t_2 + \cdots + t_n = 1$ we have $f(t_1x_1 + t_2x_2 + \cdots + t_nx_n) \leq t_1f(x_1) + t_2f(x_2) + \cdots + t_nf(x_n)$ for all $x_1 \leq x_2 \leq \cdots \leq x_n \in I$.
(*Hint*: The definition gives you the case $n = 2$.) EOE.

In this context, the whole point of this business is to derive inequalities for the values of functions (which are always hard to come by).

Example

For $x, y > 0$ consider the following definitions of a mean value:

There is the **arithmetic mean** defined as $\frac{x+y}{2}$ and there is the **geometric mean** defined as $\sqrt{xy}$. Is there a relationship between them?

In fact, we have $\sqrt{xy} \leq \frac{1}{2}(x+y)$.

To prove this, it suffices to show that $\log \sqrt{xy} \leq \log\left(\frac{1}{2}(x+y)\right)$.

Note that $\sqrt{xy} = (xy)^{\frac{1}{2}}$. Then $\log(\sqrt{xy}) = \frac{1}{2}(\log x + \log y)$.

Suppose we knew that $\log x$ was concave (a.k.a concave down). Then we could conclude that

$$\log\left(\frac{1}{2}x + \frac{1}{2}y\right) \geq \frac{1}{2}\log(x) + \frac{1}{2}\log(y) = \frac{1}{2}(\log(x) + \log(y)) = \log\sqrt{xy}$$

as needed.

Using the exercise above, one can modify this argument to show that for $x_1, x_2, \ldots, x_n > 0$ we always have

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

EOE.

How could we possibly determine whether or not a function is convex or concave? The general idea is the Horse Race Theorem.

To be convex, the graph of a function must always stay below the secant between two points in the graph:

$$f(t) \leq f(a) + \frac{f(b) - f(a)}{b - a}(t - a)$$

for all $t \in [a, b]$.

Theorem (Convexity Theorem)

Suppose f is continuous on an interval I and differentiable on its interior I°. Then f is convex on I iff f' is monotone increasing on I°. f is concave on I iff f' is monotone decreasing on I°. If f' is strictly monotone increasing/decreasing on I°, then f is strictly convex/concave EOT.

Proof. We will prove the criterion for convexity. Applied to $-f$ the criterion for concavity follows.

Suppose f is convex, and let $a < b \in I^\circ$. We must show that $f'(a) \leq f'(b)$.

Now $\frac{f(b)-f(a)}{b-a} \geq \frac{f(t)-f(a)}{t-a}$ for all $t \neq a$. This shows that $\lim\limits_{t \to a^+} \frac{f(t)-f(a)}{t-a} = f'(a) \leq \frac{f(b)-f(a)}{b-a}$.

Similarly, $f(t) \leq f(b) + \frac{f(a)-f(b)}{b-a}(b-t)$, and so

$$\frac{f(b) - f(t)}{b - t} \geq \frac{f(b) - f(a)}{b - a}$$

for all $t \in [a, b)$, and so $f'(b) = \lim_{t \to b^-} \frac{f(b)-f(t)}{b-t} \geq \frac{f(b)-f(a)}{b-a} \geq f'(a)$.

Now suppose f' is increasing on I°. We must show that f is convex. Let $a < b \in I$. Then for any $x \in (a, b)$ we must have $\frac{f(x)-f(a)}{x-a} = f'(c_1)$ and $\frac{f(b)-f(x)}{b-x} = f'(c_2)$ for some $c_1 \in (a, x) \subseteq I^\circ$ and some $c_2 \in (x, b) \subseteq I^\circ$. In particular $c_1 < c_2$ so $f'(c_1) \leq f'(c_2)$.

Thus, for all $x \in (a, b)$, $\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(x)}{b-x}$

Note $x - a, b - x > 0$, and so

$$(b - x)\big(f(x) - f(a)\big) \leq (x - a)(f(b) - f(x))$$

or

$$(b - a)f(x) \leq (b - x)f(a) + (x - a)f(b)$$

Let $s = \frac{x-a}{b-a}$, then $x = (1 - s)a + sb$, so $b - x = (1 - s)b - (1 - s)a$, and $x - a = -sa + sb$.

The above then becomes

$$(b - a)f(x) \leq (1 - s)(b - a)f(a) + s(b - a)f(b)$$

or, as $b - a > 0$

$$f\big((1 - s)a + sb\big) \leq (1 - s)f(a) + sf(b)$$

for all $s \in (0,1)$. QED.

Note that we have shown that if f' is strictly increasing then f is strictly convex.

Corollary
Suppose f is defined and continuous on an interval I and twice differentiable on I°. Then

1. f is convex iff $f'' \geq 0$ on I°.
2. f is concave iff $f'' \leq 0$ on I°.

If $f'' > 0$ then f is strictly convext. If $f'' < 0$ then f is strictly concave.

Proof. $f'' \geq 0$ is equivalent to f' increasing on I°. By the theorem, this is equivalent to f convex. The same reasoning applied to $-f$ gives the second assertion.

We also know that if $f'' > 0$ on I° then f' is strictly increasing on I°, showing that f is strictly convex. Again, the same reasoning applied to $-f$ shows that if $f'' < 0$ then f is strictly concave.

Example

1. e^x is strictly convex.
2. $\log x$ is strictly concave on $I = (0, \infty)$. Indeed, the second derivative is $-\frac{1}{x^2} < 0$.
3. If f is a polynomial function of odd degree > 2, then f cannot be concave or convex everywhere. Indeed, f'' is a polynomial of odd degree and hence changes its sign somewhere.

Definition

Let f be defined on an interval I and $c \in I^\circ$ an interior point. Then c is called an **inflection point** if there is $\delta > 0$ such that f is convex on $(c-\delta, c]$ and concave on $[c, c+\delta)$ or vice versa f is concave on $(c-\delta, c]$ and convex on $[c, c+\delta)$. EOD.

To find inflection points, the following lemma is often useful.

Lemma
Let f be twice differentiable at an inflection point c. Then $f''(c) = 0$.

Proof. f is twice differentiable at c, so there is $\delta > 0$ such that both f' is defined on $(c-\delta, c+\delta)$ and f' is increasing on $(c-\delta, c]$ and decreasing on $[c, c+\delta)$ (or vice versa) by the Convexity Theorem. But this means f' has a local maximum at c (or a local minimum). In either case $f''(c) = 0$ by Lemma 5.2.2. QED.

Example
Let $f(x) = (1+x)\sqrt{1-x^2}$ defined on $[-1,1]$. Then f is differentiable (in fact, smooth) on $(-1,1)$ with

$$f'(x) = \frac{-2x^2 - x + 1}{\sqrt{1-x^2}}$$

$$f''(x) = \frac{2x^3 - 3x - 1}{(1-x^2)\left(\sqrt{1-x^2}\right)}$$

We find $f'(x) = 0$ for $x_1 = \frac{1}{2}$ (The numerator of f' has a second root namely -1, but the formula given for f' is not defined at -1.)
The numerator of $f''(x)$ has roots -1, $x_2 = \frac{1-\sqrt{3}}{2}$, and $\frac{1+\sqrt{3}}{2}$, of which only x_2 is in $(-1,1)$.
Note that $f''(x) > 0$ for $x < x_2$ and $f'' < 0$ for $x > x_2$.
We conclude the following f is (strictly) convex on $[-1, x_2]$, strictly concave $[x_2, 1]$, and has a local maximum at x_1. x_2 is the only inflection point.
It is elementary to check that $\lim_{x\to 1} f'(x) = -\infty$. This suggests that f is not differentiable at $x = 1$.
Indeed, as the limit exists, one can show that f cannot be differentiable at 1 (as the limit would necessarily be equal to that limit, as in the Lemma in 5.6).
On the other hand, L'Hôpital's Rule shows that $\lim_{x\to -1} f'(x) = 0$, and so f is differentiable there. EOE.

Exercise
Let $f: [a,b] \to \mathbb{R}$ be a convex function and suppose f is twice differentiable on I and that $f'' > 0$ on I (so f is strictly convex on I).
Show that if $x_0 \in I$ is arbitrary then for all $x \neq x_0 \in I$ we have

$$f(x) > f(x_0) + f'(x_0)(x - x_0)$$

In other words, the graph of f is above all of its tangents. If $f'' < 0$, then $f(x) < f(x_0) + f'(x_0)(x - x_0)$ for all $x \neq x_0 \in I$.

5.8 Some complements

5.8.1 The Intermediate Value Theorem for derivatives

We have discussed this in a live meeting. We know that continuous functions satisfy IVT.

On the other hand, we know that derivatives need not be continuous (see the last example in 5.5 above). Nevertheless derivatives are special, as they also satisfy the IVT.

Theorem (Intermediate Value Theorem for Derivatives; IVTD)
Let $f: [a, b] \to \mathbb{R}$ be a differentiable function. Suppose $f'(a) \neq f'(b)$. Then for each y between $f'(a)$ and $f'(b)$, there is $c \in (a, b)$ with $f'(c) = y$. EOT.

Proof. We first show that if $f'(a) < 0 < f'(b)$, then there is $c \in (a, b)$ such that $f'(c) = 0$. If $f'(a) < 0$, then we know that f does not have a local minimum at a: there is $\delta > 0$ such that $f(x) < f(a)$ for all $x \in (a, a + \delta)$. Similarly, there is (a possibly different) $\delta > 0$ such that $f(x) < f(b)$ for all $x \in (b - \delta, b)$. (See 5.2.1.)

f is continuous so by MMP, f attains its minimum, which is necessarily at a point $c \in (a, b)$. But then $f'(c) = 0$.

Now suppose $f'(a) < y < f'(b)$. Then consider the differentiable function $g(x) = f(x) - yx$ on $[a, b]$. $g'(a) < 0 < g'(b)$, so there is c such that $g'(c) = 0 = f'(c) - y$.

In case $f'(b) < f'(a)$ apply the above to $-f$. QED.

Example
There is no function $f: \mathbb{R} \to \mathbb{R}$ with $f' = H$, where H is the Heaviside function. EOE.

Definition
Let I be an interval with boundary $a < b$. A function f on I is called **piecewise continuous** if there are $x_1 < x_2 < \cdots < x_n \in I$ such that f is continuous on each open subinterval (x_i, x_{i+1}) where $x_0 = a$ and $x_{n+1} = b$, and if for all $i = 1,2, \dots, n$, the one-sided limits $\lim\limits_{x \to x_i^{\pm}} f(x)$ exist (and are finite).

Example
If a piecewise continuous function f is not continuous, then there is no function F such that $F' = f$.
To see this, suppose $x_1, x_2, \dots, x_n$ are as in the definition. As f is not continuous $n > 0$. We may also assume f is not continuous at x_1. Shrinking I, we may assume $I = (a, b)$ with $x_1 \in I$ and $x_2 > b$.
Then $\lim\limits_{x \to x_1^+} f(x) \neq f(x_1)$ or $\lim\limits_{x \to x_1^-} f(x) \neq f(x_1)$ (otherwise, f is continuous at x_1 and therefore continuous).
In the first case, we first assume $L = \lim\limits_{x \to x_1^+} f(x) < f(x_1)$. Then there is $\delta > 0$ such that $f(x) < \frac{f(x_1)+L}{2}$ for all $x \in (x_1 - \delta, x_1)$. If there is F with $F' = f$, then $F' = f$ also for the restriction of F to $(x_1 - \delta, x_1]$. There is no $c \in \left(x_1 - \frac{\delta}{2}, x_1\right)$ such that $F'(c) = f(c) = \frac{f(x_1)+L}{2}$, even though $F'\left(x_1 - \frac{\delta}{2}\right) < \frac{f(x_1)+L}{2} < F'(x_1)$. This is a contradiction.
The arguments in the remaining cases ($L > f(x_1)$), and $\lim\limits_{x \to x_1^-} f(x) \neq f(x_1)$ are too similar to warrant repetition. EOE.

5.8.2 The Horse Race Theorem*

Theorem
Let $I = [a, b]$ and f, g be continuous functions on I, differentiable on (a, b).
Suppose
1. $f(a) \geq f(b)$
2. $f' \geq g'$ on (a, b)

Then $f(b) \geq g(a)$. If $f' > g'$ on (a, b) then $f(b) > g(b)$.

Proof. The function $g = f - g$ is monotone increasing on $[a, b]$ as $h' \geq 0$ on (a, b). Then $h(b) \geq h(a) \geq 0$. If $f' > g'$, then $h' > 0$ and hence h is strictly increasing and $h(b) > 0$. QED.

Example
Consider $f(x) = \exp x$ and $g(x) = 1 + x$ on $[0, b]$ (for any $b > 0$). Then as f is strictly monotone and $f' = f$, we find that $f' > 1 = g'$. As $f(0) = g(0) = 1$, we have $\exp(x) > 1 + x$ for all $x > 0$ and consequently also $x > \log(1 + x)$.

5.8.3 Newton's Method

It is a well known theorem in algebra that there is no "formula" to compute roots of polynomials of degree 5 or higher. It is of course a philosophical question what exactly is meant by a formula. Since we have effective methods to compute nth roots, products, and sums of real numbers, a formula in this context is understood as a composition of these operations (and their inverses: powers, subtractions, quotients). There are fomulas for polynomial equations of degree 4 or less (albeit nasty in degrees 3 and 4).

But given the fact that a (general) real number can never be fully "known" (in the sense that we can never write down all its digit in a decimal expansion), having a formula is not really useful anyway. A numerical method, that is, a method that produces an approximation to the desired result to arbitrarily fine precision is all we can hope for anyway. Think about it for a second: while we all know that the solutions for $x^2 = 2$ are $x_1 = \sqrt{2}$ and $x_2 = -\sqrt{2}$, these are mere symbols. In fact, the symbol $\sqrt{2}$ precisely means "the positive solution of $x^2 = 2$". In that sense we could just introduce more symbols and say A_f is the smallest solution of $f(x) = 0$. The fact that $\sqrt{2}$ is a useful symbol rests on the fact that we have efficient algorithms to compute it to arbitrarily precision. The fact that A_f is not a useful symbol rests on the fact that in general there may not be any efficient algorithm. But the point here is that this is not exactly a theoretical distinction, but rather an arbitrary choice: we are comfortable with using symbols for solutions of certain quadratic equations. We are not comfortable with using symbols for solutions of arbitrary quintic equations.

A naïve approach to find a root of a function f is the following: choose x_0, compute $f(x_0)$ and $f'(x_0)$. If $f(x_0) = 0$ you are done. Otherwise, you want to move "closer" to a root of f. If you know nothing about f other than the two values $f(x_0)$ and $f'(x_0)$, then a reasonable approach is to go in the direction of the tangent of the graph of f at $(x_0, f(x_0))$ that brings you closer to 0.

The functional equation of the tangent is $t(x) = f(x_0) + f'(x_0)(x - x_0)$

In other words: if $f(x_0) > 0$ and $f'(x_0) < 0$, then t has a root $x_1 > x_0$. If $f'(x_0) < 0$, then t has a root $x_1 < x_0$. The inequalities are reversed if $f(x_0) < 0$.

One can now repeat the process with x_1 in place of x_0, and so on. This results in a sequence $x_0, x_1, x_2, \ldots$

Lemma
Suppose f is defined and differentiable at each x_n constructed above, and $f'(x_n) \neq 0$ for all n. Suppose x_n converges to $x_\infty \in \mathbb{R}$ and f' is defined and continuous at x_∞ and $f'(x_\infty) \neq 0$. Then $f(x_\infty) = 0$. EOL.

Proof. Let $t_n(x) = f(x_n) + f'(x_n)(x - x_n)$. Then $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$. As f' is defined and continuous at x_∞, f is also continuous at x_∞, and we have $x_\infty = \lim_{n\to\infty} x_{n+1} = x_\infty - \frac{f(x_\infty)}{f'(x_\infty)}$. This is only possible if $f(x_\infty) = 0$. QED.

This procedure is known as Newton's Method. There are criteria to make sure that f satisfies the conditions of the lemma. For example if $[a, b]$ is a given interval on which f is twice differentiable and $f'' \geq 0$ or $f'' \leq 0$, and f' has no zeros, then the method works, provided $f(a) < 0 < f(b)$ or $f(b) < 0 < f(a)$ (so we know that f has a zero by the IVT).

Even if f does not satisfy such strict restrictions you would have to be rather unlucky to randomly hit a point where $f'(x_n) = 0$. But see for example (Heuser, *Lehrbuch der Analysis Teil 1*, 10th edition, 1993, p 407): the function $f(x) = -x^4 + 6x^2 + 11$. If you start with $x_0 = 1$, then $x_{2n} = 1$ and $x_{2n+1} = -1$ for all $n > 0$. Indeed, $f'(x) = -4x^3 + 12x$. So $f(1) = 16$ and $f'(1) = 8$. So $x_1 = 1 - 2 = -1$. Next, $f(-1) = 16$ and $f'(-1) = -8$. So $x_2 = -1 + 2 = 1$. And then the pattern repeats. In this case Newton's Method does not work. Of course, for "reasonable" functions, the probability that you hit a starting point x_0 with such a problem is zero, provided you choose your starting point randomly.

5.8.4 Loose points

In these notes I is always an interval.

Remark
We have seen that if $I = [a, b]$ and f continuous on I, and differentiable on (a, b) then f is monotone increasing iff $f' \geq 0$ on (a, b). If f happens to be differentiable at a or b, and is increasing, say, is it still true that $f'(a) \geq 0$? Indeed, it is. For if $f'(a) < 0$, we know that there is $\delta > 0$ such that $f(x) < f(a)$ whenever $x > a$ and $|x - a| < \delta$. This contradicts that f is strictly monotone increasing. The same applies to $f'(b)$ if it is defined. So for functions that are differentiable on the entire interval $[a, b]$ the statement that $f' \geq 0$ for monotone increasing f and $f' \leq 0$ for monotone decreasing f still holds (on $[a, b]$). And the converse is obviously also true. There is a similar version of this argument for half-open intervals. EOR.