Contents

1 Introduction: A Running Example ................................................................. 3
  1.1 Geometric interpretation ........................................................................... 5
  1.2 The derivative ......................................................................................... 6

2 LIMITS: classification and properties .......................................................... 6
  2.1 Classification of Limits ............................................................................ 6
  2.2 Properties of Limits .................................................................................. 8
  2.3 Examples .................................................................................................. 9

3 DERIVATIVES ................................................................................................. 11
  3.1 Visualizing the derivative ......................................................................... 12
  3.2 Rules for computation of Derivatives ....................................................... 16
    3.2.1 Constants, sums and differences ......................................................... 16
    3.2.2 Power Rule (positive integer case) ..................................................... 17
    3.2.3 Product Rule ...................................................................................... 18
    3.2.4 Quotient Rule ..................................................................................... 19
    3.2.5 Power Rule (negative integer power case) ....................................... 20
1 **Introduction: A Running Example**

You are a running coach. Your athlete is doing laps around the Chestnut Hill Reservoir and you want to know how fast she’s running. Actually, you are the kind of coach who cares about details, so you want to know *exactly* how fast she’s running.

Dragging your heel across the path, you draw a line $L$ in the dirt, in order to measure her speed as she crosses this line $L$. Just as she crosses $L$ you click the stopwatch. Right after she crosses $L$ you click the stopwatch again, observing her position at this slightly later time. You measure this distance of this later position from $L$ while she continues on the next lap.

Let’s say she was $d$ feet beyond $L$ at your second click. And say your second click was $h$ seconds after she crossed your line $L$. Assuming her speed did not change very much between your two clicks, her speed while crossing $L$ was approximately $d/h$ feet per second.
If her speed was actually constant between the two clicks, it would be exactly \( \frac{d}{h} \). But she is human, so her speed is not constant.

You are not satisfied because you want to know exactly how fast she’s running when she crosses the line \( L \). If you had clicked faster, say at an interval \( h_2 < h \), you would also have a different distance \( d_2 < d \), and \( \frac{d_2}{h_2} \) would be a better approximation to her actual speed while crossing \( L \). But there has to be some time interval between the clicks. Otherwise you have just one click, which tells you nothing because she travels 0 feet in 0 seconds no matter how fast she’s running. So your \( h \) must be positive, but smaller and smaller.

To tell this story in mathematical language, let \( f(t) \) be the distance, measured in feet, that she has covered for the first \( t \) seconds of the run. Say she crossed \( L \) the first time when \( t = 12 \), and you clicked again \( h \) seconds later. So \( d = f(12 + h) - f(12) \), and her approximate speed is

\[
\frac{f(12 + h) - f(12)}{h}.
\]

If you click faster, with shorter and shorter intervals between clicks, you get a better approximation, so her actual speed at time \( t = 12 \) is what the numbers \( (f(12 + h) - f(12))/h \) are getting close to as \( h \) gets close to zero. We write this as

\[
\text{Actual Speed at 12 seconds} = \lim_{h \to 0} \frac{f(12 + h) - f(12)}{h}.
\]

Just as we cannot make just one click, we cannot set \( h = 0 \) right away, since that would give \( (f(12) - f(12))/0 = 0/0 \), which makes no sense. Let’s call this the “0/0 ambiguity”. Usually it takes some work to make the 0/0 ambiguity go away, and thereby uncover the behavior of some ratio like \( (f(12 + h) - f(12))/h \) as \( h \) gets close to zero.

For example, say \( f(t) = t^2 \). Then

\[
\frac{f(12 + h) - f(12)}{h} = \frac{(12 + h)^2 - (12)^2}{h} = \frac{144 + 24h + h^2 - 144}{h} = 24 + h.
\]

Now the 0/0 ambiguity is gone. As \( h \) gets close to zero, \( 24 + h \) gets close to 24. Therefore, her actual speed at time \( t = 12 \) is

\[
\lim_{h \to 0} \frac{f(12 + h) - f(12)}{h} = \lim_{h \to 0} (24 + h) = 24 \text{ ft/sec}.
\]

**Question:** If you rode a bicycle alongside your runner, with a speedometer on the bike that measures speed in terms of wheel radius and RPM’s, you again get only an approximate value of her (and your) speed. How does his approximation differ from the one described above?
1.1 Geometric interpretation

Your runner’s average speed

\[ \frac{f(12 + h) - f(12)}{h} \]

in the time interval \([12, 12+h]\) is also the slope of the secant line\(^1\) through the points \((12, f(12))\) and \((12 + h, f(12 + h))\) on the graph of \(f(t)\). As \(h\) gets close to zero, the secant line approaches the tangent line\(^2\) and the slope of the secant line approaches the slope of the tangent line. So the runner’s exact speed when \(t = 12\) equals the slope of the tangent line to the graph of \(f(t)\) at the point \((12, f(12))\).

---

\(^1\)from the Latin *secare*, meaning “to cut”.

\(^2\)from the Latin *tangere*, meaning “to touch”.

1.2 The derivative

There was nothing special about the time $t = 12$; your runner has a speed at any time $t$. Just as the distance travelled up to time $t$ is the function $f(t)$, the speed at time $t$ is a new function $f'(t)$, called the derivative of $f(t)$.

At any time $t$, the value of the derivative $f'(t)$ is given by the limit of averages, or limit of slopes:

$$f'(t) = \lim_{h \to 0} \frac{f(t + h) - f(t)}{h}.$$ 

For example, if $f(t) = t^2$, then

$$f'(t) = \lim_{h \to 0} \frac{(t + h)^2 - t^2}{h} = \lim_{h \to 0} \frac{2th - h^2}{h} = \lim_{h \to 0} (2t - h) = 2t.$$ 

That is, if $f(t) = t^2$ then $f'(t) = 2t$.

2 LIMITS: classification and properties

The derivative is defined as a limit, so we must understand limits before we can better understand derivatives. Not all limits are as simple as the one we just computed.

First we must define what limits are. Our definition will be precise enough to understand and solve all of the calculus problems in this course, but will be informal enough to be easily grasped.

A limit is a value that a function is heading towards. If $F$ is our function and $L$ is the value, we say that

$$\lim_{h \to 0} F(h) = L$$

if the values of $F(h)$ get close and stay close to $L$ as $h$ gets close to zero but is not zero. Sometimes we also express this by writing

$$F(h) \to L, \quad \text{as} \quad h \to 0.$$ 

Note that $h$ can get close to zero from either side, positive or negative. In the limit above we mean the values of $F(h)$ get close to $L$ as $h \to 0$ from either side. Very rarely we’ll have $h \to 0$ from the positive side only or negative side only. We’ll specify this by writing $h \to 0^+$ or $h \to 0^-$, respectively.

2.1 Classification of Limits

In practice, there are five kinds of limits.
1. **Constant limits.** If \( F, \) is a constant function, say \( F(h) = C \) for all \( h, \) then it is not varying at all, so we have
\[
\lim_{h \to 0} F(h) = C.
\]

2. **Trivial limits.** For these limits you can just plug in \( h = 0 \) without doing anything else; you get
\[
\lim_{h \to 0} F(h) = F(0)
\]
right away. Examples:
\[
\lim_{h \to 0} (h + 2) = 0 + 2 = 2, \quad \lim_{h \to 0} (2x + h) = 2x + 0 = 2x.
\]
Note in the second example that \( x \) is regarded as a constant, because only \( h \) is varying in the limit.

3. **Easy Limits.** These are limits that can be reduced to trivial limits with a little bit of algebra or properties of limits (which we'll discuss shortly). For example, we have seen that if \( f(x) = x^2, \) then
\[
\lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h} = \lim_{h \to 0} (2x + h) = 2x.
\]

4. **Hard Limits.** These are the most interesting limits; they cannot be easily converted to trivial limits, but require ingenuity or more powerful tools. Examples:
\[
\lim_{h \to 0} \frac{\sin h}{h}, \quad \lim_{h \to 0} (1 + h)^{1/h}.
\]
In these limits there is a tension between different parts of the function \( F(h). \) The 0/0 ambiguity is one such tension, as in the limit of \( (\sin h)/h. \) Here the numerator is making the function get small, while the denominator makes the function get big.

The limit of \( (1 + h)^{1/h} \) has another kind of tension. Here \( 1 + h \to 1, \) and any power of 1 is 1. On the other hand, \( 1/h \to \infty \) (see below), and if \( a > 1 \) then \( a^{1/h} \to \infty. \) However, both \( (1 + h) \) and \( 1/h \) are moving simultaneously, and it turns out that \( (1 + h)^{1/h} \) does get close to a definite number, which is slightly larger than 1. We'll find out what happens to limit later in the course.

5. **Infinite Limits.** In this limits \( F(h) \) gets and stays larger or smaller without bound. Example:
\[
\lim_{h \to 0^+} \frac{1}{h} = \infty, \quad \lim_{h \to 0^-} \frac{1}{h} = -\infty.
\]

6. **Limits that do not exist.** Here \( F(h) \) does not tend towards any fixed number or \( \pm \infty. \) Example:
\[
\lim_{h \to 0} \sin \left( \frac{1}{h} \right).
\]
This function oscillates between ±1 faster and faster as \( h \) approaches zero. The picture shows only part of the graph. In fact this function cannot be completely graphed on any computer because no matter how small the pixels are, close to the origin the function crosses the \( x \)-axis infinitely many times within one pixel width.

### 2.2 Properties of Limits

Familiar properties of numbers transfer to become properties of limits. To list these properties, let \( F(h) \), \( G(h) \) be two functions defined for small numbers \( h \), except possibly for \( h = 0 \). Assume that \( L, M \) are finite values such that the limits

\[
\lim_{h \to 0} F(h) = L, \quad \lim_{h \to 0} G(h) = M
\]

all exist. Note that these must be finite limits for the properties below to hold.
Arithmetic Properties.

\[
\lim_{h \to 0} (F(h) + G(h)) = L + M, \quad \lim_{h \to 0} (F(h) - G(h)) = L - M
\]

\[
\lim_{h \to 0} (F(h) \cdot G(h)) = L \cdot M, \quad \lim_{h \to 0} \frac{F(h)}{G(h)} = \frac{L}{M},
\]

where in the last quotient we require that \( M \neq 0 \). As a special case of the product and constant properties, if \( c \) is a constant, we have

\[
\lim_{h \to 0} cF(h) = c \lim_{h \to 0} F(h) = cL.
\]

You can always pull a constant out of a limit.

Inequality Property.

Assume \( F(h) \leq G(h) \) for all \( h \) where the two functions are defined, then

\[
\lim_{h \to 0} F(h) \leq \lim_{h \to 0} G(h).
\]

Finally, we come to the most useful and important property of limits. Here one function is squeezed between two other functions which both have the same limit \( L \). Then the middle function also has limit \( L \). This is called the Squeeze Law. Here is its statement in more detail.

**Squeeze Law.** Suppose \( F(h), G(h), K(h) \) are three functions defined for numbers \( h \) near zero, such that the following two conditions hold:

1. \( F(h) \leq G(h) \leq K(h) \) for all \( h \),

2. \( \lim_{h \to 0} F(h) = L = \lim_{h \to 0} K(h) \).

Then

\[
\lim_{h \to 0} G(h) = L.
\]

In practice, \( F(h) \) and \( K(h) \) will have trivial limits, but \( G(h) \) will have a hard limit. So the Squeeze law helps us compute hard limits. (see example 3 below).

### 2.3 Examples

1. First, a trivial limit.

\[
\lim_{h \to 0} (3h + 5) = \lim_{h \to 0} (3h) + \lim_{h \to 0} 5 = 3 \lim_{h \to 0} h + 5 = 5.
\]
So you just end up plugging $h = 0$ into $F(h) = 3h + 5$.

2. Now for a nontrivial but easy limit:

$$\lim_{h \to 0} \frac{h^2 + h}{h^2 - h} = \lim_{h \to 0} \frac{h + 1}{h - 1} = \frac{0 + 1}{0 - 1}.$$  

Here we could not directly evaluate $(h^2 + h)/(2 - h)$ at $h = 0$, but we used algebra to make it a trivial limit and then used the quotient property.

3. Now for a hard limit:

$$\lim_{h \to 0} h \sin \left(\frac{1}{h}\right) = ?$$

Note that we cannot use the product property, since, as we have seen,

$$\lim_{h \to 0} \sin \left(\frac{1}{h}\right) \text{ does not exist.}$$

However, note that

$$-1 \leq \sin \left(\frac{1}{h}\right) \leq 1.$$  

If $h > 0$ we can multiply by $h$ to get

$$-h \leq h \sin \left(\frac{1}{h}\right) \leq h.$$  

If $h < 0$ we can multiply by $h$ to get (reversing the inequalities)

$$h \leq h \sin \left(\frac{1}{h}\right) \leq -h.$$  

Therefore in all cases we have

$$-|h| \leq h \sin \left(\frac{1}{h}\right) \leq |h|.$$  

We can use the Squeeze Law with

$$F(h) = -|h|, \quad G(h) = |h|.$$  

These have trivial limits

$$\lim_{h \to 0} F(h) = 0 = \lim_{h \to 0} G(h).$$  

From the Squeeze Law, we conclude that

$$\lim_{h \to 0} h \sin \left(\frac{1}{h}\right) = 0.$$  

In this picture you can see the squeezing. Again it is only a partial picture, since the complete graph cannot be drawn.
We will use the Squeeze Law to compute most of the important hard limits used in this course.

**Remark:** Our discussion of limits has been informal, intended to give a working understanding only. This will be enough for this course. However, to go more deeply into calculus, it is necessary to have a more precise understanding of limits. Students who are curious about this deeper perspective and learn more about it in the appendix (p.501-517) in Lang.

Assuming these foundational notions about limits, we will prove everything from now until we get to the deeper theorems of Calculus.

### 3 DERIVATIVES

Now that we learned something about limits, we return to derivatives.

First a slight change in notation. Usually \( t \) denotes time, but our independent variable could represent many quantities, such as a distance or an angle or an amount of something. So we usually call the independent variable \( x \), instead of \( t \).
Thus, the derivative of a function \( f(x) \) is the new function
\[
f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h},
\]
provided the limit exists. If the limit does indeed exist, we say \( f'(x) \) is **differentiable** at \( x \).

In can happen that the limit (1) does not exist. In that case we say that \( f(x) \) is not differentiable at \( x \).

The quotient on the right side of (1) is called the **difference quotient**. Other notations for the derivative \( f'(x) \) are
\[
f' = \frac{df}{dx} = \dot{f}(x).
\]

If we have a specific formula, like \( f(x) = x^2 \) or \( f(x) = \sin x \), we often write \( f'(x) = (x^2)' \) or \( f'(x) = (\sin x)' \).

The limit formula (1) is the official definition of the derivative. It may seem intimidating at first. We will develop some handy rules for computing derivatives that allow you to sometimes bypass the official definition (1). But throughout the course we will from time to time encounter new derivatives that will not be computable using our rules; in these cases we will return to (1) to compute the derivative, often using the Squeeze Law.

First, let us try to visualize the derivative.

### 3.1 Visualizing the derivative

As in our runner example, if \( x \) is time and \( f(x) \) is the distance covered by a moving object at time \( x \), then \( f'(c) \) is the speed of the object at time \( c \).

In any context, \( f'(c) \) is the exact rate of change of \( f(x) \) with respect to \( x \) at the point \( x = c \).

Sometimes we are happy just to know the sign of \( f'(c) \). If \( f'(c) > 0 \) then \( f(x) \) is increasing at \( c \) (our runner is at least going forward). If \( f'(c) < 0 \) then \( f(x) \) is decreasing at \( c \) (oh no she’s going backwards). If \( f'(c) = 0 \) then anything can happen (she stopped and now could head in either direction). Geometrically, \( f'(c) \) is the slope of the tangent line to the graph of \( f \) at the point \( (c, f(c)) \). The tangent line is horizontal exactly when \( f'(c) = 0 \).

We will illustrate this right now, and will explain the computations a bit later.

**Example 1:** Take
\[
f(x) = x^3 - 6x^2 + 11x - 4.
\]

It turns out that the derivative is
\[
f'(x) = 3x^2 - 12x + 11.
\]
Using the quadratic formula, we find that $f'(x)$ has zeros at $x = 2 \pm \frac{1}{\sqrt{3}}$, and that $f'(x)$ is positive, negative, positive for

$$x < 2 - \frac{1}{\sqrt{3}}, \quad 2 - \frac{1}{\sqrt{3}} < x < 2 + \frac{1}{\sqrt{3}}, \quad 2 + \frac{1}{\sqrt{3}} < x.$$
Not all examples are so easy to visualize.

**Example 3:** Take

\[
f(x) = \begin{cases} 
  x^2 \sin(1/x) & \text{for } x \neq 0 \\
  0 & \text{for } x = 0.
\end{cases}
\]

Let us examine the graph near \(x = 0\). None of the fancy techniques to come will help us compute \(f'(0)\), for example it doesn’t help to know \((\sin x)'\). On the bright side, this means we can actually compute \(f'(0)\) right now, straight from the definition via the difference quotient:

\[
f'(0) = \lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} \frac{h^2 \sin(1/h)}{h} = \lim_{h \to 0} h \sin(1/h) = 0.
\]

Recall we computed this last limit when learning the Squeeze Law. Our comments there about \(x \sin(1/x)\) apply here as well: the graph of \(f(x)\) around 0 cannot be physically drawn, either on paper or a computer screen. Nevertheless we have been able to determine that the tangent line to the graph of \(f(x)\) at \(x = 0\) is the \(x\)-axis.
In our previous examples we could determine whether $f(x)$ is increasing or decreasing on either side of a critical point. In this example, is $f(x)$ increasing or decreasing on either side of the critical point $x = 0$?

This reasonable-sounding question is actually too naive. We will later see that if you take any number $c > 0$, say $c = 10^{-18}$, there are infinitely many critical points of $f(x)$ in the interval $(0, c)$, and $f'(x)$ changes sign infinitely often in $(0, c)$.

This example shows that we cannot always rely on our geometric intuition when analyzing
3.2 Rules for computation of Derivatives

We now embark on several rules for computing derivatives in many cases.

3.2.1 Constants, sums and differences

There are two things to remember about constants:

1. The derivative of a constant function is zero. That is \((c)' = 0\) if \(c\) is a constant function.
   Proof:
   \[
   (c)' = \lim_{h \to 0} \frac{c - c}{h} = \lim_{h \to 0} \frac{0}{h}.
   \]
   Note this is not a 0/0 ambiguity. For in the limit \(h\) is getting close to zero, but is not zero.

2. You can pull a constant out of the derivative. That is, if \(f\) is differentiable at \(x\) and \(c\) is a constant then we have
   \[
   (cf(x))' = cf'(x).
   \]
   Proof:
   \[
   (cf(x))' = \lim_{h \to 0} \frac{cf(x + h) - cf(x)}{h} = \lim_{h \to 0} \frac{c(f(x + h) - f(x))}{h} = c \lim_{h \to 0} \frac{f(x + h) - f(x)}{h},
   \]
   using the constant rule for limits.

For example, \((2x^2)' = 2(x^2)' = 2 \cdot 2x = 4x\).

**Warning:** You must remember what kind of object the letter stands for. If \(c\) is a nonzero constant and we are differentiating with respect to a variable such as \(x\), then then \((c^2)'\) is **not** equal to \(2c\). In fact \((c^2)' = 0\), because \(c^2\) is also a constant.

Next we have the rules
\[
(f + g)'(x) = f'(x) + g'(x), \quad \text{and} \quad (f - g)'(x) = f'(x) - g'(x)
\]
For any two differentiable functions \(f(x)\) and \(g(x)\), as follows easily from the corresponding rules for limits.
3.2.2 Power Rule (positive integer case)

You can differentiate positive integral powers of \( x \) like this:

- \((x)^{'} = x\).
- \((x^2)^{'} = 2x\).
- \((x^3)^{'} = 3x^2\).
- \((x^n)^{'} = nx^{n-1}\) for any positive integer power \( n \) of \( x \).

**Proof:** Since we don’t know much about derivatives yet, we’ll have to use the definition of the derivative via the difference quotient:

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

\[
(x)^{'} = \lim_{h \to 0} \frac{(x + h) - x}{h} = \lim_{h \to 0} \frac{h}{h} = \lim_{h \to 0} 1 = 1
\]

\[
(x^2)^{'} = \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h} = \lim_{h \to 0} \frac{2xh + h^2}{h} = \lim_{h \to 0} (2x + h) = 2x
\]

\[
(x^3)^{'} = \lim_{h \to 0} \frac{(x + h)^3 - x^3}{h} = \lim_{h \to 0} \frac{3x^2h + 3xh^2 + h^3}{h} = \lim_{h \to 0} (3x^2 + 3xh + h^2) = 3x^2.
\]

For a general positive integer \( n \) we have the **binomial expansion**

\[
(x + h)^n = x^n + \binom{n}{1} x^{n-1} h + \binom{n}{2} x^{n-2} h^2 + \binom{n}{3} x^{n-3} h^3 + \cdots + \binom{n}{n-1} x h^{n-1} + h^n,
\]

where \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \) and the **factorial** function \( n! = 1 \cdot 2 \cdot 3 \cdots n \). In particular \( \binom{n}{1} = n \). For us, the point of the binomial expansion is that

\[
\frac{(x + h)^n - x^n}{h} = nx^{n-1} + (\text{terms involving } h),
\]

so

\[
\lim_{h \to 0} \frac{(x + h)^n - x^n}{h} = nx^{n-1}.
\]

**QED**

Now you can differentiate any polynomial, for example,

\[ (3x^3 - 4x^2 + 3x - 1)^{'} = 9x^2 - 8x + 3. \]
3.2.3 Product Rule

By analogy with sums, it is natural to assume that \((f \cdot g)'(x) = f'(x) \cdot g'(x)\), but this is false. For example:
\[
(x \cdot x)' = (x^2)' = 2x, \quad \text{but} \quad (x)' \cdot (x)' = 1 \cdot 1 = 1.
\]

The correct Product Rule is as follows:

Assume \(f'(x)\) and \(g'(x)\) exist. Then \((f \cdot g)'(x)\) exists and is given by the formula
\[
(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x).
\]

**Proof:** From the definition of derivative we have
\[
(f \cdot g)'(x) = \lim_{h \to 0} \frac{(f \cdot g)(x + h) - (f \cdot g)(x)}{h} = \lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x)g(x)}{h}.
\]

Now we want to somehow factor out the difference quotients for \(f'(x)\) and \(g'(x)\), but it is not clear how to do this, so we seem to be stuck. However, we can use a trick: We will subtract and add \(f(x + h)g(x)\) to the numerator and denominator and then we will be able to factor. Here we go:
\[
(f \cdot g)'(x) = \lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x)g(x)}{h} \quad \text{(from above)}
= \lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x + h)g(x) + f(x + h)g(x) - f(x)g(x)}{h} \quad \text{(the trick)}
= \lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x + h)g(x)}{h} + \lim_{h \to 0} \frac{f(x + h)g(x) - f(x)g(x)}{h} \quad \text{(separating middle sum)}
= \left(\lim_{h \to 0} f(x + h)\right) \lim_{h \to 0} \frac{g(x + h) - g(x)}{h} + \lim_{h \to 0} f(x + h) - f(x) \quad \text{(factoring)}
= \left(\lim_{h \to 0} f(x + h)\right) g'(x) + g(x) \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \quad \text{(definition of derivatives, which exist by assumption)}
\]

We’ll be done with the proof of the product rule, once we show that
\[
\lim_{h \to 0} f(x + h) = f(x).
\]

Since \(f(x)\) does not depend on \(h\), we have \(\lim_{h \to 0} f(x) = f(x)\). Therefore we must show that
\[
\lim_{h \to 0} f(x + h) = \lim_{h \to 0} f(x).
\]

By the properties of limits, this amounts to showing that
\[
\lim_{h \to 0} (f(x + h) - f(x)) = 0.
\]
But
\[
\lim_{h \to 0} (f(x + h) - f(x)) = \lim_{h \to 0} \left( \frac{f(x + h) - f(x)}{h} \right) \cdot h
\]
\[
= \lim_{h \to 0} \left( \frac{f(x + h) - f(x)}{h} \right) \cdot \left( \lim_{h \to 0} h \right) \quad \text{(because both limits exist)}
\]
\[
= f'(x) \cdot 0 = 0,
\]
as desired. We have shown that equation (2) is true, and therefore the product rule is true.
QED

We will use the product rule on a daily basis. For now we just do one very useful example.

Example: Differentiate the function \( f(x) = \sqrt{x} \). First note that \( x = f(x) \cdot f(x) \), so by the product rule we have
\[
(x)' = f'(x)f(x) + f(x)f'(x) = 2f(x)f'(x).
\]
Since \( x' = 1 \), we can solve for \( f'(x) \) and get \( f'(x) = 1/(2f(x)) \). That is,
\[
(\sqrt{x})' = \frac{1}{2\sqrt{x}}.
\]

3.2.4 Quotient Rule

This will be easier, since the hard work as been done when we proved the product rule. First, a special, but useful case:

If \( f'(x) \) exists and \( f(x) \neq 0 \), then \((1/f)'(x) \) exists and we have
\[
\left( \frac{1}{f} \right)'(x) = -\frac{f'(x)}{f(x)^2},
\]
(3)

Proof: Since
\[
f \cdot \frac{1}{f} = 1,
\]
and 1 is a constant function, we have
\[
\left( f \cdot \frac{1}{f} \right)' = 0.
\]

On the other hand, by the product rule, we have
\[
\left( f \cdot \frac{1}{f} \right)'(x) = f'(x) \left( \frac{1}{f(x)} \right) + f(x) \left( \frac{1}{f(x)} \right)',
\]
so
\[
f'(x) \left( \frac{1}{f(x)} \right) + f(x) \left( \frac{1}{f(x)} \right)' = 0.
\]
solving for \( \left( \frac{1}{f(x)} \right)' \) gives the result.

Now we can prove the quotient rule:

\[
(\frac{f}{g})' (x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}
\]

(quotient rule)

**Proof:** We write \( f/g \) as \( f \cdot (1/g) \) and apply the product rule:

\[
(\frac{f}{g})' (x) = \left( f \cdot \frac{1}{g} \right)' (x)
\]
\[
= f'(x) \frac{1}{g(x)} + f(x) \cdot \left( \frac{1}{g} \right)' (x)
\]
\[
= f'(x) \frac{1}{g(x)} + f(x) \cdot \left( -\frac{g(x)}{g(x)^2} \right)
\]
\[
= \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2},
\]

as claimed. QED

### 3.2.5 Power Rule (negative integer power case)

You can differentiate negative integral powers of \( x \) like this:

- \((x^{-1})' = -x^{-2}\).
- \((x^{-2})' = -2x^{-3}\).
- \((x^{-3})' = -3x^{-4}\).
- \((x^{-n})' = -nx^{-n-1}\) for any negative integer power \(-n\) of \( x \).

**Proof:** This is now easy, because we have some tools

\[
(x^{-n})' = \left( \frac{1}{x^n} \right)' = -\frac{(x^n)'}{x^{2n}}
\]
\[
= -\frac{n \cdot x^{n-1}}{x^{2n}}
\]
\[
= -\frac{n \cdot x^{-n-1}}{x^{2n}}
\]

as claimed. QED
Note that both the positive and negative integer powers are differentiated in the same way:

For any integer \( n \), whether positive or negative, we have

\[
(x^n)' = nx^{n-1}
\]

integer power rule

### 3.2.6 Chain Rule

The derivative \( f'(x) \) measures how \( f \) changes with respect to changes in the variable \( x \).

Often functions depend on other functions. For example suppose

\( x = \) annual rainfall,

\( g(x) = \) amount of grass, which depends on the rainfall,

\( f(g) = \) number of antelope, which eat the grass,

\( \ell(f) = \) number of lions, which eat the antelopes.

Since the antelopes eat the grass which depends on rain, the number of antelopes is really a function of rainfall: the number of antelope is \( f(g(x)) \), which can also be written as \( (f \circ g)(x) \). Likewise the number of lions is \( \ell(f(g(x))) \).

The chain rule expresses the derivative of the last function with respect to \( x \) in terms of the derivatives of the intermediate functions. In the \( df/dx \) notation the chain rule is simply:

\[
\frac{df}{dx} = \frac{df}{dg} \frac{dg}{dx}, \quad \frac{d\ell}{dx} = \frac{d\ell}{df} \frac{df}{dg} \frac{dg}{dx}.
\]

This looks easy, but can get a little confusing in practice. It is better to use the \( f'(x) \) notation. In this form, the chain rule is as follows.

\[
(f \circ g)'(x) = f'(g(x)) \cdot g'(x)
\]

chain rule (4)

In words: Differentiate the outside function, plug in the inside function, then multiply by the derivative of the inside function.

**Example 1:** Differentiate the function \( F(x) = (3x^2 + 5x + 7)^{10} \).

Here \( F(x) = f(g(x)) \), where \( f(g) = g^{10} \) and \( g(x) = 3x^2 + 5x + 7 \). By the chain rule we have

\[
F'(x) = \left((3x^2 + 5x + 7)^{10}\right)' = 10(3x^2 + 5x + 7)^9 \cdot (6x + 5).
\]

**Example 2:** Differentiate the function \( F(x) = \sqrt{x^2 + 4} \).
Here \( F(x) = f(g(x)) \), where \( f(g) = \sqrt{g} \) and \( g(x) = x^2 + 4 \). Recall that \((\sqrt{x})' = 1/2\sqrt{x}\). By the chain rule we have
\[
F'(x) = \frac{1}{2\sqrt{x^2 + 4}} \cdot (2x) = \frac{x}{\sqrt{x^2 + 4}}.
\]

**Example 3:** Differentiate the function \( F(x) = \sqrt{1 + \sqrt{1 + \sqrt{x}}} \).

Here \( F(x) = \ell(f(g(x))) \), where \( \ell(f) = 1 + \sqrt{f} \) and \( f(g) = 1 + \sqrt{g} \) and \( g(x) = 1 + \sqrt{x} \). By the chain rule we have
\[
F'(x) = \frac{1}{2\sqrt{1 + \sqrt{1 + \sqrt{x}}} \cdot \left(1 + \sqrt{1 + \sqrt{x}}\right)'}
= \frac{1}{2\sqrt{1 + \sqrt{1 + \sqrt{x}}} \cdot \frac{1}{2\sqrt{1 + \sqrt{x}}} \cdot (1 + \sqrt{x})'}
= \frac{1}{4\sqrt{1 + \sqrt{1 + \sqrt{x}}} \cdot \sqrt{1 + \sqrt{x}} \cdot \frac{1}{2\sqrt{x}}}
= \frac{1}{8\sqrt{1 + \sqrt{1 + \sqrt{x}}} \cdot \sqrt{1 + \sqrt{x}} \cdot \frac{1}{\sqrt{x}}}
\]

### 3.2.7 Inverse functions

Recall that \( f(x) \) and \( g(x) \) are called **inverse functions** if \( f(g(x)) = x \) and \( g(f(x)) = x \). For this to make sense, \( f \) must be defined at all values of \( g \), and vice versa. Note this is an inverse in the sense of *composition*, not multiplication. The multiplicative inverse \( 1/f(x) \) is completely different.

If \( f(x) \) and \( g(x) \) are inverse functions, one can compute the derivative of \( g(x) \) in terms of \( f(x) \), as follows.
\[
\boxed{g'(x) = \frac{1}{f'(g(x))}} \quad \text{derivative of inverse function}
\]

**Proof:** Differentiate both sides of the equation \( f(g(x)) = x \), using the chain rule on the left side. We find
\[
f'(g(x)) \cdot g'(x) = 1.
\]
Dividing both sides by \( f'(g(x)) \) gives the asserted formula for \( g'(x) \). QED

### 3.2.8 Proof of chain rule

We first give the precise statement of the Chain Rule:
If $x$ is a fixed number such that $g'(x)$ and $f'(g(x))$ both exist then $(f \circ g)'(x)$ exists and

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x).$$

**Proof:** We have to examine the limit

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

Set $y = g(x)$. The idea is to write $g(x+h)$ as $y+k$, where $k \to 0$, so that we'll have the difference quotient for $f'(y)$.

By assumption, the limit

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

exists. We can express this as

$$\lim_{h \to 0} \left( \frac{g(x+h) - g(x)}{h} - g'(x) \right) = 0.$$  

Since $x$ is fixed, we can think of the term in big parentheses as a function of $h$, which we’ll call $\gamma(h)$. So

$$\gamma(h) = \frac{g(x+h) - g(x)}{h} - g'(x),$$

and $\lim_{h \to 0} \gamma(h) = 0$.

We have $g(x + h) = g(x) + h(\gamma(h) + g'(x))$. According to our plan, we set $y = g(x)$ and $k = h(\gamma(h) + g'(x))$. Note that $y$ is fixed and $k \to 0$ as $h \to 0$. Now we’re ready to take the limit of the difference quotient for $f \circ g$ at $x$.

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

$$= \lim_{h \to 0} \frac{f(y+k) - f(y)}{h}$$

$$= \lim_{h \to 0} \frac{f(y+k) - f(y)}{k} \cdot \frac{k}{h}$$

$$= \lim_{h \to 0} \frac{f(y+k) - f(y)}{k} \cdot \lim_{h \to 0} (\gamma(h) + g'(x))$$

$$= f'(y) \cdot g'(x)$$

$$= f'(g(x)) \cdot g'(x),$$

as desired. **QED**

### 3.3 Implicit Differentiation

This is used when you have a relation between two variables, say $x$ and $y$, and you cannot (or do not want to) solve for one variable in terms of the other.
Example 1: Tangent line to a circle

For this we don’t need calculus. If our circle $C$ has radius $r$ and is centered at $(0,0)$, then its equation is $x^2 + y^2 = r^2$.

At a point $(x, y)$ on $C$, the tangent vector to $C$ is perpendicular to the line from $(0, 0)$ to $(x, y)$, so the tangent vector has slope $-x/y$.

Let us find this slope again, now using implicit differentiation. We differentiate both sides of the equation $x^2 + y^2 = r^2$ with respect to $x$. We know that $(x^2)' = 2x$. But note that $y$ is a(n implicit) function of $x$, so we have use the chain rule to see that $(y^2)' = 2yy'$, where $y' = dy/dx$.

Thus, differentiating both sides, we get

$$2x + 2yy' = 0.$$ 

This gives $y' = -x/y$ for the tangent slope, as before.

Example 2: An ellipse

Let $C$ be the curve with equation

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

First, what kind of curve is it? It is some kind of conic, therefore either an ellipse, hyperbola or parabola. If we try to solve for $y$ using the quadratic formula, we get

$$y = \frac{1}{2}(-x \pm \sqrt{4 - 3x^2}).$$

Since $y$ must be real, we must have $4 - 3x^2 \geq 0$, or $|x| \leq 2/\sqrt{3}$. If we had instead solved for $x$, we would have also found $|y| \leq 2/\sqrt{3}$. Hence the coordinates of the points on the curve are bounded, which means $C$ must be an ellipse.

Now we’ll find the slopes of its tangent lines. Implicitly differentiating, using the product rule for $xy$ and the chain rule for $y^2$, we get

$$2x + y + xy' + 2yy' = 0.$$ 

This leads to the formula

$$y' = -\frac{2x + y}{x + 2y}$$

for the slope of the tangent line to $C$ at the point $(x, y)$.

For example, where is the tangent line horizontal? This happens when $y' = 0$, which means $y = -2x$. But don’t forget that $x^2 + xy + y^2 = 1$, since $(x, y)$ is on $C$. Solving these two equations gives $x = \pm 1/\sqrt{3}$ and therefore $y = \mp 2/\sqrt{3}$. So the tangent line is horizontal at the two points $\left(1/\sqrt{3}, -2/\sqrt{3}\right), \left(-1/\sqrt{3}, 2/\sqrt{3}\right)$. Likewise the tangent line is vertical at the two points $\left(2/\sqrt{3}, -1/\sqrt{3}\right), \left(-2/\sqrt{3}, 1/\sqrt{3}\right)$. Here is a picture of the ellipse $C$ and these four tangent lines.
Example 3: A lemniscate

Fix two points $P, Q$ in the plane of distance $2d$ apart. The Lemniscate for $P, Q$ is the curve $L$ consisting of all points from which the product of the distances to $P$ and $Q$ is $d^2$. Note that the midpoint $O$ of $PQ$ is on $L$. If $P$ and $Q$ are the locations of planets of equal mass, then a satellite starting at $O$ will travel on $L$ without using any power. \(^3\)

Let us make $L$ the $x$-axis and choose $P = (1, 0)$ and $Q = (-1, 0)$, so that $d = 1$. The product of the distances of a point $R = (x, y)$ to $P$ and

\(^3\)A variant of the lemniscate, for unequal masses with the moon at $P$ and the earth at $Q$, was used during the Apollo missions to orbit and eventually land on the moon. On the lemniscate, gravity does all the work, so the early Apollo missions in the 1960s used this “free-return trajectory”, in case anything went wrong. But by 1970 NASA was apparently more confident; Apollo 13 was not on a free-return trajectory. Then suddenly, while outbound 200,000 miles from earth, one of its oxygen tanks exploded. This killed the main electrical power sources, and there was no longer enough power to simply turn the damaged craft around. Amazingly, the astronauts survived. A key part of the rescue involved finding the free-return trajectory (on the fly, in outer space), which was the lemniscate on which Apollo 13 could exploit the gravitational forces of the moon and earth to continue away, go around the moon and get back near earth, conserving all power for minimal life support and the final rocket burn to exit the lemniscate and re-enter our atmosphere at the exact angle between bouncing off or burning up. Missing the lemniscate would have sent the three astronauts into a fatal orbit around the moon, the earth, or both.
Q must equal 1, so we have \([(x-1)^2 + y^2] \cdot [(x+1)^2 + y^2] = 1\). Simplifying, this becomes

\[
(x^2 + y^2)^2 = 2(x^2 - y^2)
\]

(equation of lemniscate).

An object travelling on the lemniscate will have its velocity vector on the tangent line to the lemniscate at all times. To find the slope of the tangent line, we implicitly differentiate:

\[
2(x^2 + y^2)(2x + 2yy') = 2(2x - 2yy'),
\]

and then solve for \(y'\). After some algebra we get

\[
y' = \frac{x(1 - x^2 - y^2)}{y(1 + x^2 + y^2)}.
\]

The tangent line is horizontal where the lemniscate intersects the unit circle \(x^2 + y^2 = 1\), which occurs at the four points

\[
\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \quad \left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \quad \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right), \quad \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right).
\]

It also appears that the tangent line is horizontal when \(x = 0\). However the equation of the lemniscate says that \(y = 0\) when \(x = 0\), so it is not yet clear what happens to \(y'\) when \(x = 0\). To investigate this, let’s return to the geometric meaning of the derivative, as a limit of secant lines. Let our secant lines be \(y = mx\), as \(m\) varies. Such a line meets the lemniscate when

\[
(x^2 + (mx)^2)^2 = 2(x^2 - (mx)^2),
\]

that is, when \(x = 0\) and when

\[
x^2 = \frac{1 - m^2}{(1 + m^2)^2}.
\]

For \(x\) to approach zero, the slope \(m\) must approach \(\pm 1\). Indeed the lemniscate has two tangent lines at \((0,0)\), of slopes \(\pm 1\). So there is no horizontal tangent line when \(x = 0\).

### 3.4 Higher derivatives

The derivative \(f'(x)\) of a function \(f(x)\) is another function, so you could take the derivative of \(f'(x)\). We’ll call this the **second derivative** of \(f(x)\), and denote it by \(f''(x)\). And you could keep going: \(f'''(x)\) is the derivative of \(f''(x)\), etc. Other notations are

\[
\frac{d^2 f}{dx^2} = f^{(2)}(x) = f''(x), \quad \frac{d^3 f}{dx^2} = f^{(3)}(x) = f'''(x), \quad \ldots, \quad \frac{d^n f}{dx^n} = f^{(n)}(x),
\]

where \(n = 1, 2, 3, \ldots\) It is also convenient to set \(f^{(0)}(x) = f(x)\).
3.5 The second derivative

The sign of \( f''(x) \) tells you how the curve is bending. If \( c \) is a point where \( f''(c) > 0 \) then the slope is increasing at \( c \), so the graph is **concave up**. If \( f''(c) < 0 \) the graph is **concave down**. If \( f''(c) = 0 \) we call \( c \) an **inflection point**, often, but not always, the function changes concavity at an inflection point.

**Example 1:** Take the function

\[
f(x) = x^3 - 6x^2 + 11x - 4.
\]

Then \( f''(x) = 6x - 12 = 6(x - 2) \) is negative, zero, positive for \( x < 2, x = 2, x > 2 \). So \( x = 2 \) is the unique inflection point and the concavity is as shown.

**Example 2:** Take the function

\[
f(x) = x^4.
\]

Then \( f''(x) = 12x^2 \) is zero at \( x = 0 \) and positive for all other \( x \). So \( x = 0 \) is the unique inflection point, but the concavity is positive on both sides of \( x \).
Example 3: If \(f(t)\) is the distance travelled at time \(t\) by a moving object, then we have seen that \(f'(t)\) is the speed. In this context \(f''(t)\) is the **acceleration**, which by definition is the rate of change of the speed. If \(f''(c) > 0\) the object is speeding up at time \(t = c\). If \(f''(c) < 0\), it is slowing down at time \(t = c\).

Often in elementary physics one encounters distance functions which are quadratic polynomials:

\[ f(t) = At^2 + Bt + C, \]

where \(A, B, C\) are constants. Here, \(f''(t) = 2A\) is a constant. Later we will see that, conversely, any function with constant second derivative is a polynomial.

### 3.6 Curvature

The second derivative is not as precise a measurement as the first derivative is.

Recall that the sign of \(f'(c)\) (assuming \(f'(c) \neq 0\)) tells you whether \(f(x)\) is increasing or decreasing when \(x = c\). This is *qualitative* information. In fact the value of \(f'(c)\) gives precise *quantitative* information: It is the slope of the tangent line to the graph at \((c, f(c))\), and therefore tells you *how fast* the function is increasing or decreasing.

The second derivative \(f''(c)\) (assuming \(f''(c) \neq 0\)) also gives qualitative information about the concavity of the graph of \(f(x)\) at \(x = c\). However, it does not give quantitative information. For example, consider the graph of \(f(x) = x^2\). We have \(f''(x) = 2\), a positive constant. Indeed, the graph is concave up everywhere, but it bends more at \(x = 0\) than elsewhere, and this change in bending is not indicated by the constant function 2.

A more precise measurement is **curvature**, which tells *by how much* the curve is bending, and in what direction. The curvature is tells you radius of the **osculating circle** \(^4\)

The osculating circle is a direct analogue of the tangent line. Recall that for the tangent line we take two points \(P = (x, f(x))\) and \(Q = (x + h, f(x + h))\). The unique line through these points is the secant line, which converges to the tangent line as \(h \to 0\).

To get the osculating circle, we take *three* points on the graph,

\[ P = (x, f(x)), \quad Q = (x + h, f(x + h)), \quad R = (x - h, f(x - h)). \]

Through any three distinct points in the plane there is a unique *circle* (which is a line if the three points are collinear).

\(^4\)From the latin *osculare* = “to kiss”. Who knew Calculus was this much fun?
Here is a picture, for \( f(x) = x^2 \), of the circle through \( P, Q, R \), where \( P = (1, 1) \) and \( h = 1 \). In this case \( Q = (2, 4) \) and \( R = (0, 0) \).

As \( h \to 0 \), the three points \( P, Q, R \) coalesce at \( P \). By definition, the osculating circle is the limit circle as \( h \to 0 \). For the example \( f(x) = x^2 \), the green circle above becomes the green osculating circle in Example 1 below.

Just as a line has a slope, so a circle has a curvature, which is defined to be the reciprocal \(1/r\) of the radius \( r\); the bigger the circle, the smaller the curvature. A line can be regarded as a circle of infinite radius, hence a line has zero curvature.

With some work (see the appendix 9), one can show that the osculating circle at \( (x, f(x)) \) has center

\[
\left( x - f'(x) \left( \frac{1 + f'(x)^2}{f''(x)} \right), f(x) + \frac{1 + f'(x)^2}{f''(x)} \right),
\]

and its radius is \( 1/|\kappa_f| \), where

\[
\kappa_f(x) = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}}
\]
curvature.

The denominator of \( \kappa_f(x) \) is positive, so \( \kappa_f(x) \) and \( f''(x) \) have the same sign, telling you whether the osculating circle is above \( (\kappa_f(x) > 0) \) or below \( (\kappa_f(x) < 0) \) the graph of \( f(x) \). However, \( \kappa_f(x) \) says more; it tells you by how much the graph is bending.
Example 1: Take the parabola \( f(x) = x^2 \). You can check that the curvature at \((x, x^2)\) is given by

\[
\kappa_f(x) = \frac{2}{(1 + 4x^2)^{3/2}},
\]

and that the center of the osculating circle at \((x, x^2)\) is \((-4x^3, 3x^2 + \frac{1}{2})\). Note that \(\kappa_f(x) \leq 2\), and \(\kappa_f(x) = 2\) only for \(x = 0\). Thus, the curvature is maximal at the vertex of the parabola, as expected. This osculating circle is shown in red below, along with the green osculating circle through \((1, 1)\).

![Graph of the parabola with marked points and osculating circles](image)

Example 2: Just as a line is its own tangent line, so a circle is its own osculating circle. Let’s check this with the graph of \( f(x) = \sqrt{a^2 - x^2} \), which is the top half of the circle \(x^2 + y^2 = a\). It is easier to compute \( f'(x) = y' \) via implicit differentiation:

\[
y' = -\frac{x}{y}, \quad y'' = -\frac{a^2}{y^3},
\]

leading to

\[
\kappa_f = \frac{-(a^2/y^3)}{(1 + (x/y)^2)^{3/2}} = -\frac{1}{a}.
\]

We get constant curvature as expected, and this is negative because the osculating circle is below the graph.
3.7 Jerk

In physics and engineering, the third derivative is called *jerk*, because that is what you feel when acceleration changes suddenly. For example, a good driver eases off the brake just as the car comes to a halt, so that the passengers won’t feel the sudden drop to zero of the constant deceleration of braking. A bad driver keeps the same constant braking force until the car stops. Everyone’s head snaps forward and they all yell “Jerk!”.

4 TRIGONOMETRY

The six trigonometric ratios have been studied for at least 4000 years, since the time of the Babylonians, being important for astronomy and later, navigation. Their upgrade from ratios to functions took place during the last 500 years, notably in the work of Euler.

Trigonometric functions cannot be constructed by applying algebraic operations to polynomials, hence they are more difficult to analyze; even computing their values is nontrivial. In this chapter we will see how the ancient view of trigonometric functions feeds into modern calculus.

4.1 The Circle

Much of mathematics, from the ancient Greeks starting around 500BC (or even the Babylonians before that), up through the development of Calculus as we know it (starting 1600 AD) was driven by the mysteries of the circle. To astronomers and philosophers like Aristotle, the circle was regarded as a divinely perfect shape that described the orbits of planets about the earth at the center.

To mathematicians (who back then were also the astronomers), it was a famous unsolved problem to find the area of a circle.

Of course there are many circles, of different radii. An early discovery, whose proof appears in Euclid’s *Elements* (about 300 BC) says that if you can find the area of just one circle, then you’ll know the area of all circles. Namely, Euclid proves:
Circles are to one another as the squares on their diameters.

What Euclid means is that if you have two circles, \( C, C' \), and you draw the smallest squares containing them, \( S, S' \) (whose side-lengths are the diameters of the corresponding circle), then

\[
\frac{\text{area } C}{\text{area } C'} = \frac{\text{area } S}{\text{area } S'}
\]

In other words, the ratio

\[
\frac{\text{area } C}{\text{area } S} = \frac{\text{area } C'}{\text{area } S'}
\]

is the same for all circles! Today we call this ratio \( \pi/4 \). That is, Euclid shows there is a fundamental constant, that we now call \( \pi \), which is given by

\[
\pi = 4 \times \frac{\text{area } C}{\text{area } S}
\]

for any circle \( C \), with circumscribing square \( S \). We know this today as the formula

\[
\text{area } C = \pi r^2,
\]

where \( r \) is the radius of \( C \). But this is not a complete formula until you know what \( \pi \) is. This means finding the area of a circle with \( r = 1 \).

The ancient Greeks could not compute \( \pi \). This became known as one of the Three Problems of Antiquity. The other two were trisecting an angle and computing cube roots, both with straightedge and compass. However, the problem of computing \( \pi \) is much deeper than these other two problems.

5The other two were trisecting an angle and computing cube roots, both with straightedge and compass. However, the problem of computing \( \pi \) is much deeper than these other two problems.
With this method, but using polygons with more sides, Archimedes (~ 200 BC) was able to show that

\[ 3 + \frac{10}{70} < \pi < 3 + \frac{11}{70}. \]

We have since gone much further, but computing \( \pi \) poses a challenge to this day.

Archimedes also related the area of a circle to its circumference. He proved:

*The area of a circle is equal to the area of a triangle with height equal to the radius of the circle and base equal to the circumference of the circle.*

\[ \text{Area of circle} = \text{Area of triangle} \]

This is none other than the formula \( c = 2\pi r \).

This was a significant result at the time because there is no discussion of circumference anywhere in Euclid; Archimedes seems to have been the first to study the lengths of curved lines. However, we can derive Archimedes’ result on the circumference from Euclid’s result on the area by using the Squeeze Law for limits.

Consider two concentric circles of radii \( r \) and \( r + h \), for some \( h > 0 \). Let \( C(r) \) and \( C(r + h) \) be the respective circumferences of the circles. We define three functions \( E(h) \), \( F(h) \), \( G(h) \), as follows. \( F(h) \) is the area between the two circles. \( E(h) \) is the area of the rectangle with height \( h \) and width \( C(x) \), and \( G(h) \) is the area of the rectangle with height \( h \) and width \( C(x + h) \). When we try to wrap these two rectangles around the circle of radius \( r \), we see that

\[ E(h) \leq F(h) \leq G(h). \]

But \( E(h) = h \cdot C(r) \) and \( G(h) = h \cdot C(r + h) \), so

\[ C(r) \leq \frac{F(h)}{h} \leq C(r + h). \]

By the Squeeze Law,

\[ C(r) = \lim_{h \to 0} \frac{F(h)}{h} = \lim_{h \to 0} \frac{\pi (r + h^2) - \pi r^2}{h} = 2\pi r. \]

So Archimedes’ formula for circumference indeed follows from Euclid’s formula for area, plus the Squeeze Law.

### 4.2 Trigonometric functions
Trigonometric functions describe the coordinates of points on a circle, in terms of an angle that remembers how you got to that point.

Take a circle of radius=1, centered at $O = (0,0)$ in the $(x,y)$-plane. Given a number $\theta \geq 0$, start at the point $A = (1,0)$ on $C$ and run counterclockwise around the circle until you have gone exactly $\theta$ units of distance along the circle, then stop, say at $B$. By definition, $\cos \theta$ and $\sin \theta$ are the $(x,y)$ coordinates of your stopping point $B$. If $\theta \leq 0$, then run clockwise to get to $(\cos \theta, \sin \theta)$.

If you went an extra lap and then stopped at the same point $B$, you will of course have the same values for the $(x,y)$ coordinates of $B$. By Archimedes’ result on the circumference, one lap of the circle is exactly $2\pi$, so if you went an extra lap you would cover a distance of $\theta + 2\pi$. Therefore,

$$\sin(\theta + 2\pi) = \sin(\theta) \quad \text{and} \quad \cos(\theta + 2\pi).$$

Now define $\tan \theta$ to be the slope of the line $\overrightarrow{OB}$, from the center to your stopping point. This means that

$$\tan \theta = \frac{\sin \theta}{\cos \theta}.$$

You would still have the same slope if you went only half-way around, so

$$\tan(\theta + \pi) = \tan(\theta).$$

The secant, cosecant and cotangent functions the reciprocals of the previous three:

$$\sec \theta = \frac{1}{\cos \theta}, \quad \csc \theta = \frac{1}{\sin \theta}, \quad \cot \theta = \frac{1}{\tan \theta}.$$

We also used “tangent” and “secant” to describe the geometry of the derivative. Actually this is consistent: Remember that $\textit{tangere} =$ “to touch” and $\textit{secare} =$ “to cut”. Using similar triangles in the above picture, we see that if $0 \leq \theta \leq \pi/2$, then

$$\cos \theta = OC, \quad \sin \theta = CB, \quad \tan \theta = AD, \quad \sec \theta = OD,$$

so the tangent is the tangent and the secant is the secant!

However, the word “sine” is a mistranslation that is too old to be fixed. The original word is the Sanskrit ($\sim 1000\text{AD}$) for “half-chord” (an accurate term, as you can see in the picture above).
This was faithfully translated into Arabic, but later was wrongly translated into Latin as *sinus*, meaning “bosom”.

We can use this picture to compute our first hard limits, which will lead to the derivatives of the trigonometric functions.

\[
\lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1
\]

**Proof:** Since

\[
\frac{\sin(-\theta)}{-\theta} = \frac{\sin(\theta)}{\theta},
\]

it is enough to prove this as \( \theta \to 0^+ \). Hence we may assume that \( 0 < \theta < \pi \). Here is the previous picture again, for reference:

We see a chain of inequalities of areas:

\[
\text{Area } OCB \leq \text{Area } OAB \leq \text{Area } OAD.
\]

The area of the sector \( OAB \) is to the area \( \pi \) of the whole circle as the angle \( \theta \) of the sector is to the angle \( 2\pi \) of the whole circle:

\[
\frac{\text{Area } OAB}{\pi} = \frac{\theta}{2\pi},
\]

so \( \text{Area } OAB = \frac{1}{2} \theta \). The areas \( OCB \) and \( OAD \) are triangles whose bases and heights are shown in the picture. We get

\[
\frac{1}{2} \cos \theta \cdot \sin \theta \leq \frac{1}{2} \theta \leq \frac{1}{2} \tan \theta.
\]
Remembering that \( \tan \theta = \sin \theta / \cos \theta \), and that \( \sin \theta > 0 \) since \( 0 < \theta < \pi \), we can divide everything by \( \frac{1}{2} \sin \theta \) without changing the inequalities. This shows that

\[
\cos \theta \leq \frac{\theta}{\sin \theta} \leq \frac{1}{\cos \theta}.
\]

As \( \theta \to 0^+ \) the point \( C \) approaches \( A \), so \( OC \) approaches the radius \( OA \). So \( \cos \theta \to 1 \). By the Squeeze Law, we have

\[
\lim_{\theta \to 0^+} \frac{\theta}{\sin \theta} = 1.
\]

Since this is nonzero, the reciprocal limit exists and equals \( 1/1 = 1 \). Thus we have proved that \( (\sin \theta)/\theta \to 1 \) as claimed. \( \text{QED} \)

This limit says that \( \sin \theta \) is closely approximated by \( \theta \), when \( \theta \) is close to zero. This is a simplification used often in Physics.

Pendulum example.

We will later use Taylor polynomials to obtain better approximations of \( \sin \theta \).

The next hard limit is

\[
\lim_{\theta \to 0} \frac{1 - \cos \theta}{\theta} = 0
\]

\textbf{Proof:} Refer to the following picture.

The triangles \( BCE \) and \( ACB \) are similar, so their corresponding side ratios are equal:

\[
\frac{CA}{CB} = \frac{CB}{CE}.
\]

This means that

\[
\frac{1 - \cos \theta}{\sin \theta} = \frac{\sin \theta}{1 + \cos \theta}.
\]
Therefore,
\[
\frac{1 - \cos \theta}{\theta} = \frac{\sin \theta}{\theta} \cdot \frac{\sin \theta}{1 + \cos \theta}.
\]
We already proved that
\[
\lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1.
\]
And
\[
\lim_{\theta \to 0} \frac{\sin \theta}{1 + \cos \theta} = \frac{0}{2} = 0
\]
as a trivial limit. From the product rule for limits, we conclude that
\[
\lim_{\theta \to 0} \frac{1 - \cos \theta}{\theta} = 0,
\]
as desired. QED

4.3 Derivatives of Trigonometric Functions

We now replace the variable \( \theta \) by our usual variable \( x \).

The three main derivatives to know are
\[
\begin{align*}
\sin x' &= \cos x, \\
\cos x' &= -\sin x, \\
\tan x' &= \sec^2 x
\end{align*}
\]

Proof: For \( (\sin x)' \) we have to use the difference quotient.
\[
(\sin x)' = \lim_{h \to 0} \frac{\sin(x + h) - \sin x}{h}.
\]
From the addition formula
\[
\sin(x + h) = \sin x \cos h + \sin h \cos x,
\]
and the two hard limits we just computed, we get
\[
(\sin x)' = \lim_{h \to 0} \frac{\sin x \cos h + \sin h \cos x - \sin x}{h}
\]
\[
= \sin x \lim_{h \to 0} \frac{\cos h - 1}{h} + \cos x \lim_{h \to 0} \frac{\sin h}{h}
\]
\[
= \sin x \cdot 0 + \cos x \cdot 1
\]
\[
= \cos x,
\]
Now using the identity $\cos x = \sin \left( \frac{\pi}{2} - x \right)$ and the chain rule, we get $(\cos x)' = -\sin x$. Finally, using the quotient rule for $\tan x = \sin x / \cos x$, we get $(\tan x)' = \sec^2 x$, as claimed.  \textbf{QED}

Notice that the derivatives of $\sin x$ and $\cos x$ go back and forth, while alternating in sign:

$$(\sin x)' = \cos x, \quad (\sin x)'' = -\sin x, \quad (\sin x)''' = -\cos x, \quad (\sin x)^{(4)} = \sin x, \quad \text{repeat} \ldots$$

$$(\cos x)' = -\sin x, \quad (\cos x)'' = -\cos x, \quad (\cos x)''' = \sin x, \quad (\cos x)^{(4)} = \cos x, \quad \text{repeat} \ldots$$

The picture at the beginning of this section shows the graph of $\cos x$ and its osculating circles when $x$ is an integer multiple of $\pi$. From the above derivatives, we see that the curvature of $\cos x$ is

$$\kappa_{\cos x} = \frac{-\cos x}{(1 + \sin^2 x)^{3/2}}.$$

When $x = n\pi$ the curvature is $-1$ if $n$ is even and $+1$ if $n$ is odd. Therefore, when $n$ is odd (respectively even) the osculating circle is a unit circle below (resp. above) the graph of $\cos x$, as shown.

The curvature is zero when $x$ is an odd multiple of $\pi/2$. At these points the osculating “circle” is just the tangent line, which is a circle of infinite radius.

The graph of $\sin x$ is the same, but shifted $\pi/2$ units to the right.

The graphs of $\sin x$ and $\cos x$ look like film threading through the spools of a movie projector. Most pictures of these graphs online or on computers are inaccurate \footnote{alas, even here: \url{http://dlmf.nist.gov/4.15}} because people use different scales on the two coordinate axes. This scrunches up the graph like an accordion, leaving no room for the osculating circles.

### 4.4 Inverse trigonometric functions

![Graph of arcsin and arctan functions](image)

The arcsine function is the inverse of the sine function. The value of $\arcsin x$ is the angle in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ whose sine is $x$. Since the sine takes values in $[-1, 1]$, the function $\arcsin x$ is only defined for $x$ in $[-1, 1]$.\footnote{alas, even here: \url{http://dlmf.nist.gov/4.15}}
Likewise the arctangent function is the inverse of the tangent function; arctan $x$ is the angle in \((-\frac{\pi}{2}, \frac{\pi}{2})\) whose tangent is $x$. Since the tangent function takes on all real values, the function arctan $x$ is defined for all real numbers $x$. It has horizontal asymptotes at $y = \pm \pi/2$.

The derivatives of these functions are given as follows.

\[
\begin{align*}
\text{arcsin} \quad & (x)' = \frac{1}{\sqrt{1 - x^2}}, \\
\text{arctan} \quad & (x)' = \frac{1}{1 + x^2}.
\end{align*}
\]

**Proof:** Recall that if $f$ and $g$ are inverse functions then $f(g(x)) = x$, so that by the chain rule we have

\[g'(x) = \frac{1}{f'(g(x))}.
\]

Taking $f(x) = \sin x$, so that $g(x) = \arcsin x$, we have $f'(g(x)) = \cos(\arctan x)$, which is the cosine of the angle whose tangent is $x$. Make a right triangle containing an angle $\theta$ whose tangent is $x$. You will find that $\cos \theta = 1/\sqrt{1 - x^2}$. A similar calculation works for $\arctan x$.

QED

The arctangent function is especially useful.

- Conversion from cartesian to polar coordinates:

\[r = \sqrt{x^2 + y^2}, \quad \theta = \arctan(y/x).
\]

- Formulas for $\pi$ (appearing shortly)

- Integration (later in the semester)

Here is a story problem using the arctangent. You are watching an airplane fly overhead at 1000 meters altitude. You keep track of the angle $\theta(t)$ at your eye between the plane and the horizon. When this angle is $\pi/4$, you find that $\theta'(t) = -.05$ radians/second. What is the speed of the airplane?

**Answer:** Let $x(t)$ be the plane’s horizontal distance from you at time $t$, and abbreviate $h = 1000$. Then $\tan(\theta(t)) = \frac{h}{x(t)}$, so

\[\theta(t) = \arctan \left( \frac{h}{x(t)} \right),
\]

so

\[\theta'(t) = \frac{1}{1 + (h/x(t))^2} \cdot \frac{-h}{x(t)^2} \cdot x'(t) = -\frac{h}{1 + (h/x(t))^2} \cdot x'(t).
\]

When $\theta(t) = \pi/4$ we have $x(t) = h$, and $\theta'(t) = -.05$, so

\[x'(t) = 2h \cdot .05 = 100 \text{ meters per second}
\]
The trigonometric functions and their inverses are our first examples of **transcendental functions**, which transcend the algebraic functions built from \( f(x) = x \) and using algebraic operations like sums and powers and roots.

Evaluating transcendental functions is difficult. For example it is clear how to evaluate an algebraic function like

\[
\frac{1}{\sqrt{1 + x^2}}
\]

You take your value of \( x \), square it, add one, take the square-root and then the reciprocal.

But how do you evaluate \( \sin x \)? For certain angles \( x \) you can use the geometry of simple triangles to compute, for example, \( \sin(\pi/3) = \sqrt{3}/2 \). But what about angles that don’t come from simple triangles, for example, what is \( \sin 1 \)?

The remarkable fact is that many transcendental functions can be **approximated** by polynomials, and this approximation can be made as accurate as we need. This how we (or the elves in your
Phone) evaluate transcendental functions. The picture above shows these approximations for \( \sin x \). The polynomials shown are

\[
\begin{align*}
x \\
x - \frac{x^3}{6} \\
x - \frac{x^3}{6} + \frac{x^5}{120} \\
x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040}
\end{align*}
\]

We could keep going if we needed more accuracy. The picture above seems to show that the last polynomial is a pretty good approximation, at least on the interval \((-\pi, \pi)\). We can test it numerically, as well. A computer pronounces that

\[
\sin 1 \approx 0.84147098\ldots
\]

(The wishy-washy equal sign \( \approx \) stands for “is approximately equal to”.) On the other hand, if we set \( x = 1 \) in the last polynomial above, we get

\[
1 - \frac{1}{6} + \frac{1}{120} - \frac{1}{5040} \approx 0.84146825\ldots
\]

The mysterious denominators actually have a pattern:

\[
6 = 1 \cdot 2 \cdot 3, \quad 120 = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5, \quad 5040 = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7.
\]

For any positive integer \( n \), we define \( n! \), called \( n \)-factorial, to be the product of all positive integers from 1 to \( n \). Thus,

\[
n! = 1 \cdot 2 \cdots (n - 1) \cdot n.
\]

With this notation, the polynomials we are using to approximate \( \sin x \) are

\[
\sin x \approx \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^k \frac{x^{2k+1}}{(2k+1)!}.
\]

We can take \( k \) as large as we wish, depending on how accurate we want our approximation to be.

The polynomials themselves have the following recipe.

Take a function \( f(x) \) that you wish to approximate. You have to know the values \( f(0), f'(0), f''(0), \ldots \) of \( f(x) \) and its derivatives at \( x = 0 \), but only at this point. Then the Taylor polynomials of \( f(x) \) are constructed as follows.

\[
p_n(x) = f(0) + f'(0) \frac{x}{1!} + f''(0) \frac{x^2}{2!} + f'''(0) \frac{x^3}{3!} + \cdots + f^{(n)}(0) \frac{x^n}{n!}
\]
Thus, the coefficient of \(x^k\) in \(p_n(x)\) is the number \(f^{(k)}(0)/k!\), for each \(k = 1, \ldots, n\).

It is plausible that \(p_n(x)\) approximates \(f(x)\) because both \(p_n(x)\) and \(f(x)\) have the same value, slope, curvature, jerk, etc. at \(x = 0\). Indeed, you can check that

\[
p_n(0) = f(0), \quad p_n'(0) = f'(0), \quad p_n''(0) = f''(0), \quad \ldots \quad p_n^{(n)}(0) = f^{(n)}(0).
\]

After this, \(p_n^{(n+1)}(0) = 0\), but \(f^{(n+1)}(0)\) may not be zero. If you want a better approximation you must choose a larger \(n\); perhaps \(p_{n+1}(x)\) will be good enough for you.

If \(f(x) = \sin x\) then its derivatives are \(\pm \sin x\) and \(\pm \cos x\), repeating in cycles of four. Thus,

\[
f^{(n)}(0) = \begin{cases} 
0 & \text{if } n \text{ is even} \\
(-1)^k & \text{if } n = 2k + 1 \text{ is odd}, 
\end{cases}
\]

as claimed in (5). Likewise one finds that the Taylor polynomials of \(\cos x\) are

\[
\cos x \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^k \frac{x^{2k}}{(2k)!}.
\]

We have not actually proved that every function \(f(x)\) can be approximated by its Taylor polynomials. We have only given evidence and heuristic reasons. We have also not said anything about the error in the Taylor approximation. These issues will be discussed later on in the semester.

In the meantime we will use Taylor polynomials to help us understand other transcendental functions, as we encounter them.

### 4.6 The arctangent formula for \(\pi/4\)

Euclid proved that the area-ratio of a circle to the square on its diameter is the same for all circles. This ratio is a fundamental constant of nature, which we now know as

\[
\frac{\text{circle}}{\text{square}} = \frac{\pi}{4}.
\]

How can we determine this fundamental constant? By “determine”, we might first try to express this constant as a rational number, that is, as a quotient of two integers. Unfortunately, \(\pi/4\) is not a rational number. Since rational numbers are the only numbers we can write down explicitly, the next best thing is to express \(\pi/4\) as a limit of rational numbers. This means finding a sequence of rational numbers that 1) gets arbitrarily close to \(\pi/4\) and 2) is predictable, so that we can actually get as close as we like. For example, a computer will tell us that

\[
\pi/4 \approx .785398163974483096...
\]
This is a sequence of rational numbers:
\[
\frac{7}{10}, \frac{78}{100}, \frac{785}{1000}, \frac{7853}{10000}, \ldots
\]
but it is not predictable. Each new digit is a complete surprise, no matter how many digits we’ve already computed. A computer is a finite machine, so it will only be able to produce finitely many such rational numbers, and will be at a loss to go further, unless we have a pattern, of some kind. Taylor polynomials produce a pattern.

The best way to determine a mysterious number is to realize that number as the value of a nice function. Instead of a needle in a haystack you now have a thread of values of the function to lead you to your mystery number. In our case, \( \pi/4 \) is the angle whose tangent is 1, so

\[
\frac{\pi}{4} = \arctan(1)
\]
is a value of the function \( \arctan(x) \). If \( \arctan(x) \) were a polynomial with rational coefficients, we could just evaluate it at \( x = 1 \). But \( \arctan(x) \) is bounded between \( \pm \pi/2 \), so it cannot possibly be a polynomial.

However, we can approximate \( \arctan x \) by Taylor polynomials having the same value, slope, curvature, jerk... at \( x = 0 \), that \( \arctan x \) has.

Recall from the previous section that the Taylor polynomials of a function \( f(x) \) are

\[
p_n(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n.
\]

For the function \( f(x) = \arctan x \), we compute

\[
\begin{align*}
f(x) &= \arctan x \quad \text{so} \quad f(0) = 0 \\
f'(x) &= \frac{1}{1+x^2} \quad \text{so} \quad f'(0) = 1 \\
f''(x) &= \frac{-2x}{(1+x^2)^2} \quad \text{so} \quad f''(0) = 0 \\
f'''(x) &= \frac{6x^2-2}{(1+x^2)^3} \quad \text{so} \quad f'''(0) = -2 \\
f^{(4)}(x) &= \frac{24x(1-x^2)}{(1+x^2)^4} \quad \text{so} \quad f^{(4)}(0) = 0 \\
f^{(5)}(x) &= \frac{24(5x^4-10x^2+1)}{(1+x^2)^5} \quad \text{so} \quad f^{(5)}(0) = 24.
\end{align*}
\]

In general one can show that

\[
f^{(n)}(0) = \begin{cases} 0 & \text{if } n \text{ is even} \\ (-1)^k(2k)! & \text{if } k = 2k + 1 \text{ is odd.} \end{cases}
\]
So the Taylor polynomials of \( \arctan x \) are given by
\[
p_{2k+1}(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^k \frac{x^{2k+1}}{2k+1}.
\]
These look like the Taylor polynomials of \( \sin x \), with the factorial signs (!) erased.

Evaluating these Taylor polynomials at \( x = 1 \) gives the approximation
\[
\frac{\pi}{4} \approx 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots.
\]

Let us check the numerical evidence for this approximation. Recall the computer pronouncement
\[
\pi/4 \approx .785398163974483096\ldots
\]

Meanwhile the polynomials tell us
\[
1 - \frac{1}{3} \approx .66666\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} \approx .8666\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \approx .72380952380\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} \approx .8349206349206\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} \approx .7440115440115\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} \approx .8209346209346\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} \approx .7542679542679\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \frac{1}{17} \approx .8130914836791\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \frac{1}{17} - \frac{1}{19} \approx .76045990473235\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \frac{1}{17} - \frac{1}{19} + \frac{1}{21} \approx .80807895235139\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \frac{1}{17} - \frac{1}{19} + \frac{1}{21} - \frac{1}{23} \approx .76460069148183\ldots
\]
\[
1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \frac{1}{17} - \frac{1}{19} + \frac{1}{21} - \cdots + \frac{1}{1001} \approx .78589716489644\ldots
\]
The sums are dancing around \( .7853981\ldots \), getting closer and closer, but taking their sweet time.

We will later prove that the sums eventually get as close as you like to \( \pi/4 \). And we do have a pattern: \( \pi/4 \) is approximated by the alternating sums of reciprocals of odd numbers. This “determines” \( \pi/4 \) according to our criteria 1) and 2) at the beginning of this section.

Clearly, however, this approximation is not efficient. It takes many terms to get reasonably close to \( \pi/4 \). This is just a first attempt\(^7\), using what we know now. We will do better, once we have more tools.

\(^7\)This approximation of \( \pi/4 \) was first discovered by Indian astronomers around 1000AD
In third grade or thereabouts, we all learned that

\[
\begin{align*}
2^1 &= 2 \\
2^2 &= 4 \\
2^3 &= 8 \\
&\vdots \\
2^n &= 2 \cdot 2 \cdots 2 \quad \text{(two times itself } n \text{ times)}.
\end{align*}
\]

This is true, but misleading. The first hint that more is afoot is the mysterious declaration that

\[2^0 = 1.\]

The real trouble starts with numbers like \(2\sqrt{2}\) or \(2\pi\). We may believe in such numbers, but how would you explain what they are? Does it make any sense to multiply two times itself \(\sqrt{2}\) times? We have enough trouble understanding \(\pi\) itself. How do we multiply two times itself \(\pi\) times?

In fact, the numbers \(2^0, 2^1, 2\sqrt{2}, 2\pi, \ldots\) are all values of a function called \(2^x\). It takes some work to define this function. And \(2^x\) is two times itself \(x\) times only when \(x\) is a positive integer. These are the only values of \(2^x\) that one can explain to third graders.

In fact one can define a function \(a^x\) for any positive number \(a\). We even have functions like \(x^x\) or \(x^{\sin x}\), defined for \(x > 0\). The path to such functions is indirect; we will start with their inverse functions.

### 5.1 Logarithms

For \(x > 0\), we define a function \(\log x\) as follows. Consider the graph of \(y = 1/t\), above the positive \(t\)-axis. If \(x \geq 1\) then \(\log x\) is the area under the graph of \(y = 1/t\) above the interval \([1, x]\). If \(0 < x < 1\) then \(\log x\) is the negative of the area under \(y = 1/t\) above \([x, 1]\). Thus, \(\log x > 0\) if \(x > 1\), \(\log x < 0\) if \(0 < x < 1\) and \(\log 1 = 0\).

The two cases can be summarized in one case. If \(0 < a < b\) then the area under the graph of \(y = 1/x\) above the interval \([a, b]\) is \(\log b - \log a\).

We will derive all properties of \(\log x\) from this definition.
5.1.1 Derivative of the logarithm function and other properties

The derivative of \( \log x \) turns out to be the function \( 1/x \) that we started with:

\[
(\log x)' = \frac{1}{x}
\]

**Proof:** Once again we must use the difference quotient

\[
(\log x)' = \lim_{h \to 0} \frac{\log(x + h) - \log(x)}{h}.
\]

If \( h > 0 \) then \( \log(x + h) - \log(x) \) is the area under \( y = 1/x \) above \([x, x + h]\). This area contains the small rectangle with base \([x, x + h]\) and height \(1/(x+h)\), and is contained in the big rectangle with base \([x, x + h]\) and height \(1/x\). Since this interval has length \( h \), we have

\[
h \cdot \frac{1}{x+h} \leq \log(x + h) - \log(x) \leq h \cdot \frac{1}{x}.
\]

If \( h < 0 \) then by the same reasoning we get the same inequality, but now of negative areas. Dividing by \( h \) and letting \( h \to 0 \), the Squeeze Law gives the desired result. QED

The logarithm function is almost, but not quite, the only function whose derivative is \( 1/x \). The precise statement is this:

\[\log x \text{ is the unique function defined for } x > 0 \text{ which satisfies the two conditions}\]

\[
(\log x)' = 1/x \quad \text{and} \quad \log 1 = 0.
\]

**(7)**

**Proof:** We have seen that \( \log x \) satisfies these two conditions. Suppose there were some other function \( f(x) \) satisfying the same two conditions: \( f'(x) = 1/x \) and \( f(1) = 0 \).
We will prove later the very useful fact that if two functions have the same derivative then they differ by a constant.

Since \( f'(x) = 1/x = (\log x)' \), this means that \( f(x) \) and \( \log x \) differ by a constant:

\[
  f(x) = \log x + C,
\]

for some constant \( C \). Setting \( x = 1 \) we get

\[
  f(1) = \log 1 + C.
\]

But we know \( \log 1 = 0 \), and \( f(1) = 0 \) by assumption, so \( C = 0 \). Therefore \( f(x) = \log x \) as claimed. \( \text{QED} \)

Many functions can be characterized by their derivative and a single value, as in (7).

Here is an example of how this characterization of \( \log x \) is useful.

*For any two positive numbers \( a, b \), we have a product formula and a quotient formula:

\[
  \log(ab) = \log(a) + \log(b), \quad \text{and} \quad \log(b/a) = \log(b) - \log(a)
\]

Proof: Consider the function

\[
  f(x) = \log(ax) - \log(a).
\]

We have \( f'(x) = a \cdot (1/ax) + 0 = 1/x \), and \( f(1) = \log a - \log a = 0 \). Therefore \( f(x) = \log x \), by (7). Setting \( x = b \), we have

\[
  \log b = f(b) = \log(ab) - \log(a),
\]

so \( \log(ab) = \log(a) + \log(b) \), as claimed in the product formula. Setting \( x = b/a \), we have

\[
  \log(b/a) = f(b/a) = \log(b) - \log(a),
\]

as claimed in the quotient formula. \( \text{QED} \)

The quotient formula means that if \( 0 < a < b \) then the area under the graph of \( y = 1/x \) above \([a, b]\) can also be expressed as \( \log(a/b) \).
5.2 Graph of logarithm function

Since \((\log x)' = \frac{1}{x} > 0\), the function \(\log x\) is always increasing. Since \((\log x)'' = -\frac{1}{x^2} < 0\), the function \(\log x\) is always concave down.

Recall that for \(x > 0\) the function \(\arctan x\) is increasing, concave down, and is bounded by \(\pi/2\). It appears from the picture above that \(\log x\) might also be bounded.

We can examine this question as follows. From the product formula, we have

\[
\log(x^2) = \log x + \log x = 2\log x, \\
\log(x^3) = \log x + \log(x^2) = 3\log x \\
\ldots \\
\log(x^n) = n\log x
\]

for any positive integer \(n\). If \(x > 1\) then \(x^n \to \infty\) as \(n \to \infty\) and \(\log x > 0\). It follows that

\[
\lim_{x \to \infty} \log x = +\infty.
\]

If \(0 < x < 1\) then \(x^n \to 0\) as \(n \to \infty\) and \(\log x < 0\). It follows that

\[
\lim_{x \to 0} \log x = -\infty.
\]

So \(\log x\) is actually unbounded in both directions.

What about the Taylor polynomials of \(\log x\)? Unfortunately, since \(\log x\) is unbounded as \(x \to 0\) we cannot compute its value or its derivatives’ value at \(x = 0\). Instead we can shift the function to the left, and consider \(\log(1 + x)\) instead. This will still allow us to approximate values of the logarithm.
For \( f(x) = \log(1 + x) \) we compute that \( f(0) = 0 \) and for \( n \geq 1: \)

\[
f^{(n)}(x) = \frac{(-1)^{n-1}(n - 1)!}{(1 + x)^n}, \quad \text{so} \quad f^{(n)}(0) = (-1)^{n-1}(n - 1)! \]

Hence the Taylor approximation of \( \log(1 + x) \) is given by the polynomials

\[
\log(1 + x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots .
\]

Here we note that the Taylor approximation seems ok but not great near \( x = 0 \), and then goes to hell for \( x > 1 \). We will explain this next semester.

### 5.3 The exponential function

Since the logarithm is unbounded on both sides, it follows that every real number is a value of the logarithm function. Since the logarithm function is increasing on \((0, \infty)\), each real number \( x \) is in fact the logarithm of a \textit{unique} positive number, which we will call \( \exp x \). That is, \( \log(\exp x) = x \). Thus, \( \exp x \) is
the inverse function of \( \log x \); we shall call \( \exp x \) the \textit{exponential function}. Note that \( \exp(x) \) is defined for all real numbers \( x \) and \( \exp(x) > 0 \) for all \( x \). Since \( \log(1) = 0 \), we have \( \exp(0) = 1 \).

From the formula for the derivative of an inverse function, we have

\[
(\exp x)' = \frac{1}{\log'(\exp x)} = \frac{1}{1/\exp x} = \exp x.
\]

Thus, \( \exp x \) is its own derivative:

\[
(\exp x)' = \exp x
\]

Just as for the logarithm, the exponential function is characterized by its derivative and a single value:

\[
\exp x \text{ is the unique function defined for all } x \text{ which satisfies the two conditions}
\]

\[
(\exp x)' = \exp x \quad \text{and} \quad \exp(0) = 1. \tag{8}
\]

**Proof:** We have seen that \( \exp x \) satisfies the two conditions.

Conversely, suppose \( f(x) \) is some function for which \( f'(x) = f(x) \) and \( f(0) = 1 \). By the quotient rule,

\[
(\frac{f(x)}{\exp x})' = \frac{(\exp x)f'(x) - (\exp x)'f(x)}{(\exp x)^2} = \frac{(\exp x)f(x) - (\exp x)f(x)}{(\exp x)^2} = 0,
\]

because \( \exp x \) and \( f(x) \) are their own derivatives. It follows that \( \frac{f(x)}{\exp x} = C \), a constant. Since \( f(0) = \exp(0) = 1 \), it follows that \( C = 1 \), so \( f(x) = \exp x \), as claimed. \textbf{QED}

Just as for the logarithm, we can use this characterization to prove:

\textit{For any two real numbers } \( a, b \), \textit{we have sum and difference formulas:}

\[
\exp(a + b) = (\exp a)(\exp b), \quad \exp(a - b) = (\exp a)/(\exp b)
\]

**Proof:** This is similar to the logarithm proof. Consider the function

\[
f(x) = \frac{\exp(a + x)}{\exp(a)}.
\]

One checks that \( f'(x) = f(x) \) and \( f(0) = 1 \). Therefore

\[
\exp x = f(x) = \frac{\exp(a + x)}{\exp(a)}.
\]
Setting \( x = b \), we get the sum formula. Keeping \( x = b \) but replacing \( a \) by \( a - b \), we get the difference formula. QED

Often we encounter other functions inside the exponential function. For example:

\[
\exp(3x), \quad \exp(x^2), \quad \exp(\sin x), \quad \ldots
\]

You can differentiate these using the chain rule:

\[
(\exp 3x)' = 3 \exp(3x), \quad (\exp(x^2))' = 2x \exp(x^2), \quad (\exp(\sin x))' = \cos x \exp(\sin x).
\]

### 5.4 The general exponential function

If \( a \) is a positive number, then we define

\[
a^x = \exp(x \log a).
\]

For example, we can now say what \( 2^x \) is:

\[
2^x = \exp(x \log 2).
\]

Note that if \( n \) is a positive integer then

\[
2^n = \exp(n \log 2) = \exp(\log 2 + \cdots + \log 2) = \exp(\log 2) \cdots \exp(\log 2) = 2 \cdots 2,
\]

just as we learned in third grade. But now we can raise 2 to any power. For example,

\[
2^{\sqrt{2}} = \exp(\sqrt{2} \log 2), \quad \text{and} \quad 2^\pi = \exp(\pi \log 2).
\]

We can even define the function

\[
x^x = \exp(x \log x).
\]

All functions involving general exponentials can be expressed in terms of the two functions \( \exp x \) and \( \log x \).

**Example 1:** Differentiate \( 2^x \).

Start by writing \( 2^x = \exp(x \log 2) \) and use the chain rule, as in the examples above. You get

\[
(2^x)' = \log 2 \cdot (\exp(x \log 2)) = (\log 2)2^x.
\]

**Example 2:** Differentiate \( 2^{\sin x} \).
Start by writing $2^{\sin x} = \exp(\sin x \log 2)$. Then

$$(2^{\sin x})' = (\log 2 \cos x)2^{\sin x}.$$  

and $x^x = \exp(x \log x)$, so

$$(x^x)' = (x \log x)' \exp(x \log x) = (1 + \log x)x^x.$$  

**Example 3:** Differentiate $x^r$, where $r$ is any constant.

Start by writing $x^r = \exp(r \log x)$. Then

$$(x^r)' = (r \log x)' \exp(r \log x) = \frac{r}{x} \cdot x^r = rx^{r-1}.$$  

Thus, have finally proved:

$$\boxed{(x^r)' = rx^{r-1}}$$ the general power rule

### 5.5 The number $e$.

The number $e$ is the value at $x = 1$ of the exponential function $\exp x$:

$$e = \exp(1)$$

Alternatively, $e$ is the unique number such that $\log e = 1$. This means that if we use $a = e$ as the base of our exponential, we have

$$e^x = \exp(x \log e) = \exp(x).$$

So $e$ is the base that makes exponential functions simplest.

So what is this number $e$? We can compute it using the same method that we used for $\pi/4 = \arctan(1)$, using Taylor polynomials for $\exp x$ instead of $\arctan x$.

Since $\exp x = (\exp x)' = (\exp x)'' = \cdots$, and $\exp(0) = 1$, the Taylor polynomials of $\exp(x)$ are

$$p_n(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}.$$
We shall later prove that this polynomial approximates the function \( \exp(x) \). Evaluating at \( x = 1 \) then gives the approximation

\[
e \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}.
\]

If we take \( n = 10 \), we get the rational number

\[
1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{10!} \approx 2.71828180114638\ldots
\]

A computer will pronounce that

\[
e \approx 2.71828182845904523536\ldots
\]

The computer uses the same method we just did, except with a larger \( n \). We could respond with

\[
1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{20!} \approx 2.71828182845904523533\ldots
\]

In contrast to the \( \pi/4 \) example, each of our approximations is slightly low. This is to be expected, since the terms in our sum are now all positive.

Here is another way to compute \( e \):

\[
e = \lim_{h \to 0} (1 + h)^{1/h}.
\]

**Proof:** Start by writing

\[
(1 + h)^{1/h} = \exp \left( \frac{1}{h} \log(1 + h) \right).
\]

Now work on the limit inside of exp. Since \( \log 1 = 0 \), we observe that

\[
\frac{1}{h} \log(1 + h) = \frac{\log(1 + h) - \log(1)}{h}
\]

is the difference quotient for \( (\log x)' \) evaluated at \( x = 1 \). Since \( (\log x)' = 1/x \) we have.

\[
\lim_{h \to 0} \frac{\log(1 + h) - \log(1)}{h} = \frac{1}{1} = 1.
\]

Therefore

\[
\lim_{h \to 0} (1 + h)^{1/h} = \exp(1) = e,
\]

as claimed. QED
5.6 Exponential Growth

We shall often write $e^x$ instead of $\exp(x)$; they are the same function.

The function $e^x$ occurs often in nature. One frequent context is when you have a quantity that varies with time, and whose rate of change is proportional to its current amount. We model this by a function $f(t)$ (where $t =$ time) whose rate of change $f'(t)$ is proportional to $f(t)$. That is,

$$f'(t) = kf(t),$$

for some constant $k$. Such a function must be of a very special type. Namely,

$$If \ f(t) = kf(t), \ where \ k \ is \ a \ constant, \ then \ f(t) = f(0) \cdot e^{kt}. $$

**Proof:** The following argument should look familiar. We compute the derivative

$$\left( \frac{f(t)}{e^{kt}} \right)' = \frac{e^{kt}f'(t) - k e^{kt}f(t)}{e^{2kt}} = \frac{e^{kt}k f(t) - k e^{kt}f(t)}{e^{2kt}} = 0.$$ 

Therefore $f(t)/e^{kt}$ must equal some constant $C$. Evaluating at $t = 0$, we find $C = f(0)$, whence the result. QED

**Example:** Carbon 14 (C14) is a rare and unstable isotope of carbon that is present in all organic materials on earth. It is created by cosmic rays hitting nitrogen in the atmosphere. C14, as with all forms of carbon, is consumed by plants during photosynthesis, and is then consumed by all living things on earth that eat plants, or that eat what eats plants.

The percentage of C14 (vs ordinary Carbon 12) in a living organism is roughly the same as in the atmosphere. When a living organism dies, it stops eating C14, so the C14 in its corpse decays.

Radioactive substances decay exponentially. This decay is measured by half-life: the number of years it takes a quantity of the substance to decay by half. C14 decays into the Nitrogen 14 from whence it came, with a half-life of 5730 years.

Assuming that the amount of C14 in the atmosphere is constant, one can then estimate the death date of some recovered piece of organic material (wood, skull...) by measuring the percentage of remaining C14. Since the decay is exponential, in any dead object the amount $C(t)$ of C14 at time $t$ is given by

$$C(t) = C(0)e^{kt},$$

where $t = 0$ is the date of death.

\[8\text{this assumption is only accurate for about 20,000 years, due to fluctuations in the cosmic rays, so C14 dating is only regarded as accurate within this window.}\]
The constant \( k \) is determined by the half-life, as follows. Having a half-life 5730 means that \( C(5730) = (1/2)C(0) \), so

\[
(1/2)C(0) = C(0)e^{k \cdot 5730}. 
\]

Cancelling \( C(0) \) and taking log of both sides, we get

\[
\log\left(\frac{1}{2}\right) = k \cdot 5730, 
\]

so

\[
k = -\frac{\log 2}{5730}. 
\]

Therefore,

\[
C(t) = C(0)e^{-t \cdot \frac{\log 2}{5730}}. 
\]

Now suppose you dig up a skull and ask how old it is. In other words, you want to know the time \( t \) since the skull’s owner died. Let’s say you find that the skull contains 10\% of the C14 found in a living skull. Therefore,

\[
\frac{1}{10}C(0) = C(t) = C(0)e^{-t \cdot \frac{\log 2}{5730}}. 
\]

Solving for \( t \) we find the skull to be

\[
t = \frac{5730 \log(1/10)}{-\log 2} = 5730 \frac{\log 10}{\log 2} \text{ years old.} 
\]

To illustrate again the use of Taylor polynomials, we will estimate the logarithms without a calculator. We have seen that the Taylor polynomials of \( \log(1 + x) \) are

\[
p_n(x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n}. 
\]

Taking \( x = 1 \) and \( n = 10 \) gives the approximation

\[
\log 2 \approx 1 - \frac{1}{2} + \frac{1}{3} - \cdots - \frac{1}{10} = \frac{1627}{2520}. 
\]

Now for \( \log 10 \). We have seen that the Taylor polynomials do not approximate \( \log x \) for \( x > 1 \). However, since \( 10 = 2^3 + 2 = 2^3(1 + \frac{1}{4}) \), we have

\[
\log 10 = \log(2^3) + \log \left(1 + \frac{1}{4}\right) = 3\log 2 + \log \left(1 + \frac{1}{4}\right). 
\]

We have already approximated \( \log 2 \). Taking \( n = 4 \) gives the approximation

\[
\log \left(1 + \frac{1}{4}\right) \approx \frac{1}{4} - \frac{1}{2 \cdot 4^2} + \frac{1}{3 \cdot 4^3} - \frac{1}{4 \cdot 4^4} = \frac{685}{3072}. 
\]

So the age of the skull is about

\[
t = 5730 \frac{\log 10}{\log 2} = 5730 \left(3 + \frac{685}{3072} \cdot \frac{2520}{1627}\right) \text{ years old.} 
\]

This rational number works out to be a little over 19,168 years. Using more precise values of the logarithms, we would get about 19034 years. Considering that C14 dating is itself an estimate, perhaps we should just call it 19,000 years old.
Some functions blow up faster than others. For example, \( \log x \to \infty \), \( x^n \to \infty \) and \( e^x \to \infty \) as \( x \to \infty \), but in practice, there are significant differences between logarithmic, polynomial and exponential growths. In this section we will analyze this hierarchy of growth rates, which can be briefly summarized as: \textit{exponential crushes power crushes logarithm.}

This means that \( e^x \) grows much faster than any power of \( x \), which in turn grows much faster than \( \log x \), as \( x \to \infty \).

The picture to the left shows five functions that blow up as \( x \to \infty \). It indicates that \( e^x \) is bigger than the powers \( x^2 \), \( x \) and \( \sqrt{x} \) which are in turn bigger than \( \log x \), but it does not speak to \textit{crushing}. That is because the picture is confined to relatively small \( x \) in order to fit on the page, precisely because \( e^x \) grows so rapidly. We will prove the growth hierarchy (crushing) by comparing \( e^x \) to its Taylor polynomials.

First we need the \textbf{Inequality Tool}:

\[
\text{If } f(0) \geq g(0) \text{ and } f'(x) \geq g'(x) \text{ for all } x \geq 0, \text{ then } f(x) \geq g(x) \text{ for all } x \geq 0.
\]

\textbf{Proof:} Let \( h(x) = f(x) - g(x) \) be the difference function. By our assumptions, \( h(0) \geq 0 \) and \( h'(x) \geq 0 \) for \( x \geq 0 \). This last means that \( h(x) \) is increasing, so \( h(x) \geq 0 \) for all \( x \geq 0 \). But that means \( f(x) \geq g(x) \), as claimed. \textbf{QED}

Take \( f(x) = e^x \) and \( g(x) = 1 + x \). Then \( f(0) = g(0) = 1 \) and \( f'(x) = e^x \), \( g'(x) = 1 \). Since \( e^x \geq 1 \) for \( x \geq 0 \), the Inequality Tool tells us that \( e^x \geq 1 + x \).

Now take \( f(x) = e^x \) and \( g(x) = 1 + x + x^2/2 \). Then \( f(0) = g(0) = 1 \) and \( f'(x) = e^x \), \( g'(x) = 1 + x \). We just proved that \( e^x \geq 1 + x \) for \( x \geq 0 \), so now the Inequality Tool tells us that

\[
e^x \geq 1 + x + \frac{x^2}{2}.
\]

Now take \( f(x) = e^x \) and \( g(x) = 1 + x + x^2/2 + x^2/3! \). Then \( f(0) = g(0) = 1 \) and \( f'(x) = e^x \), \( g'(x) = 1 + x + x^2/2 \). We just proved that \( e^x \geq 1 + x + x^2/2 \) for \( x \geq 0 \), so now the Inequality
$e^x$ is bigger than its Taylor polynomials

Tool tells us that

$$e^x \geq 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!},$$

continuing in this way we conclude that

$$e^x \geq 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} \quad \text{for every integer } n \geq 1.$$

Since $n$ could be any positive integer, we could even go one step further:

$$e^x \geq 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!} \quad \text{for every integer } n \geq 1.$$

Dividing by $x^n$, we get,

$$\frac{e^x}{x^n} \geq \frac{1}{x^n} + \frac{1}{x^{n-1} \cdot 1!} + \frac{1}{x^{n-2} \cdot 2!} + \cdots + \frac{1}{n!} + \frac{x}{(n+1)!} \quad \text{for every integer } n \geq 1.$$

As $x \to \infty$ the terms with $x$ in their denominator go to zero and the last term goes to infinity. So we have

$$\lim_{x \to \infty} \frac{e^x}{x^n} = \infty.$$
Reciprocally, this says that
\[
\lim_{x \to \infty} \frac{x^n}{e^x} = 0
\] (9)

At first sight this limit is a contest between \(x^n\) and \(e^x\); as \(x \to \infty\), the numerator \(x^n\) is trying to make \(x^n/e^x\) large, while the denominator \(e^x\) is trying to make \(x^n/e^x\) small. We have just proved that \(e^x\) wins. That is what we mean by \(e^x\) crushing powers as \(x \to \infty\).

Note we are not saying that \(e^x\) is always bigger than \(x^n\); this is only true for sufficiently large \(x\). To illustrate with \(n = 3\), we see that \(x^3/e^x > 1\) near \(x = 3\), but that \(x^3/e^x \to 0\) as \(x \to \infty\).

As for powers vs. logarithms, it is the power function \(x^n\) which wins:
\[
\lim_{x \to \infty} \frac{\log x}{x^n} = 0 \quad \text{for every } n > 0.
\] (10)

Proof: We are asking about the behavior of \((\log x)/x^n\) for large \(x\). If \(x = e^y\) and \(y\) is large, then \(x\) is certainly large. So
\[
\lim_{x \to \infty} \frac{\log x}{x^n} = \lim_{y \to \infty} \frac{\log e^y}{(e^y)^n} = \lim_{y \to \infty} \frac{y}{e^{ny}} = \frac{1}{n} \lim_{y \to \infty} \frac{ny}{e^{ny}}.
\]

If \(ny = z\) and \(y\) is large, then \(z\) is certainly large. As a special case of (9), we know that \(z/e^z \to 0\) as \(z \to \infty\). Therefore, \((\log x)/(x^n) \to 0\) as \(x \to \infty\), as claimed. QED

In this proof we did something new with limits: we transformed a limit as \(x \to \infty\) into a limit as \(y \to \infty\) via the change from \(x \to e^y\). This useful technique can be applied in other situations. For example, We can transform a limit as \(x \to 0\) into a limit as \(y \to \infty\) via the change from \(x \to e^{-y}\); this might allow us to use our hierarchy to compute limits as \(x \to 0\).

Example 1: Compute
\[
\lim_{x \to 0} x \log x.
\]
This is another hard limit, of the form \(0 \cdot (-\infty)\). We let \(x = e^{-y}\), so that \(x \to 0\) as \(y \to \infty\). Since \(x \log x = e^{-y}(-y)\), we have
\[
\lim_{x \to 0} x \log x = - \lim_{y \to \infty} \frac{y}{e^y} = 0,
\]
since exponentials crush powers.
6 DEEPER THEOREMS OF CALCULUS

6.1 The existence theorems for continuous functions

First, what exactly is a “continuous function”? Is it a function whose graph has no holes in it? Is it a function whose graph can be drawn without lifting your pencil off the paper? In fact neither of these conditions make any sense if you think about them carefully.

First, every physical graph has holes in it. The computer has holes between pixels, the chalk and pencil-lead have vast empty spaces in the atoms of the chalk and pencil-lead molecules.

Second, we have seen that the graphs of some functions cannot be physically drawn either by computer or by human. One example is the function

\[ f(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases} \]

which actually is continuous, even though we cannot see the graph to check if it has any holes.

We will clear this up with the following precise definition of “continuous function”.

First, we must specify both the function \( f(x) \) and a closed interval \([a, b]\) on which \( f(x) \) is defined.

So suppose \( f(x) \) is a function defined on a closed interval \([a, b]\).

**Definition** We say that \( f(x) \) is **continuous** on \([a, b]\) if

\[
\lim_{h \to 0} f(x + h) = f(x),
\]

for every \( x \) in \([a, b]\).

If \( x = a \) the limit is as \( h \to 0^+ \). If \( x = b \) we take the limit as \( h \to 0^- \).

If \( f(x) \) is defined on the whole real number line \((-\infty, \infty)\) it may be continuous on some closed intervals but not others. If \( f(x) \) is continuous on every closed interval we say it is **continuous everywhere**.

**Examples:** The function

\[ f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases} \]

\(^9\)Recall that \([a, b]\) consists of all the numbers \( x \) such that \( a \leq x \leq b \). We will also use open intervals \((a, b)\), consisting of all \( x \) such that \( a < x < b \), and half-open intervals \([a, b), (a, b]\), defined similarly.
is continuous on every closed interval not containing 0.

The function

$$g(x) = \begin{cases} 
\sin(1/x) & \text{if } x \neq 0 \\
0 & \text{if } x = 0,
\end{cases}$$

is continuous everywhere.

### 6.2 The Max/Min Theorem

Here is our first deep theorem on continuous functions.

**MAX/MIN THEOREM**

If $f(x)$ is continuous on $[a, b]$ then there are numbers $c$ and $d$ in $[a, b]$ such that

$$f(c) \leq f(x) \leq f(d)$$

for all numbers $x$ in $[a, b]$.

The number $f(c)$ is the minimum value of $f(x)$ on $[a, b]$ and $f(d)$ is the maximum value.

The Max/Min Theorem guarantees that a continuous function $f(x)$ has a maximum and minimum value on $[a, b]$, but it does not tell us how to find these values. However, if $f(x)$ is also differentiable on $(a, b)$, then the max and min will occur either at one of the endpoints $a, b$ or at a point $c$ in $(a, b)$ where $f'(c) = 0$. Also $f(c)$ is a maximum if $f''(c) < 0$ and is a minimum if $f'(c) > 0$.

**Example 1:** Take the function $f(x)$ defined on the interval $[0, 1]$ by

$$f(x) = \begin{cases} 
x \log x & \text{if } x > 0 \\
0 & \text{if } x = 0.
\end{cases}$$

For $x$ in $(0, 1]$ the limit $\lim_{h \to 0} f(x + h)$ is a trivial limit and equals $f(x)$. Also, you showed in hw 6 F that $\lim_{h \to 0^+} h \log h = 0$. Since $f(0) = 0$, this means $f(x)$ is continuous at 0 as well. Therefore, $f(x)$ is continuous on the whole closed interval $[0, 1]$. The Max/Min theorem guarantees that $f(x)$ has a maximum and minimum value on $[0, 1]$, but does not tell us how to find these values. Since $f(0) = 0 = f(1)$, we only know that the minimum is $\leq 0$ and the maximum is $\geq 0$. 

60
However, in this example, the function is also differentiable on $(0, 1)$, and if the max or min occurs in $(0, 1)$ then it must be at a point $c$ where $f'(c) = 0$. Moreover, $f(c)$ is a max if $f''(c) < 0$ and is a min if $f''(c) > 0$.

We find that

$$f'(x) = 1 + \log x, \quad f''(x) = 1/x$$

for $x$ in $(0, 1)$. Since $f'(x) = 0$ only for $x = 1/e$, and $f''(1/e) > 0$, it follows that $f(1/e) = -1/e$ is the minimum value of $f(x)$ on $[0, 1]$. Thus $1/e$ is the point $c$ asserted to exist by the Max/Min theorem. What about $d$? Since $\log x < 0$ for $x$ in $(0, 1)$ it follows that the maximum value of $f(x)$ on $[0, 1]$ is 0, so $d$ could be either 0 or 1.

Note finally that it was crucial that the interval be closed. The function $f(x) = x \log x$ has no maximum on $(0, 1)$ (though it does have a minimum on $(0, 1)$, which is still $1/e$). Worse, if the interval is $(0, 1/e)$, the function $x \log x$ has no minimum or maximum. So you must have a closed interval $[a, b]$, in order to apply the Max/Min theorem.

![Graph of x log x](image)

**Example 2:** $f(x) = x^2 - \sin x$ on the interval $[0, \pi/2]$. We note that $f(0) = 0$ and $f(\pi/2) = \pi^2/4 - 1 > 0$. So the minimum value must be $\leq 0$ and the maximum must be $\geq \pi^2/4 - 1$.

We again take advantage of the fact that $f(x)$ is differentiable. We compute $f'(x) = 2x - \cos x$. Looking at the graphs of $y = 2x$ and $y = \cos x$, it appears that the graphs cross exactly once at some point $c$ in $(0, \pi/2)$. 

61
So \( f'(c) = 0 \). And since \( f''(x) = 2 + \sin x \) is always positive, it must be that \( f(c) \) is the minimum value of \( f(x) \) on \([0, \pi/2]\). However, the Max/Min theorem does not tell us what \( c \) is.

**Example 3:** Our use of derivatives in Examples 1 and 2 took advantage of some extra property (differentiability) enjoyed by the functions \( x \log x \) and \( x^2 - \sin x \). In fact,

> there are continuous functions that have no derivatives at any point at all.

Such functions were discovered in the mid 19th century by Bolzano and Weierstrass. Here is a picture of one of Weierstrass’ functions.
To write down the formula for $f(x)$ you start with $\sin x$, then add $\frac{1}{2} \sin 2x$, then add $\frac{1}{4} \sin 4x$ to that, and keep going:

$$f(x) = \sin x + \frac{1}{2} \sin 2x + \frac{1}{4} \sin 4x + \frac{1}{8} \sin 8x + \cdots$$

You can see a drunken version of $\sin x$ in the above graph. These kinds of infinite sums are called *Fourier Series*. They are used to describe many kinds of periodic phenomena, such as molecular vibrations, planetary orbits and heartbeats.

Anyway, Weierstrass proved that his function $f(x)$, crazy as it is, is still continuous on the whole real line. Hence by the Max/Min theorem, it has a minimum and maximum value on any closed interval $[a,b]$. But we cannot find these maxima and minima by differentiating, because $f'(x)$ does not exist for any $x$.

### 6.3 The Intermediate Value Theorem

You may have noticed in Example 2 above that I was hesitant to say that the graphs of $y = 2x$ and $y = \cos x$ actually cross in $[0, \pi/2]$, even though the graph indicates that they do. After all, at $x = 0$ the second function is bigger, and at $x = \pi/2$ the first function is bigger, so is it not obvious that at some point in between the graphs of the functions must cross?

Let us examine this issue in a simpler example: $f(x) = x^2$ on the interval $[0, 2]$. Since $f(0) = 0$ and $f(2) = 4$, the values of the function must pass through 2 on their way from 0 to 4, right? Of course they do, for indeed $f(\sqrt{2}) = 2$. This is obvious from the picture below, right?

[Link to the animation showing the intermediate value theorem](roy/general/weierstrass/weier.htm) which also has a striking animation zooming in showing how microscopically jagged the function is, as it must be, since it has no tangent line anywhere.
In this picture, the $x$ and $y$-axis are shown as real number lines.

Now I will draw the graph showing the $x$ and $y$ axes as rational number lines. Rational numbers are fractions $n/m$ where $n,m$ are integers. They are the only numbers you can know exactly or write down completely. Now between any two rational numbers, say $r$ and $s$, there is a third rational number $(r + s)/2$ which is the midpoint between $r$ and $s$. This means there are no holes in the rational number line, right? So the rational number line looks just like the real number line, so when we draw the graph of $y = x^2$ with only rational numbers on the $x$ and $y$ axes, the picture looks exactly the same:
However, this time the graph does not cross the line \( y = 2 \), because the required \( x \)-coordinate (namely \( x = \sqrt{2} \)) is not rational. \(^{11}\) It just skips right over it, even if we cannot see exactly where the skip starts and ends.

So there must be invisible holes in the rational numbers, after all. This is disturbing, for who is to say that the real numbers do not have invisible holes as well? Maybe even \( \sqrt{2} \) is a hole. What is \( \sqrt{2} \) anyway? They tell us it is a positive number whose square is 2. How do we know there is such a number? Has anyone ever seen it? Can you tell me what it is exactly? Your computer may tell you that

\[
\sqrt{2} = 1.41421356237309504880168872420969807856967187537694807317667973799
\]

but this is not exactly right, because this long decimal is a rational number, whereas we know that \( \sqrt{2} \), if it exists, cannot be rational. If you square this long decimal you will be slightly less than 2, which is below the possible skip that we’re worried about, so it does not reassure us at all.

Fortunately, the Intermediate Value Theorem does reassure us, by showing that the graph of a continuous function on a closed real interval \([a, b]\) does not have any invisible holes. The exact statement is this:

\(^{11}\)If \( \sqrt{2} \) were rational it would be a quotient of two integers, say \( \sqrt{2} = n/m \). This would mean that \( 2m^2 = n^2 \), but 2 divides \( n^2 \) an even number of times and 2 divides \( 2m^2 \) an odd number of times, so it cannot be that \( 2m^2 = n^2 \). This contradiction shows that \( \sqrt{2} \), if it exists, cannot be a rational number.
INTERMEDIATE VALUE THEOREM

If \( f(x) \) is a continuous function on a closed (real) interval \([a, b]\) and if \( d \) is any number between \( f(a) \) and \( f(b) \), then there is a number \( c \) in \([a, b]\) such that \( f(c) = d \).

The number \( d \) is the "intermediate value".

The Intermediate Value Theorem (IVT) is used to show that certain equations have solutions. You can usually set things up so that \( d = 0 \) in the IVT. Sometimes the solutions are famous numbers, as we’ll now see in a series of examples.

**Example 3a:** Consider the function \( f(x) = x^2 - 2 \). This function is continuous everywhere. We have \( f(1) = -2 < 0 \) and \( f(2) = 2 > 0 \). Apply the IVT on the interval \([1, 2]\) with \( d = 0 \). Then by the IVT there is a number \( c \) in \([1, 2]\) such that \( f(c) = 0 \). Thus, \( c^2 = 2 \) and \( c \) is positive, so \( c = \sqrt{2} \). So, whew: despite our doubts, there does in fact exist a real number whose square is 2.

**Example 3b:** Consider the function \( f(x) = \sin x \) on the interval \([\pi/2, 3\pi/2]\). The function is continuous everywhere and takes values \(+1\) and \(-1\) at the endpoints of the interval. By the IVT there is a number \( c \) in \([\pi/2, 3\pi/2]\) such that \( f(c) = 0 \). In fact, \( c = \pi \).

**Example 3c:** Consider the function \( f(x) = \log x - 1 \) on the interval \([1, 3]\), where \( f(x) \) is continuous. We have \( f(1) = -1 < 0 \) and \( f(3) = \log 3 - 1 > \log e - 1 = 0 \). By the IVT there is a number \( c \) in \([1, 3]\) such that \( f(c) = 0 \). In fact \( c = e \).

Sometimes the solution whose existence is guaranteed by the IVT is not a famous number and does not have a name.

**Example 4:** Show that the equation \( 2x = \cos x \) has a solution in the interval \([0, 1]\).

Let \( f(x) = 2x - \cos x \). This function is continuous everywhere. We know that \( f(0) = -1 < 0 \) and \( f(1) = 2 - \sin(1) > 0 \). In the IVT we take \( d = 0 \). Then by the IVT there is a number \( c \) in \([0, 1]\) such that \( f(c) = 0 \). Thus, \( x = c \) is a solution of the equation \( 2x = \cos x \). Our task is accomplished, but at this point we have no idea what \( c \) is. Later we’ll find ways to approximate \( c \) to arbitrary precision.

The IVT has some surprising consequences.

**Example 5:** There are two antipodal points on the earth having the exact same temperature. (Assuming temperature varies continuously.)
Proof: Choose any two antipodal points \( p \) and \( q \) on the earth. If they have the same temperature, we’re done. Suppose they don’t have the same temp, say temp at \( p > \) temp at \( q \). Make a great circle \( C \) around the earth passing through \( p \) and \( q \). Move \( p \) and \( q \) in the same direction along \( C \), always remaining antipodal. Let \( f(\theta) \) be the temp at \( p \) minus the temp at \( q \) after we’ve moved the points \( \theta \) radians around \( C \). Then \( f(\theta) \) is a continuous function on \([0, \pi]\), by our assumption that temp varies continuously. We have \( f(0) > 0 \) because initially \( p \) is warmer than \( q \). When \( \theta = \pi \) the points have reversed, so \( f(\pi) < 0 \). By the IVT there is some \( c \) in \([0, \pi]\) for which \( f(c) = 0 \). Therefore the antipodal points at \( \theta = c \) have the same temperature.

6.4 About proofs of the Deep Theorems

The Max/Min and Intermediate Value Theorems are deep because they depend on a subtle property that real numbers have, but that rational numbers do not have, called the “Least Upper Bound” property. To explain it requires first explaining exactly what a real number is. These issues are beyond the scope of our course (see MATH3310 or MATH3311), so we will not prove these two Deep Theorems. But we will prove all of their consequences for differentiable functions.

7 BASIC THEOREMS ABOUT DIFFERENTIABLE FUNCTIONS

We embark on a journey through several closely related theorems that rely on the Deep Theorems just discussed. All of these new theorems will be about a function \( f(x) \) which is continuous on \([a, b]\) and differentiable on \((a, b)\). Each theorem will assert the existence of some point \( c \) in \((a, b)\) such that something is true about \( f'(c) \). The theorems will not tell us what \( c \) is, but they will still be very useful.

7.1 Rolle’s Theorem

All of the theorems will follow from this first one:

Rolle’s Theorem

Suppose \( f(x) \) is continuous on \([a, b]\) and differentiable on \((a, b)\). Assume also that \( f(a) = f(b) = \)
0. Then there is a real number \( c \) in \((a, b)\) such that \( f'(c) = 0 \).

**Proof:** The condition that \( f'(x) \) is differentiable on \((a, b)\) means that the limit

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

exists (and is finite) for all \( x \) in \((a, b)\). The proof breaks into three cases.

I. Suppose \( f(x) = 0 \) for all \( x \) in \([a, b]\). Then for every \( x \) in \((a, b)\) and \( h > 0 \) we have \( f(x + h) - f(x) = 0 - 0 = 0 \), so the above limit is zero and we have \( f'(c) = 0 \) for any \( c \) in \((a, b)\).

II. Suppose \( f \) takes a positive value somewhere on \([a, b]\). By the Max/Min theorem, there is a point \( c \) in \([a, b]\) such that \( f(x) \leq f(c) \) for all \( x \) in \([a, b]\). Since \( f(x) \) takes a positive value somewhere on \([a, b]\), we must have \( f(c) > 0 \). Note that \( c \neq a \) and \( c \neq b \) since \( f(a) = f(b) = 0 \). Therefore \( c \) is in \((a, b)\).

Since \( f(c) \) is the maximum value of \( f \) on \([a, b]\), we have \( f(c + h) \leq f(c) \), for every \( h > 0 \) such that \( c + h \) is in \((a, b)\). so

\[
f(c + h) - f(c) \leq 0.
\]

Now if \( h \to 0 \) from the positive side (meaning \( h > 0 \)) we have \( (f(c + h) - f(c))/h \leq 0 \), so that

\[
f'(c) = \lim_{h \to 0} \frac{f(c + h) - f(c)}{h} \leq 0.
\]

And if \( h \to 0 \) from the negative side (meaning \( h < 0 \)) we have \( (f(c + h) - f(c))/h \geq 0 \), so that

\[
f'(c) = \lim_{h \to 0} \frac{f(c + h) - f(c)}{h} \geq 0.
\]

The only way that \( f'(c) \) can be both \( \geq 0 \) and \( \leq 0 \) is if \( f'(c) = 0 \).

III. Suppose \( f \) takes a negative value somewhere on \([a, b]\). By the Max/Min theorem, there is a point \( d \) in \([a, b]\) such that \( f(x) \geq f(d) \) for all \( x \) in \([a, b]\). Since \( f(x) \) takes a negative value somewhere on \([a, b]\), we must have \( f(d) < 0 \), so that as above \( d \) is in \((a, b)\).

Since \( f(d) \) is the minimum value of \( f \) on \([a, b]\), we have \( f(d + h) \geq f(d) \) for every \( h > 0 \) such that \( d + h \) is in \((a, b)\). so

\[
f(d + h) - f(d) \geq 0.
\]

Now if \( h \to 0 \) from the positive side we have \( (f(d + h) - f(d))/h \geq 0 \), so that

\[
f'(d) = \lim_{h \to 0} \frac{f(d + h) - f(d)}{h} \geq 0.
\]

And if \( h \to 0 \) from the negative side we have \( (f(d + h) - f(d))/h \leq 0 \), so that

\[
f'(d) = \lim_{h \to 0} \frac{f(d + h) - f(d)}{h} \leq 0.
\]
Again, the only way that \( f'(d) \) can be both \( \geq 0 \) and \( \leq 0 \) is if \( f'(d) = 0 \).

Therefore in all cases we have shown that \( f' \) is equal to zero at some point in \((a,b)\). \textbf{QED.}

It may seem surprising that Rolle’s theorem is useful, even though it does not tell us what \( c \) is.

**Example 1:** Show that the polynomial \( x^3 + x + 1 \) has exactly one real root.

Let \( f(x) = x^3 + x + 1 \). We note that \( f(0) = 1 \) and \( f(-1) = -1 \). By the IVT, \( f \) has a zero somewhere in \((-1,0)\). We must show there are no other zeros. If there were other zeros, there would be a pair of numbers \( a, b \) such that \( f(a) = f(b) = 0 \). By Rolle’s Theorem there would be a number \( c \) in \((a,b)\) such that \( f'(c) = 0 \). But \( f'(x) = 3x^2 + 1 \) is always positive, hence is never zero. So there can be no such \( c \), hence there can be no pair of numbers \( a, b \) for which \( f(a) = f(b) = 0 \). Therefore \( f(x) \) has only one real root.

### 7.2 The Mean-Value Theorem

The Mean-Value-Theorem (MVT) says that the slope of the secant line is equal to the slope of at least one tangent line.

The MVT is actually equivalent to Rolle’s theorem, but it looks different. What’s happening is that both theorems examine the same phenomenon from different angles (literally).

---

**Mean-Value Theorem**

*Suppose \( f(x) \) is continuous on \([a,b]\) and differentiable on \((a,b)\). Then there is a real number \( c \) in \((a,b)\) such that*

\[
f'(c) = \frac{f(b) - f(a)}{b - a}.
\]

---

**Proof:** The idea is to apply Rolle’s theorem to the function \( g(x) \) which measures the vertical distance from the graph of \( f \) to the secant line. This secant line has slope

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

and is defined by the equation

\[
y - f(a) = m(x - a).
\]

So the secant line is the graph of the function

\[
\ell(x) = m(x - a) + f(a).
\]
The main thing about $\ell(x)$ that $\ell(a) = f(a)$ and $\ell(b) = f(b)$. Now let

$$g(x) = f(x) - \ell(x).$$

Then $g(a) = f(a) - \ell(a) = 0$ and $g(b) = f(b) - \ell(b) = 0$. By Rolle’s theorem (applied to $g(x)$) there is a real number $c$ in $(a, b)$ such that $g'(c) = 0$.

Since $g(x) = f(x) - \ell(x)$, we have $g'(x) = f'(x) - \ell'(x)$. And since $\ell(x)$ is a line with slope $m$, we have $\ell(x) = m$ for all $x$. Therefore,

$$g'(c) = f'(c) - m = 0.$$

This means

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

as we wanted to prove. QED

**Example:** The Pennsylvania Turnpike Commission (PTC) uses the MVT to give speeding tickets based on the time stamp on your toll ticket. Here $f(t) =$ the distance you have travelled from the entry toll booth at time $t$, $a = 0$ and $b$ is the time you arrive at the exit toll booth, so that

$$\frac{f(b) - f(a)}{b - 0} = \frac{f(b)}{b} = \frac{L}{b},$$

where $L$ is the distance between entry and exit. If $L/b$ is greater than the speed limit, then by the MVT, the PTC knows that at some time $c$ your speed $f'(c) = L/b$ exceeded the speed limit, so you get a ticket.

You will note that the proof of the MVT was much simpler than the proof of Rolle’s theorem. With Rolle’s in hand, we just needed the trick of defining the function $g(x)$, which essentially rotated the MVT picture into the Rolle’s picture.

On the other hand, the MVT implies a better version of Rolle’s theorem, which we’ll call

---

**Rolle’s Theorem v.2**

*Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on $(a, b)$. Assume also that $f(a) = f(b)$. Then there is a real number $c$ in $(a, b)$ such that $f'(c) = 0$.***

---

**Proof:** This follows immediately from the MVT, because $f(b) - f(a) = 0$. QED.

The only difference between this and the original Rolle’s is that we no longer need to assume that $f(a) = f(b) = 0$, just that $f(a) = f(b)$.
7.2.1 First consequence of the MVT: Solving differential equations

The first consequence allows you to determine a function from its derivative.

**Consequence 1:** If \( f'(x) = 0 \) for all \( x \) in \((a, b)\) then \( f(x) \) is constant on \((a, b)\).

**Proof:** Let \( x_1 \) and \( x_2 \) be any two points in \((a, b)\). For some \( c \) in \((x_1, x_2)\) we have

\[
\frac{f(x_1) - f(x_2)}{x_1 - x_2} = f'(c),
\]

by the MVT applied to the interval \((x_1, x_2)\). Since \( c \) is in \((a, b)\) and \( f' \) is zero at every point in \((a, b)\) it follows that \( f'(c) = 0 \). This means \( f(x_1) = f(x_2) \). Since \( x_1 \) and \( x_2 \) were any two points in \((a, b)\), it follows that \( f(x) \) is constant on \((a, b)\). QED.

We used this earlier, for example, to show that if \( f'(x) = kf(x) \), where \( k \) is a constant, then \( f(x) = f(0)e^{kx} \). Indeed, recall that the condition \( f' = kf \) implies that \((f/e^{kx})' = 0\), so by the first consequence of the MVT, we have \( f/e^{kx} = C \) for some constant \( C \). Setting \( x = 0 \) we find that \( C = f(0) \), which implies that \( f(x) = f(0)e^{kx} \).

7.2.2 Second consequence of the MVT: Proving inequalities

The second consequence allows you to prove inequalities between functions.

**Consequence 2:** Suppose \( f(x) \) and \( g(x) \) are differentiable for \( x > a \). Assume that for \( x > a \) we have

\[
f(a) \geq g(a) \quad \text{and} \quad f'(x) \geq g'(x).
\]

Then \( f(x) \geq g(x) \) for all \( x > a \).

**Proof:** Let \( h(x) = f(x) - g(x) \). Then \( h(a) \geq 0 \) and \( h'(x) \geq 0 \) for all \( x > a \). Choose any number \( b \) such that \( b > a \) and apply the MVT to \( h(x) \) on the interval \((a, b)\). By the MVT there is a number \( c \) in \((a, b)\) such that

\[
\frac{h(b) - h(a)}{b - a} = h'(c).
\]

Since \( c > a \), we have \( h'(c) > 0 \). Therefore \( h(b) - h(a) \geq 0 \). Since \( h(a) \geq 0 \), it follows that \( h(b) \geq 0 \). This means that \( f(b) \geq g(b) \). Since \( b \) was any number \( > a \), it follows that \( f(x) \geq g(x) \) for all \( x > a \). QED.

We used this earlier, to show that \( e^x \geq 1 + x + x^2/2 \), then from this that \( e^x \geq 1 + x + x^2/2 + \cdots + x^n/n! \), and so on, so that

\[
e^x \geq 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}
\]
for any integer \( n \geq 0 \). In turn, this inequality allowed us to show that exponential growth crushes polynomial growth.

## 7.3 Cauchy’s Mean-Value Theorem

Augustin-Louis Cauchy was a 19th century French mathematician involved in understanding the subtle aspects of Calculus that earlier mathematicians such as Newton and Leibniz had overlooked.

Cauchy realized that the Mean-Value Theorem is really about two functions, but one of them is hidden in the formulation of the MVT that we’ve just seen. Here is Cauchy’s version the MVT, which we’ll call the CMVT:

### Cauchy’s Mean-Value Theorem

Suppose \( f(x) \) and \( g(x) \) are two functions, each continuous on \([a, b]\) and differentiable on \((a, b)\). Then there is a real number \( c \) in \((a, b)\) such that

\[
g'(c)[f(b) - f(a)] = f'(c)[g(b) - g(a)].
\]

### Proof:

If \( g(b) = g(a) \), then by Rolle’s v.2 there is a number \( c \) in \((a, b)\) such that \( g'(c) = 0 \). Then the equation in the CMVT holds because both sides are zero.

Assume \( g(b) \neq g(a) \). Let

\[
m = \frac{f(b) - f(a)}{g(b) - g(a)}
\]

and consider the function \( h(x) = f(x) - mg(x) \). Then you can check that

\[
h(b) - h(a) = f(b) - f(a) - m[g(b) - g(a)] = 0.
\]

By Rolle’s v.2 again there is a number \( c \) in \((a, b)\) such that \( h'(c) = 0 \). This means that

\[
0 = h'(c) = f'(c) - mg'(c),
\]

or

\[
\frac{f'(c)}{g'(c)} = m = \frac{f(b) - f(a)}{g(b) - g(a)}.
\]
Cross-multiplying, we get

\[ g'(c)[f(b) - f(a)] = f'(c)[g(b) - g(a)], \]

as desired. QED.

To visualize the CMVT, imagine bug travelling on the \((x, y)\)-plane, such that at time \(t\) the bug is at the point \((g(t), f(t))\). So the bug starts at the point \((g(a), f(a))\) and finishes at the point \((g(b), f(b))\).

The number \(m\) in the proof of the CMVT is the slope of the straight line path from \((g(a), f(a))\) to \((g(b), f(b))\). The number \(f'(c)/g'(c)\) is the slope of the tangent line to the path of the bug at time \(t = c\). The CMVT says that at some point, the bug is travelling in the same direction as the straight line path.

If \(g(x) = x\), we get the old MVT. This \(x\) is the hidden function. In essence, Cauchy asked "why can’t we use another function besides \(x\)?”.

This turns out to be a very good idea. For the CMVT also has consequences that are even more amazing than those for the MVT.

### 7.4 The error in the Taylor expansion

If \(f(x)\) is continuous on \([a, b]\) and differentiable on \((a, b)\) then according to the MVT there is a real number \(c\) in \((a, b)\) such that

\[ \frac{f(b) - f(a)}{b - a} = f'(c). \]

This could be written as

\[ f(b) = f(a) + f'(c) \cdot (b - a). \]

This is the \(n = 0\) case of

Taylor’s Formula with Error Term
Take any integer \( n \geq 0 \). Suppose \( f \) and its first \( n \) derivatives \( f', f'', \ldots, f^{(n)} \) are continuous on \([a, b]\) and differentiable on \((a, b)\). Then there is a real number \( c \) in \((a, b)\) such that

\[
f(b) = f(a) + \frac{f'(a)}{1!} (b - a) + \frac{f''(a)}{2!} (b - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!} (b - a)^n + \frac{f^{(n+1)}(a)}{(n+1)!} (b - a)^{n+1}.
\]

**Proof:** This proof is a bit tricky. We will apply the CMVT to the two functions

\[
F(x) = f(b) - \left[ f(x) + \frac{f'(x)}{1!} (b - x) + \frac{f''(x)}{2!} (b - x)^2 + \cdots + \frac{f^{(n)}(x)}{n!} (b - x)^n \right], \quad \text{and} \quad G(x) = \frac{1}{(n+1)!}(b - x)^{n+1}.
\]

The CMVT says there is a number \( c \) in \((a, b)\) such that

\[
G'(c) \cdot [F(b) - F(a)] = F'(c) \cdot [G(b) - G(a)].
\]  

This will be the \( c \) that we want. Note that the equation we want to prove may be written as

\[
F(a) = \frac{f^{(n+1)}(c)}{(n+1)!} (b - a)^{n+1}.
\]  

We also note that \( F(b) = G(b) = 0 \), and that

\[
G'(x) = -\frac{1}{n!} (b - x)^n
\]

is nonzero at \( c \) since \( c \neq b \). Therefore equation (12) may be written (after cancelling minus signs on both sides) as

\[
F(a) = \frac{F'(c)}{G'(c)} \cdot G(a).
\]  

We now compute \( F'(x) \). In the formula for \( F(x) \), each term in the square brackets is of the form

\[
\frac{f^{(k)}(x)}{k!} (b - x)^k.
\]

We use the product rule and differentiate the \((b - x)\) factor first, to get

\[
\left( \frac{f^{(k)}(x)}{k!} (b - x)^k \right)' = -\frac{f^{(k)}(x)}{(k - 1)!} (b - x)^{k-1} + \frac{f^{(k+1)}(x)}{k!} (b - x)^k.
\]

The first term here will cancel the second term from

\[
\left( \frac{f^{(k-1)}(x)}{(k - 1)!} (b - x)^{k-1} \right)'.
\]
So each term in the square brackets will cancel except the last one which has no next term to cancel it. Thus, we get

\[ F'(x) = 0 - \left[ f'(x) - f'(x) + \frac{f''(x)}{1!}(b-x) - \frac{f''(x)}{2!}(b-x)^2 + \cdots + \frac{f^{(n+1)}(x)}{n!}(b-x)^n \right] = -\frac{f^{(n+1)}(x)}{n!}(b-x)^n. \]

Comparing with the formula above for \( G'(x) \), we find that

\[ F'(x) = f^{(n+1)}(x)G'(x). \]

Putting \( x = c \) here, we get

\[ \frac{F'(c)}{G'(c)} = f^{(n+1)}(c). \]

Since

\[ G(a) = \frac{1}{(n+1)!}(b-a)^{n+1}, \]

equation (14), which we have proved, becomes equation (13), which is the equation we needed to prove. QED.

That was a tough calculation. You don’t need to know the proof of the CMVT, but let’s see why the CMVT is useful.

First, we will only use the Taylor approximation when \( a = 0 \), and we will write \( b = x \). So we will only use the following form of the Taylor approximation:

\[ f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + f'''(0)\frac{x^3}{3!} + \cdots + f^{(n)}(0)\frac{x^n}{n!} + f^{(n+1)}(c)\frac{x^{n+1}}{(n+1)!} \]

for some \( c \) between 0 and \( x \). This last term is the Error:

\[ E_n = f^{(n+1)}(c)\frac{x^{n+1}}{(n+1)!}. \]

You will typically be given an acceptable error, and your task is to find \( n \) such that \( |E_n| \) is less than this acceptable error.

The problem is that you don’t know \( c \). You can overcome this by finding an explicit number \( M \) such that \( |f^{(n+1)}(c)| \leq M \) no matter what \( n \) and \( c \) might be. This is often easier than it sounds, if you remember that \( c \) is between 0 and \( x \).

Once you have found \( M \), then

\[ |E_n| \leq \frac{Mx^{n+1}}{(n+1)!}, \]

and it only remains to find \( n \) for which \( \frac{Mx^{n+1}}{(n+1)!} \) is less than the acceptable error.
Example 1: Compute $e$ with an error of no more than 1/10.

Our function here is $f(x) = e^x$, and the acceptable error is 1/10.

Since $f^{(n+1)}(x) = e^x$, the error is

$$E_n = \frac{e^c}{(n+1)!},$$

where $c$ is some number in the interval $(0, 1)$. Since $e < 3$ and $c < 1$ we have $e^c < 3$. Thus, we take $M = 3$. This means that no matter what $n$ and $c \in (0, 1)$ are, we always have $|f^{(n+1)}(c)| < 3$.

Thus, we have

$$|E_n| \leq \frac{3}{(n+1)!}. \quad (15)$$

We are required to have $|E_n| < 1/10$. This means we need $(n + 1)! > 30$. So $n = 4$ will do, because $5! = 120 > 30$. Thus

$$e \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!},$$

to within $3/120 = 1/40$, which is better than required.

Suppose the stakes are raised and we need to have the error less than 1/100. The bound (15) shows that $(n + 1)! > 300$ will work. This can be achieved with $n = 5$, because $6! = 720 > 300$. Thus,

$$e \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!},$$

to within an error of $3/720 = 1/240$, which is again better than required.

In this way we can now compute $e$ to any desired degree of accuracy.

Example 2: Compute $\sin(1)$ to within three decimal places.

We take $f(x) = \sin x$, and look at the error formula. Since $|f^{(n+1)}(c)| \leq 1$ and $x = 1$, we have

$$|E_n| \leq \frac{1}{(n+1)!}.\quad (12)$$

We want $|E_n| < 1/1000$. Since $7! = 5040 > 1000$, we can take $n = 6$. Since $f^{(6)}(0) = 0$, there is no $x^6$ term in $\sin x$, so

$$\sin(1) \approx 1 - \frac{1}{3!} + \frac{1}{5!},$$

to within $1/5040$, better than required.\footnote{In fact, $1/5040 \approx .0001984$, while $\sin(1) - [1 - \frac{1}{3!} + \frac{1}{5!}] \approx .0001956$.}

Thus, we can now compute $\sin(1)$ to to any desired degree of accuracy.

Example 3: Compute $\log 2$ to within $1/10$.

In fact, $1/5040 \approx .0001984$, while $\sin(1) - [1 - \frac{1}{3!} + \frac{1}{5!}] \approx .0001956$.\footnote{In fact, $1/5040 \approx .0001984$, while $\sin(1) - [1 - \frac{1}{3!} + \frac{1}{5!}] \approx .0001956$.}
We take \( f(x) = \log(1 + x) \) because \( \log x \) is not defined, hence certainly not differentiable at \( x = 0 \). We want to approximate \( f(1) \). We find that

\[
f^{(n+1)}(x) = (-1)^n \frac{n!}{(1 + x)^{n+1}},
\]

Hence the error \( E \) in the approximation for \( x = 1 \) has the bound

\[
|E| \leq \frac{n!}{(1 + c)^{n+1}} \cdot \frac{1}{(n + 1)!} = \frac{1}{(1 + c)^{n+1}} \cdot \frac{1}{n + 1} \leq \frac{1}{n + 1}
\]
since \( 0 < c < 1 \). We want \( 1/(n + 1) < 1/10 \), so take \( n = 10 \). Thus,

\[
\log 2 \approx 1 - \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \cdots + \frac{1}{10},
\]
to within 1/11. \(^{13}\)

**Example 4:** Compute \( \sin(61^\circ) \) to three decimal places.

First we convert to radians. \( 60^\circ = \pi/3 \) radians and \( 1^\circ = \pi/180 \) radians. For brevity, let’s write \( p = \pi/180 \). We go back to the original Taylor formula with remainder (with \( a \) and \( b \)) and take \( a = \pi/3(= 60^\circ) \) and \( b = a + p(= 61^\circ) \). Note that \( b - a = p \).

We want to approximate \( \sin(b) \). The error after the \( n^{th} \) term has absolute value

\[
\left| \frac{f^{(n+1)}(c)}{(n + 1)!} p^{n+1} \right| \leq \frac{p^{n+1}}{(n + 1)!},
\]

which we want to be < 1/1000. Now \( p = \pi/180 < 1/10 \), so \( p^3 < 1/1000 \) and we can take \( n = 2 \).

We have

\[
f(a) = \sin(\pi/3) = \sqrt{3}/2, \quad f'(a) = \cos(\pi/3) = 1/2, \quad f''(a) = -\sin(\pi/3) = -\sqrt{3}/2.
\]

The Taylor approximation at \( n = 2 \), accurate to within 1/1000, is

\[
\sin b \approx \frac{\sqrt{3}}{2} + \frac{p}{2} - \frac{\sqrt{3}p^2}{4}.
\]

### 7.5 L’Hôpital’s Rule

This is another application of Cauchy’s Mean-Value Theorem, which enables us to compute certain hard limits.

---

\(^{13}\)In fact, 1/11 \approx .090909, while \( \log 2 - [1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} - \cdots + \frac{1}{10}] \approx .04751. \)
Let \( f(x) \) and \( g(x) \) be functions defined on an interval \([a, b]\) and let \( c \) be a point in \([a, b]\). Assume that \( f(x) \) and \( g(x) \) are continuous on \([a, b]\) and differentiable on \((a, b)\), and that the following conditions hold.

1. \( f(c) = g(c) = 0 \);
2. \( g'(x) \neq 0 \) for all \( x \) in \([a, b]\) except possibly \( x = c \);
3. \( \lim_{x \to c} f'(x)/g'(x) = L \), where \( L \) is finite or infinite.

Then
\[
\lim_{x \to c} \frac{f(x)}{g(x)} = L.
\]
(The limits are one-sided limits if \( c = a \) or \( c = b \).)

**Proof:** Let \( h \) be any nonzero number such that \( c + h \) is in \([a, b]\). Note that \( g(c + h) \neq 0 \) for any such \( h \). For otherwise, we would have \( g(c) = g(c + h) = 0 \), so Rolle’s theorem would imply that \( g'(d) = 0 \) for some \( d \) between \( c \) and \( c + h \). But we are assuming that \( g'(x) \) is only zero when \( x = c \). So indeed, \( g(c + h) \neq 0 \).

The CMVT implies that there is a number \( c_h \) between \( c \) and \( c + h \) such that
\[
(f(c + h) - f(c))g'(c_h) = (g(c + h) - g(c))f'(c_h).
\]
We are given that \( f(c) = g(c) = 0 \), so in fact we have
\[
f(c + h)g'(c_h) = g(c + h)f'(c_h).
\]
Since \( g'(c_h) \neq 0 \) and \( g(c + h) \neq 0 \), we can cross-divide and get
\[
\frac{f(c + h)}{g(c + h)} = \frac{f'(c_h)}{g'(c_h)},
\]
for any number \( h \) such that \( c + h \) is in \( I \).

Since \( c_h \) is between \( c \) and \( c + h \), we have \( c_h \to c \) as \( h \to 0 \). Therefore, taking the limit as \( h \to 0 \), we get
\[
\lim_{h \to 0} \frac{f(c + h)}{g(c + h)} = \lim_{h \to 0} \frac{f'(c_h)}{g'(c_h)} = L.
\]
Since
\[
\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{h \to 0} \frac{f(c + h)}{g(c + h)},
\]
our result is proved. QED
There is also a version of L'Hôpital's rule for limits as \( x \to \pm \infty \). We will just prove it for \( x \to +\infty \).

### L'Hôpital's Rule, v.2

Let \( f(x) \) and \( g(x) \) be functions defined on an interval \([a, \infty)\), where \( a > 0 \). Assume that \( f(x) \) and \( g(x) \) are continuous on \([a, \infty)\), differentiable on \((a, \infty)\) and that the following conditions hold:

1. \( \lim_{x \to \infty} f(x) = \lim_{x \to \infty} g(x) = 0; \)
2. \( g'(x) \neq 0 \) for all \( x > a; \)
3. \( \lim_{x \to \infty} \frac{f'(x)}{g'(x)} = L \), where \( L \) is finite or infinite.

Then

\[ \lim_{x \to \infty} \frac{f(x)}{g(x)} = L. \]

**Proof:** We will transform this situation back into the version we have already proved. Define functions \( F(x) \) and \( G(x) \) on the interval \([0, 1/a]\) as follows:

\[ F(x) = \begin{cases} 
  f(1/x) & \text{if } x \neq 0 \\
  0 & \text{if } x = 0; 
\end{cases} \]

\( G(x) \) is defined similarly, using \( g(1/x) \). These functions are continuous on \((0, 1/a]\) and differentiable on \((0, 1/a)\), by the chain rule. Further, by condition 1 we have

\[ \lim_{x \to 0^+} F(x) = \lim_{x \to 0^+} G(x) = 0, \]

so \( F(x) \) and \( G(x) \) are actually continuous on \([0, 1/a]\). We may apply L'Hôpital's rule v.1 to these functions with the point \( c = 0 \), and use the chain rule to get

\[ \lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{x \to 0^+} \frac{F(x)}{G(x)} = \lim_{x \to 0^+} \frac{F'(x)}{G'(x)} = \lim_{x \to 0^+} \frac{-x^{-2} f'(1/x)}{-x^{-2} g'(1/x)} = \lim_{x \to 0^+} \frac{f'(1/x)}{g'(1/x)} = \lim_{x \to \infty} \frac{f'(x)}{g'(x)} = L, \]

as claimed. QED

Both versions of L'Hôpital's rule are true if condition 1 is replaced by

\[ \lim_{x \to \infty} f(x) = \lim_{x \to \infty} g(x) = \pm \infty. \]
This is proved by replacing $f(x)$ and $g(x)$ by $1/f(x)$ and $1/g(x)$. We leave this to the reader.

Many of our earlier hard limit calculations can be done more easily using L’Hopital’s rule.

**Example 1:**

$$\lim_{x \to \infty} \frac{x}{e^x}.$$  

Here $f(x) = x$ and $g(x) = e^x$ are differentiable everywhere and $g'(x)$ is never zero. We apply version 2 with any interval $[a, \infty)$ where $a > 0$, and get

$$\lim_{x \to \infty} \frac{x}{e^x} = \lim_{x \to \infty} \frac{1}{e^x} = 0.$$  

Here “$= \lim_{\text{LH}}$” means that we used L’Hôpital at this step.

We can use L’Hôpital’s rule over and over again, provided that at each stage all the conditions are met. So for any positive integer $n$ we have

$$\lim_{x \to \infty} \frac{x^n}{e^x} = \lim_{x \to \infty} \frac{n x^{n-1}}{e^x} = \lim_{x \to \infty} \frac{n(n-1) x^{n-2}}{e^x} = \cdots = \lim_{x \to \infty} \frac{n!}{e^x} = 0.$$  

Thus we have reproved our earlier result that $e^x$ crushes polynomials.

**Example 2:**

$$\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\cos x}{1} = 1.$$  

Here we used the fact that $(\sin x)' = \cos x$, but to compute this derivative we first had to prove the hard limit $\lim_{x \to 0} (\sin x)/x = 1$ via geometry. So this example is actually circular reasoning: in order to use L’Hôpital’s rule to prove a hard limit, we needed to have proved the hard limit already!

**Example 3:** Sometimes you just have to be persistent:

$$\lim_{x \to 0} \frac{3 \sin x - \sin 3x}{x - \sin x} = \lim_{x \to 0} \frac{3 \cos x - 3 \cos 3x}{1 - \cos x} = \lim_{x \to 0} \frac{-3 \sin x + 9 \sin 3x}{\sin x} = \lim_{x \to 0} \frac{-3 \cos x + 27 \cos 3x}{\cos x} = 24.$$  

**Example 4:** L’Hôpital’s rule gives the wrong answer if the conditions are not met. Consider

$$\lim_{x \to \infty} \frac{x + \sin x}{x} \neq \lim_{x \to \infty} \frac{1 + \cos x}{1} \text{ does not exist.}$$  

But the original limit is easy and does exist:

$$\lim_{x \to \infty} \frac{x + \sin x}{x} = \lim_{x \to \infty} \frac{1 + (\sin x)/x}{1} = 2.$$  

The problem here is that condition 3 fails, so L’Hôpital’s rule is invalid in this situation.

**Example 5:** Sometimes all the conditions are met, but L’Hôpital’s rule gets stuck. Consider

$$\lim_{x \to \infty} \frac{\sin x}{\cosh x} = \lim_{x \to \infty} \frac{\cosh x}{\sinh x} = \lim_{x \to \infty} \frac{\sinh x}{\cosh x}.$$  

80
We are back where we started. In fact this is an easy limit.

\[
\lim_{x \to \infty} \frac{\sinh x}{\cosh x} = \lim_{x \to \infty} \frac{e^x - e^{-x}}{e^x + e^{-x}} = \lim_{x \to \infty} \frac{1 - e^{-2x}}{1 + e^{-2x}} = \frac{1 - 0}{1 + 0} = 1.
\]

**Example 6:** You can use L'Hôpital's rule on a product of the form \(0 \cdot \infty\), by turning it into a quotient. Here is a limit we did earlier:

\[
\lim_{x \to 0^+} x \log x = \lim_{x \to 0^+} \frac{\log x}{1/x} = \lim_{x \to 0^+} \frac{1/x}{-1/x^2} = \lim_{x \to 0^+} (-x) = 0.
\]

8 OPTIMIZATION

The Max/Min theorem was the main foundation for our theoretical results in the previous section. Now we will use it to solve practical problems which ask for a quantity to be optimized subject to certain constraints.

We will express what is to be optimized by a function \(f(x)\), defined on a closed interval \([a, b]\), where it is continuous, and differentiable on \((a, b)\). The Max/Min Theorem says that \(f(x)\) has a minimum value and a maximum value on \([a, b]\). These values either occur at the endpoints \(a, b\) or at points \(x\) where \(f'(x) = 0\).

So the strategy is to find all points where \(f'(x) = 0\), evaluate \(f(x)\) at these points, and also evaluate \(f(a)\) and \(f(b)\). The maximum is the largest of these values and the minimum is the smallest of these values.

The strategy is very simple, once you have \(f(x)\). But finding \(f(x)\) is usually the part which requires the most thought.

**Example 1:** How can you make a rectangular pen of maximal area using 100 feet of fencing?

Let \(x, y\) be the dimensions of the pen. We want to maximize the area \(xy\), subject to the constraint that the perimeter \(2x + 2y = 100\). Solving for \(y\) we get \(y = 50 - x\), so the area is

\[
f(x) = x(50 - x),
\]

defined on the interval \([0, 50]\). At the endpoints we have \(f(0) = f(50) = 0\), corresponding to the pen being all length or all width. We compute \(f'(x) = 50 - 2x\), which equals zero at \(x = 25\). We evaluate \(f(25) = (25)^2\); this is the maximum area.

The constraint on the perimeter is not important. For any perimeter \(P\) the maximal rectangular pen would be a square of side \(P/4\) and area \(P^2/16\).
Going beyond our original instructions to build a rectangular pen, suppose we built a circular pen with the same perimeter. Then we would do slightly better. If \( r \) is the radius then the perimeter is \( P = 2\pi r \), so \( r = P/2\pi \) and the area is

\[
\pi r^2 = \pi \frac{P^2}{4\pi^2} = \frac{P^2}{4\pi} > \frac{P^2}{16}
\]

since \( \pi < 4 \). In fact a circle is the biggest pen you can make, but we cannot prove this in a beginning calculus course; it is a theorem in the *Calculus of Variations*.

The square is the most symmetric of rectangles and the circle is the most symmetric plane figure of all. The above results illustrate a general principle:

*the optimal configuration is the most symmetric one.*

However, the constraint can make it difficult to find the most symmetric configuration.

**Example 2:** What is the maximal area of two identical rectangular pens side-by-side, sharing a common wall, that you can make with 100 feet of fencing?

Let \( x \) be the length of the common wall, and let \( y \) be the other side of each rectangle. Then the total area of the two rectangles is \( 2xy \), and the given perimeter is

\[
3x + 4y = 100.
\]

Solving for \( y = 25 - (3/4)x \), the area is

\[
f(x) = 2x \left(25 - \frac{3}{4}x\right) = 50x - \frac{3}{2}x^2
\]

defined on the interval \([0, 100/3]\). Again the endpoints give zero area, so the maximum occurs when

\[
f'(x) = 50 - 3x = 0,
\]

so \( x = 50/3 \) and \( y = 25/2 \). The symmetry here is that for any perimeter \( P \), the maximum occurs when \( 3x = 4y = P/2 \).

**Example 3:** There are \( n \) apple trees in an orchard. Each tree produces \( a \) apples. For each additional tree planted in the orchard, the output per tree drops by \( d > 0 \) apples. How many trees should be planted or cut down to maximize the number of apples produced by the orchard?

Let \( x \) be the number of additional trees planted (or cut down, if \( x < 0 \)). Then we have \( n + x \) trees, each producing \( a - dx \) apples. So the total number of apples will be

\[
f(x) = (n + x)(a - dx) = na + (a - nd)x - dx^2.
\]
Note that $x \geq -n$ since we can’t cut down more trees than we already have, so $x + n \geq 0$. This means that $f(x)$ is positive only if $-n < x < a/d$. Thus, our interval is $[-n, a/d]$, with the endpoints giving zero apples. The derivative $f'(x) = a - nd - 2dx$ is zero when $x = (a - nd)/2d$. Note that $x$ is negative if $a < nd$, meaning that we have too many trees. It could happen that $(a - nd)/2d$ is not an integer. In this case choose the nearest integers $x', x''$ on either side of $(a - nd)/2d$, find the larger of $f(x')$ and $f(x'')$ and plant (or cut down) either $x'$ or $x''$ trees, accordingly.

**Example 4:** Two points $P$ and $Q$ lie offshore from a straight mainland. It is required to lay an underwater cable from $P$ to the mainland, thence to $Q$. How do we make the cable as short as possible?

Let the mainland shore be the $x$-axis, and let $P = (a, b)$, $Q = (c, d)$, with $b, d > 0$. The cable will go from $P$ to a point $X = (x, 0)$ on the mainland, thence to $Q$. We want to minimize the sum of distances $PX + XQ$, which is given by the function

$$f(x) = \sqrt{(x-a)^2 + b^2} + \sqrt{(x-c)^2 + d^2}.$$ 

Clearly $x$ is between $a$ and $c$. Setting $f'(x) = 0$, we get

$$(x-a)\sqrt{(x-c)^2 + d^2} = -(x-c)\sqrt{(x-a)^2 + b^2}.$$ 

Squaring both sides, and solving for $x$, we get

$$x = \frac{ad + bc}{b + d}.$$ 

**Example 6:** Find the point on the graph of $y = x^3$ where the curvature is maximal.

Recall that the curvature is given by

$$\kappa(x) = \frac{y''}{(1 + (y')^2)^{3/2}} = \frac{6x}{(1 + 9x^4)^{3/2}}.$$ 

To maximize $\kappa(x)$ is equivalent to maximizing

$$f(x) = \frac{x^2}{(1 + 9x^4)^3},$$ 

for $x \geq 0$. Here there is no closed interval. However, $f(x) \to 0$ as $x \to \infty$, so it has a maximum on $[0, \infty)$. Solving $f'(x) = 0$ we find that

$$x = (45)^{-1/4}.$$
Appendix: Curvature and the Osculating Circle

In this section we will derive the formula for Curvature used in section 3.6. We will need (and will prove) a difference-quotient formula for the second derivative.

Take three points on the graph of \( y = f(x) \):

\[ P = (a, f(a)), \quad Q = (a + h, f(a + h)), \quad R = (a - h, f(a - h)), \]

and let \( \overrightarrow{PQ} \) and \( \overrightarrow{RP} \) be the vectors from \( Q \) to \( P \) and from \( P \) to \( R \). \(^{14}\) Thus,

\[ \overrightarrow{PQ} = h(1, \Delta_+), \quad \overrightarrow{RP} = h(1, \Delta_-), \]

where \( \Delta_+ \) and \( \Delta_- \) are the difference quotients

\[ \Delta_+ = \frac{1}{h}(f(a + h) - f(a)), \quad \Delta_- = \frac{1}{h}(f(a) - f(a - h)). \]

Next, let \( M, N \) be the midpoints of the segments \( PQ, RP \). Explicitly,

\[ M = P + \frac{1}{2} \overrightarrow{PQ} = (a + \frac{h}{2}, f(a) + \frac{h}{2} \Delta_+), \quad N = P - \frac{1}{2} \overrightarrow{RP} = (a - \frac{h}{2}, f(a) - \frac{h}{2} \Delta_-). \]

The center \( O \) of the circle through \( P, Q, R \) is the intersection of the perpendicular bisectors of \( PQ \) and \( PR \). Thus \( O \) is determined by the conditions

\[ \overrightarrow{OM} \perp \overrightarrow{PQ}, \quad \overrightarrow{ON} \perp \overrightarrow{RP}. \]

Setting \( O = (x, y) \), these become the system of equations:

\[ x + y \Delta_+ = a + \frac{h}{2} + f(a) \Delta_+ + \frac{h}{2} \Delta_+^2, \]
\[ x + y \Delta_- = a - \frac{h}{2} + f(a) \Delta_- - \frac{h}{2} \Delta_-^2. \]

Solving this for \( x \) and \( y \) we get

\[ x = a - \frac{h}{2} \frac{\Delta_+ + \Delta_-}{\Delta_+ - \Delta_-} - \frac{h}{2} \frac{\Delta_+ \Delta_- (\Delta_+ + \Delta_-)}{\Delta_+ - \Delta_-}, \]
\[ y = f(a) + \frac{h}{\Delta_+ - \Delta_-} + \frac{h}{2} \frac{\Delta_+^2 + \Delta_-^2}{\Delta_+ - \Delta_-}. \]

Using the second derivative difference quotient (see next section)

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

\(^{14}\)See Lang pp. 531-537 for background on vectors.
and letting $h \to 0$ we get

$$x \to a - \frac{f'(a)}{f''(a)} - \frac{f'(a)^3}{f''(a)}, \quad y \to f(a) + \frac{1}{f''(a)} + \frac{f'(a)^2}{f''(a)}.$$ 

So the center $C = \lim_{h \to 0} (x, y)$ of the osculating circle is $P + \overrightarrow{PC}$, where

$$\overrightarrow{PC} = \left( \frac{1 + f'(a)^2}{f''(a)} \right) (-f'(a), 1),$$

and the radius of curvature is the length

$$||\overrightarrow{PC}|| = \frac{(1 + f'(a)^2)^{3/2}}{|f''(a)|}.$$ 

This proof uses the second derivative difference quotient, whose validity requires certain conditions on $f(x)$. We investigate this next.
The second derivative difference quotient

The second derivative \( f''(x) \) is of course a derivative of a derivative, hence is a limit of limits. In fact for nice functions (read on) it can be expressed as a single limit:

\[
f''(a) = \lim_{h \to 0} \frac{f(a+h) - 2f(a) + f(a-h)}{h^2}. \tag{16}
\]

The internet seems to contain only false proofs fo formula (16), but at least on Wikipedia we find the following cautionary example: the function defined by

\[
f(x) = \begin{cases} 
\frac{x}{|x|} & \text{if } x \neq 0 \\
0 & \text{if } x = 0
\end{cases}
\]

has the right hand limit of (16) equal to zero for \( a = 0 \), but \( f(x) \) is not even continuous at \( a = 0 \), so \( f''(0) \) does not exist.

From this example we conclude that, unlike the ordinary difference quotient which defines \( f'(a) \), the limit (16) cannot be used as the definition of \( f''(a) \). In order for (16) to hold at all, the function \( f(x) \) must clearly satisfy further conditions (that would exclude the above example).

This is a problem, because we have used (16) to find the center of the osculating circle to the graph of \( f(x) \). Indeed, in our previous notation

\[
\Delta_+ = \frac{1}{h}[f(a+h) - f(a)], \quad \Delta_- = \frac{1}{h}[f(a) - f(a-h)],
\]

the formula (16) is equivalent to

\[
\lim_{h \to 0} \frac{\Delta_+ - \Delta_-}{h} = f''(a),
\]

which we used. Note that

\[
\lim_{h \to 0} \Delta_+ = f'(x) = \lim_{h \to 0} \Delta_-,
\]

so this is a 0/0 situation (the false proofs overlook this), so we have to be careful. According to our classification of limits, (16) is a Hard Limit!

The further conditions on \( f \) that we need are that \( f''(a) \) exists (obviously, in order for (16) to make any sense at all) and moreover that \( f'' \) is continuous on an open interval containing \( a \). Assuming these two conditions hold for \( f \), we can now prove (16).

We start with the second degree Taylor approximation with remainder:

\[
f(a + h) = f(a) + hf'(a) + \frac{h^2}{2}f''(c_h)
\]

for some number \( c_h \) between \( a \) and \( a + h \). This holds under our conditions on \( f(x) \), see Lang p. 427-437. Also, we write \( c_h \) to indicate that \( c_h \) depends on \( h \). Likewise we have

\[
f(a - h) = f(a) - hf'(a) + \frac{h^2}{2}f''(c'_h)
\]

86
for some number $c'_h$ between $a$ and $a - h$. Now move $f(a)$ to the left sides of both equations, add them and divide by $h^2$. We get

\[
\frac{f(a + h) - 2f(a) + f(a - h)}{h^2} = \frac{1}{2} [f''(c_h) + f''(c'_h)].
\] (17)

The unknown numbers $c_h$ and $c'_h$ lie in the interval $[a - h, a + h]$, so as $h \to 0$, we have $c_h \to a$ and $c'_h \to a$. Since $f''$ is assumed to be continuous at $a$, we have

\[
\lim_{h \to 0} f''(c_h) = f''(a) = \lim_{h \to 0} f''(c'_h).
\]

Hence as $h \to 0$, the right side of (17) goes to $f''(a)$. Thus we have proved formula (16), under the assumptions that $f''$ exists and is continuous on an open interval containing $a$. 